ISSN 1032-3813
ISBN 0 7326 0776 0

MONASH UNIVERSITY

AUSTRALIA

A SMALL SAMPLE VARIABLE SELECTION PROCEDURE

Catherine Scipione Forbes
Maxwell L. King
Alan Morgan

Working Paper 15/95
September 1995

DEPARTMENT OF ECONOMETRICS

A Small Sample Variable Selection Procedure

Catherine Scipione Forbes, Maxwell L. King
and Alan Morgan *
Department of Econometrics, Monash University
Clayton, Victoria 3168 Australia

Abstract

This paper presents an information criteria based model selection procedure (called FIC) for choosing the variables to be used in a linear regression. The penalty function is based on sums of critical values from particular F-distributions which are related to the small sample probabilities of incorrectly including additional regressors. Results from a Monte Carlo simulation study demonstrate that the performance of this new procedure is competitive with other asymptotically motivated procedures, while providing the practitioner with controls over the desired small sample probabilities of correct selection. An alternative, somewhat simpler selection criterion based on an asymptotic distribution is presented and compared to the finite sample criterion. Conditions for strong consistency of this variable selection procedure based on an approximate penalty function are presented.

Keywords: Information criteria; F-statistic; penalty functions; critical values.

1 Introduction

Often in econometrics a set of regressor variables is available and the practitioner must decide which subset of regressors (and possibly lagged regressors) to include in a linear

*This research was supported by an Australian Research Council grant. A previous version of this paper was presented at the 1995 Econometrics Conference at Monash. Special thanks to Don Poskitt for early comments and references.

regression model based on how well the each of the models appears to fit the observed data. In practice, several different techniques for choosing a set of regressor variables are available, ranging from stepwise procedures to penalized likelihood methods. The choice of the set of variables will depend crucially on the particular method of selection used. Consequently variable selection procedures have an important role to play in econometric modeling.

Variable selection in linear regression is a special case of the more general model selection problem. For a recent review of model selection procedures, see Fox (1995). The concensus in the model selection literature seems to favor information criteria (IC) procedures in which the maximized log-likelihood values, each adjusted by subtracting a penalty term related to the number of unknown parameters in each model, are compared and the model with the largest adjusted log-likelihood value is selected. The question is what the form of the penalty function should be. Many IC penalty functions have been suggested in the literature, with AIC (Akaike 1973) and BIC (Schwartz 1978) being two of the most widely used IC model selection procedures in practice. Discussion of the asymptotic properties of these procedures has been one of the main directions of this literature due partly to the fact that almost all of the IC procedures have been derived using asymptotic justifications. See in particular Hannan and Quinn (1979), Geweke and Meese (1981) and Sawa (1978). However, it is important that a model selection procedure have good small sample properties as well. Improvements on existing procedures have been proposed, see for example Sugiura (1978) and Hurvich and Tsai (1989, 1991). These authors propose a corrected AIC procedure, denoted AIC_c, which is designed to improve the small sample bias of AIC. In the hypothesis testing literature, the probabilities of wrongly choosing the null and alternative hypotheses are controlled and optimized, respectively, to achieve desired small sample properties. King, Forbes and Morgan (1995) considered controlling probabilities of correct selection using Monte Carlo techniques in the general model selection problem. Our aim in this paper is to develop a variable selection procedure for the multiple linear regression problem that has a good small sample justification and one that utilizes our wealth of experience with controlling probabilities as in hypothesis testing.

For our problem, we assume for theoretical reasons that given r possible regressors, not necessarily all 2^r models need be considered. For example, the practitioner may not wish to have higher order lagged variables included in a chosen model without the presence of

Table 1: Penalty functions p_j for model M_j

Akaike's AIC	$k_j + 1$
Schwartz's BIC	$\frac{(k_j+1)}{2} \ln n$
Hannan & Quinn's HQ	$(k_j + 1) \ln \ln n$
Theil's $\bar{R}^2$	$-\frac{n}{2} \ln (n - k_j)$

lower order lags of the same variable. For this reason the traditional stepwise procedure will not be an appropriate model selection approach. To define our variable selection procedure to choose between the m possible linear regression models, we follow an IC approach. This approach is favored because, as Pötscher (1991) noted, minimizing such an IC amounts to testing each model against all other models by means of a standard likelihood ratio test, and selecting that model which is accepted against all other models. The critical values of the tests are determined by the penalty function values. Let L_j denote the maximized log-likelihood function and p_j the penalty associated with model M_j. Then the chosen set of regressors is that set associated with model M_j such that

$$L_j - p_j > L_i - p_i \qquad \text{for all } i \neq j. \tag{1}$$

The choice of the penalty function is the primary concern of this paper. As the penalties are related to the critical values of a likelihood ratio test, our approach is to define them by controlling probabilities of incorrect selection, as is done in classical hypothesis testing. Some other suggestions from the literature for the form of the penalty function are given in Table 1, where k_j represents the number of regressors in model M_j and n is the sample size. Note that we are following Fox (1995) in representing Theil's $\bar{R}^2$ as an IC.

The plan of the paper is as follows. In Section 2 we consider the simple nested regression model selection problem and present our underlying ideas. Section 3 generalizes this notion to the variable selection problem and compares our finite sample penalty term to an approximate penalty. We also present a justification of this approximate penalty function based on quasi maximum likelihood (QML) arguments and prove conditions for strong consistency of this approximate procedure. Section 4 presents a comparison of our asymptotically derived procedure and some existing IC selection procedures by inverting the argument previously used to generate the penalty function. Section 5 presents the results of a Monte Carlo simulation study of choosing regressors in three different regression designs when the model is both correctly and incorrectly specified. Some concluding remarks are made in Section 6.

2 A simple nested regression setting

We consider first the simpler problem of choosing between m nested regression models $M_1 \subset M_2 \subset \ldots \subset M_m$. Here we assume that model M_j is of the form

$$Y = x_1\beta_1 + x_2\beta_2 + \ldots x_j\beta_j + U \tag{2}$$

where Y is an $n \times 1$ vector of observations on the dependent variable, x_i is an $n \times 1$ vector of regressor variables and β_i is a scalar regression coefficient associated with x_i. Thus model M_j contains $k_j = j$ unknown regression coefficients. In addition we assume that U is an $n \times 1$ vector of independent and identically distributed normal random variables with mean zero and unknown variance σ^2. The maximized log-likelihood function for model M_j is

$$L_j = -\frac{n}{2}\left[\ln\frac{\hat{u}_j'\hat{u}_j}{n} + \ln 2\pi + 1\right] \tag{3}$$

where $\hat{u}_j$ is the ordinary least squares residual vector under the assumption of model M_j.

In the case where models M_{j-1} and M_j alone are under consideration, the IC selection procedure would choose model M_j over model M_{j-1} if

$$L_j - p_j > L_{j-1} - p_{j-1}, \tag{4}$$

or equivalently, if

$$\frac{(\hat{u}_{j-1}'\hat{u}_{j-1} - \hat{u}_j'\hat{u}_j)}{(\hat{u}_j'\hat{u}_j)/(n-k_j)} > (n-k_j)\left[\exp\frac{2}{n}(p_j - p_{j-1}) - 1\right]. \tag{5}$$

However, we know from standard econometric theory (see, for example, Greene, 1990, p. 214) that the left hand side of (5) has an $F_{1,n-k_j}$ distribution when model M_{j-1} is 'true'. Thus, we can choose to define the relative penalty $p_j - p_{j-1}$ by setting the probability of incorrectly selecting model M_j when the lower dimensional model M_{j-1} is 'true' to be α_j. That is, we find $p_j - p_{j-1}$ such that

$$P_{M_{j-1}}(L_j - p_j > L_{j-1} - p_{j-1}) = \alpha_j. \tag{6}$$

Let f_{α,ν_1,ν_2} denote the upper α^{th} percentile of the F distribution with ν_1 numerator and ν_2 denominator degrees of freedom. Solving (5) for the relative penalty based on a fixed α_j yields

$$p_j - p_{j-1} = \frac{n}{2} \ln \left[\frac{1}{(n-k_j)} f_{\alpha_j,1,(n-k_j)} + 1 \right]. \quad (7)$$

Because our model selection procedure only depends on the differences in p_j values, not their actual values, we can set $p_1 = 0$. Equation (7) then provides a recursion formula for the remaining $m-1$ penalties $p_2, \ldots, p_m$ for the nested regression model selection problem. Notice that the values of α_j need not all be equal so that, for example, the probability of choosing M_2 over M_1 need not be the same as the probability of choosing M_3 over M_2.

2.1 A nested regression example

To demonstrate construction of the penalty function, we present a simple nested regression example. We consider choosing between the following five nested models

$$M_0 : Y = U \quad (8)$$
$$M_1 : Y = x_1\beta_1 + U \quad (9)$$
$$M_2 : Y = x_1\beta_1 + x_2\beta_2 + U \quad (10)$$
$$M_3 : Y = x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + U \quad (11)$$
$$M_4 : Y = x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + x_4\beta_4 + U \quad (12)$$

where all of the x_i are $n \times 1$ vectors and $n = 30$. Table 2 shows the resulting penalties and the corresponding AIC, BIC, HQ and $\bar{R}^2$ penalties, each adjusted so that the first penalty is zero. The fourth column corresponds to all $\alpha_j = .10$ for $j = 0, \ldots, 4$, and the seventh column corresponds to $\alpha_1 = \alpha_2 = .10$ and $\alpha_3 = \alpha_4 = .05$.

3 The general variable selection problem

In this section we consider the case where there are r regressors and any (although not necessarily all) of the $m = 2^r$ models corresponding to each combination of regressors may be under consideration. We seek to generalize the previous method for defining penalties

Table 2: Nested regression relative penalties ($n = 30$)

k_j	α_j	$p_j - p_{j-1}$	p_j	α_j	$p_j - p_{j-1}$	p_j	AIC	BIC	HQ	$\bar{R}^2$
0	-	-	0	-	-	0	0	0	0	0
1	.10	1.42	1.42	.10	1.42	1.42	1	1.70	1.22	0.51
2	.10	1.48	2.90	.10	1.48	2.90	2	3.40	2.44	1.03
3	.10	1.53	4.43	.05	2.17	5.07	3	5.10	3.67	1.58
4	.10	1.59	6.02	.05	2.26	7.33	4	6.80	4.90	2.15

in the nested regression case to this more complicated one. To do this, we first consider the case where $r = 2$ so that we are comparing the $m = 2^2$ models

$$\begin{aligned} M_1 : Y &= U && (13)\\ M_2 : Y &= x_1\beta_1 + U && (14)\\ M_3 : Y &= x_2\beta_2 + U && (15)\\ M_4 : Y &= x_1\beta_1 + x_2\beta_2 + U. && (16) \end{aligned}$$

In this case, each of x_1 and x_2 are $n \times 1$ vectors and β_1 and β_2 are scalars. We are interested in determining penalties p_i, for $i = 1, \ldots, 4$, by generalizing our results from the nested regression setting. To do this, we first notice that no matter whether we are comparing models M_1 with M_2 or M_1 with M_3, the relative penalties have the same form, namely

$$p_2 - p_1 = \frac{n}{2} \ln \left[\frac{1}{(n-1)} f_{\alpha_{12},1,(n-1)} + 1 \right] \tag{17}$$

and

$$p_3 - p_1 = \frac{n}{2} \ln \left[\frac{1}{(n-1)} f_{\alpha_{13},1,(n-1)} + 1 \right], \tag{18}$$

so that if $\alpha_{12} = \alpha_{13}$, where

$$\alpha_{ij} = P_{M_i}(L_j - p_j > L_i - p_i), \tag{19}$$

then $p_2 = p_3$. Similarly, regardless as to whether we are comparing M_2 with M_4 or M_3 with M_4, the relative penalties

$$p_4 - p_2 = \frac{n}{2} \ln \left[\frac{1}{(n-2)} f_{\alpha_{24},1,(n-2)} + 1 \right] \tag{20}$$

and

$$p_4 - p_3 = \frac{n}{2} \ln \left[\frac{1}{(n-2)} f_{\alpha_{34},1,(n-2)} + 1 \right], \tag{21}$$

are again the same if $\alpha_{34} = \alpha_{24}$. Thus, if $n = 30$ and $\alpha_{ij} = \alpha = .10$, for all (i,j) pairs $(1,2),(1,3),(2,4)$ and $(3,4)$, the resulting penalties for each model are $p_1 = 0$, $p_2 = p_3 = 1.42$, and $p_4 = 2.90$. Generalizing this to larger values of r, we see that the penalty for model M_j with k_j regressors is

$$p_j = \frac{n}{2} \sum_{i=1}^{k_j} \ln \left[\frac{1}{(n-i)} f_{\alpha_i,1,(n-i)} + 1 \right]. \tag{22}$$

This is the penalty function we recommend for all regressor selection problems. We call the IC procedure using the penalty function in (22) FIC as its derivation comes from F distributions. A number of comments are in order.

First, the penalties require a decision by the user on the values of α_i. These values can be interpreted as the probability of wrongly choosing a model that is larger in dimension by one regressor when the true model has $i-1$ regressors and only these two regressions are under consideration. These values might be set to the same value for $i = 1, \ldots, k_j$, or could be allowed to change with the number of regressors.

Secondly, it is clear that our proposed penalty assumes a natural build-up of the model one dimension at a time. Of course there can easily be problems in which dimensions change by more than one unit at a time because variables are considered in blocks. For example, suppose that M_{j-1} has 3 regressors less than model M_j which has k_j regressors. Then one could consider choosing between penalties constructed such that

$$p_j - p_{j-1} = \frac{n}{2} \ln \left[\frac{1}{n-k_j} f_{\alpha_j,3,(n-k_j)} + 1 \right] \tag{23}$$

or

$$p_j - p_{j-1} = \frac{n}{2} \sum_{i=k_{j-1}+1}^{k_j} \ln \left[\frac{1}{n-i} f_{\alpha_i,1,(n-i)} + 1 \right]. \tag{24}$$

Our approach is based on using (24) rather than (23). One of the advantages in using this penalty function is that if suddenly there is a change in the set of models under consideration, then all existing calculations do not need to be redone. All that is required are the maximized log-likelihood values for the new models plus their associated penalties. In the above example, use of our penalty function would not require any recalculations if

Table 3: $\text{FIC}_{\alpha=.10}$ penalty functions

sample size n	number of regressors k_j in model 1	2	3	4	5	6	7	8	9	10
20	1.46	3.01	4.64	6.39	8.25	10.25	12.41	14.76	17.33	20.17
40	1.41	2.85	4.33	5.85	7.42	9.04	10.70	12.42	14.19	16.03
60	1.39	2.80	4.24	5.70	7.19	8.70	10.25	11.83	13.43	15.07
80	1.38	2.77	4.19	5.62	7.08	8.55	10.04	11.55	13.09	14.65
100	1.37	2.76	4.16	5.58	7.01	8.46	9.92	11.40	12.89	14.40
120	1.37	2.75	4.14	5.55	6.97	8.40	9.84	11.30	12.77	14.25
140	1.37	2.74	4.13	5.53	6.94	8.36	9.79	11.23	12.68	14.14
160	1.37	2.74	4.12	5.51	6.92	8.33	9.74	11.17	12.61	14.06
180	1.36	2.76	4.12	5.50	6.90	8.30	9.71	11.13	12.56	14.00
200	1.36	2.73	4.11	5.49	6.88	8.28	9.69	11.10	12.52	13.95
∞	1.35	2.71	4.06	5.41	6.76	8.12	9.47	10.82	12.17	13.53

an intermediate model was introduced into the set of models under consideration. This may not be the case using penalties derived using (23) where in fact several penalties may need to be adjusted due to the introduction of an intermediate model. In addition, the penalties defined using (23) need not be uniquely defined. However, as a consequence of using (24), models with the same number of parameters will have the same, uniquely defined penalty function.

Tables 3 and 4 display the penalty function for the variable selection problem for varying sample sizes and number of regressors. As can easily be seen, if many regressors are in the model and the sample size is small, the penalty associated with the model is high, so that the value of the maximized log-likelihood function would necessarily need to be very high relative to models with a smaller number of regressors in order for that model to be chosen. However, as the sample size increases, the relative penalties get smaller, so that the model with the higher maximized log-likelihood is more likely to be selected.

It should be noted that others have recognized the potential of controlling probabilities of incorrect selection in the likelihood ratio tests. In particular, Hosoya (1984) defines an alternative procedure (also called FIC) for chosing a model based on setting probabilities of incorrect selection. The differences appear to be in the manner in which the likelihood ratio tests are employed (Hosoya advocates a sequential approach in a nested setting rather than a penalized likelihood approach) and the way in which probabilities are controlled

Table 4: $\text{FIC}_{\alpha=.05}$ penalty functions

sample size	number of regressors k_j in model									
n	1	2	3	4	5	6	7	8	9	10
20	2.07	4.27	6.59	9.07	11.71	14.56	17.62	20.96	24.61	28.64
40	2.00	4.04	6.15	8.32	10.54	12.83	15.19	17.63	20.15	22.75
60	1.97	3.97	6.01	8.09	10.20	12.36	14.55	16.79	19.07	21.40
80	1.96	3.94	5.95	7.98	10.05	12.14	14.26	16.40	18.58	20.79
100	1.95	3.92	5.91	7.92	9.95	12.01	14.08	16.18	18.31	20.45
120	1.95	3.91	5.88	7.88	9.89	11.92	13.97	16.04	18.13	20.23
140	1.94	3.90	5.87	7.85	9.85	11.87	13.89	15.94	18.00	20.08
160	1.94	3.89	5.85	7.83	9.82	11.82	13.84	15.86	17.91	19.96
180	1.94	3.88	5.84	7.81	9.79	11.79	13.79	15.81	17.84	19.87
200	1.94	3.88	5.84	7.80	9.78	11.76	13.76	15.76	17.78	19.81
∞	1.92	3.84	5.76	7.68	9.60	11.52	13.45	15.37	17.29	19.21

(Hosoya controls marginal probablities of incorrect selection). We believe our procedure has an advantage in that we define a single model selection procedure based on controlling probabilities that have a simple interpretation for each model, and hence will be easier for the practitioner to determine.

3.1 Approximate penalties

We have demonstrated the construction of an IC penalty function based on controlling probabilities of correct selection for a fixed sample size n. The penalty function in (22) was derived using the knowledge that

$$(n-k_j)\left[\exp\frac{2}{n}(L_j - L_{j-1}) - 1\right] = \frac{(\hat{u}'_{j-1}\hat{u}_{j-1} - \hat{u}'_j\hat{u}_j)}{(\hat{u}'_j\hat{u}_j)/(n-k_j)} \tag{25}$$

has an $F_{1,n-k_j}$ distribution when model M_{j-1} is true and $k_j - k_{j-1} = 1$. However, this distribution only holds due to the assumption of normally distributed errors, U. In fact we can use the same principles to derive an approximate penalty function based on asymptotic arguments in the case when finite sample normality need not hold. We want to find the relative penalties $p_j - p_{j-1}$ such that

$$P_{M_{j-1}}(L_j - L_{j-1} > p_j - p_{j-1}) = \alpha_j \tag{26}$$

under less restrictive assumptions. Note that the likelihood ratio (LR) statistic $2(L_j - L_{j-1})$ has an asymptotic chi-squared distribution when model M_{j-1} is 'true', under the assumption of normal errors. Note also that due to the QML results, such as those given in Gourieroux and Monfort (1993) and references therein, that even when the assumed error distribution is not 'true', but still belongs to a broader class of distributions, the LR statistic retains its asymptotic chi-squared distribution. What is required is that the conditional mean vector and covariance matrix satisfy

$$\begin{aligned} \mathrm{E}[Y \mid X_{j-1}] &= X_{j-1}\beta \\ \mathrm{Var}(Y \mid X_{j-1}) &= \mathrm{Var}(U) = \sigma^2 \mathrm{I}_n, \end{aligned}$$

respectively, where X_{j-1} represents the $n \times k_{j-1}$ matrix of regressors appropriate to model M_{j-1}, and that the distribution of U belongs to the quadratic exponential family. Under this QML setup, any quasi-likelihood function resulting from the quadratic exponential family could be used, including that given in (3) which arises from the assumption of normal errors. Using these QML results for the LR statistic, we can calculate approximate critical values for any given α_j desired and, assuming $k_j - k_{j-1} = 1$, the resulting penalty function is defined by

$$p_j = \frac{1}{2} \sum_{i=1}^{k_j} \chi_1^2(\alpha_i), \tag{27}$$

where $\chi_1^2(\alpha)$ is the upper α^{th} quantile of the chi-squared distribution with one degree of freedom. Note that if all $\alpha_i = \alpha$, then

$$p_j = \frac{k_j}{2} \chi_1^2(\alpha). \tag{28}$$

We denote the procedure based on the penalty function above as QFIC. The final row of Tables 3 and 4 illustrate these approximate values for all $\alpha = .10$ and $\alpha = .05$, respectively. It is worthwhile noting that for moderate sample sizes and moderate numbers of unknown regression coefficients, the small sample penalties are quite similar to the approximate penalties. However, for smaller sample sizes, the relative penalty functions $p_j - p_{j-1}$ are not all equal and are increasing as k_j increases, indicating there may be something to gain from using the small sample penalties when the assumption of normal errors can be justified. In both cases, the penalties must be compared with the value of the maximized (quasi) log-likelihood values to determine if the difference in penalty functions is significant.

3.2 Strong consistency of the QFIC procedure

Much of the variable selection literature has been concerned with the consistency properties of various selection criteria. Our main view is that the user may feel comfortable in selecting a 'significance' level for the variable selection procedure due to experience in the particular setting. That is, the practitioner may have justification for selecting a particular value for α, the (approximate) probability of incorrectly adding an additional regressor. Still, it may be desirable for theoretical reasons to determine conditions under which the QFIC variable selection procedure does have strong consistency properties. We establish the following result.

Theorem 3.1 *The QFIC procedure based on (28) with $\alpha = \alpha(n)$, a function of the sample size n, is strongly consistent if both*

$$(i)\ \lim_{n\to\infty} \frac{-\log \alpha(n)}{n} = 0$$

and

$$(ii)\ \lim_{n\to\infty} \frac{-\log \alpha(n)}{\log\log n} = \infty$$

hold.

Proof: To prove the theorem we first prove the following lemma.

Lemma 3.2 *For any $f : I\!N \to I\!R$ with $\lim_{n\to\infty} f(n) = \infty$, we have*

$$\lim_{n\to\infty} \frac{\chi_1^2(\alpha(n))}{f(n)} = \infty \qquad \textit{iff} \qquad \lim_{n\to\infty} \frac{-\log \alpha(n)}{f(n)} = \infty.$$

Proof: The proof of this lemma follows along the same lines as the proof of Pötscher's (1983) Theorem 5.8 where it is established that

$$\exp(-2\chi_1^2(\alpha(n))) \le \alpha(n) \le \exp(-\frac{1}{2}\chi_1^2(\alpha(n)))$$

for large n. Simple algebra yields that, for large n,

$$\frac{\chi_1^2(\alpha(n))}{f(n)} \le \frac{-2\log \alpha(n)}{f(n)} \le \frac{4\chi_1^2(\alpha(n))}{f(n)}, \tag{29}$$

and hence if

$$\lim_{n\to\infty} \frac{\chi_1^2(\alpha(n))}{f(n)} = \infty,$$

then

$$\lim_{n\to\infty} \frac{-\log \alpha(n)}{f(n)} = \infty$$

also. Similarly, assuming

$$\lim_{n\to\infty} \frac{-\log \alpha(n)}{f(n)} = \infty,$$

then (29) implies that

$$\lim_{n\to\infty} \frac{\chi_1^2(\alpha(n))}{f(n)} = \infty$$

also, and the lemma is proved. □

Now, to establish the proof of the theorem, from Nishi (1988, Theorem 4) we have that QFIC is strongly consistent if both

$$\text{(i}'\text{)} \qquad \lim_{n\to\infty} \frac{\chi_1^2(\alpha(n))}{n} = 0$$

and

$$\text{(ii}'\text{)} \qquad \lim_{n\to\infty} \frac{\chi_1^2(\alpha(n))}{\log\log n} = \infty$$

hold. To demonstrate that (i) holds if and only if (i′) holds, we refer to Pötscher (1983, Theorem 5.8), which states that for any $f : \mathbb{N} \to \mathbb{R}$ with $\lim_{n\to\infty} f(n) = \infty$, we have

$$\lim_{n\to\infty} \frac{\chi_1^2(\alpha(n))}{f(n)} = 0 \qquad \text{iff} \qquad \lim_{n\to\infty} \frac{-\log \alpha(n)}{f(n)} = 0.$$

Thus, taking $f(n) = n$ we have that (i′) holds if and only if (i) holds. That is,

$$\lim_{n\to\infty} \frac{\chi_1^2(\alpha(n))}{n} = 0 \qquad \text{iff} \qquad \lim_{n\to\infty} \frac{-\log \alpha(n)}{n} = 0.$$

Using our Lemma 3.2, and taking $f(n) = \log\log n$, we have that (ii′) holds if and only if (ii) holds. Thus,

$$\lim_{n\to\infty} \frac{\chi_1^2(\alpha(n))}{\log\log n} = \infty \qquad \text{iff} \qquad \lim_{n\to\infty} \frac{-\log \alpha(n)}{\log\log n} = \infty,$$

and the theorem is proven. □

As a final note, if our variable selection procedure was intended to choose models with error distributions that were known and nonnormal, the above approximate penalty may also be correct under usual regularity conditions such as those found in Godfrey (1988).

Under these alternative assumptions, although the penalties would be the same as those defined by (27), our theory would suggest using the maximized log-likelihood values arising from the nonnormal distributions, rather than those arising from the assumption of normal errors.

4 Effective probabilities for other IC methods

In the two previous Sections, we described a method of choosing relative penalties based on controlling probabilities of correctly increasing the number of regressors. In this Section, we invert the argument to uncover the corresponding marginal probabilities of correctly increasing the number of regressors for other existing IC procedures.

For any IC procedure, (5) can be used to determine the α level corresponding to the relative penalty $p_j - p_{j-1}$. To do this, we determine the upper α^{th} percentile of the $F_{1,n-k_j}$ distribution corresponding to the value of

$$f_{\alpha,1,n-k_j} = (n - k_j)\left[\exp \frac{2}{n}(p_j - p_{j-1}) - 1\right]. \tag{30}$$

Table 5 displays the effective α levels for varying sample sizes and numbers of regressors included in a linear regression for the $\text{QFIC}_{\alpha=.10}$, $\text{QFIC}_{\alpha=.05}$, AIC, BIC, HQ and $\bar{R}^2$ procedures. These effective α levels were calculated under the assumption of independent normal errors. The effective α levels for the approximate $\text{QFIC}_{\alpha=.10}$ and $\text{QFIC}_{\alpha=.05}$ procedures demonstrate that for finite samples the probability of adding an unnecessary regressor is larger than α and increasing as k_j increases for fixed sample size n, and decreases to α as n increases for fixed k_j. Thus, relative to the finite sample FIC procedure, the QFIC procedure will be more likely to choose models with more regressors.

Note that, following the dimension consistency results from Akaike (1970), Hannan and Quinn (1979) and Geweke and Meese (1981) among others, we expect the effective α levels for BIC and HQ to decrease to zero, while those for AIC and $\bar{R}^2$ stay bounded above zero. Note the differences in the rates at which these effective α levels appear to decrease, with BIC reducing to small effective α levels for very small sample sizes, whereas the effective α levels of $\bar{R}^2$ remain quite large, regardless of the sample size.

It is interesting to look closer at Theil's $\bar{R}^2$ and the effective α level relationship as

Table 5: Effective α values for $\text{QFIC}_{\alpha=.10}$, $\text{QFIC}_{\alpha=.05}$, AIC, BIC, HQ and $\bar{R}^2$ procedures

		effective α_j					
n	j	$\text{QFIC}_{\alpha=.10}$	$\text{QFIC}_{\alpha=.05}$	AIC	BIC	HQ	$\bar{R}^2$
15	1	0.119	0.063	0.179	0.118	0.180	0.334
30		0.109	0.056	0.168	0.072	0.127	0.326
45		0.106	0.054	0.164	0.055	0.108	0.323
60		0.105	0.053	0.163	0.046	0.097	0.321
15	2	0.133	0.073	0.196	0.133	0.197	0.336
30		0.116	0.061	0.176	0.077	0.134	0.326
45		0.110	0.057	0.169	0.058	0.112	0.323
60		0.108	0.055	0.166	0.048	0.100	0.321
15	3	0.150	0.086	0.215	0.149	0.216	0.337
30		0.122	0.065	0.184	0.083	0.141	0.326
45		0.115	0.060	0.174	0.061	0.116	0.323
60		0.111	0.057	0.170	0.050	0.103	0.322
15	4	0.169	0.101	0.236	0.168	0.237	0.339
30		0.130	0.071	0.192	0.089	0.149	0.327
45		0.119	0.063	0.180	0.064	0.121	0.323
60		0.114	0.059	0.174	0.052	0.106	0.322
15	5	0.191	0.118	0.260	0.190	0.261	0.341
30		0.137	0.077	0.201	0.096	0.157	0.327
45		0.124	0.066	0.185	0.068	0.126	0.323
60		0.117	0.062	0.178	0.054	0.110	0.322
15	6	0.216	0.140	0.286	0.215	0.287	0.343
30		0.146	0.083	0.211	0.103	0.166	0.327
45		0.129	0.070	0.191	0.071	0.130	0.323
60		0.121	0.064	0.182	0.056	0.112	0.322
15	7	0.245	0.165	0.317	0.244	0.318	0.347
30		0.155	0.090	0.221	0.110	0.175	0.328
45		0.134	0.074	0.197	0.075	0.136	0.324
60		0.124	0.067	0.186	0.058	0.116	0.322
15	8	0.278	0.196	0.351	0.278	0.352	0.351
30		0.164	0.097	0.231	0.118	0.185	0.328
45		0.139	0.078	0.203	0.079	0.141	0.324
60		0.128	0.069	0.190	0.061	0.120	0.322

follows. From Table 1 we know that for $\bar{R}^2$

$$p_j - p_{j-1} = \frac{n}{2}\log\left(\frac{n-k_{j-1}}{n-k_j}\right). \tag{31}$$

Thus, using (30),

$$\begin{aligned} f_{\alpha,1,n-k_j} &= (n-k_j)\left\{\exp\frac{2}{n}\left[\frac{n}{2}\log\left(\frac{n-k_{j-1}}{n-k_j}\right)\right]-1\right\} && (32)\\ &= k_j - k_{j-1} = 1. && (33)\end{aligned}$$

This corresponds to the well-known result (see Griffiths, Hill and Judge, 1993, p. 343,) that the value of $\bar{R}^2$ will increase with the inclusion of an additional regressor with coefficient β_{k_j} if and only if the absolute value of the $t-$statistic for the hypothesis $H_0 : \beta_{k_j} = 0$ is greater than one. This is obviously due to the fact that if T has a $t-$distribution with ν degrees of freedom, then T^2 has an $F-$distributution with one numerator and ν denominator degrees of freedom.

5 Monte Carlo study

This Section reports the results of Monte Carlo studies conducted to assess the small sample performance of our new F-statistic based IC procedures (FIC and QFIC) under both the correctly specified normal error distribution and also under a few misspecified nonnormal error distributions. Our objective is to select the correct model from among a group of plausible models and as such we calculated the Monte Carlo probabilities of correct selection (based on $N = 2000$ repetitions) and compared them to those of AIC, BIC, HQ and $\bar{R}^2$, for sample sizes of $n = 32, 64$ and 128.

Three linear regression designs were used, each with a different number of possible regressors. We first present a brief description of the three designs used. In each case the values of $\sigma^2 = 1.0$ and $\beta_0 = 1.0$ were used as the results are invariant to the values of the scale and shift parameters.

Design I: Full model $Y = \beta_0 + x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + x_4\beta_4 + U$ with

x_1 : Quarterly seasonally adjusted Australian disposable income beginning 1959(4) in millions of dollars

x_2 : x_1 lagged by one quarter

x_3 : Final consumption expenditure (private) beginning 1959(4) in millions of dollars

x_4 : x_3 lagged by one quarter.

The six possible models considered were:

$$M_1 : Y = \beta_0 + x_1\beta_1 + U$$
$$M_2 : Y = \beta_0 + x_3\beta_3 + U$$
$$M_3 : Y = \beta_0 + x_1\beta_1 + x_3\beta_3 + U$$
$$M_4 : Y = \beta_0 + x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + U$$
$$M_5 : Y = \beta_0 + x_1\beta_1 + x_3\beta_3 + x_4\beta_4 + U$$
$$M_6 : Y = \beta_0 + x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + x_4\beta_4 + U.$$

Parameter specifications:

n=32	$\beta_i = 9, 27$
n=64	$\beta_i = 3, 9$
n=128	$\beta_i = 0.5, 1.5.$

Design II: Full model $Y = \beta_0 + x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + x_4\beta_4 + x_5\beta_5 + U$ with

x_1 : Quarterly Australian consumer price index beginning 1948(4)

x_2 : x_1 lagged by one quarter

x_3 : Seasonal dummy variable for 4th quarter

x_4 : Seasonal dummy variable for 1st quarter

x_5 : Seasonal dummy variable for 2nd quarter.

The four possible models considered were:

$$M_1 : Y = \beta_0 + x_1\beta_1 + U$$
$$M_2 : Y = \beta_0 + x_1\beta_1 + x_2\beta_2 + U$$
$$M_3 : Y = \beta_0 + x_1\beta_1 + x_3\beta_3 + x_4\beta_4 + x_5\beta_5 + U$$
$$M_4 : Y = \beta_0 + x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + x_4\beta_4 + x_5\beta_5 + U.$$

Parameter specifications:

n=32	$\beta_i = 0.6, 1.8$
n=64	$\beta_i = 0.4, 1.2$
n=128	$\beta_i = 0.2, 0.6.$

Design III: Full model $Y = \beta_0 + x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + U$ with

x_1 : Australian household populations in 1966 and 1971 (see Williams and Sams, 1981)

x_2 : Number of Australian households in 1966 and 1971

x_3 : Australian household headship ratios in 1966 and 1971 (proportion of people in any given population category who are heads of household).

The six possible models considered were:

$M_1 : Y = \beta_0 + x_1\beta_1 + U$

$M_2 : Y = \beta_0 + x_2\beta_2 + U$

$M_3 : Y = \beta_0 + x_1\beta_1 + x_2\beta_2 + U$

$M_4 : Y = \beta_0 + x_1\beta_1 + x_3\beta_3 + U$

$M_5 : Y = \beta_0 + x_2\beta_2 + x_3\beta_3 + U.$

$M_6 : Y = \beta_0 + x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + U.$

Parameter specifications:

n=32	$\beta_i = 1, 3$
n=64	$\beta_i = 0.3, 0.9$
n=128	$\beta_i = 0.25, 0.75.$

In the case where the errors were misspecified, we generated the true errors according various distributions using a uniform random number generator and a generalized four parameter Tukey-λ transformation (Ramberg *et al*, 1979). The transformation equation is

$$R(p) = \lambda_1 + [p^{\lambda_3} - (1-p)^{\lambda_4}]/\lambda_2, \qquad 0 \le p \le 1,$$

where p is a uniform pseudo-random number and λ_i, $i = 1, \ldots, 4$ are four nonzero transformation parameters. This allows the generation of pseudo-random variates with specified values of the first four moments. Table 6 reports the values of the λ values used in our study, along with the corresponding measures of skewness and kurtosis. All of the resulting distributions have zero mean and unit variance.

5.1 Results

Tables 7, 8 and 9 contain the probabilities of correct selection for each of designs I, II and III, respectively, under the correctly specified normal errors model. Notice that the probabilities of correct selection have been broken down according to each model, and that averages over all sample sizes and model choices are given in the margins. We first compare FIC with AIC, BIC, HQ and $\bar{R}^2$. The results indicate that the $\text{FIC}_{\alpha=.10}$ procedure has probabilities of correct selection that fall somewhere between those of AIC and BIC, and are quite close to those of the HQ procedure, whereas the $\text{FIC}_{\alpha=.05}$ procedure more closely resembles BIC than the others. This is not surprising due to the similarities of the range of effective α-levels corresponding to these other procedures. Indeed, both $\text{FIC}_{\alpha=.10}$ and $\text{FIC}_{\alpha=.05}$ are quite competitive with the other IC procedures for each model, and in fact $\text{FIC}_{\alpha=.10}$ has a slightly higher average probability of correct selection over all of the designs and sample sizes we considered. What distinguishes the FIC procedure from the others is the simplicity by which it can be explained and the control that the practitioner has over the probabilities of error made at each stage.

We have also included the procedures $\text{QFIC}_{\alpha=.10}$ and $\text{QFIC}_{\alpha=.05}$. The penalty functions for these procedures are even simpler to calculate, and we see from the Monte Carlo simulation results that overall the procedures are quite competitive with the finite sample FIC procedures and the other IC procedures considered. The difference between the finite sample and the approximately derived QFIC procedures appears to be that the finite sample FIC does well more uniformly across models, regardless of the number of regressors, whereas the QFIC procedures slightly favor larger models. This corresponds to the discussion in Section 4 relating to Table 5 which shows the larger than nominal effective α levels for the QFIC procedures. In an average over all models considered, however, QFIC appears to do about as well as the finite sample FIC procedure.

The Monte Carlo simulations using the misspecified nonnormal errors from the Tukey-λ transformations indicated in Table 6 gave similar results to the correctly specified case

Table 6: Parameters for Tukey-λ transformation

λ_1	λ_2	λ_3	λ_4	Skewness	Kurtosis
-0.3790	-0.0562	-0.0187	-0.3880	1	6
0.0000	-0.3203	-0.1359	-0.1359	0	9

reported above and as such we have not included the tables here. The results of these simulations are available upon request from the authors. Thus, we conclude that the FIC and QFIC procedures of model selection are quite robust to misspecified error distributions in the linear regression setting. For these reasons, we believe variable selection using FIC and QFIC offer competitive tools for the practitioner, and ones that are easily understood and implemented.

6 Concluding remarks

We have presented two related IC variable selection procedures for choosing the regressors in a linear regression procedure. The penalties calculated allow for the practitioner to determine the magnitude of the probability of incorrectly adding a regressor to the true model. The resulting procedures perform competitively against existing methods, without seeming to overly favor models with fewer or greater numbers of regressors. In addition, by inverting the argument used to derive the finite sample FIC penalty function, some insight is gained into the probabilities of incorrectly adding a regressor for other IC methods.

The approximate penalty function was also derived based on similar principles. The penalty function for the QFIC is easier to calculate than the finite sample procedure and also has a quasi-maximum likelihood justification to support it. Both procedures are easy to implement and seem to do quite well in the Monte Carlo simulation studies conducted. The finite sample procedure offers the practitioner more control over the probability of incorrectly adding a regressor when the assumption of normal errors is justified, while the approximate procedure may be more robust to nonnormal errors.

References

[1] Akaike, H. (1970) 'Statistical predictor identification,' *Annals of the Institute of Statistical Mathematics,* **22**, 203-217.

[2] Akaike, H. (1973) 'Information theory and an extention of the maximum likelihood principle,' in B. N. Petrov and F. Csaki, eds., *Second International Symposium on Information Theory,* Budapest: Akademiai Kiado, 267-281.

[3] Fox, K. J. (1995) 'Model selection criteria: A reference source,' unpublished manuscript.

[4] Geweke, J. and Meese, R. (1981) 'Estimating regression models of finite but unknown order,' *International Economic Review,* **22**, 55-70.

[5] Godfrey, L. G. (1988) *Misspecification Tests in Econometrics: The Lagrange Multiplier Principle and other Approaches,* Cambridge, New York.

[6] Gourieroux, C. and Monfort, A. (1993) 'Pseudo-Likelihood Methods,' in G. S. Maddala, C. R. Rao and H. D. Vinod, eds., *Handbook of Statistics, Vol. 11,* Elsevier Science Publishers, 335-363.

[7] Greene, W. H. (1990) *Econometric Analysis,* Macmillan, New York.

[8] Griffiths, W. E., Hill, R. C. and Judge, G. G. (1993) *Learning and Practicing Econometrics,* John Wiley, New York.

[9] Hannan, E. J. and Quinn, B. G. (1979) 'The determination of the order of an autoregression,' *Annals of Statistics*, **8**, 1081-1081.

[10] Hosoya, Y. (1984) 'Information criteria and tests for time-series models,' in O. D. Anderson, ed., *Time Series Analysis: Theory and Practice 5,* North-Holland, 39-52.

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

[12] Hurvich, C. M. and Tsai, C-L. (1991) 'Bias of the corrected AIC criterion for underfitted regression and time series models,' *Biometrika,* **78**, 499-509.

[13] King, M. L., Forbes, C. S. and Morgan, A. (1995) 'Improved small sample model selection procedures,' paper presented at the 1995 World Congress of the Econometric Society, Tokyo, Japan.

[14] Nishi, R. (1988) 'Maximum likelihood principle and model selection when the true model is unspecified,' *Journal of Multivariate Analysis,* **27**, 392-403.

[15] Pötscher, B. M. (1983) 'Order estimation in ARMA-models by Lagrangian multiplier tests,' *The Annals of Statistics,* **11**, 872-885.

[16] Pötscher, B. M. (1991) 'Effects of model selection on inference,' *Econometric Theory,* **7**, 163-185.

[17] Ramberg, J. S., Dudewicz, E. J., Tadikamalla, P. R. and Mykytka, E. F. (1979) 'A probability distribution and its uses in fitting data,' *Technometrics,* **21**, 201-215.

[18] Sawa, T. (1978) 'Information criteria for discriminating among alternative regression models,' *Econometrica,* **46**, 1273-1291.

[19] Schwartz, G. (1978) 'Estimating the dimension of a model,' *Annals of Statistics,* **6**, 461-464.

[20] Sugiura, N. (1978) 'Further analysis of the data by Akaike's information criterion and the finite corrections,' *Communications in Statistics: Theory and Methods,* **7**, 13-26.

[21] Williams, P. and Sams, D. (1981), 'Household headship in Australia: Further developments to the IMPACT Project's econometric model of household headship', IMPACT Project Research Centre Working Paper, No: BP-26, pp. 91-96.

Table 7: Monte Carlo probabilities of correct selection for Design I

		M_1	M_2	M_3	M_4	M_5	M_6	Average
$n = 32$	$FIC_{\alpha=.10}$	0.803	0.786	0.523	0.746	0.494	0.371	0.620
	$FIC_{\alpha=.05}$	0.875	0.857	0.474	0.714	0.432	0.264	0.603
	AIC	0.662	0.650	0.504	0.717	0.508	0.505	0.591
	BIC	0.830	0.817	0.505	0.737	0.480	0.355	0.621
	HQ	0.740	0.723	0.521	0.728	0.511	0.450	0.612
	$\bar{R}^2$	0.435	0.427	0.391	0.636	0.487	0.623	0.500
	$QFIC_{\alpha=.10}$	0.765	0.748	0.516	0.735	0.507	0.426	0.616
	$QFIC_{\alpha=.05}$	0.849	0.838	0.487	0.732	0.461	0.325	0.615
$n = 64$	$FIC_{\alpha=.10}$	0.839	0.822	0.682	0.719	0.356	0.318	0.622
	$FIC_{\alpha=.05}$	0.911	0.893	0.679	0.697	0.267	0.222	0.611
	AIC	0.716	0.705	0.626	0.710	0.418	0.423	0.600
	BIC	0.912	0.894	0.675	0.695	0.264	0.224	0.611
	HQ	0.835	0.818	0.676	0.720	0.362	0.329	0.623
	$\bar{R}^2$	0.435	0.433	0.458	0.628	0.444	0.577	0.496
	$QFIC_{\alpha=.10}$	0.822	0.805	0.674	0.722	0.372	0.343	0.623
	$QFIC_{\alpha=.05}$	0.900	0.884	0.682	0.702	0.284	0.248	0.617
$n = 128$	$FIC_{\alpha=.10}$	0.819	0.810	0.690	0.860	0.684	0.657	0.753
	$FIC_{\alpha=.05}$	0.897	0.889	0.685	0.876	0.639	0.574	0.760
	AIC	0.711	0.708	0.636	0.806	0.676	0.725	0.710
	BIC	0.926	0.917	0.653	0.868	0.594	0.516	0.746
	HQ	0.847	0.841	0.697	0.868	0.678	0.633	0.761
	$\bar{R}^2$	0.418	0.421	0.450	0.671	0.604	0.819	0.564
	$QFIC_{\alpha=.10}$	0.811	0.802	0.685	0.858	0.678	0.667	0.750
	$QFIC_{\alpha=.05}$	0.890	0.883	0.688	0.877	0.644	0.586	0.761
Average	$FIC_{\alpha=.10}$	0.820	0.806	0.632	0.775	0.511	0.449	0.665
	$FIC_{\alpha=.05}$	0.894	0.879	0.613	0.762	0.446	0.353	0.658
	AIC	0.696	0.688	0.589	0.744	0.534	0.551	0.634
	BIC	0.889	0.876	0.611	0.767	0.446	0.365	0.659
	HQ	0.808	0.794	0.631	0.772	0.517	0.471	0.665
	$\bar{R}^2$	0.429	0.427	0.433	0.645	0.512	0.673	0.520
	$QFIC_{\alpha=.10}$	0.799	0.785	0.625	0.771	0.519	0.479	0.663
	$QFIC_{\alpha=.05}$	0.879	0.868	0.619	0.770	0.463	0.387	0.664

Table 8: Monte Carlo probabilities of correct selection for Design II

		M_1	M_2	M_3	M_4	Average
$n = 32$	$\text{FIC}_{\alpha=.10}$	0.483	0.021	0.188	0.527	0.305
	$\text{FIC}_{\alpha=.05}$	0.682	0.015	0.165	0.345	0.302
	AIC	0.296	0.025	0.177	0.682	0.295
	BIC	0.528	0.018	0.180	0.502	0.307
	HQ	0.376	0.023	0.181	0.626	0.301
	$\bar{R}^2$	0.145	0.022	0.167	0.765	0.275
	$\text{QFIC}_{\alpha=.10}$	0.413	0.021	0.181	0.599	0.303
	$\text{QFIC}_{\alpha=.05}$	0.586	0.015	0.173	0.450	0.306
$n = 64$	$\text{FIC}_{\alpha=.10}$	0.489	0.023	0.194	0.513	0.305
	$\text{FIC}_{\alpha=.05}$	0.659	0.019	0.166	0.352	0.299
	AIC	0.335	0.026	0.189	0.636	0.297
	BIC	0.660	0.018	0.163	0.355	0.299
	HQ	0.477	0.022	0.190	0.529	0.304
	$\bar{R}^2$	0.158	0.022	0.178	0.745	0.275
	$\text{QFIC}_{\alpha=.10}$	0.454	0.024	0.190	0.547	0.304
	$\text{QFIC}_{\alpha=.05}$	0.617	0.020	0.170	0.397	0.301
$n = 128$	$\text{FIC}_{\alpha=.10}$	0.658	0.038	0.178	0.285	0.290
	$\text{FIC}_{\alpha=.05}$	0.826	0.029	0.116	0.139	0.277
	AIC	0.486	0.045	0.209	0.430	0.292
	BIC	0.898	0.020	0.078	0.075	0.268
	HQ	0.716	0.036	0.158	0.236	0.286
	$\bar{R}^2$	0.208	0.040	0.221	0.630	0.275
	$\text{QFIC}_{\alpha=.10}$	0.641	0.039	0.183	0.303	0.291
	$\text{QFIC}_{\alpha=.05}$	0.807	0.029	0.123	0.154	0.278
Average	$\text{FIC}_{\alpha=.10}$	0.543	0.027	0.187	0.442	0.300
	$\text{FIC}_{\alpha=.05}$	0.722	0.021	0.149	0.279	0.293
	AIC	0.372	0.032	0.192	0.583	0.295
	BIC	0.695	0.018	0.140	0.311	0.291
	HQ	0.523	0.027	0.176	0.463	0.297
	$\bar{R}^2$	0.170	0.028	0.188	0.713	0.275
	$\text{QFIC}_{\alpha=.10}$	0.503	0.028	0.185	0.483	0.300
	$\text{QFIC}_{\alpha=.05}$	0.670	0.021	0.155	0.334	0.295

Table 9: Monte Carlo probabilities of correct selection for Design III

		M_1	M_2	M_3	M_4	M_5	M_6	Average
$n = 32$	$\text{FIC}_{\alpha=.10}$	.840	0.815	0.891	0.612	0.575	0.578	0.718
	$\text{FIC}_{\alpha=.05}$	0.911	0.902	0.932	0.590	0.541	0.480	0.726
	AIC	0.706	0.677	0.813	0.602	0.568	0.671	0.673
	BIC	0.871	0.851	0.906	0.604	0.563	0.555	0.725
	HQ	0.775	0.755	0.852	0.608	0.572	0.632	0.699
	$\bar{R}^2$	0.483	0.458	0.673	0.544	0.526	0.768	0.575
	$\text{QFIC}_{\alpha=.10}$	0.804	0.777	0.869	0.608	0.574	0.611	0.707
	$\text{QFIC}_{\alpha=.05}$	0.887	0.881	0.922	0.597	0.551	0.531	0.728
$n = 64$	$\text{FIC}_{\alpha=.10}$	0.819	0.786	0.722	0.315	0.249	0.153	0.507
	$\text{FIC}_{\alpha=.05}$	0.909	0.861	0.699	0.251	0.181	0.093	0.499
	AIC	0.711	0.685	0.704	0.357	0.301	0.242	0.500
	BIC	0.915	0.866	0.694	0.245	0.178	0.092	0.498
	HQ	0.817	0.785	0.721	0.315	0.249	0.157	0.507
	$\bar{R}^2$	0.475	0.472	0.617	0.381	0.350	0.413	0.451
	$\text{QFIC}_{\alpha=.10}$	0.803	0.769	0.720	0.323	0.257	0.168	0.507
	$\text{QFIC}_{\alpha=.05}$	0.897	0.851	0.701	0.262	0.193	0.105	0.501
$n = 128$	$\text{FIC}_{\alpha=.10}$	0.835	0.793	0.807	0.396	0.337	0.229	0.566
	$\text{FIC}_{\alpha=.05}$	0.920	0.886	0.821	0.338	0.263	0.139	0.561
	AIC	0.736	0.707	0.769	0.428	0.382	0.317	0.557
	BIC	0.954	0.927	0.814	0.291	0.221	0.099	0.551
	HQ	0.869	0.835	0.813	0.374	0.306	0.196	0.566
	$\bar{R}^2$	0.500	0.478	0.650	0.434	0.407	0.501	0.495
	$\text{QFIC}_{\alpha=.10}$	0.824	0.789	0.805	0.399	0.342	0.239	0.566
	$\text{QFIC}_{\alpha=.05}$	0.915	0.880	0.820	0.341	0.270	0.146	0.562
Average	$\text{FIC}_{\alpha=.10}$	0.831	0.798	0.806	0.441	0.387	0.320	0.597
	$\text{FIC}_{\alpha=.05}$	0.913	0.883	0.817	0.393	0.328	0.237	0.595
	AIC	0.718	0.690	0.762	0.462	0.417	0.410	0.576
	BIC	0.913	0.881	0.805	0.380	0.320	0.249	0.591
	HQ	0.820	0.791	0.795	0.432	0.376	0.328	0.591
	$\bar{R}^2$	0.486	0.469	0.647	0.453	0.427	0.561	0.507
	$\text{QFIC}_{\alpha=.10}$	0.810	0.778	0.798	0.443	0.391	0.339	0.593
	$\text{QFIC}_{\alpha=.05}$	0.899	0.871	0.814	0.400	0.338	0.261	0.597