GARCH-Type models and Performance of Information Criteria

Farrukh Javed and Panagiotis Mantalos

Department of Statistics, Lund University, Box 743, S-220 07 Lund, Sweden

Abstract:

GARCH models have been gaining popularity since the last two decades probably because of their ability to capture non-linear dynamics in the real life data which we often observe especially in financial markets. This paper discusses the relative ability of some common information criteria (AIC, AICc, SIC and HQ) using their probability of correct selection, as a measure of performance, in the presence of GARCH effect. The investigation has been performed using Monte Carlo simulation of conditional variance GARCH processes with 6 different kinds of DGPs including $GARCH(p,q)$ for $p=1,2$ and $q=1,2$, GARCH (1,1)-Leverage, and GARCH(1,1)-Spillover. All these models are further simulated with different parameter combinations to study the possible effect of volatility structures on these information criteria. We noticed an impact of volatility structure of time series on the performance of these criteria.

Keywords: GARCH, Volatility, Leverage, Spillover.

1. Introduction:

In financial time series, modeling real data needs proper attention and suitable model selection is required to better understand the structure of the candidate data which ultimately helps in better forecasting. Because these selected models are later used for policy making whether in finance or economics. The reason for this care is the nonlinear dynamics present in such data. For financial data, it is sometime obvious to find volatility clusters in the candidate data. Volatility clustering refers to the observation that “large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes” (Mandelbrot, 1963). In other words, volatility depends more on recent past values than distant past values. Many research papers have been presented to discuss the random behavior of volatility.

In the initial empirical research of finance, researchers were mainly focused on the linear relationship of data and most of the modeling was done by the popular Box and Jenkins approach. In ARMA framework, current value of the series of interest is written as a linear function of its own lagged values and current and past noise process or innovations. These innovations are further assumed to be independent and identically distributed or, technically speaking, homoskedastic in nature. Because of this property, these models are not suitable for non-linear financial and monetary problems. One of the challenges for econometricians is to specify how this non-linear information can be used to forecast the mean and variance of the return of any asset or portfolio conditioned on the past information. Most of the early empirical work on finance used equally weighted average of the squared residuals of most recent returns as the method for modeling variance. But as we mentioned before, volatility of asset returns moves in clusters depending on recent past values more than distant past values, this phenomenon seems unrealistic in this case and allowed a suitable approach to be introduced.

Finally, the model is developed to capture such behavior namely Autoregressive Conditional Heteroskedasticity (ARCH), originally proposed by Robert Engle in his celebrated paper in 1982, which arise as a good framework to discuss such issues and later its generalized version GARCH, proposed by Bollerslev (1986). The original model is:

$$y_t = \varepsilon_t h_t, \qquad \varepsilon_t \sim N(0,1), \quad t=1,...,n$$

$$h_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 \qquad y_t \mid \psi_t \sim N(0, h_t)$$

$$\textit{where}$$

$$q \geq 0, \ \alpha_0 > 0, \ \alpha_i \geq 0, \ \ i = 1,...,q$$

Where y_t is the dependent variable, ε_t is the disturbance term and ψ_t is the information set available at time t, here p shows the lag length of ARCH model and $\alpha's$ are the vector of unknown parameters. In the variance equation, Engle (1986) allowed the best weights ($\alpha's$) to be estimated for the data for forecasting the variance.

And GARCH(p,q) model is:

$$y_t = \varepsilon_t h_t, \qquad \varepsilon_t \sim N(0,1), \quad t=1,...,n$$
$$h_t^2 = \alpha_0 + \sum_{i=1}^{p} \beta_i h_{t-i}^2 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 \qquad \varepsilon_t \mid \psi_t \sim N(0, h_t)$$
$$where$$
$$p \geq 0,\ q \geq 0,\ \alpha_0 > 0,\ \alpha_i \geq 0, \quad i = 1,...,q$$
$$\beta_i \geq 0, \quad i=1,...,p$$

For $p = 0,$ the process reduces to the ARCH(q) process, and for $p = q = 0,$ ε_t is simply a white noise.

There is a plethora of models which have been proposed by researchers (e.g., Nelson (1991), Zakoian (1994), Engle and Bollerslev (1986), Baillie et al (1996) among others) subsequent to the work of Engle (1982) to discuss different dynamical structure related to the temporal variation in financial data according to need. In ARCH (q) process the conditional variance is specified as a linear function of past values but the GARCH (p,q) process also allows the conditional variances to depend linearly on the past behavior of squared errors. Such specification can be thought of as adaptive learning mechanism. Under GARCH framework, the shocks persist according to an autoregressive moving average (ARMA) structure of the squared residuals of the process. Empirically GARCH model is proved to be the useful mean for capturing the momentum of conditional variance in financial data.

Despite the fact that many models have been proposed belonging to the ARCH family (e.g. Bera and Higgins (1993), Bollerslev et. al. (1994) and Diebold and Lopez (1995) for a survey), existing studies, almost unanimously, agreed on the performance of standard GARCH (1,1) model rather than attempting to determine the 'appropriate' lag values. The reason can possibly be the attitude of researchers that first lag of conditional variance is sufficient to capture all the volatility clustering present in the data. This supposition can be misguided if the real data has high order variance structure so care should be taken to insure the true specification of volatility structure for the candidate data. Because mis-specified variance models (say, GARCH) can influence the forecast accuracy (Brooks and Burke (1998)). We tried to simulate higher order GARCH process as well, to study if there are any such misspecifications.

The purpose of this paper is to study the ability of the traditional information criteria (IC) to identify the real DGP in the presence of GARCH effects. For mean equation, Akaike (Akaike, 1973) and Schwarz (Schwarz, 1978) information criteria are quite popular when choosing the true model among alternative specifications. But when the variance equation is jointly estimated, there accuracy has always been ambiguous. These traditional criteria seem inappropriate for ARCH type errors because they focus on the first moment of the data (Pagan and Schwert, 1990). There have been several modifications proposed by different researchers (Brooks and Burke, 2003, Hughes et al, 2004 and Bierens, H. J., 2006) to incorporate ARCH effect in the IC. We simulated GARCH (p,q) error models for p=1,2 and q=1,2. These models are further simulated at different parameter combinations keeping in mind the volatile behavior of financial time series. The reason for such study is to investigate the possible effect of such series on the previously proposed data selection methods. A model with possible leverage effect is also simulated at two different leverage levels. The rest of the paper proceeds as follows. Section 2 explains the brief description about model selection criteria used in this paper. Section 3 deals with the theory and modeling method of spillover because causality in variance is also an important factor when modeling financial time series across different markets. Section 4 presents the estimated results and discusses their implications. Section 5 finally summarizes the finding and comments.

2. Model Selection Criteria:

All the inference and evaluation of real life data depends on the true specification of model. Choosing most 'adequate' model is the essence of data analysis, which ultimately returns with good forecasting results. This leads to the importance of model selection criteria for assessing the 'goodness' of a specified model. Finding optimal dimension of a model that will fit a candidate data has always been favorite for researchers. The number of parameters in the model plays an important role in both analysis and forecasting. Because the addition of unnecessary lags reduces the sum of squares of estimated residuals and the forecasting performance as well.

According to Akaike (1973), a model should be evaluated on the basis of good results when it is used for prediction. We know the prediction error study through residual analysis is one of the popular method used for model validation. He suggested method for evaluating model in terms of Kullback–Leibler information, which is based on the concept of closeness between generic distribution $g(x)$ defined by the model and the true distribution $f(x)$ besides the more commonly used method of simply minimizing the prediction error. Kullback-Leibler information $I(f,g)$ is the information lost when model $g(x)$ is used to approximate $f(x)$, this is defined for continuous functions as the integral

$$I(f,g) = \int f(x) \ \ln\left[\frac{f(x)}{g(x \mid \theta)}\right] dx$$

Here it is obvious that the best model loses least information among others in the set.

Four different, popularly known, methods are used for our purpose. The one most popular is Akaike's (1973) information criterion. The Akaike Information Criterion (AIC) has commonly been used and significantly known method in the model selection for decades in a wide variety of fields for analyzing actual data. It was designed to be an asymptotically unbiased estimate of the Kullback-Leibler index of the fitted model relative to the true model. For example, in the special case of least squares (LS) estimation with normally distributed errors AIC can be expressed as:

$$AIC = -2\log\left(\hat{\sigma}^2\right) + 2k$$

where $\hat{\sigma}^2$ is the estimated model error variance,

'k' is the number of free parameters in the model,

Bierens, H.J. (2006) recommended the following modification, if the model includes ARCH type errors.

$$AIC = -2\log\left(\hat{\sigma}^2\right) + 2(k) - 1 - \log(2\pi)$$

Akaike is the measure of goodness-of-fit (likelihood) of an estimated statistical model. In AIC, the estimation is divided into two sets and the compromise takes place between the log maximized likelihood i.e., $-2\log(\hat{\theta})$ (the lack of fit component) and k (the penalty component) which increases with the number of parameters and it is used to prevent overfitting. Panelizing likelihood in the model is an attempt to select a more parsimonious model.

Existing studies showed that AIC is not consistent and hence does not lead to the choice of the correct model, with high probability, in large samples. Shibata (1976) showed through empirical evidences that AIC has the tendency to choose models which are over-parameterized. Various modifications have been produced to overcome this lack of consistency. Schwarz (1978) developed a consistent criterion for models defined in terms of their posterior probability (Bayesian approach) which is given as:

$$SIC = -2\log\left(\hat{\sigma}^2\right) + k\log\left(n\right)$$

where $\hat{\sigma}^2$ is the estimated model error variance,

k is the number of free parameters in the model,

n is the number of observations,

In ARCH context, this form will look like;

$$SIC = -2\log\left(\hat{\sigma}^2\right) + \left(k\right)*\log\left(n\right) - 1 - \log\left(2\pi\right)$$

From the above formulation, it can be seen that SIC is used for models estimated by using the maximum likelihood method and that the criterion is obtained under the condition that the sample size n is made sufficiently large. Burnham and Anderson (2002, 2004) provided convincing arguments in favor of AIC (or AICc) over SIC.

The general belief about an information criterion is that it provides a measure of information that strikes a balance between the measure of goodness of fit and parsimonious specification of the model. Various information criteria differ in how to strike this balance. Therefore the chosen model would be the one which minimizes the value of the information criterion, so that a model with larger number of lags would only be selected if the minimized value of log-likelihood outweighed the value of the penalty term.

Hannan-Quinn (1979) and Hannan (1980) proposed another consistent criterion for the order of an autoregression based on law of iterated logarithm, which has the quality for penalty function to decrease as fast as possible for a strongly consistent estimator, as the sample size increases. The criterion is given by

$$HQ = -\log\left(\hat{\sigma}^2\right) + 2k\log\left(\log\left(n\right)\right)$$

where $'n'$ is the number of observations.

In ARCH context;

$$HQ = -\log\left(\hat{\sigma}^2\right) + 2\left(k\right)*\log\left(\log\left(n\right)\right) - 1 - \log\left(2\pi\right)$$

In 1989, Hurvich and Tsai found that SIC, which was modified from AIC, is not asymptotically efficient. So they suggested the bias corrected version of AIC, known as Akaike's Information Corrected Criterion or AICc, which is given as:

$$AICc = -2\log\left(\hat{\sigma}^2\right) + 2\frac{kn}{n-k-1}$$

The correction is of particular use when the sample size is small, or when the number of fitted parameters is relatively large to sample size $'n'$.

In ARCH context, we use the following form for AICc;

$$AICc = -2\log\left(\hat{\sigma}^2\right) + \left\{\frac{2kn}{n-k-1}\right\} - 1 - \log(2\pi)$$

In small sample, AIC and AICc may perform better towards selecting the true DGP than HQ and SIC. As, the former two criteria are designed for minimizing the forecast error variance, models based on AIC and AICc may produce superior forecasting results although they may not select the models correctly (e.g. Lutkepohl 2005).

3. Volatility Spillover:

Detection of causality in variance has been gained more attention in recent decades across various financial asset price movements. By causality we mean causality in the Granger (1969, 1980) sense. Another equally important issue is the persistence in the variance of a random variable, which has been evolving through time. Persistence in variance refers to the conditional variance. However, recently increasing interest is shown in the causation in conditional variance across various financial asset price movements. The causation in variance between a pair of economic time series is popularly known as Volatility Spillover.

Many of the empirical and theoretical studies, on the relationship between different assets, concentrate on the spillover between economies and within economy. The study of volatility spillover helps understanding how the information is transmitted across assets and markets and how market assimilates new information. According to Ross (1989), volatility is often related to the rate of information flow. Absence of volatility spillover implies that one large shock increases the volatility only in that specific asset or market while in the existence of spillover it could increases the volatility not only in its own assets or market but also in other assets or markets (e.g. Hong, Y, 2001).

The model for the spillover in a GARCH process is considered as:

$$h_{1t} = \gamma + \psi_1 h_{1t-1} + \phi_1 \varepsilon_{1t-1}^2 + \zeta_{12} h_{2t-1} + \varphi_{12} \varepsilon_{2t-1}^2$$

To investigate the performance of Information Criteria in the presence of volatility spillover, a set of Monte Carlo experiment with different sample size (n =500, 1000, 2000) is executed and the results are summarized in Table 5.

4. Monte Carlo Methodology:

The model we considered, for our purpose, is $y_t = \varepsilon_t h_t, \quad \varepsilon_t \sim N(0,1), \ t = 1,...,n$ with

$$\begin{aligned} &h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} \\ &with \ \ \alpha_1 = 0.089, \ \beta_1 = 0.85 \\ &\qquad \alpha_1 = 0.15, \ \beta_1 = 0.80 \\ &\qquad \alpha_1 = 0.15, \ \beta_1 = 0.75 \end{aligned} \qquad GARCH(1,1)$$

$$\begin{aligned} &h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \beta_1 h_{t-1} \\ &with \ \ \alpha_1 = 0.15, \ \alpha_2 = 0.15 \ \beta_1 = 0.65 \\ &\qquad \alpha_1 = 0.25, \ \alpha_2 = 0.15 \ \beta_1 = 0.55 \end{aligned} \qquad GARCH(1,2)$$

$$\begin{aligned} &h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} + \beta_2 h_{t-2} \\ &with \ \ \alpha_1 = 0.1, \ \beta_1 = 0.45, \ \beta_2 = 0.40 \\ &\qquad \alpha_1 = 0.22, \ \beta_1 = 0.40, \ \beta_2 = 0.35 \end{aligned} \qquad GARCH(2,1)$$

$$h_t = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \alpha_2\varepsilon_{t-2}^2 + \beta_1 h_{t-1} + \beta_2 h_{t-2}$$

with $\alpha_1 = 0.10,\ \alpha_2 = 0.1,\ \beta_1 = 0.42,\ \beta_2 = 0.35$ and

$\alpha_1 = 0.10,\ \alpha_2 = 0.12,\ \beta_1 = 0.38,\ \beta_2 = 0.38$ $\quad GARCH(2,2)$

$$\ln(h_t) = \alpha_0 + \alpha_1\left(\frac{|\varepsilon_{t-1}|}{\sqrt{h_{t-1}}} - \sqrt{\frac{\pi}{2}}\right) + \beta_1\ln(h_{t-1}) + \lambda_1\left(\frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}}\right) \qquad GARCH-Leverage$$

with $\alpha_0 = 0.1,\ \alpha_1 = 0.089,\ \beta_1 = 0.85,\ \lambda_1 = -0.4$ and

$\alpha_0 = 0.1,\ \alpha_1 = 0.15,\ \beta_1 = 0.80,\ \lambda_1 = -0.8$

$$h_{it} = \gamma + \alpha_i\varepsilon_{it-1}^2 + \beta_i h_{it-1} + \phi_i\varepsilon_{jt-1}^2 + \psi_i h_{jt-1}$$

with $\alpha_i = 0.15,\ \beta_i = 0.35,\ \phi_i = 0.12,\ \psi_i = 0.35$ and $\quad GARCH-Spillover$

$\alpha_i = 0.1,\ \beta_i = 0.42,\ \phi_i = 0.12,\ \psi_i = 0.33$

We simulated a total of 10 different models, including $GARCH(1,1)$ at three different frequency levels (Low, Medium, and High), $GARCH(1,2), GARCH(2,1)$ nd $GARCH(2,2)$, GARCH-Leverage, and GARCH-Spillover at different parameter values for sample sizes n=500,1000 and 2000 . In the next step, simulated models are fitted to the $GARCH(p,q)$ type models with $p = 1,2$ and $q = 1,2$ to check the performance of information criteria in favor of true data generating process (DGP).

For each combination, a sample size of n observations are generated with n=500, 1000 and 2000. The minimum sample size of 500 observations is selected because the detection of true ARCH effects is less probable for samples fewer than 300 and around 500 observations is desired for fitting ARCH models (McClain et al, 1996). Large sample sizes (1000 and 2000) are chosen to approximately represent heavy set of real life financial data and also to observe the structure more accurately. Frequency levels are defined according to pre-assigned initial values of parameters.

Then we used four different and popularly known information criteria to check the performance of simulation which are AIC, SIC, HQ and AICc. The appropriate information criterion is one which most frequently selects the model used to generate the data. Thus the optimality of model selection is defined as how often does the criterion select the most appropriate model among all combinations. We estimated all 13 combinations of models and evaluate the criteria, and the results are summarized in the next chapter.

We used 1000 replications for each model to study the performance. All the simulations are performed using the GAUSS 8.0 random number generator.

5. Results

6.1 GARCH(1,1):

The focus of this paper is on the relative ability of the model selection criteria to detect GARCH type data generating process. The first of the models to be considered is $GARCH(1,1)$ which is used to generate n different random numbers. A set of $GARCH(p,q)$ with $p = 1,2$ and $q = 1,2$ are then fitted to this data and with the help of four information criteria the optimal model is identified. The process is repeated 1000 times and the result is summarized in Table 1.1 where the highlighted area shows the estimated probability of correct selection.

Table 1.1 shows, using the probability of correct selection as a measure of performance, that SIC and HQ outperform both the AIC and AICc in each case considered and the general performance of these model selection criteria is slightly lower in the presence of GARCH effect. SIC selected the correct model as an optimal with the performance rate of almost 96% for $n = 500$ followed by HQ with 85.5% while the AIC and AICc managing to correctly identify the generating process almost 71% and 72% of the times respectively. Here we observed the consistency of SIC and HQ towards the correct model. For n=1000, the performance rate for SIC is once again higher than others with 96.6%. Here again HQ performed the second best criterion with 87.9 selection rate of true DGP, followed by AIC with 74.2% of times which is slightly greater than the performance rate at n=500. The performance of AICC is much closer to AIC at this stage. Further, we notice that increase in sample observation affects the criterion performance positively. Like, at n=2000, the same criterion SIC chose the correct model with highest performance rate of 98% and HQ remained the next best with 90.7% performance level.

On the other side, the performance of these criteria toward selecting other models is quite low and likely to be ignored. Here AIC and AICc selected $GARCH(1,2)$ as optimal model with some level of performance compared to SIC and HQ. For n=500, AIC selected the same model 14% of the times While AICc also performed nearly close to AIC with 13.7% for n=500. Similarly with increased sample size like n=2000, the performance rate for AIC and AICc decreased slightly with 12.2%.

As we know the structure of financial time series varies with respect to volatility, keeping this in mind, we simulated some other GARCH (1,1) models with different parameter combinations to study the impact of such series on information criteria. The results for one such model is presented in Table 1.2. We observed a similar pattern of correct selection for all the considered criteria. Both SIC and HQ outperform AIC and AICc with high rate of performance. Notice that in this model, we marginally increase the clustering parameter α with the sum indicates process near integration, but we observed it doesn't affect the information criteria. In the last case, another combination is simulated with a slight lesser value of β and same value of α , to examine if it effects the performance rate or not. Here the process is slightly away from integration. Table 1.3 summarizes the results and again we observed a consistence performance of SIC and HQ with high rate of selection among all the considered criteria.

As we mentioned earlier, Akaike Information Criteria (AIC) is biased towards selecting higher order $GARCH(p,q)$ models and their performance is not satisfactory for small sample size which we satisfied through our study. We conclude SIC is observed as best among all criteria followed by HQ in $GARCH(p,q)$ model selection and continued to exhibit excellent model selection abilities.

Table:1.1 GARH(1,1) $(\alpha_1 = 0.089,\ \beta_1 = 0.85\ )$

Obs_500	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	**0.708**	0.140	0.125	0.027
SIC	**0.958**	0.018	0.023	0.001
HQ	**0.855**	0.064	0.076	0.005
AICc	**0.717**	0.137	0.120	0.026
Obs_1000	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	**0.713**	0.128	0.134	0.025
SIC	**0.966**	0.020	0.014	0
HQ	**0.879**	0.064	0.051	0.006
AICc	**0.715**	0.128	0.133	0.024

Obs_2000	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	**0.739**	0.122	0.117	0.022
SIC	**0.980**	0.009	0.011	0
HQ	**0.907**	0.050	0.040	0.003
AICc	**0.741**	0.122	0.116	0.021

Table:1.2 GARH(1,1) $\left(\alpha_1 = 0.15,\ \beta_1 = 0.80\right)$

Obs_500	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	**0.726**	0.121	0.126	0.027
SIC	**0.961**	0.015	0.024	0
HQ	**0.865**	0.062	0.068	0.005
AICc	**0.731**	0.120	0.124	0.025
Obs_1000	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	**0.726**	0.116	0.137	0.021
SIC	**0.961**	0.021	0.018	0
HQ	**0.876**	0.058	0.062	0.004
AICc	**0.730**	0.115	0.135	0.020
Obs_2000	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	**0.741**	0.124	0.114	0.021
SIC	**0.975**	0.015	0.010	0
HQ	**0.909**	0.048	0.040	0.003
AICc	**0.741**	0.124	0.114	0.021

Table:1.3 GARH(1,1) $\left(\alpha_1 = 0.15,\ \beta_1 = 0.75\right)$

Obs_500	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	**0.738**	0.108	0.133	0.021
SIC	**0.962**	0.015	0.023	0
HQ	**0.867**	0.056	0.074	0.003
AICc	**0.742**	0.108	0.131	0.019
Obs_1000	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	**0.712**	0.127	0.140	0.021
SIC	**0.969**	0.018	0.013	0
HQ	**0.879**	0.066	0.052	0.003
AICc	**0.714**	0.127	0.139	0.020
Obs_2000	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	**0.736**	0.125	0.121	0.018
SIC	**0.985**	0.005	0.010	0
HQ	**0.909**	0.045	0.045	0.001
AICc	**0.737**	0.125	0.120	0.018

5.2 GARCH(2,1):

Table 2.1 presents the results for the estimation procedure with slightly higher dimension where a GARCH(2,1) is simulated to study the effect of increasing GARCH parameters in the model. For n=500, none of the criteria correctly identify the true DGP and choose the lowest possible GARCH(1,1) model. The average probability of underfitting for SIC remains high followed by HQ, while two Akaike's specifications show little tendency towards the selection of true DGP with around 20% of times. A change in their performance is observed with the increase in sample observation. For n=2000, AIC and AICc select the correct model almost 44% of the times but their level of performance is still unsatisfactory. While SIC and HQ remain consistent with their choice of GARCH(1,1) as the correct model. Another process, with a change in parameter combination, is simulated and the results are contained in Table 2.2. In this case, we noticed an impact of volatility structure on the performance of these proposed criteria. Initially at small sample size, AIC and AICc detect the true data with almost 38% of times but their performance gets better with the increase in sample size. Both the Akaike's specification managed to select the right model almost 72.5% of times for n=2000. Similarly this change is noticed for SIC and HQ and they chose the right model 29% and 56% of times respectively. Here we observed a high probability for HQ towards the correct model but SIC remains consistent with GARCH(1,1). As we mentioned earlier, we observed here the volatile structure of data do impact the performance information criteria but a large sample size is needed to properly capture such behaviour. For small sample size, these criteria could mislead the selection of correct model specially with higher dimension.

Table 2.1: GARCH(2,1) $(\alpha_1 = 0.1,\ \beta_1 = 0.45,\ \beta_2 = 0.40)$**:**

Obs_500	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	**0.759**	0.04	0.198	0.003
SIC	**0.976**	0.003	0.021	0
HQ	**0.901**	0.01	0.088	0.001
AICc	**0.765**	0.039	0.193	0.003
Obs_1000	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	**0.649**	0.046	0.299	0.006
SIC	**0.947**	0.003	0.050	0
HQ	**0.831**	0.015	0.153	0.001
AICc	**0.653**	0.045	0.297	0.005
Obs_2000	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	**0.526**	0.023	0.437	0.014
SIC	**0.923**	0.001	0.076	0
HQ	**0.735**	0.008	0.257	0
AICc	**0.526**	0.023	0.437	0.014

Table 2.2: GARCH(2,1) $(\alpha_1 = 0.22,\ \beta_1 = 0.40,\ \beta_2 = 0.35)$**:**

Obs_500	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	**0.598**	0.016	0.383	0.003
SIC	**0.936**	0.002	0.062	0
HQ	**0.79**	0.005	0.204	0.001
AICc	**0.609**	0.015	0.373	0.003
Obs_1000	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	0.463	0.017	**0.501**	0.019
SIC	**0.868**	0.002	0.130	0
HQ	**0.661**	0.006	0.333	0
AICc	0.467	0.017	**0.498**	0.018
Obs_2000	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)

AIC	0.222	0.006	**0.728**	0.044
SIC	**0.711**	0	0.289	0
HQ	0.429	0.002	**0.560**	0.009
AICc	0.225	0.006	**0.725**	0.044

5.3 GARCH(1,2):

Next we proceed for some different combination of GARCH-type model and study the effect of traditional IC's in the presence of such models. None of these criteria scored perfectly towards selecting the correct model for n=500. AIC and AICc manage to outperform SIC and HQ but with small probability. But these two criteria continued to show an improvement in performance with sample size, detecting the correct data generating process almost 78% of times. There is an improvement observed for HQ and SIC as well and they could manage to detect the true process 69% and 49% of times for n=1000 respectively. For larger sample observations, all of these criteria unanimously chose the true DGP with high performance rate. Here, we again noticed an impact of sample observations defining the true process. For small sample observations, SIC and probably HQ can mislead the analyst to the wrong process or the one with lower dimension. We consider another process with the same dimension but with different parameter combination. The similar pattern of performance is observed in this case but SIC is found more consistent towards GARCH (1,1) even with sample size of 1000 observations. While the two Akaike's specification and HQ furnished a level of performance consistent with the earlier case. A negligible change in performance of these criteria is observed for both these processes.

Table 3.1: GARCH(1,2) $(\alpha_1 = 0.15,\ \alpha_2 = 0.15\ \beta_1 = 0.65\ )$**:**

Obs_500	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	0.34	**0.592**	0.007	0.061
SIC	**0.759**	0.238	0	0.003
HQ	**0.549**	0.435	0.004	0.012
AICc	0.346	**0.587**	0.007	0.06
Obs_1000	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	0.138	**0.783**	0.002	0.077
SIC	**0.508**	0.492	0	0
HQ	0.295	**0.691**	0	0.014
AICc	0.138	**0.783**	0.002	0.077
Obs_2000	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	0.025	0.886	0	0.089
SIC	0.228	0.766	0	0.006
HQ	0.082	0.896	0	0.022
AICc	0.025	0.886	0	0.089

Table 3.2: GARCH(1,2) $(\alpha_1 = 0.25,\ \alpha_2 = 0.15\ \beta_1 = 0.55\ )$**:**

Obs_500	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	0.387	**0.571**	0.004	0.038
SIC	**0.781**	0.21	0	0.001
HQ	**0.607**	0.381	0.003	0.009
AICc	0.393	**0.566**	0.004	0.037
Obs_1000	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	0.215	**0.722**	0.004	0.059
SIC	**0.622**	0.377	0.001	0
HQ	0.398	**0.591**	0.001	0.010

AICc	0.215	**0.722**	0.004	0.059
Obs_2000	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	0.062	**0.853**	0	0.085
SIC	0.380	**0.616**	0	0.004
HQ	0.151	**0.830**	0	0.019
AICc	0.062	**0.854**	0	0.084

5.4 GARCH(2,2):

Finally, among the set of GARCH-type variance models, a slightly higher dimension of model is simulated with two different set of parameters to study the dynamical nature of GARCH model. The models are simulated at n=500, 1000, and 2000 sample observations and results are contained in Table 4. Table 4.1 presents the estimated results for GARCH (2,2) process with $(\alpha_1 = 0.10,\ \alpha_2 = 0.1,\ \beta_1 = 0.42,\ \beta_2 = 0.35)$ set of parameters. None of the criterion were able to detect the higher order GARCH(2,2) process as optimal and chose GARCH(1,1) as the right model with high rate of performance. There is a shift in model selection noticed for the two Akaike's criteria with the increase in sample size more towards GARCH(1,2) with almost 36% rate of performance and almost 15% for GARCH(2,2). But for the next model with different parameter combinations, we observed marginally better performance and both the Akaike's specification could manage to correctly identify the process almost 22% of times while the two remaining criteria are found consistent with the earlier findings. We noticed a change in performance for HQ but it failed to detect the true DGP similar like SIC . For this case, we also observed an impact of parameter combination or technically speaking the volatility structure of the candidate data the performance of these criteria. We proved through simulation that as we increase α and correspondingly decrease the effect of β in the GARCH(p,q) process, these information criteria are heading towards the choice of right model with high performance rate.

Table 4.1: GARCH(2,2) $(\alpha_1 = 0.1\ \alpha_2 = 0.1\ \beta_1 = 0.42,\ \beta_2 = 0.35)$ **:**

Obs_500	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	**0.629**	0.261	0.046	0.064
SIC	**0.923**	0.069	0.007	0.001
HQ	**0.802**	0.160	0.022	0.016
AICc	**0.634**	0.259	0.045	0.062
Obs_1000	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	**0.560**	0.328	0.023	0.089
SIC	**0.921**	0.074	0.003	0.002
HQ	**0.766**	0.205	0.010	0.019
AICc	**0.563**	0.325	0.023	0.089
Obs_2000	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	**0.461**	0.368	0.018	0.153
SIC	**0.895**	0.100	0.002	0.003
HQ	**0.723**	0.239	0.006	0.032
AICc	**0.462**	0.368	0.018	0.152

Table 4.2: GARCH(2,2) $(\alpha_1 = 0.10,\ \alpha_2 = 0.12,\ \beta_1 = 0.38,\ \beta_2 = 0.38)$ **:**

Obs_500	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	**0.644**	0.229	0.017	0.11
SIC	**0.937**	0.058	0.002	0.003
HQ	**0.83**	0.139	0.005	0.026
AICc	**0.651**	0.229	0.014	0.106

Obs_1000	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	**0.465**	0.381	0.016	0.138
SIC	**0.873**	0.124	0	0.003
HQ	**0.686**	0.272	0.006	0.036
AICc	**0.469**	0.380	0.015	0.136
Obs_2000	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	0.333	**0.434**	0.009	0.224
SIC	**0.836**	0.157	0.001	0.006
HQ	**0.582**	0.343	0.002	0.073
AICc	0.336	**0.433**	0.009	0.222

5.5 GARCH Effect with Leverage:

In this sub-section we considered some of the exotic variants of the GARCH-type models. EGARCH(p,q) model is introduced by Nelson(1991) to capture asymmetric behavior of the market's response to shocks. Table 5 presents the simulation results for GARCH(1,1)-Leverage models. We simulated this process with two predefined leverage parameter λ=-0.4 and -0.8 so that we can study its effect on the performance rate of the traditional information criteria towards the true data generating process. First we consider the same GARCH(1,1) process studied earlier without leverage effect with $(\alpha_1 = 0.089,\ \beta_1 = 0.85,\ \lambda_1 = -0.4)$ combination of parameter constraints. The choice of same process will help us comparing any change because of possible leverage effect with the previous findings. In this case, we observed the findings are quite similar to the one without leverage effect and it seems leverage parameter did not affect the performance of these criteria. But for the next model with slightly bigger value of λ, a marginally inferior change is noticed for the two Akaike's criteria but the other two SIC and HQ procedures consistently exhibit the superior model selection performance comparably, managing to correctly identified the process 95% and 85% of the times respectively at n=500. All these criteria are gaining more potential towards selecting the correct model with sample size.

Table 5.1: $(\alpha_1 = 0.089,\ \beta_1 = 0.85,\ \lambda_1 = -0.4)$

Obs_500	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	**0.635**	0.077	0.144	0.144
SIC	**0.973**	0	0.019	0.008
HQ	**0.86**	0.027	0.074	0.039
AICc	**0.651**	0.076	0.14	0.133
Obs_1000	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	**0.67**	0.068	0.131	0.131
SIC	**0.985**	0.001	0.011	0.003
HQ	**0.895**	0.017	0.064	0.024
AICc	**0.675**	0.067	0.129	0.129
Obs_2000	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	**0.698**	0.066	0.112	0.124
SIC	**0.995**	0	0.005	0
HQ	**0.934**	0.009	0.042	0.015
AICc	**0.699**	0.065	0.113	0.123

Table 5.2: $(\alpha_1 = 0.15,\ \beta_1 = 0.80,\ \lambda_1 = -0.8)$

Obs_500	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	**0.623**	0.066	0.129	0.182
SIC	**0.954**	0.006	0.028	0.012
HQ	**0.85**	0.03	0.067	0.053
AICc	**0.638**	0.066	0.127	0.169
Obs_1000	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)

AIC	**0.67**	0.072	0.125	0.133
SIC	**0.982**	0.001	0.01	0.007
HQ	**0.919**	0.015	0.037	0.029
AICc	**0.678**	0.066	0.123	0.133
Obs_2000	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	**0.678**	0.068	0.101	0.153
SIC	**0.989**	0	0.004	0.007
HQ	**0.925**	0.011	0.038	0.026
AICc	**0.68**	0.067	0.101	0.152

6.6 GARCH Effect with Spillover:

Finally in Table 6, we summarized the simulation results for GARCH(1,1) with spillover. We simulated two different process with a change in parameters. First of the model considered, we restrict β to be same for both the series. All the criteria showed their tendency towards GARCH(1,1) as a correct model. SIC and HQ remained the best procedures with a high performance rate , scoring 88% and 76% accuracy respectively, but Akaike's procedures could able to detect GARCH(2,1) and GARCH(2,2) but with low probability (22% and 15% respectively) for n=500. For large sample sizes, their performance accuracy increased towards GARCH(1,1) as the true DGP and correspondingly decreased in favor of GARCH(2,1) and GARCH(2,2). For another model considered with different value of β , a similar pattern is observed and the performance of these information criteria is found largely unaltered.

So as we said before, $GARCH(1,1)$ is proved to be sufficient in capturing volatility by the practitioners, here also our simulation results are providing sound evidence to accept this model as the optimal among higher order GARCH processes.

Table 6.1: $(\alpha_i = 0.15,\ \beta_i = 0.35,\ \phi_i = 0.12,\ \psi_i = 0.35)$

Obs_500	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	**0.558**	0.067	0.224	0.151
SIC	**0.888**	0.002	0.093	0.017
HQ	**0.763**	0.02	0.163	0.054
AICc	**0.568**	0.064	0.225	0.143
Obs_1000	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	**0.581**	0.083	0.189	0.147
SIC	**0.915**	0.004	0.075	0.006
HQ	**0.798**	0.018	0.14	0.044
AICc	**0.588**	0.079	0.187	0.146
Obs_2000	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	**0.587**	0.078	0.177	0.158
SIC	**0.927**	0.001	0.069	0.003
HQ	**0.841**	0.017	0.123	0.019
AICc	**0.588**	0.077	0.177	0.158

Table 6.2: $(\alpha_i = 0.1,\ \beta_i = 0.42,\ \phi_i = 0.12,\ \psi_i = 0.33)$

Obs_500	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	**0.53**	0.097	0.233	0.14
SIC	**0.87**	0.004	0.114	0.012
HQ	**0.751**	0.017	0.189	0.043
AICc	**0.539**	0.094	0.234	0.133

Obs_1000	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	**0.543**	0.085	0.206	0.166
SIC	**0.901**	0.002	0.089	0.008
HQ	**0.799**	0.015	0.148	0.038
AICc	**0.551**	0.083	0.203	0.163
Obs_2000	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)	GARCH(2,2)
AIC	**0.547**	0.102	0.194	0.157
SIC	**0.899**	0.002	0.098	0.001
HQ	**0.808**	0.02	0.141	0.031
AICc	**0.55**	0.102	0.193	0.155

Conclusion

Finding an optimal model for any time series is one of the goals of analysis so that one can get good forecasting results with less prediction error. The model with minimum possible lags is also given priority to get parsimony and avoid unnecessary penalties. Several measures have been proposed for selection of a model which are in some sense optimal for conditional mean equations but in presence of GARCH effect, their performance is not well-understood. As we know from past research that GARCH(1,1) is proved to be sufficient in capturing volatility by the practitioners, a blind faith on GARCH(1,1) process, in every case, can result with high prediction errors. We simulated 13 different GARCH-type models to study the merits of various criteria in the role of GARCH-type model selection and this investigation is done by a huge set of Monte Carlo experiment. We simulated higher order GARCH process to study the behavior of these selection procedures in the presence of such processes. Our simulation results also provide some insight about this wrong belief. For higher order dimensions, specially for GARCH(2,2) process, these criteria chose the wrong model for small sample sizes. Among all information criteria, SIC and HQ performed well with high probability of selecting the true DGP in the presence of GARCH effect for lower dimension. But for higher dimension two Akaike's specification are found consistent in the selection of higher order GARCH process. We noticed an impact of volatility structure of time series on the performance of these criteria which can be misled if not taken seriously.

References

Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. *Proc. 2nd International Symposium on Information Theory, In: B N Petrov and F Csaki (Eds.)* Akademiai Kiado, Budapest, pp 267–81.

Akaike, H. (1974) A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, **AC-19**, 716-23.

Bera, A.K. and Higgins, M. L. (1993) ARCH Models: Properties, Estimation and Testing, *Journal of Economic Surveys*, **4**, 305-362.

Bierens, H.J. (2006)**,** *Information Criteria and Model Selection*, lecture notes.

Black, F. (1976) Studies of stock price volatility changes, *Proc. Meeting of the Business and Economics Statistics Section*, pp 177–81. (Washington, DC: American Statistical Association)

Bollerslev, T. (1986) Generalized autoregressive conditional heteroscedasticity, *Journal of Econometrics*, **31**, 307-27.

Brooks, C. And Burke, S.P. (2003) Information Criteria for GARCH model selection, *The European Journal of Finance*, **9**, 557-580.

Burnham, K. P. and Anderson D. R. (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (2nd ed.) Springer, New York.

Burnham K. P. and Anderson D.R. (2004) Multimodel inference : Understanding AIC and BIC in model selection, *Sociological Methods & Research,* **33,** 261–304.

Diebold, F. X. And Lopez, J. A. (1995) Modelling Volatility dynamics, *Macroeconometrics: Developments, Tensions and Prospects* ed K. D. Hoover (Boston, MA: Kluwer).

Engle, R. F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica,* **50,** 987–1007.

Glosten, L., Jagannathan, R. and Runkle, D. (1993) On the relation between the expected value and the volatility of nominal excess return on stocks, *Journal of Finance*, **48**, 1779-1801.

Granger, C. J. (1969) Investigating Causal Relationships by Econometrics Models and Cross Spectral Methods, *Econometrica,* **37**, 425-435.

Granger, C. W. J. (1980) Testing for causality: A personal viewpoint, *Journal of Economic Dynamics and Control*,**2**, 329-352.

Hannan, E.J. and B.G. Quinn (1979) The determination of the order of an autoregression, *Journal of the Royal Statistical Society, Series B*, **41**, 190-5.

Hughes, A. W. (2004) Selection of order of an ARCH model, *Economics Letters*, **83**, 269-275.

Hurvich, C. M. and Tsai, C. L, (1989) Regression and time series model selection in small samples, *Biometrika,* **76,** 297–307.

Nelson, D. B. (1991) Conditional heteroscedasticity in asset returns: a new approach, *Econometrica,* **59,** 347–70.

Pagan, A. G. and Schwert, W. (1990) Alternative models for conditional stock volatility, *Journal of Econometrics,* **45,** 267–90.

Ross, Stephen A. (1989) Information and Volatility: The No-Arbitrate Martingale Approach to Timing and Resolution Irrelevancy, *Journal of Finance,* **44**, 1-17.

Schwarz, G. (1978) Estimating the dimension of a model, *Annals of Statistics*, **6**, 461-4.

Schwert, William. (1989) Why Does Stock Market Volatility Change over Time? *Journal of Finance* **44**, 1115-1153.

Shibata, R. (1976) Selection of the Order of an Autoregressive Model by Akaike's Information Criterion, *Biometrika,* **63**, 117-26.

Zaköian, J.M. (1994) Threshold heteroscedastic models, *Journal of Economic Dynamics and Control,* ***18,*** 931–95.