Australian Journal of Basic and Applied Sciences, 3(4): 3524-3530, 2009
ISSN 1991-8178

A Comparison of Akaike, Schwarz and R Square Criteria for Model Selection Using Some Fertility Models

[1]Ezra Gayawan and [2]Reuben A. Ipinyomi

[1]Department of Mathematical Sciences, Redeemer's University, Redemption City, Nigeria
[2]Department of Statistics, University of Ilorin, Ilorin, Nigeria

Abstract: This paper compares the relative performance of Akaike information criterion (AIC), Schwarz information criterion and the adjusted R^2 ($\overline{R^2}$) when they are applied to a crucial and difficult task of selecting model with best fit when real data are used. The three criteria are used to select model among the four fertility models applied to age specific fertility data from some African and European countries. Overall, models chosen by AIC and SIC differ in 5%; AIC and $\overline{R^2}$ in 33% while the SIC and $\overline{R^2}$ differ in 38%. It was discovered that SIC leads to models with lower parameters than AIC and $\overline{R^2}$ will do. Where different models were chosen, a comparison of the ability of the competing models to reproduce the empirical data revealed that models chosen by $\overline{R^2}$ were able to reproduce the data more but they are often the most complex.

Key words: Akaike information criterion; Schwarz information criterion; adjusted R^2; fertility data; maximum likelihood

INTRODUCTION

Studies involving modeling are often faced with two major problems; the problems of model identification and parameter estimation. The identification problem is usually the most difficult task. Once the functional form of the model has been identified, the problem of estimating the parameters can be easily solved. It is necessary to determine the ability of a model to reproduce the empirical data generating it and if it can be used to predict future values before deciding if it is the best for a given dataset.

The present study is concerned with the ability of a model to reproduce an empirical dataset. A model is chosen by studying the data values in a fit set. The model can be made to fit data increasingly well by increasing its number of parameters thereby making it more complex. Many of the parameters in a complex model may however, simply be accounting for random noise in the data and where the model is to be used for predictive purposes; it may predict future values quite poorly because of its noisy terms (Hibon, 1984). The problem then is that of deciding the model with appropriate number of parameters that best explains a series. This problem is usually encountered virtually everywhere. As Rao and Wu (2001) explained, in linear regression analysis, it is of interest to select the right number of nonzero regression parameters. In the analysis of time series, it is essential to know the true orders of an ARMA model. In problems of clustering, it is important to find out the number of clusters. In the signal detection, it is necessary to determine the number of true signals and the list continues. This problem is often solved using various model selection techniques.

Model selection can be based on hypothesis testing or through the use of information criteria. Hypothesis testing procedure is constructed based on a sequence of hypothesis tests. Assuming there are an order in the set of candidate models $\{M_i, i=1,2,...\}$ such that M_i is preferable to M_{i+1}, a sequence of hypothesis H_{io}: M_i *holds true* versus H_{ia}: M_{i+1} holds true, i = 1, 2, …, can be tested sequentially. Once H_{i0} is accepted, the test

Corresponding Author: Ezra Gayawan, Department of Mathematical Sciences, Redeemer's University, Redemption City, Nigeria.
E-mail: ezrascribe@yahoo.com

procedure stops and the model M_i is selected. A test procedure however do not penalize for over parameterization (Sclove, 1987). Information criteria on the other hand are selection criterions which balance model fit and its complexity. The Akaike information criterion (AIC) (Akaike, 1974) and Schwarz information criterion (SIC) (Schwarz, 1978) are two objective measures of a model's suitability which takes those considerations into account. They defer in terms of penalty attached to increasing the number of parameters in a model. Also, a commonly used technique in this regard is the adjusted coefficient of multiple determination ($\overline{R^2}$).

This paper compares the relative performances of AIC, SIC, and $\overline{R^2}$ by fitting some fertility models to age specific fertility data from African and European countries and then applying the three criteria to select models with best fit.

Models and the Selection Criteria:

The models considered in this study are selected from models developed and well discussed in literatures for capturing the pattern of age specific fertility rate (ASFR). The ASFR is the number of birth per thousand women in a given age or age group in a population at a particular period usually, a year. The distribution of ASFR has a typical shape common in all human populations through time. It begins with a minimum placed at the beginning of reproductive age interval and then rises until it attains a maximum somewhere in the 30's then it declines again to level off near age 50. Countries however, show differences with respect to the age where the fertility rate reaches a maximum and the variable speed with which the maximum is approached from the beginning and is the passed to reach the end of the fertility span (Peristera and Koistaki, 2007). Various models have been proposed to capture the patterns and they all have been proved to provide good fit for the population they are proposed to model. Among the available models, we have selected the following four for this study: the Hadwiger function (Hadwiger, 1940); the Gamma model (Hoem *et al*, 1981); the Beta model (Hoem *et al*, 1981) and Model 2 (Peristera and Koistaki, 2007). The models have three, four, five and six parameters respectively and hence, their choice for this study.

In a sequence, we review the models mentioned. The Hadwiger function by Hadwiger (1940) is expressed as

$$f(x) = \frac{ab}{c}\left(\frac{c}{x}\right)^{\frac{3}{2}} exp\left\{-b^2\left(\frac{c}{x}+\frac{x}{c}-2\right)\right\}$$

where x is the age of the mother at birth, $a,$ $b,$ and $c,$ the parameters to estimate are demographically interpreted as $a,$ the total fertility rate. c is the mean age of motherhood, the parameter b determines the height of the curve, while the term $\frac{ab}{c}$ is related to the maximum age-specific fertility rate (or modal age-specific fertility rate).

The Gamma function by Hoem *et al* (1981) is given by

$$f(x) = R\frac{1}{\Gamma(b)c^b}(x-d)^{b-1}exp\left\{-\left(\frac{x-d}{c}\right)\right\} \qquad x > d$$

where, d represents the lower age at childbearing, while the parameter R determines the level of fertility. Hoem *et al* (1981) substituted the parameters b and c by the mode $m,$ the mean μ and the variance σ^2 of the density function, as they have no direct demographic interpretation, thus

$$c = \mu - m \qquad \text{and} \qquad b = \frac{\mu - d}{c} = \frac{\sigma^2}{c^2}$$

The Beta function also by Hoem *et al* (1981) is given by the formula

$$f(x) = R\frac{\Gamma(A+B)}{\Gamma(A)\Gamma(B)}(\beta-\alpha)^{-(A+B-1)}(x-\alpha)^{A-1}(\beta-x)^{B-1} \qquad \alpha < x < \beta$$

The parameters are related to the mean μ and the variance σ^2 through the relations

$$B = \left\{\frac{(\mu-\alpha)(\beta-\mu)}{\sigma^2} - 1\right\}\left(\frac{\beta-\mu}{\beta-\alpha}\right) \quad \text{and} \quad A = B\left(\frac{\mu-\alpha}{\beta-\mu}\right)$$

The parameters α and β are interpreted as lower and upper age limit of fertility. The parameter *R* determines the level of fertility.

In order to capture the recent distorted fertility curves characterized by excess early-age fertility noticed in the fertility pattern of countries like Ireland, UK and the US, Peristera and Kostaki (2007) proposed the following two terms model and referred to it as Model 2:

$$f(x) = c_1 exp\left\{-\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right\} + c_2 exp\left\{-\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right\}$$

where $f(x)$ is the ASFR at age x of the mother, c_1, c_2 are the total fertility rates of the first and second hump respectively, μ_1, μ_2 are the mean ages of the two subpopulations, the one with earlier fertility and the other with fertility at later ages, while σ_1, σ_2 reflect the variances of the two humps.

In deciding on a model that fits the ASFR curve well among a number of models, demographers often plot the empirical and estimated fertility data and then eyeball the curves to decide on the one that best captures the pattern. Alternatively, the model with the least value of the minimization criterion usually, the sum of square residual (SSR) used in fitting the models is considered to be the best without considering the complexity of the models caused by additional parameters. Various objective model selection criteria have however been proposed in the literatures. The major and most widely used criteria have been surveyed and their theoretical and practical features extensively discussed by de Gooijer *et al.* (1985). Among them are the AIC (Akaike, 1974), SIC (Schwarz, 1978) and $\overline{R^2}$.

The AIC, SIC and the $\overline{R^2}$ are criteria that involve adjusting the SSR by one factor or another usually, the sample size (n) and the number of parameters (p) in the model, to create an index of fit of an equation. These criteria can be used to decide if the improved fit caused by an additional variable is worth the decreased degrees of freedom and increased complexity caused by the addition. Hossian (2002) described the AIC and SIC among the information criteria that are most viable and useful as they are based on choosing the model that maximized the log of the likelihood function minus a penalty term.
The AIC and SIC are derived from the following expression

$$IC(p) = nln\hat{\sigma}^2(p) + p[f(n)]$$

Where $\hat{\sigma}^2$ is the estimated variance which depends upon n and p. $p[f(n)]$ can be interpreted as the penalty function for increasing the parameters in the model. Different choice of $f(n)$ give different information criteria. The AIC results from setting $f(n)$=2, and will therefore be equal to

$$AIC = nln\hat{\sigma}^2 + 2p$$

Koehler and Murphree (1988) considered the AIC as being derived by considering the principles of maximum likelihood and of negative entropy and that the criterion estimates twice the negative entropy and is therefore asymptotically unbiased as n increases. Thus, the model having minimum AIC should have minimum prediction set error, at least asymptotically. The procedure has been used to identify models by Tong (1977), Ozaki (1977) and Larimore and Mehra (1985) among others.

The AIC procedure has however been criticized because it is inconsistent and tends to over fit models. A criterion is said to be order consistent if, as the sample size increases, the criterion is minimized at the true order with a probability which approaches unity. Geweke and Meese (1981) showed this for regression models, Shibata (1976) for autoregressive models and Hannan (1982) for ARMA model.

The SIC results from assuming $f(n) = In(n)$, and will therefore be equal to

$$SIC = nln\hat{\sigma}^2 + pln(n)$$

Schwart (1978) derived the SIC as the Bayes solution to the problem of model identification and is therefore referred to as the Bayes Information criterion (BIC). Romero (2007) affirmed that SIC deals with the problem of inconsistency noticed in AIC. Asymptotically, the SIC is minimized at the model order having the highest posterior probability. Akaike (1977) had shown that the SIC can be more successful than the AIC in estimating the degree of a polynomial regression model and in estimating the order of an autoregressive model.

$\overline{R^2}$ is derived by adjusting the coefficient of multiple determination (R^2). The additions of explanatory variables to a model often increase the value of R^2 even when the additional variables have no explanatory power. Montgomery and Morrison (1973) showed that R^2 is a (positively) biased estimator of the true coefficient of determination. Additionally, the coefficient does not penalize the likelihood function for having additional variables. Hence, adjusting R^2 for n and p yields $\overline{R^2}$. Barton (1962) showed that although $\overline{R^2}$ is also positively biased, it is consistent since it converges to the true coefficient of multiple determination as the sample size increases. Since $\overline{R^2}$ is consistent as n increases, and that it takes account of the number of parameters in a model, it is useful for comparing the fit of specifications that differ in the addition of more parameters (Johnston and Dinardo, 1997). A model with a higher R^2 would be preferred if and only if the $\overline{R^2}$ is higher too.

The $\overline{R^2}$ statistics necessary for model selection, is defined as

$$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k}$$

Romero (2007) related $\overline{R^2}$ to the maximum likelihood estimator of σ^2.

Methodology:

The merits of the criteria have been extensively debated mainly in the context of their asymptotic properties by Shibata (1976), Gweeke and Meese (1981) and Hannan (1980, 1982) or their model choices in experiments where the data are simulated and the generating process is known (see Akaike (1977), Bhansali and Downham (1977) and Sneek (1984)). In this study, however, the data are real and the true generating process is unknown. Although Sneek (1984) and Koehler and Murphree (1988) also compared the criteria on real data, their applications were on time series data and their studies were limited to AIC and SIC.

The four models displayed in Section 2 were fitted to a series of ASFR data for some African and European countries. Fertility data of 15 African countries and those of the urban and rural dwellers separated for each country were used for the African case. In the case of the European countries, dataset for 12 countries were used. For each country, data were obtained for 3 successive years, 2004, 2005 and 2006 for the study. Dataset for the African countries were directly obtained from published reports of Demographic and Health Survey (DHS) which are available online from ORC Macro's website [http://www.measuredhs.com]. These data are based on a five-year age group from age 15 to 49, hence, there are seven point data for each case.

The empirical data for the European countries were obtained from Eurostat New Cronos database accessibly online at [http://www.esds.ac.uk/international/support/user_guides/eurostat/Cronoswe.asp]. Unlike the African case, European fertility data are based on a single-year age of the mother also from age 15 to 49. Therefore, there are 35 data points for each case.

Table 1: Summary of dataset and differences in model selection

Category of Data	No. of Dataset	AIC and SIC		AIC and $\overline{R^2}$		SIC and $\overline{R^2}$	
		No	Percentage	No	Percentage	No	Percentage
All data	81	4	5	27	33	31	38
All Africa	45	0	0	24	53	24	53
All European	36	4	11	3	8	7	19
African	15	0	0	5	33	5	33
African Urban	15	0	0	10	67	10	67
Africa Rural	15	0	0	9	60	9	60
European 2004	12	2	17	0	0	2	17
European 2005	12	1	8	2	17	3	25
European 2006	12	1	8	1	8	2	17

In order to fit the models, it is generally accepted that the most efficient procedure is to use weighted least squares with weights equal to the reciprocals of the variances of the empirical fertility rates. However, Hoem (1976) and Hoem at al (1981) pointed out that a weighted estimating procedure would give too much attention to the low fertility ages in the tails and especially to the higher ones in the upper tail, while giving too little attention to the high fertility ages in the middle, and thus is not desirable. Therefore, for fitting the models, a nonlinear unweighted least-square procedure was adopted. This was accomplished by minimizing the sum of square residual (SSR) by means of the inbuilt *solver* in MS Excel. The SSR is given by

$$\sum_x (f_x - \hat{f}_x)^2$$

where f_x is the empirical fertility rate at age x and $\hat{f}_x$ is the estimate of f_x.

Having fitted the models, the fitted errors for all models were computed. Where the criteria choose different models for a dataset, the mean absolute percentage error (MAPE) for the competing models were computed. This allows us to judge the performance of the competing models in terms of their ability to reproduce the empirical data. The MAPE is given as

$$MAPE = \left[\frac{1}{n}\sum_{i=1}^{n}(|f_x - \hat{f}_x|)\right] \times 100$$

Results:

One of the significant findings of this study is that the AIC and SIC chose same models in almost all the fertility data used. Much difference occur when they are compared with the $\overline{R^2}$ Table 1 presents the number and percentage of times for which the criteria in pairs, chose different models. Overall, the models chosen by AIC and SIC differ in only 5%, AIC and $\overline{R^2}$ differ in 33% while SIC and $\overline{R^2}$ differ in 38%. The AIC and SIC consistently chose same model in all the 45 series of the African data. In contrast, for same dataset, each of the two criteria differ with the $\overline{R^2}$ in 24 ways given 53%. Unlike in the African dataset, the AIC and SIC differ in 11%, AIC and $\overline{R^2}$ in 8% while the SIC and $\overline{R^2}$ differ in 19% in the European data set considered.

Table 2 lists all the African countries for which the criteria chose different models. Those for the European countries are presented separately in Table 3 because in the African case, AIC and SIC consistently chose same models for all the cases. Also presented in the Tables are the values of the SSR at the exist of the iteration procedure in estimating the parameters, the number of parameters of the selected models and the values of the

Table 2: Results of models selected by the three criteria for data from African countries

Data Source	AIC & SIC				$\overline{R^2}$			
	Model	K	$SSR.10^6$	MPAE	Model	K	$SSR.10^6$	MPAE
Cameroon	Gamma	4	451.75	0.68	Beta	5	234.38	0.52
Cameroon Rural	Hadwiger	3	1875.2	1.49	Beta	5	891.45	0.93
Burkina Faso	Hadwiger	3	1124.3	1.08	Beta	5	319.29	0.56
Burkina Faso Urban	Hadwiger	3	831.1	0.99	Beat	5	316.03	0.61
Burkina Faso Rural	Hadwiger	3	1299.0	1.19	Beta	5	384.51	0.61
Chad Urban	Hadwiger	3	3043.34	1.84	Beta	5	384.92	0.97
Egypt Urban	Gamma	4	90.72	0.29	Beta	5	54.83	0.23
Egypt Rural	Gamma	4	39.75	0.18	Beta	5	20.77	0.14
Ethiopia Rural	Hadwiger	3	839.34	0.84	Beta	5	411.09	0.67
Ghana	Hadwiger	3	302.56	0.52	Beta	5	136.12	0.31
Ghana Urban	Hadwiger	3	248.12	0.53	Beta	5	101.96	0.35
Ghana Rural	Hadwiger	3	671.46	0.73	Beta	5	337.47	0.55
Kenya Rural	Hadwiger	3	1360.44	1.27	Beta	5	566.94	0.74
Morocco Urban	Gamma	4	240.69	0.51	Beta	5	134.85	0.37
Namibia	Hadwiger	3	637.75	0.73	Beta	5	235.74	0.46
Namibia Urban	Hadwiger	3	1302.33	1.00	Beta	5	558.32	0.80
Niger Urban	Hadwiger	3	2231.01	1.60	Beta	5	989.08	1.06
Nigeria Urban	Gamma	4	499.05	0.69	Beta	5	266.01	0.55
Rwanda Urban	Hadwiger	3	879.21	0.97	Gamma	4	813.89	0.87
Rwanda Rural	Hadwiger	3	2759.01	1.64	Model 2	6	439.0	0.65
Senegal	Gamma	4	517.96	0.80	Beta	5	265.8	0.57
South Africa Urban	Gamma	4	138.85	0.39	Beta	5	86.91	0.30
South Africa Rural	Hadwiger	3	839.33	0.84	Beta	5	411.09	0.67
Zambia Rural	Hadwiger	3	2892.54	1.63	Beta	5	972.17	0.99

MPAE. It can be observed from Tables 2 and 3 that while the $\overline{R^2}$ procedure consistently chose models with minimum SSR but higher parameters, AIC and SIC chose models with lower number of parameters. Hence, the MPAE which evaluates the ability of the chosen model to reproduce the empirical data is consistently lower for the models chosen by $\overline{R^2}$ than those chosen by the other two criteria. Results in Table 3 show that AIC and $\overline{R^2}$ chose same models for the European cases except in two places Bulgaria 2006 and Denmark 2005. In those two places, AIC and SIC chose the same models. Like before, the models chosen by $\overline{R^2}$ have its SSR values lower than any other model and hence, they are more able to reproduce the empirical data as indicated by the MPAE values even though they are more complex than those chosen by the other criteria.

Table 3: Results of models selected by the three criteria for data from European countries

Data Source	AIC				SIC				$\overline{R^2}$			
	Model	K	$SSR.10^6$	MPAE	Model	K	$SSR.10^6$	MPAE	Model	K	$SSR.10^6$	MPAE
Australia 2004	Model 2	5	162.02	0.20	Gamma	4	206.71	0.20	Model 2	5	162.02	0.16
Bulgaria 2004	Model 2	5	491.59	0.17	Gamma	4	630.01	0.25	Model 2	5	491.59	0.17
Bulgaria 2005	Beta	4	597.44	0.21	Beta	5	597.44	0.21	Model 2	5	558.03	0.18
Bulgaria 2006	Model 2	5	624.07	0.16	Beta	5	713.18	0.23	Model 2	5	624.07	0.16
Denmark 2005	Beta	4	239.44	0.24	Beta	5	239.44	0.24	Model 2	5	234.42	0.21
Norway 2005	Model 2	5	158.83	0.17	Beta	5	179.32	0.17	Model 2	5	158.83	0.15
Norway 2006	Beta	4	131.00	0.13	Beta	5	131.00	0.13	Model 2	5	125.94	0.13

Also, a close examination of the models chosen by the AIC and SIC in the European dataset with larger sample size, indicate that those chosen (where there are differences as presented in Table 3) by AIC have higher number of parameters than those chosen by SIC. This imply that SIC has more ability to penalize the inclusion of additional parameters than does AIC and $\overline{R^2}$ and hence, AIC and $\overline{R^2}$ will tend to over fit models than SIC.

Conclusions:

This study has shown using empirical data that $\overline{R^2}$ will consistently choose complex models than AIC and SIC. By this, it will lead to over parameterization than the other two selection procedures. Comparing the performances of AIC with SIC, it was observed that the former chose complex models than the latter. This supports the results of Sneek (1984), Neftci (1982) and Anne and Murphree (1988). Furthermore, this study reveals that models chosen by $\overline{R^2}$ will be able to reproduce the empirical dataset than those chosen by AIC and SIC. It is however important to stress at this point that the findings of this study and in combination of the principle of simplicity argue in favour of using SIC rather than AIC and $\overline{R^2}$ for model selection.

REFERENCES

Akaike, H., 1974. A new look at the statistical identification model. IEEE Trans. Auto Control, 19: 716-723.

Akaike, H., 1977. Canonical correlation analysis of time series and the use of the information criterion. *In system Identification: Advances and Case studies* (eds R. K. Mehra and D, G. Laintiotis), 27-96. New York: Academic Press.

Barten, A.P., 1962. Note on unbiased estimation of the squares multiple correlation coefficient. Statistic Neerlandica, 16(2): 151-163.

Bhansali, R.J. and D.Y. Downham, 1977. Some properties of the order of an autogressive model selected by generalization of Akaike's FRE criterion. Biometrika, 64: 547-551.

de Gooijer, J.G., B. Abraham, A. Gould and L. Robinson, 1985. Methods for determining the order of an autoregressive-moving average process: a survey. Int. Statist. Rev., 53: 301-329.

Gweke, J.F. and R.A. Meese, 1981. Estimating regression models of finite but unknown order. Int. Econom Rev., 22: 55-70.

Hadwiger, H., 1940. Eine analytische Reprodutions-funktion fur biologische Gesamtheiten. Skandinavisk Aktuarietidskrift, 23: 101-113.

Hannan, E.J., 1982. Testing for autocorrelation and Akaike's criterion. *In Essays in Statistical Science* (eds J. Gani and E. J. Hannan) 403-412. Sheffield: Applied Probability Trust.

Hibon, M., 1984. Naïve, moving average, exponential smoothing and regression methods. In the Accuracy of Major Time Series Methods, Chichester: Wiley.

Hoem, J.M., D. Madsen, J.L. Nielsen, E. Ohlsen, H.O. Hansen, B. Rennermalm, 1981. Experiments in modeling recent Danish fertility curves. Demography, 18: 231-244.

Hossian, M.Z., 2002. Modified Akaike information criterion for statistical model selection. Pak. J. Statist., 18(3): 383-393.

Johnston, J. and J. Dinardo, 1997. Econometric Methods. McGraw-Hill/Irwin.

Koehler, A.B. and E.S. Murphree, 1988. A Comparison of the Akaike and Schwarz criteria for selecting model order. Appl. Statist., 37(2): 187-195.

Larimore, W.E. and R.K. Mehra, 1985. The problem of overfitting data. Byte., 10: 167-180.

Montgomery, D. and D. Morrison, 1973. A note on adjusting R^2. Journal of Finance, 28: 1009-1013.

Neftci, S.N., 1982. Specification of economic time series models using Akaike's criterion. J. Amer. Statist. Ass., 77: 537-540.

Ozaki, T., 1977. On the order determination of ARIMA models. Applied Statistics, 26: 290-301.

Peristera, P. and A. Kostaki, 2007. Modeling fertility in modern populations. Demographic Research, 16(6): 141- 194.

Rao, C.R. and Y. Wu, 2001. Institute of Mathematical Statistics, Lecture note-monograph series, 38: 1-64.

Romero, A.A., 2007. A note on the use of $\overline{R^2}$ in model selection. Working paper No. 62; Dept. of Economics, College of William and Mary.

Schwarz, G., 1978. Estimating the dimension of a model. Ann Statist., 6: 461-464.

Sclove, S.T., 1987. Application of model selection criteria to some problems in multivariate analysis. Psychometrika, 52(3): 333-343.

Shibata, R., 1976. Selection of the order of an autoregressive model by Akaike's information criterion. Biometrika, 63: 117-126.

Sneek, J.M., 1984. *Modelling Procedures for Univariate Time Series.* Amsterdam: Free University Press.

Tong, H., 1977. Some comments on the Canadian lynx data. Journal of the Royal Statistical Society A., 140: 432-436.