SELECTING THE ORDER OF AN ARCH MODEL

Anthony W. Hughes*, Maxwell L. King**
and Kwek Kian Teng**

*Adelaide University and **Monash University

Working Paper 99-1

SCHOOL OF ECONOMICS
Adelaide University
SA 5005 AUSTRALIA

Selecting the Order of an ARCH Model

Anthony W. Hughes

Department of Economics,

University of Adelaide,

Adelaide, SA, 5005, Australia

Maxwell L. King and Kwek Kian Teng

Department of Econometrics and Business Statistics

Monash University

Clayton, Vic, 3168, Australia

1. Introduction and Motivation

Financial time series data has often been found to exhibit volatility clustering. By this we mean that there is often time dependence in the second moment of the error term of models that is of the form of a lag structure. One model that has been developed to capture this behaviour, and one which is very popular amongst practitioners, is the autoregressive conditional heteroskedasticity (ARCH) model, originally proposed by Engle (1982). The general form of this specification is

$$\begin{aligned} y_t &= x_t'\beta + \varepsilon_t \\ \varepsilon_t &\sim N(0,\sigma_t^2) \qquad (1) \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{i-1}^2 \end{aligned}$$

where y_t is the dependent variable of interest, x_t is a vector containing non-stochastic explanatory variables, ε_t is the disturbance term, $\theta = (\beta', \alpha_0, \alpha_1, \cdots, \alpha_q)'$ is a vector of unknown parameters and q is the order of the ARCH model (i.e. an ARCH(q) model). This paper is concerned with the selection of the appropriate value for q.

The standard method of selecting q, that was first suggested by Bollerslev (1986), is to use a sequence of two-sided Lagrange multiplier (LM) tests. This method involves testing the hypotheses $H_0 : \alpha_i = 0$ *vs.* $H_1 : \alpha_i \neq 0$ for $i = 1,2,\cdots$ in turn until the null hypothesis cannot be rejected, the final i value represents the estimate of q. One improvement that has been made to this testing procedure is to utilise the fact that every α_i must be non-negative in order to ensure the non-negativity of σ_t^2. The hypotheses therefore become $H_0 : \alpha_i = 0$ *vs.* $H_1 : \alpha_i > 0$. Lee and King (1993) showed that the corresponding one-sided LM test and a one-sided locally most mean powerful (LMMP) test both have superior power properties when compared to the two-sided LM test.

The main problem associated with choosing q by sequential testing methods such as these is that the sequence used can impact upon the final selection. For example, suppose that the true process is ARCH(4). It is possible that a rejection will not occur at an early point in the sequence, leading to the selection of an ARCH(1) or ARCH(2) model, without ever testing the significance of the ARCH(4) parameter. A better approach, which avoids the problems caused by sequential testing, involves the use of model selection criteria such as Akaike's (1973) AIC and Schwarz's (1978) BIC. The use of these criteria involves the estimation of all models under consideration and selecting the model with the largest penalised maximised log-likelihood. For a detailed discussion of the benefits of using model selection criteria as opposed to sequential tests, see Granger, King and White (1995).

Recently, Hughes and King (1994, 1998) developed a version of AIC which takes account of the non-negativity of parameters. This criterion, which is called one-sided AIC (OSAIC), is an asymptotically unbiased estimate of the Kullback-Leibler information where one-sided information is present. Like with AIC, OSAIC embodies the assumption that all candidate models are correct in the sense that the true process can be obtained by restricting parameters contained in the model equal to zero. It involves selecting the model for which

$$OSAIC = L(\theta^+ | y) - \sum_{s=0}^{p} w(p,s)(k - p + s)$$

is a maximum, where $L(\theta^+ | y)$ is the inequality constrained maximised log-likelihood, k is the number of parameters contained in the model, p is the number of inequality constrained parameters and $w(p,s)$ is a series of probability weights determined by the probability that s out of p inequality constraints will not be enforced in the model, assuming that the true process lies where all constraints hold. For a detailed discussion of these weights, see Gupta (1963), Kudô (1963) or Gouriéroux, Holly and Monfort (1982). The values of the weights are related to the asymptotic correlation between the unconstrained maximum likelihood estimates. As the information matrix for ARCH(q) models is diagonal, assuming that all ARCH parameters are zero (see Demos and Sentana (1998)), the asymptotic correlations are zero making the weights relatively easy to calculate. The weights and the corresponding penalties are given in Table 1.

TABLE 1

Weights and Penalties Associated with ARCH Models

MODEL						**OSAIC PENALTY**
ARCH(1)	$w(1,0)$ 0.5	$w(1,1)$ 0.5				0.5
ARCH(2)	$w(2,0)$ 0.25	$w(2,1)$ 0.5	$w(2,2)$ 0.25			1.0
ARCH(3)	$w(3,0)$ 0.125	$w(3,1)$ 0.375	$w(3,2)$ 0.375	$w(3,3)$ 0.125		1.5
ARCH(4)	$w(4,0)$ 0625	$w(4,1)$ 0.25	$w(4,2)$ 0.375	$w(4,3)$ 0.25	$w(4,4)$ 0625	2.0

OSAIC embodies the notion that the risk associated with the incorrect inclusion of parameters for which additional non-sample information is available, is less than if such information is unavailable, i.e. the additional penalty for non-negative parameters is less than that for unconstrained parameters. For this reason, we feel that OSAIC is a theoretically more appropriate criterion for the selection of q than is AIC or BIC. In the next section, we compare the small sample performance of OSAIC, AIC and BIC for the problem of ARCH lag order selection by Monte Carlo simulation.

2. The Monte Carlo Experiment

The Monte Carlo experiments involve generating 2000 data sets of size 60 and 200 from the following process

$$y_t = 0.2 + e_t$$
$$e_t \sim N(0, s_t^2)$$
$$s_t^2 = 0.1 + a_1 e_{i-1}^2 + a_2 e_{i-2}^2 + a_3 e_{i-3}^2 + a_4 e_{i-4}^2.$$

The values used for the a_i s are given by Table 2. We then select between the classical linear regression model (i.e. all ARCH parameters are zero) and ARCH(1) to ARCH(4), where all ARCH models are estimated by inequality constrained maximum likelihood.

TABLE 2

The Parameters Used in the Monte Carlo Experiments

Correct Model	**DGP Number**	a_1	a_2	a_3	a_4
ARCH(1)	**1**	0.2	0	0	0
ARCH(1)	**2**	0.5	0	0	0
ARCH(2)	**3**	0.2	0.12	0	0
ARCH(2)	**4**	0.5	0.4	0	0
ARCH(3)	**5**	0.2	0.15	0.1	0
ARCH(3)	**6**	0.5	0.2	0.15	0

The benefit of using data generating processes of various types is that we can assess the performance of the criteria across a wide variety of model selection situations. Due to fact that the same inequality constrained parameter estimates are used in the calculation of each criterion, and since

$$penalty(OSAIC) < penalty(AIC) < penalty(BIC)$$

for n > 7, OSAIC is more likely to select models of larger dimension than AIC and BIC is more likely to select models of smaller dimension than AIC. This means that where the smallest model in the choice set is the correct model, BIC will outperform the other criteria; conversely, where the largest model in the choice set is correct, OSAIC will perform best. For this reason we do not consider cases where the CLRM or the ARCH(4) model are correct.

3. Results and Discussion

The results of the Monte Carlo experiments are contained in Tables 3 and 4. We present the estimated probability of selecting each model, in each case the highlighted area shows the estimated probability of correct selection. We also include summary statistics where we calculate the average probability of correct selection, the average probability of underfitting (i.e. selecting a model of lesser dimension than the correct model) and the average probability of overfitting (i.e. selecting a model of larger dimension than the true model).

From the tables, using the probability of correct selection as a measure of performance, we can see that for $n = 60$ OSAIC outperforms both AIC and BIC in every considered case. For $n = 200$, AIC performs marginally better than OSAIC where the ARCH(1) model is correct, but OSAIC outperforms AIC for all DGPs larger than ARCH(1). Overall, the average probability of correct selection for OSAIC is higher than for AIC. Both OSAIC and AIC outperform BIC for both sample sizes and for every considered DGP.

TABLE 3

Estimated Probabilities of Selecting Various ARCH models, n = 60

DGP No.	1			2			3			4		
	OSAIC	AIC	BIC	OSAIC	AIC	BIC	OSAIC	AIC	BIC	OSAIC	AIC	BIC
CLRM	0.33	0.48	0.67	0.47	0.64	0.82	0.22	0.38	0.57	0.39	0.56	0.77
ARCH1	0.54	0.48	0.33	0.40	0.32	0.18	0.33	0.34	0.31	0.31	0.29	0.19
ARCH2	0.06	0.02	0.00	0.06	0.02	0.00	0.34	0.25	0.12	0.20	0.12	0.04
ARCH3	0.04	0.01	0.00	0.04	0.01	0.00	0.08	0.03	0.00	0.06	0.02	0.00
ARCH4	0.03	0.01	0.00	0.03	0.01	0.00	0.04	0.01	0.00	0.04	0.01	0.00
DGP No.	5			6						OVERALL AVES		
	OSAIC	AIC	BIC	OSAIC	AIC	BIC				OSAIC	AIC	BIC
CLRM	0.17	0.30	0.51	0.33	0.51	0.73				Correct Selection		
ARCH1	0.19	0.23	0.25	0.23	0.24	0.18				0.33	0.25	0.13
ARCH2	0.23	0.21	0.14	0.20	0.15	0.06				Underfitting		
ARCH3	0.34	0.23	0.09	0.18	0.08	0.02				0.57	0.72	0.87
ARCH4	0.07	0.03	0.00	0.05	0.01	0.00				Overfitting		
										0.10	0.03	0.00

TABLE 4

Estimated Probabilities of Selecting Various ARCH models, n = 200

DGP No.	1			2			3			4		
	OSAIC	AIC	BIC	OSAIC	AIC	BIC	OSAIC	AIC	BIC	OSAIC	AIC	BIC
CLRM	0.04	0.08	0.24	0.15	0.25	0.54	0.01	0.03	0.13	0.08	0.16	0.44
ARCH1	0.79	0.85	0.75	0.68	0.69	0.46	0.15	0.23	0.40	0.31	0.40	0.41
ARCH2	0.10	0.05	0.01	0.09	0.04	0.01	0.68	0.68	0.46	0.48	0.39	0.14
ARCH3	0.05	0.02	0.00	0.05	0.01	0.00	0.10	0.04	0.01	0.09	0.04	0.00
ARCH4	0.04	0.01	0.00	0.03	0.01	0.00	0.05	0.02	0.00	0.05	0.01	0.00
DGP No.	5			6						OVERALL AVES		
	OSAIC	AIC	BIC	OSAIC	AIC	BIC				OSAIC	AIC	BIC
CLRM	0.00	0.01	0.07	0.03	0.10	0.34				Correct Selection		
ARCH1	0.02	0.05	0.19	0.12	0.20	0.33				0.63	0.60	0.39
ARCH2	0.13	0.21	0.31	0.34	0.37	0.25				Underfitting		

ARCH3	0.73	0.68	0.43	0.42	0.29	0.08		0.23	0.35	0.61
ARCH4	0.11	0.05	0.01	0.09	0.03	0.00		**Overfitting**		
								0.14	0.06	0.01

AIC, and especially BIC, both have high average probabilities of underfitting the true process relative to that of OSAIC. The average probability of overfitting the true process is consequently higher for OSAIC relative to the other criteria. It could be argued that, whilst ideally we want the probability of correct selection to be as high as possible, if errors are made we prefer to overfit than underfit. The reason for this is that overfitting leads to an efficiency loss in inference whereas underfitting leads to omitted variable biases and invalidation of the distribution theory of inferential procedures. Since OSAIC has a higher average probability of correct selection and lower average probability of underfitting relative to the other criteria, the small sample properties of OSAIC relative to AIC and BIC are superior.

References

Akaike, H. (1973) Information Theory and an Extension of the Maximum Likelihood Principle. *Second International Symposium on Information Theory, B.N. Petrov and F. Csaki (Eds.)*. Akademiai Kiado, Budapest, p.267-281.

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

Engle R.F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica, 50, p.987-1007.

Gourieroux, C., Holly, A. and Monfort, A. (1982). Likelihood Ratio Test, Wald Test and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters. *Econometrica*, **50**, p.63-80.

Granger, C., King, M.L. and White, H. (1995). Testing Economics Theories and the Use of Model Selection Criteria. *Journal of Econometrics*, **67**, p.173-187.

Gupta, S.S. (1963). Probability Integrals of Multivariate Normal and Multivariate t. *Annals of Mathematical Statistics*, **34**, p.792-838.

Hughes, A.W. and King, M.L. (1994). One-sided Model Selection Procedures. *Paper presented at the Australasian Meeting of the Econometric Society, Armidale, NSW.*

Hughes, A.W. and King, M.L. (1998). Model Selection Using AIC in the Presence of One-sided Information. *Mimeo., Department of Econometrics and Business Statistics, Monash University.*

Kudo, A. (1963). A Multivariate Analogue of the One-sided Test. *Biometrika*, **50**, p.403-418.

Lee, J.H.H. and King, M.L. (1993). A Locally Most Mean Powerful Based Score Test for ARCH and GARCH Regression Disturbances. Journal of Business and Economic Statistics, 11, p. 17-27.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Mathematical Statistics*, **42**, p.1003-1009.