Faculty Working Papers

UNBIASED DECISION RULE FOR THE CHOICE OF REGRESSION MODELS

Takamitsu Sawa and Kei Takeuchi

#400

College of Commerce and Business Administration
University of Illinois at Urbana-Champaign

FACULTY WORKING PAPERS

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

May 19, 1977

UNBIASED DECISION RULE FOR THE CHOICE OF REGRESSION MODELS

Takamitsu Sawa and Kei Takeuchi

#400

Unbiased Decision Rule for the Choice of Regression Models*

Takamitsu Sawa** and Kei Takeuchi***

May 11, 1977

Prepared Under

NATIONAL SCIENCE FOUNDATION GRANT SOC 76-22232

*The original version of this paper was presented at the Meeting of the Biometric Society in Boston in August 1977.

**Department of Economics, University of Illinois and Kyoto University.

***Faculty of Economics, University of Tokyo.

Abstract

The unbiased decision rule, based on the conventional F-statistic, is proposed for the choice of the most adequate model from a given set of nested alternative regression models. Mallows' (1973) idea is adapted to measure the adequacy of each model. The biases of several customarily used criteria for the choice of regression models, including the AIC, Mallows' C_p and $\bar{R}$, are discussed. The critical point for the unbiased decision is numerically computed and tabulated. Also, the approximate formula for the unbiased critical point is derived.

1. Introduction

There has been a vast amount of literature dealing with procedures of selecting the most adequate regression model from a given set of alternatives. This fact illustrates that the choice of the most appropriate subset of regressors is one of the most difficult as well as bothersome problems in practical regression analysis. To compare several alternative models, the most frequently used as well as most primitive criterion may be the residual sum of squares

$$(1.1) \qquad Q_p = \underset{\sim}{Y}' \left[\underset{\sim}{I} - \underset{\sim}{X} (\underset{\sim}{X}'\underset{\sim}{X})^{-1} \underset{\sim}{X}'\right] \underset{\sim}{Y} = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

for a regression model: $E(Y) = X\beta$, $V(Y) = \sigma^2 I$, or an equivalent expression, the multiple correlation coefficient

$$(1.2) \qquad R = \left[1 - \frac{Q_p}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}\right]^{1/2}$$

where $\underset{\sim}{Y}$ is an n-dimensional vector random variable; $\underset{\sim}{X}$ is an n x p matrix of fixed constants; σ^2 is a scalar unknown; $\underset{\sim}{\beta}$ is a p x 1 vector of unknowns; $\underset{\sim}{I}$ is an n x n identity matrix; Y_i's are components of $\underset{\sim}{Y}$; $\bar{Y}$ is the mean of Y_i's; $\hat{Y}_i$'s are interpolated values. These criteria have an apparent shortcoming in that they do not weigh the costs of successive increases in the parameter space against the improvements in model fit. To overcome this, we usually make an adjustment such as

$$(1.3) \qquad \bar{R}^2 = 1 - (n - 1)(1 - R^2)/(n - p).$$

$\bar{R}$ is commonly called the multiple correlation coefficient adjusted for the degrees of freedom. As a descriptive measure of the goodness of

fit, $\bar{R}$ has a sound intuitive appeal. So far, however, no persuasive reasons have been given to justify the use of $\bar{R}$. That is, it is not clear why $\bar{R}$ should be preferred to any other possible adjustments to R.

Mallows (1973) proposed another criterion called the C_p statistic

$$(1.4) \qquad C_p = Q_p + 2p\,\hat{\omega}^2$$

where the first term on the right hand side is the residual sum of squares, p is the number of regressors, and $\hat{\omega}^2$ is an appropriately chosen unbiased estimate for the true variance of the Y_i's. In fact it is an open question how to obtain $\hat{\omega}^2$, but it should be noted that $\hat{\omega}^2$ is independent of p and common factor in C_p to all alternative models. The true distribution of $\underset{\sim}{Y}$ is assumed to be $N(\underset{\sim}{\mu}, \omega^2 \underset{\sim}{I})$. The C_p statistic has been derived as an unbiased estimate for the mean squared error

$$(1.5) \qquad W_p = E\,||\,\underset{\sim}{Y}^0 - \underset{\sim}{\hat{Y}}\,||^2 = \sum_{i=1}^{n} E\,(Y_i^0 - \hat{Y}_i)^2$$

of the least squares prediction $\underset{\sim}{\hat{Y}} = \underset{\sim}{X}\,(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\,\underset{\sim}{X}'\underset{\sim}{Y}$ for $\underset{\sim}{Y}_0$ which is distributed as $N(\underset{\sim}{\mu}, \omega^2\underset{\sim}{I})$ independently of $\underset{\sim}{Y}$.

Based on a different and more profound reasoning, Akaike (1974) suggested a procedure for model identification, which is called the Akaike Information Criterion (AIC). It is defined as minus twice the maximized likelihood function plus twice the number of parameters in the likelihood function. This assumes that the likelihood function is well defined by each of the alternative models. Given a set of alternative models, we choose the one that gives the smallest AIC. The decision rule based on the AIC is termed the minimum AIC (MAIC)

procedure. This procedure has an advantage of being applicable to any statistical problem as long as each of the alternative models well defines the likelihood function. If we apply the MAIC procedure to a linear normal regression model, we arrive at the asymptotically equivalent decision rule as Mallows' C_p.

One of the present authors (Sawa (1976)) has recently developed another information criterion for the choice of regression models. The criterion is called the BIC and has been derived, following Akaike (1973), from the Kullback-Leibler measure of the mean information for discrimination between two alternative models. However, the difference between the AIC and the BIC is quite pronounced for the sample size typically dealt with in practical data analysis. Primarily it stems from the difference existing between Akaike's and Sawa's views on the true model.[1/]

The purpose of the present paper is to propose another criterion for the choice of regression models. The proposed decision rule is called the unbiased decision rule; unbiased in the sense that it leads us to the correct decision with probability higher than .5.

If the alternative models are nested, all the decision rules described above, are reduced to a decision based on a magnitude of the observed F statistic. Therefore, they can be compared in terms of implied critical points for the preliminary F test of the null hypothesis that a subset of coefficients be zero. Some readers may feel that it is useless to study the preliminary test any more because the resultant estimator has been proved to be inadmissible. To avoid this

criticism in advance, we point out that what we are proposing is not an estimation procedure but a procedure for model identification. More precisely, in the present context we aim to develop a procedure for identifying the most adequate model from a given set of alternatives rather than estimating unknown parameters involved in a given true model.

2. Unbiased Critical Point

We begin by postulating a linear normal regression model

$$(2.1) \qquad \underset{\sim}{Y} \sim N\ (\underset{\sim}{X}\beta,\ \sigma^2 \underset{\sim}{I})$$

for an n-dimensional vector random variable $\underset{\sim}{Y}$, the true distribution of which is

$$(2.2) \qquad \underset{\sim}{Y} \sim N\ (\underset{\sim}{\mu}, \omega^2 \underset{\sim}{I}),$$

where $\underset{\sim}{X}$ is an n x p full rank matrix of known constants; $\underset{\sim}{\beta}$ is a p x 1 vector of unknown parameters; $\underset{\sim}{I}$ is an n x n identity matrix; $\underset{\sim}{\mu}$ is an n x 1 vector of unknowns; ω^2 and σ^2 are unknown positive constants.

The regression model (2.1) assumes that the mean vector $\underset{\sim}{\mu}$ belongs to a linear subspace spanned by the columns of $\underset{\sim}{X}$. The least squares estimate $\hat{\underset{\sim}{\beta}} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{Y}$ has a normal distribution with mean $\beta_0 = (X'X)^{-1}X'\mu$ and variance $\omega^2(X'X)^{-1}$.[2/]

Let us consider two nested alternative regression models

$$(2.3) \qquad M_1: \quad Y \sim N\ (X_1\beta_1,\ \sigma_1^2\ I)$$

$$(2.4) \qquad M_2: \quad Y \sim N\ (X_1\beta_1 + X_2\beta_2,\ \sigma_2^2\ I)$$

where X_1 and X_2 are, respectively, n x p and n x q matrices. Without loss of generality we may assume that X_1 and X_2 are orthogonal. The

regression coefficients β_1 and β_2 are

$$(2.5) \qquad \beta_1 = (X_1'X_1)^{-1} X_1'\mu, \ \beta_2 = (X_2'X_2)^{-1} X_2'\mu.$$

Mallows' statistic for each of the models is given by

$$(2.6) \qquad C_p = Q_p + 2\, p\, \hat{\omega}^2$$

for M_1 and

$$(2.7) \qquad C_{p+q} = Q_{p+q} + 2\,(p+q)\,\hat{\omega}^2$$

for M_2, where Q_p and Q_{p+q} are the residual sums of squares for M_1 and M_2, respectively; $\hat{\omega}^2$ is a certain unbiased estimate of ω^2. Without additional information, a reasonable estimate of ω^2 would be the unbiased estimate $Q_{p+q}/(n-p-q)$ of σ_2^2. Then Mallows' decision rule, based on whether $C_p < C_{p+q}$ or $C_p > C_{p+q}$, is equivalent to a decision based on whether F is less than or greater than two, where

$$(2.8) \qquad F = \frac{(Q_p - Q_{p+q})/q}{Q_{p+q}/(n-p-q)}$$

is the conventionally used F statistic to test the null-hypothesis that M_1 is true. Hence Mallows' C_p decision rule is equivalent to a preliminary F test with a constant critical point 2 irrespective of the degrees of freedom.

If the simpler model M_1 is "true", F has a central F distribution with q and n-p-q degrees of freedom; if the more complex model M_2 is true, it has a singly noncentral F distribution, with noncentrality parameter.

$$(2.9) \qquad \lambda_1 = \underset{\sim}{\mu}'\underset{\sim}{X}_2 \, (\underset{\sim}{X}_2'\underset{\sim}{X}_2)^{-1} \, \underset{\sim}{X}_2' \, \underset{\sim}{\mu}/\omega^2$$

and q and n-p-q degrees of freedom. If the true distribution of $\underset{\sim}{Y}$ is $N(\underset{\sim}{\mu}, \omega^2 \underset{\sim}{I})$ and hence M_1 and M_2 are both incorrect, F has a doubly noncentral F distribution with q and n-p-q d.f. and noncentrality parameters λ_1 and

$$(2.10) \quad \lambda_2 = \underset{\sim}{\mu}' \left[\underset{\sim}{I} - \underset{\sim}{X} (\underset{\sim}{X}'\underset{\sim}{X})^{-1} \underset{\sim}{X}'\right] \underset{\sim}{\mu}/\omega^2$$

where $\underset{\sim}{X} = (\underset{\sim}{X}_1, \underset{\sim}{X}_2)$.

The MAIC procedure leads us to a preliminary F test with a varying critical point: that is, we choose M_1 if

$$(2.11) \quad F \leq \left[\exp\left(\frac{2q}{n}\right) - 1\right] \frac{(n-p-q)}{q}$$

and choose M_2 otherwise. The AIC critical point, given in (2.11), varies with n, p, and q. It approaches 2 from below as n becomes large.

Sawa's (1977) MBIC procedure is somewhat more complicated. It implies the following decision rule: choose M_1 if

$$(2.12) \quad n \log W - 2(p+2) W + 2W^2 + 2(p+q+1) < 0$$

or <u>vice versa</u>, where

$$(2.13) \quad W = \frac{Q_{p+q}}{Q_p} = \left[1 + \frac{q}{n-p-q} F\right]^{-1}$$

Although the BIC critical point cannot be explicitly written as a function of n, p, and q, (2.12) is equivalent to the inequality that F is less than some constant depending on n, p, and q. Asymptotically, the critical point approaches 2, and hence the BIC is equivalent to the AIC.[3/]

It is also straightforward to see that the decision based on $\bar{R}$ is also equivalent to the preliminary F test with a constant critical point equalling one.

In what follows we confine our discussion to a class of decision rules based on the F statistic: choose M_1 if $F \leq c$ and choose M_2 if $F > c$, where c is nonnegative and varies with n, p, and q.

Following Mallows (1973), we choose

$$(2.14) \quad W_p = E \left|\left| \underset{\sim}{Y}_0 - \hat{\underset{\sim}{Y}} \right|\right|^2$$

as a risk function of the model M_1, where $\underset{\sim}{Y}_0$ is an n-dimensional random vector, independent of $\underset{\sim}{Y}$ but distributed with the same mean $\underset{\sim}{\mu}$ and variance-covariance matrix $\omega^2 \underset{\sim}{I}$ as $\underset{\sim}{Y}$, and $\hat{\underset{\sim}{Y}} = \underset{\sim}{X} (\underset{\sim}{X}'\underset{\sim}{X})^{-1} \underset{\sim}{X}'\underset{\sim}{Y}$. By virtue of the independence assumed between $\underset{\sim}{Y}$ and $\underset{\sim}{Y}_0$, we have

$$(2.15) \quad W_p = \underset{\sim}{\mu}' \left[\underset{\sim}{I} - \underset{\sim}{X}_1 (\underset{\sim}{X}_1'\underset{\sim}{X}_1)^{-1} \underset{\sim}{X}_1'\right] \underset{\sim}{\mu} + (n + p)\, \omega^2$$

$$= [\lambda_1 + \lambda_2 + (n + p)]\, \omega^2$$

Similarly, the risk function of the model M_2 is

$$(2.16) \quad W_{p+q} = \underset{\sim}{\mu}' \left[\underset{\sim}{I} - \underset{\sim}{X} (\underset{\sim}{X}'\underset{\sim}{X})^{-1} \underset{\sim}{X}'\right] \underset{\sim}{\mu} + (n+p+q)\, \omega^2$$

$$= [\lambda_2 + (n+p+q)]\, \omega^2$$

If $\hat{\omega}^2$ is an unbiased estimate of ω^2 independent of Y, then we can unbiasedly estimate W_p and W_{p+q} by C_p and C_{p+q}, because $E(Q_p) = (\lambda_1 + \lambda_2 + n - p)\, \omega^2$ and $E(Q_{p+q}) = (\lambda_2 + n - p - q)\, \omega^2$.

Based on the Mallows' risk, we say that the simpler model M_1 is better than M_2 if $W_p < W_{p+q}$ and <u>vice versa</u>. That is, we choose M_1 if

$$(2.17) \quad \lambda_1 \leq q$$

and choose M_2 if

$$(2.18) \quad \lambda_1 > q$$

Now we are in a position to define the <u>unbiasedness</u> of a decision rule.

Definition: A decision rule based on the F statistic with critical point c is said to be unbiased if

(2.19) $\Pr(F < c \mid \lambda_1 \leq q) \geq .5$

and

(2.20) $\Pr(F > c \mid \lambda_1 > q) \geq .5.$

If either of the above conditions is not fulfilled, the decision rule is said to be biased.

Noting that λ_1 is the noncentrality parameter of the numerator of F and also that the F distribution is continuous, we realize that the above conditions are equivalent to a single equality

(2.21) $\Pr(F < c \mid \lambda_1 = q) = .5$

which implies that the probability of selecting M_1 (or M_2) is .5 when M_1 and M_2 are indifferent.

The unbiased critical point c satisfying (2.21) depends on the noncentrality parameter λ_2 of the denominator as well as the degrees of freedom. The parameter λ_2 measures the discrepancy of M_2 from the true distribution. Since there is no way of estimating λ_2, we confine ourselves to the case when $\lambda_2 = 0$, i.e., when the more complicated model M_2 is true. The unbiased critical points are computed and tabulated in Table 2.1 for various values of q and n p-q. To solve the equation (2.21) we made use of the algorithm developed by Mudholkar, et al. (1976) for Gram-Charlier series approximation of the noncentral F distribution. The approximation is accurate enough for the present purpose. To compare the result with the conventional preliminary F test with a fixed significance level we computed the probability of F exceeding the unbiased critical point under the null-hypothesis that M_1 is true. The results are tabulated in Table 2.2.

The following points are observed from these tables. First, the unbiased critical point increases with q and decreases with the degrees of freedom. The MAIC and Mallows' procedures are biased toward the simpler model; i.e., they are favorable to the model with less parameters unless we have the extremely small degrees of freedom associated with the large number of additional parameters. Secondly, the implied significance level varies over a fairly wide range. This may imply that a decision rule with fixed level of significance may not be recommended as far as unbiasedness is concerned. Thirdly, the decision rule based on $\bar{R}$ is nearly unbiased when $q = 1$. Therefore, a decision rule based on $\bar{R}$ has a desirable property of unbiasedness if the decision is concerned with one additional variable.

3. Approximate Formula

In this section we will derive the approximate formula for the unbiased critical point. Our decision has been based on the statistic F defined by (2.8), the distribution of which is the doubly noncentral F with q and ν (= n-p-q) degrees of freedom and noncentrality parameters λ_1 and λ_2. If we apply Patnaik's (1949) central χ^2 approximation to each of the noncentral χ^2 variates, we may approximate

$$(3.1) \qquad \frac{1 + \rho_2}{1 + \rho_1} F, \text{ where } \rho_1 = \frac{\lambda_1}{q} \text{ and } \rho_2 = \frac{\lambda_2}{\nu}$$

by a central F distribution with degrees of freedom

$$(3.2) \qquad \nu_1 = \frac{q(1 + 2\rho_1)}{(1 + \rho_1)^2}, \quad \nu_2 = \frac{\nu(1 + 2\rho_2)}{(1 + \rho_2)^2}.$$

Now we have an approximate equality

$$(3.3) \qquad P(F \leq c) \simeq P\left[U = \left(\frac{\chi_1^2}{\nu_1}\right)^{1/3} - c^{1/3}\left(\frac{1+\rho_2}{1+\rho_1}\right)^{1/3}\left(\frac{\chi_2^2}{\nu_2}\right)^{1/3} < 0\right].$$

where χ_1^2 and χ_2^2 are independent χ^2 variate ρ with ν_1 and ν_2 degrees of freedom, respectively. The cubic root of the χ^2 variate is approximately normally distributed, and hence we approximate the distribution of U by the normal distribution. The approximate mean of U is given by

$$(3.4) \qquad \mu_U \simeq \left(1 - \frac{2}{9\nu_1}\right) - c^{1/3}\left(\frac{1+\rho_2}{1+\rho_1}\right)^{1/3}\left(1 - \frac{2}{9\nu_2}\right)$$

where the error of approximation is of order $O(\nu_1^{-2})$ and $O(\nu_2^{-2})$. Since the median of the normal distribution is identical with the mean, the right hand side of (3.3) is approximately equal to 0.5 when $\mu_U = 0$, i.e.,

$$(3.5) \qquad c = \left\{\frac{1 - 2/(9\nu_1)}{1 - 2/(9\nu_2)}\right\}^3$$

$$= \frac{2}{1+\rho_2}\left\{\frac{1 - 1/(6q)}{1 - 2(1+2\rho_2)/[9\nu(1+\rho_2)^2]}\right\}^3$$

The value of c given by (3.5) is monotone decreasing in ρ_2, and when $\rho_2 = 0$ it is equal to

$$c = 2\left\{\frac{1 - 1/(6q)}{1 - 2/(9\nu)}\right\},$$

which may be regarded as the approximate formula for the unbiased critical points tabulated in Table 2.1.

4. Conclusion

Various decision rules have been developed and used for the model selection in practical regression analysis. Some of them are based on intuitive reasoning, while others on sophisticated theoretical frameworks.

In any case, as long as alternative models are nested, all the decision rules, considered in this paper, reduce to a decision based on the magnitude of the conventionally used F statistic. The difference among them is, therefore, described in terms of "implied" critical points.

To compare various decision rules we have introduced a certain criterion of optimality, which is termed the unbiasedness of a decision rule. The critical point for the F test that leads us to the unbiased decision has been calculated for various combinations of the degrees of freedom and the number of variables. It turned out that the decision rules which are customarily used are more or less biased; that is, the MAIC and Mallows' rules are biased toward the simpler model, whereas the minimum $\bar{R}$ rule is rather biased toward the more complex model.

Table 2.1. Unbiased Critical Point

ν/q	1	2	3	4	5	6	7	8	9	10
5	1.261	1.754	1.932	2.022	2.076	2.113	2.139	2.159	2.174	2.187
6	1.232	1.714	1.888	1.975	2.028	2.064	2.089	2.108	2.123	2.135
7	1.213	1.687	1.858	1.944	1.995	2.030	2.055	2.073	2.088	2.099
8	1.198	1.667	1.835	1.920	1.971	2.005	2.029	2.048	2.062	2.074
9	1.187	1.651	1.819	1.903	1.953	1.986	2.010	2.028	2.042	2.054
10	1.178	1.639	1.805	1.889	1.938	1.972	1.995	2.013	2.027	2.038
11	1.171	1.629	1.795	1.877	1.927	1.960	1.983	2.001	2.015	2.026
12	1.165	1.621	1.786	1.868	1.917	1.950	1.973	1.991	2.004	2.015
13	1.160	1.614	1.778	1.860	1.909	1.942	1.965	1.982	1.996	2.007
14	1.156	1.608	1.772	1.854	1.902	1.935	1.958	1.975	1.989	1.999
15	1.152	1.603	1.766	1.848	1.896	1.929	1.952	1.969	1.982	1.993
16	1.149	1.599	1.762	1.843	1.891	1.923	1.946	1.963	1.977	1.988
17	1.147	1.595	1.758	1.839	1.887	1.919	1.942	1.959	1.972	1.983
18	1.144	1.592	1.754	1.835	1.883	1.915	1.938	1.955	1.968	1.979
19	1.142	1.589	1.751	1.831	1.879	1.911	1.934	1.951	1.964	1.975
20	1.140	1.586	1.748	1.828	1.876	1.908	1.931	1.948	1.961	1.971
21	1.138	1.584	1.745	1.825	1.873	1.905	1.928	1.945	1.958	1.968
22	1.137	1.581	1.743	1.823	1.871	1.902	1.925	1.942	1.955	1.966
23	1.135	1.579	1.740	1.821	1.868	1.900	1.923	1.939	1.952	1.963
24	1.134	1.577	1.738	1.818	1.866	1.898	1.920	1.937	1.950	1.961
25	1.133	1.576	1.737	1.817	1.864	1.896	1.918	1.935	1.948	1.959
26	1.131	1.574	1.735	1.815	1.862	1.894	1.916	1.933	1.946	1.957
27	1.130	1.573	1.733	1.813	1.861	1.892	1.915	1.931	1.944	1.955
28	1.129	1.571	1.732	1.812	1.859	1.891	1.913	1.930	1.943	1.953
29	1.128	1.570	1.731	1.810	1.858	1.889	1.912	1.928	1.941	1.952
30	1.128	1.569	1.729	1.809	1.856	1.888	1.910	1.927	1.940	1.950
40	1.122	1.561	1.720	1.800	1.847	1.878	1.900	1.917	1.930	1.940
50	1.118	1.556	1.715	1.794	1.841	1.872	1.894	1.911	1.924	1.934
60	1.116	1.552	1.711	1.790	1.837	1.868	1.890	1.907	1.920	1.930
70	1.114	1.550	1.709	1.788	1.834	1.865	1.888	1.904	1.917	1.927
80	1.113	1.548	1.707	1.786	1.832	1.863	1.885	1.902	1.915	1.925
90	1.112	1.547	1.705	1.784	1.831	1.862	1.884	1.900	1.913	1.923
100	1.111	1.546	1.704	1.783	1.830	1.861	1.883	1.899	1.912	1.922
200	1.107	1.541	1.699	1.777	1.824	1.855	1.877	1.893	1.906	1.916
300	1.106	1.539	1.697	1.776	1.822	1.853	1.875	1.891	1.904	1.914
400	1.106	1.538	1.696	1.775	1.821	1.852	1.874	1.890	1.903	1.913
500	1.105	1.538	1.696	1.774	1.821	1.852	1.873	1.890	1.902	1.913
600	1.105	1.537	1.695	1.774	1.820	1.851	1.873	1.889	1.902	1.912
700	1.105	1.537	1.695	1.774	1.820	1.851	1.873	1.889	1.902	1.912
800	1.105	1.537	1.695	1.773	1.820	1.851	1.873	1.889	1.902	1.912
900	1.105	1.537	1.695	1.773	1.820	1.851	1.872	1.889	1.901	1.912
1000	1.105	1.537	1.695	1.773	1.820	1.850	1.872	1.889	1.901	1.911

Table 2.2. Significance Level Implied by the Unbiased Critical Point

ν/q	1	2	3	4	5	6	7	8	9	10
5	0.313	0.265	0.243	0.230	0.221	0.215	0.210	0.206	0.203	0.201
6	0.309	0.258	0.233	0.218	0.207	0.200	0.194	0.190	0.186	0.183
7	0.307	0.253	0.225	0.208	0.196	0.188	0.181	0.176	0.172	0.169
8	0.306	0.249	0.219	0.201	0.188	0.178	0.171	0.165	0.161	0.157
9	0.304	0.246	0.214	0.194	0.181	0.170	0.163	0.156	0.151	0.147
10	0.303	0.243	0.210	0.189	0.175	0.164	0.155	0.149	0.143	0.138
11	0.302	0.241	0.207	0.185	0.169	0.158	0.149	0.142	0.136	0.131
12	0.302	0.239	0.204	0.181	0.165	0.153	0.144	0.136	0.130	0.125
13	0.301	0.238	0.201	0.178	0.161	0.149	0.139	0.131	0.125	0.119
14	0.300	0.237	0.199	0.175	0.158	0.145	0.135	0.127	0.120	0.115
15	0.300	0.235	0.197	0.172	0.155	0.142	0.131	0.123	0.116	0.111
16	0.300	0.234	0.195	0.170	0.152	0.139	0.128	0.119	0.112	0.106
17	0.299	0.234	0.194	0.168	0.150	0.136	0.125	0.116	0.109	0.103
18	0.299	0.233	0.193	0.166	0.148	0.133	0.122	0.113	0.106	0.101
19	0.299	0.232	0.191	0.165	0.146	0.131	0.120	0.111	0.103	0.097
20	0.298	0.231	0.190	0.163	0.144	0.129	0.118	0.108	0.101	0.094
21	0.298	0.231	0.189	0.162	0.142	0.127	0.116	0.106	0.098	0.092
22	0.298	0.230	0.188	0.161	0.141	0.126	0.114	0.104	0.096	0.090
23	0.298	0.230	0.187	0.160	0.139	0.124	0.112	0.102	0.094	0.088
24	0.298	0.229	0.187	0.159	0.138	0.123	0.111	0.101	0.093	0.086
25	0.297	0.229	0.186	0.158	0.137	0.121	0.109	0.099	0.091	0.084
26	0.297	0.228	0.185	0.157	0.136	0.120	0.108	0.098	0.089	0.082
27	0.297	0.228	0.185	0.156	0.135	0.119	0.106	0.096	0.088	0.081
28	0.297	0.228	0.184	0.155	0.134	0.118	0.105	0.095	0.087	0.080
29	0.297	0.227	0.183	0.154	0.133	0.117	0.104	0.094	0.085	0.078
30	0.297	0.227	0.183	0.154	0.132	0.116	0.103	0.093	0.084	0.077
40	0.296	0.225	0.179	0.148	0.126	0.109	0.095	0.084	0.075	0.068
50	0.295	0.224	0.177	0.145	0.122	0.104	0.090	0.079	0.070	0.062
60	0.295	0.223	0.175	0.143	0.120	0.101	0.087	0.076	0.066	0.058
70	0.295	0.222	0.174	0.142	0.118	0.099	0.085	0.073	0.064	0.056
80	0.295	0.222	0.173	0.141	0.116	0.098	0.083	0.071	0.062	0.054
90	0.295	0.221	0.173	0.140	0.115	0.096	0.082	0.070	0.060	0.052
100	0.294	0.221	0.172	0.139	0.114	0.095	0.080	0.068	0.059	0.051
200	0.294	0.220	0.170	0.136	0.110	0.091	0.075	0.063	0.053	0.045
500	0.294	0.219	0.169	0.134	0.109	0.089	0.074	0.061	0.051	0.043
400	0.294	0.219	0.169	0.134	0.108	0.088	0.073	0.060	0.050	0.042
500	0.294	0.219	0.168	0.133	0.108	0.088	0.072	0.060	0.050	0.041
600	0.294	0.219	0.168	0.133	0.107	0.087	0.072	0.059	0.049	0.041
700	0.294	0.219	0.168	0.133	0.107	0.087	0.072	0.059	0.049	0.041
800	0.294	0.219	0.168	0.133	0.107	0.087	0.071	0.059	0.049	0.041
900	0.294	0.219	0.168	0.133	0.107	0.087	0.071	0.059	0.049	0.040
1000	0.294	0.219	0.168	0.133	0.107	0.087	0.071	0.059	0.048	0.040

Footnotes

1. See Sawa (1976).
2. In Sawa (1976) β_0 is denoted by the pseudo-true value of β.
3. For a small sample typically dealt with in practical data analysis, the difference between the AIC and BIC critical points is far from negligible.

References

Akaike, H. (1974) "A New Look at the Statistical Model Identification," IEEE Transactions on Automatic Control, Vol. AC-19, No. 6, pp. 716-723.

Aty, A. S. H. (1954) "Approximate Formulae for the Percentage Points and the Probability Integral of the Non-central χ^2 Distribution," Biometrika, 41, 538-540.

Mallows, C. L. (1973) "Some Comments on C_p," Technometrics, Vol. 15, pp. 661-675.

Mudholkar, G. S., Y. P. Chaubey and Ching-Chuong Lin (1976) "Some Approximations for the Noncentral-F Distribution," Technometrics, Vol. 18, No. 351-358.

Patnaik, P. B. (1949) "The Noncentral χ^2 and F-Distributions and their Applications," Biometrika, 36, 202-232.

Sawa, T. (1976) "Information Criteria for the Choice of Regression Models".