MONASH

WP 15/98

ISSN 1440-771X
ISBN 0 7326 1052 4

MONASH UNIVERSITY

ANCORA · IMPARO

Model Selection When a Key Parameter is Constrained to be in an Interval

Md. Zakir Hossain and Maxwell L. King

Working Paper 15/98
October 1998

DEPARTMENT OF ECONOMETRICS
AND BUSINESS STATISTICS

Model Selection When a Key Parameter is Constrained to be in an Interval

Md. Zakir Hossain and Maxwell L. King

Department of Econometrics and Business Statistics, Monash University
Clayton, Victoria 3168, Australia

Abstract

This paper considers the construction of model selection procedures based on choosing the model with the largest maximised log-likelihood minus a penalty, when key parameters are restricted to be in a closed interval. The approach adopted is based on King *et al.*'s (1995) representative models method with the use of the parametric bootstrap to handle nuisance parameters. The method is illustrated by application to two model selection problems in the context of Box-Cox transformations and the linear regression model. Simulation results for both problems indicate that the new procedure clearly dominates existing procedures in terms of having higher probabilities of correctly selecting the true model.

Keywords: Box-Cox transformations; Controlled information criterion; Nuisance parameters; Parametric bootstrap; Regression model.

1. Introduction

Often in statistics and particularly in disciplines such as econometrics which require non-experimental data sets to be modelled, we are forced to use the data to make a choice of model specification from a range of possibilities. One approach often used is to apply a series of pairwise hypothesis tests to help make the choice. As Granger *et al.* (1995) and others have observed, this has a number of limitations. An alternative and increasingly popular approach favoured by Granger *et al.* is to use a model selection procedure based on an information criterion (IC). Typically this involves choosing the model specification with the largest maximised log-likelihood function minus a penalty which, among other things, reflects the number of estimated parameters in the model. Unfortunately, there is little agreement about what this penalty should be. For example, for Akaike's (1973) IC (AIC) it is q, for Schwartz's (1978) Bayesian IC (BIC) it is $q\log(n)/2$, while for Hannan and Quinn's (1979) IC it is $q\log(\log(n))$ where q is the number of estimated parameters in the model and n is the sample size.

While there is a large and growing literature that gives different suggestions on what this penalty should be in a range of situations, very little has been written on how it should be calculated in the presence of inequality restrictions on important parameters. This is in contrast to the vast and growing literature on testing in the presence of such restrictions. It is somewhat surprising given that, as Pötscher (1991) points out, using an IC procedure amounts to testing each model against all other models by means of likelihood ratio tests and selecting that model which is accepted against all other models where the critical values are determined by the penalty terms of the criterion. Just as the likelihood ratio test requires different critical values when used in the presence of inequality retrictions, one would expect IC procedures to have different penalties when the restrictions involve parameters in dispute.

Hughes and King (1994, 1998) and Hughes (1997) were perhaps the first to propose model selection procedures that use one-sided information on parameters in dispute. Based on asymptotic arguments, they developed a class of one-sided AIC procedures called OSAIC. Hossain and King (1998) (also see Hossain (1998)) modified AIC for the case where some parameters of interest are restricted to an interval and proposed a new procedure called PAIC. They also suggested an analogous modification to BIC which they called PBIC. A range of simulation experiments were conducted involving the Box-Cox (1964) transformation (BCT) on the dependent variable in the linear regression model, in which the Box-Cox parameter, λ, is restricted to be in the range [0,1]. $\lambda = 0$ gives the log-linear model while $\lambda = 1$ gives the classical linear model. They found the probabilities of correct selection (PCS) for this set of three competing models vary significantly from one design matrix to another, being undesirably small in some cases while being extremely large in other cases. Consequently, the idea of controlling PCS or alternatively, minimising the variation in PCS seems worthy of investigation.

In the context of hypothesis testing, it is well known that critical values are calculated in order to control the trade-off between the size and power of the test by fixing the PCS for the null hypothesis model. King *et al.* (1995) argued that this principle can be usefully applied to the problem of model selection. They proposed a general model selection procedure in which the penalties are calculated by controlling the probabilities of selection in such a way that no one model is unnecessarily favoured. To do this, they suggested two approaches; one called the common model approach which involves controlling probabilities of selection for a common minimal model (usually with all parameters under dispute set to zero) and the other called the representative fixed points approach. The latter involves deciding on representative parameter values for each of the models and determining the penalties to make the PCS for these representative models equal. The philosophy behind this approach is similar to that for point optimal testing (see King (1987)). King *et al.* also presented a general algorithm for calculating penalties for both approaches using simulation methods.

The aim of this paper is to investigate whether King *et al.'s* fixed point approach can be successfully applied to selection problems in which parameters of interest are restricted to a closed interval. In particular we focus on selection problems involving the BCT regression model in order to see whether we can improve on Hossain and King's (1998) PAIC and PBIC procedure in small samples.

The plan of the paper is as follows. In Section 2 we illustrate our method of applying King *et al.*'s representative fixed points approach by discussing its application to a simple BCT regression model. The use of the parametric bootstrap to handle nuisance parameters is explained. The section closes with an outline of the algorithm for calculating the penalties. The illustration is extended to a more complicated model selection problem involving two restricted parameters and nine possible models in Section 3. A detailed algorithm for calculating the penalties in this case is presented. Two Monte Carlo experiments, designed to compare the small sample properties of the new procedure with those of existing procedures for the model selection problems outlined in Sections 2 and 3, are reported in Section 4. We find that the new procedure clearly dominates existing procedures in terms of having higher PCS. Some concluding remarks are made in the final section.

2. The Proposed Procedure for One Parameter Restriction

In this section, we illustrate our approach by discussing its application to a simple model selection problem in the context of the BCT regression model

$$y_t(\lambda) = \beta_0 + \beta_1 x_t + u_t, \qquad t = 1,\ldots,n, \tag{1}$$

where β_0 and β_1 are regresssion coefficients, x_t is a single nonstochastic regressor and $y_t(\lambda)$ is defined as

$$\begin{aligned} y_t(\lambda) &= \frac{y_t^\lambda - 1}{\lambda} \quad \text{when } 0 < \lambda < 1, \\ &= \log(y_t) \quad \text{when } \lambda = 0, \\ &= y_t - 1 \quad \text{when } \lambda = 1. \end{aligned} \tag{2}$$

u_t is the disturbance term which, in theory, has a truncated distribution and therefore cannot be normally distributed. We assume that the truncation effect is insignificant and that

$$u_t \sim IN(0,\sigma^2), \qquad t = 1,\ldots,n.$$

From (1) and (2) we obtain three possible models, the first (denoted model-1) being

$$y_t = \beta_0^* + \beta_1 x_t + u_t,$$

when $\lambda = 1$, where $\beta_0^* = \beta_0 + 1$. The second model (model-2) is

$$\log(y_t) = \beta_0 + \beta_1 x_t + u_t$$

and the third model (model-3) is

$$(y_t^{\lambda} - 1)/\lambda = \beta_0 + \beta_1 x_t + u_t$$

where $0 < \lambda < 1$. Observe that λ is our parameter of interest which we assume is restricted to lie in the interval [0,1]. The literature on the BCT does discuss the possibility of $\lambda < 0$ and $\lambda > 1$. Many researchers assume $\lambda \in [-1,1]$ although Spitzer (1978) pointed out that this model can result in poor forecasts when $\lambda \in [-1,0]$. Econometricians often have to decide whether to use a log-linear or a classical linear regression model. The BCT model with $\lambda \in (0,1)$ provides a range of models that cover the parameter space between these two alternative models. These arguments, taken together, make a case for sometimes restricting λ to the range [0,1]. Our challenge is to find an effective model selection procedure for the situation in which one may wish to impose this restriction and therefore choose between Model-1, Model-2 and Model-3.

Denote the log-likelihood function for the i^{th} model by $L_i(\theta)$ where θ is the parameter vector. $L_i(\theta)$ is well-known for i = 1,2 and $\theta = (\beta_o^*, \beta_1, \sigma^2)'$ for model-1 and $\theta = (\beta_0, \beta_1, \sigma^2)'$ for model-2 can be estimated via ordinary least squares. For model-3, $\theta = (\beta_0, \beta_1, \sigma^2, \lambda)'$ and

$$L_3(\theta) = \frac{-n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{n}\left(y_t(\lambda) - \beta_0 - \beta_1 x_t\right)^2 + (\lambda - 1)\sum_{t=1}^{n}\log(y_t) \ .$$

Let $L_i(\hat{\theta})$ denote the maximized value of $L_i(\theta)$. Then, because model-1 and model-2 involve 3 parameters while model-3 involves 4 parameters, AIC (BIC) chooses the model corresponding to the largest value of

$$L_1(\hat{\theta})-3, \qquad L_2(\hat{\theta})-3, \qquad L_3(\hat{\theta})-4$$
$$\left(L_1(\hat{\theta})-1.5\log(n), \quad L_2(\hat{\theta})-1.5\log(n), \quad L_3(\hat{\theta})-2\log(n)\right).$$

In contrast, Hossain and King's (1998) PAIC (and PBIC) chooses the model corresponding to the largest value of

$$L_1(\hat{\theta})-3, \qquad L_2(\hat{\theta})-3, \qquad L_3(\hat{\theta})-3.5$$
$$\left(L_1(\hat{\theta})-1.5\log(n), \quad L_2(\hat{\theta})-1.5\log(n), \quad L_3(\hat{\theta})-1.75\log(n)\right).$$

The first step in applying King *et al.'s* (1995) representative fixed points approach involves choosing representative parameter values for each of the three models. Given that the model selection problem boils down to choosing between $\lambda = 1$ (model-1), $\lambda = 0$ (model-2) or $0 < \lambda < 1$ (model-3); it is not hard to select $\lambda = 1$, $\lambda = 0$ and $\lambda = 0.5$ as the representative values of λ for the respective models. For each model, there are three nuisance parameters namely β_0 (or equivalently β_0^* for model-1), β_1 and σ^2. Our suggestion for fixing values for these parameters is to employ the solution used in the parametric bootstrap. This involves calculating maximum likelihood estimates of the nuisance parameters using the data at hand and then using these estimates, together with the representative value of λ, to generate a large number of simulated data sets for each model. Just as the Monte Carlo method can be used to find critical values that solve the size equation in hypothesis testing, these sets of simulated data samples can be used to find penalty function values that make the simulated probabilities of correctly selecting each model (at its representative parameter values) equal. By making the probabilities of correct selection equal, no particular model is favoured.

Let p_i be the penalty associated with $L_i(\hat{\theta})$ so that model-i is chosen if

$$L_i(\hat{\theta}) - p_i > L_j(\hat{\theta}) - p_j \qquad \text{for} \quad j = 1,\ 2,\ 3 \text{ and } j \neq i. \tag{3}$$

Because the comparisons made in (3) depend only on penalty differences, namely $p_j - p_i$, we can set $p_1 = 0$ without loss of generality. The details of our approach for finding the remaining penalty values, p_2 and p_3, are as follows.

Given the data (n observations of y_t) to be used to decide between the three models, we first estimate θ under each of model-1, model-2 and model-3. The respectively estimated values of β_0, β_1 and σ^2 together with $\lambda = 1$ for model-1, $\lambda = 0$ for model-2 and $\lambda = 0.5$ for model-3, are used to generate 500 simulated sets of samples of size n for each of the three models. The maximized log-likelihood, $L_i(\hat{\theta})$, for each of the three competing models is calculated for each simulated sample of size n. This results in nine sets of 500 values of $L_i(\hat{\theta})$. The nine sets are made up of three sets (one for each true model) of three (one for each log-likelihood fitted). For each set of three fitted log-likelihoods, one represents the true model. This can be used to estimate the probability, for any value of p_2 and p_3, of choosing model-i when it is true based on (3) for i = 1, 2, 3. Denote these estimated probabilities which depend on the values of p_2 and p_3 used, by

$$P_1(p_2, p_3),\ P_2(p_2, p_3) \text{ and } P_3(p_2, p_3)$$

for when model-1, model-2 and model-3 are respectively the true model.

The penalties p_2 and p_3 are then found by an iterative method so that they solve

$$P_1(p_2, p_3) = P_2(p_2, p_3) = P_3(p_2, p_3) \ . \tag{4}$$

Observe that (4) represents two equations in two unknowns, namely p_2 and p_3. To solve (4) in the Monte Carlo experiments outlined in Section 4, we used the following iterative method. Choose possible values for p_2 and p_3. Calculate $P_1(p_2, p_3)$, $P_2(p_2, p_3)$ and $P_3(p_2, p_3)$ for these values. First adjust p_3 to make

$$P_1(p_2, p_3) = P_3(p_2, p_3) \ . \tag{5}$$

If $P_1(p_2, p_3) > P_3(p_2, p_3)$, then reduce p_3 otherwise increase p_3 until (5) holds (for a specified level of tolerance). Then check if

$$P_1(p_2, p_3) = P_2(p_2, p_3) \ . \tag{6}$$

If $P_1(p_2, p_3) > P_2(p_2, p_3)$ then reduce p_2 otherwise increase p_2. Each time p_2 is changed, one has to find the value of p_3 that makes (5) hold. This process is repeated until both (5) and (6) hold. We have then found the required p_2 and p_3 values which can be used for the original data and (3) to choose a model.

We call our new procedure controlled IC (CIC).

3. Our Proposed Procedure for Two Parameter Restrictions

In this section, we discuss the application of CIC to a more complicated model selection problem that involves two restricted parameters and nine competing models. The larger number of candidate models makes the procedure more time consuming and difficult to apply. We shall assess in Section 4 whether this extra effort is worthwhile.

Consider the following two parameter version of a BCT regression model,

$$y_t(\lambda_1) = \beta_0 + \beta_1 x_t(\lambda_2) + u_t, \qquad t = 1,\dots,n, \tag{7}$$

where it is assumed that $u_t \sim IN(0,\sigma^2)$. The notation used in (7) is as defined in (1) with $y_t(\lambda_1)$ given by (2) with $\lambda = \lambda_1$ and $x_t(\lambda_2)$ defined by (2) with $\lambda = \lambda_2$ and $y_t = x_t$. Again, for the reasons discussed in Section 2, we restrict λ_1 and λ_2 to the closed interval [0,1].

For this model, there are nine possible combinations of boundary/non-boundary values of the two key parameters, λ_1 and λ_2. For these nine possibilities, we obtain nine models to choose between. They are

M1 : $y_t = \beta_0 + \beta_1 x_t + u_t$,

M2 : $y_t = \beta_0 + \beta_1 \log(x_t) + u_t$,

M3 : $\log(y_t) = \beta_0 + \beta_1 x_t + u_t$,

M4 : $\log(y_t) = \beta_0 + \beta_1 \log(x_t) + u_t$,

M5 : $y_t = \beta_0 + \beta_1 \left(x_t^{\lambda_2} - 1\right)/\lambda_2 + u_t$, $\qquad 0 < \lambda_2 < 1$, M6 : $\log(y_t) = \beta_0 + \beta_1\left(x_t^{\lambda_2} - 1\right)/\lambda_2 + u_t$, $\quad 0 < \lambda_2 < 1$,

M7 : $\left(y_t^{\lambda_1} - 1\right)/\lambda_1 = \beta_0 + \beta_1 x_t + u_t$, $\quad 0 < \lambda_1 < 1$,

M8 : $\left(y_t^{\lambda_1} - 1\right)/\lambda_1 = \beta_0 + \beta_1 \log x_t + u_t$, $\quad 0 < \lambda_1 < 1$,

M9 : $\left(y_t^{\lambda_1} - 1\right)/\lambda_1 = \beta_0 + \beta_1\left(x_t^{\lambda_2} - 1\right)/\lambda_2 + u_t$, $\quad 0 < \lambda_1 < 1$, $0 < \lambda_2 < 1$, $t = 1,\ldots,n$.

Of these nine models, the first four contain three parameters each and therefore have AIC and BIC penalties of 3 and 1.5log(n), respectively. Because no restricted parameters are involved, the penalties are the same for Hossain and King's (1998) PAIC and PBIC, respectively. On the other hand, the second four models each contain four parameters, one of which is restricted to the open interval (0,1). The usual AIC and BIC penalties for these models are 4 and 2log(n), respectively, while PAIC and PBIC have penalties of 3.5 and 1.75log(n), respectively. The last model, M9, has five parameters, two of which are restricted to be in the range (0,1). The penalties for PAIC and PBIC are now

$$4 + \left(\sin^{-1}\rho_{12}\right)/\pi \text{ and } \left[2 + \left(\sin^{-1}\rho_{12}\right)/(2\pi)\right]\log(n),$$

respectively, where ρ_{12} is the correlation between the maximum likelihood estimates of λ_1 and λ_2 from M9.

Again the first step in applying King *et al.*'s (1995) representative fixed points approach involves choosing representative parameter values for each of the nine models. In this case, our model selection problem boils down to choosing between $(\lambda_1, \lambda_2) = (1,1), (1,0), (0,1), (0,0), \left(1, \lambda_2^*\right), \left(0, \lambda_2^*\right), \left(\lambda_1^*, 1\right), \left(\lambda_1^*, 0\right), \left(\lambda_1^*, \lambda_2^*\right)$, respectively, for

M1 - M9, where λ_1^* and λ_2^* are both restricted to be in the open interval (0,1). Obvious representative values for (λ_1, λ_2) for models M1 - M9 are (1,1), (1,0), (0,1), (0,0), (1, 0.5), (0,0.5), (0.5,1), (0.5,0) and (0.5,0.5), respectively.

Let $L_i(\varphi)$ be the log-likelihood for model Mi, $i = 1,\ldots,9$, and let φ be its associated parameter vector. If $L_i(\hat{\varphi})$ denotes the maximized value of $L_i(\varphi)$ and p_i denotes the penalty associated with $L_i(\hat{\varphi})$, then our procedure involves setting $p_1 = 0$ and finding $p_2,\ldots,p_9$. Then model Mi is chosen if

$$L_i(\hat{\varphi}) - p_i > L_j(\hat{\varphi}) - p_j \quad \text{for } j = 1,\ldots,9 \text{ and } j \neq i \, . \tag{8}$$

Our procedure finds values for $p_2,\ldots,p_9$ as follows.

Given the n observations of y_t to be used to choose between the nine models, we first obtain maximum likelihood estimates of φ for each of M1, M2, ..., M9. For each model, the estimated values of β_0, β_1 and σ^2 together with the corresponding representative values of λ_1 and λ_2 are used to generate 200 simulated samples of size n. The maximized log-likelihood, $L_i(\hat{\varphi})$, for each of the nine competing models is calculated for each simulated sample of size n. We therefore end up with 81 $(= 9 \times 9)$ sets of 200 values of $L_i(\hat{\varphi})$. We have nine sets (one for each true model) of nine (one for each fitted log-likelihood). These 81 sets of 200 values of $L_i(\hat{\varphi})$ can be used to estimate the probability of correctly choosing model Mi based on (8) for any values of $p_2,\ldots,p_9$. Denote these estimated probabilities by

$$P_i(p_2,\ldots,p_9) \text{ when model M}i \text{ is true,} \qquad i = 1,\ldots,9 \, .$$

The penalties $p_2,\ldots,p_9$ are then found by an iterative method (similar to that outlined in Section 2) so that they solve

$$P_1(p_2,\ldots,p_9) = P_2(p_2,\ldots,p_9) = \ldots = P_9(p_2,\ldots,p_9) \ . \tag{9}$$

Note that (9) is eight equations in eight unknowns, which has to be solved to find $p_2,\ldots,p_9$. It can be solved using a generalization of the iterative method outlined in Section 2 as follows:

1. Choose possible values for $p_2,\ldots,p_9$.

2. Calculate $P_i(p_2,\ldots,p_9), \quad i = 1,\ldots,9$.

3. Check (to a specified tolerance) if

$$P_1(p_2,\ldots,p_9) = P_9(p_2,\ldots,p_9) \ . \tag{10}$$

   If (10) holds, proceed to step 4, otherwise if $P_1(p_2,\ldots,p_9) - P_9(p_2,\ldots,p_9)$ is positive (negative), reduce (increase) p_9 and return to step 2.

4. Check if

$$P_1(p_2,\ldots,p_9) = P_8(p_2,\ldots,p_9) \ . \tag{11}$$

   If (11) holds, proceed to step 5, otherwise if $P_1(p_2,\ldots,p_9) - P_8(p_2,\ldots,p_9)$ is positive (negative), reduce (increase) p_8 and return to step 2.

5. Check if

$$P_1(p_2,\ldots,p_9) = P_7(p_2,\ldots,p_9) \ . \tag{12}$$

If (12) holds, proceed to step 6, otherwise if $P_1(p_2,\ldots,p_9)-P_7(p_2,\ldots,p_9)$ is positive (negative), reduce (increase) p_7 and return to step 2.

6. Check if

$$P_1(p_2,\ldots,p_9)=P_6(p_2,\ldots,p_9) \, . \tag{13}$$

.
.
.
.
.

10. Check if

$$P_1(p_2,\ldots,p_9)=P_2(p_2,\ldots,p_9) \, . \tag{14}$$

If (14) holds, then the required values of $p_2,\ldots,p_9$ have been found, otherwise if $P_1(p_2,\ldots,p_9)-P_2(p_2,\ldots,p_9)$ is positive (negative), reduce (increase) p_2 and return to step 2.

4. Monte Carlo Simulation

In order to evaluate the small sample performance of the new procedure, CIC, and compare it with the performance of AIC, BIC, PAIC and PBIC, we conducted two Monte Carlo experiments. The first experiment involved the simple selection problem discussed in Section 2 and the second the more complicated problem of Section 3. Note that for each iteration of a Monte Carlo experiment, CIC requires the bootstrap method to be applied in order to find the appropriate penalties for that set of simulated data. In other words, we simulated the application of the simulation method used to find the penalty values. As might be imagined, this required considerable computational time, although the fact that it was able to be simulated does demonstrate that the computation required in order to choose between nine models is not prohibitive.

4.1 Experimental Design

For the first experiment of choosing between model-1, model-2 and model-3, the regressor x_t was generated from the AR(1) process

$$x_t = \rho x_{t-1} + v_t, \qquad v_t \sim IN(0, \sigma_v^2), \qquad t = 1, \ldots, n, \tag{15}$$

with $\sigma_v^2 = 1.0$ and four different values of ρ, namely ρ = 0, 0.25, 0.95, 1.0. The values of x_t were generated once from (15) and then held fixed from iteration to iteration. The four values of ρ result in four different design matrices denoted X1, X2, X3 and X4, respectively. This choice of artificially generated regressors was influenced by Engle *et al.*'s (1985) Monte Carlo study. The four ρ values cover four different types of regressor, namely white noise, low autoregressive, high autoregressive and random walk. This covers a range of different types of economic data.

Data was simulated from each of model-1, model-2 and model-3 with $\beta_0 = 10.0$, $\beta_1 = 1.0$ and $\sigma^2 = 1.0$. In the case of model-3, $\lambda = 0.5$ was always used as the representative value of λ in the application of CIC, but in the comparison, data was simulated for a range of λ values, namely λ = 0.1, 0.3, 0.5, 0.7, 0.9. The experiment was conducted for four different sample sizes which were n = 20, 50, 100, 500. 1000 replications were used except when calculating probabilities of selection for CIC for which 200 replications were used. As noted in Section 2, 500 bootstrap samples were used in each replication in order to calculate the penalties for CIC.

As already observed, the second experiment of simulating the problem of choosing between M1 - M9, is extremely computationally intensive for CIC. We were therefore forced to reduce the extent of the second experiment compared to the first. We only used two design matrices, X2 and X4, two samples sizes, $n = 20, 100$, and for the application of CIC, 200 bootstrap samples. Data was simulated for each of M1 - M9 with $\beta_0 = 10.0$, $\beta_1 = 1.0$ and $\sigma^2 = 1.0$. In the case of M7 - M9, data was simulated for each of λ_1 = 0.25, 0.5, 0.75 and for M5, M6 and M9, it was simulated for each of λ_2 = 0.25, 0.5, 0.75.

4.2 Results

The simulations involved estimating the probabilities of choosing each of the competing models for each of the data generating processes and each of the selection procedures. This is a large number of probabilities in total, especially when in the second experiment, nine competing models are involved. We have therefore concentrated on PCS. These are presented in Tables 1 and 2 for the first experiment and in Tables 3 and 4 for the second experiment. Full results may be found in Hossain (1998) or may be obtained from the authors on request.

4.2.1 Results of the First Experiment

An obvious feature of the results is how they differ for X1 and X2 compared to X3 and X4. For the former design matrices, model-3 seems to be a difficult model to correctly select (with BIC never able to detect it when $\lambda = 0.9$) while for X3 and X4, model-3 is correctly selected by all procedures almost always.

As is well known, BIC favours models with fewer parameters while AIC favours models with larger numbers of parameters. This pattern is clearly evident in a comparison between AIC and BIC and also between PAIC and PBIC. There are times that BIC achieves large probabilities of correctly selecting model-1 and model-2, at the expense of, for some λ values, relatively low probabilities of correctly selecting model-3. Through looking at the results for X1 and X2, one would conclude that BIC is poorly balanced, although the same conclusion would be difficult to draw from the results for X3 and X4. For X1 and X2, PAIC perhaps provides the most balanced results of the existing procedures (AIC, BIC, PAIC and PBIC) with PBIC and BIC being best for smaller and larger sample sizes, respectively, in the case of X3 and X4.

The other obvious feature of the results is how CIC tends to dominate all other procedures. It almost always has the highest probability of correct selection. The only exceptions occur for model-3, $\lambda = 0.1$ for X1 and X2 and in some circumstances for $\lambda = 0.3$. These exceptions are obviously a consequence of it always correctly selecting model-2 ($\lambda = 0$) when it is true. In many circumstances, there is a large improvement in the probability of correct selection when one moves from an existing procedure to CIC, particularly for larger λ values when model-3 is true for either X1 or X2. It should also be noted that, unlike the other procedures, for X3 and X4 CIC has a perfect record at selecting the correct model for all sample sizes except $n = 20$.

4.2.2 Results of the Second Experiment

Again we see considerable contrast between the results for X2 and X4. For X2, models M5 - M9 appear to be extremely difficult to detect, especially in the case of $n = 20$, while for X4, PCS for these models are a lot higher especially for $n = 100$.

The importance of making good use of knowledge of inequality restrictions in calculating appropriate penalties is clearly seen for X2 when $n = 20$. For models M5 - M9, AIC and BIC are almost always unsuccessful at choosing the correct model. PAIC and PBIC are marginally better in this regard. A similar pattern is observed for X2 when $n = 100$, although for M7 and perhaps M8, some PCS are more respectable.

For design matrix X4, the results for the existing procedures tend to follow expected patterns although in some cases (for example M9), there are some unexpected results.

Again a feature of the results is that CIC dominates the other procedures in terms of almost always having the highest PCS. The only exceptions occur for M7 and M8 when $\lambda_1 = 0.5$ in the case of X2 and for M9 and (λ_1, λ_2) = (0.25,0.5) when $n = 20$ and (λ_1, λ_2) = (0.25,0.25) when $n = 100$ in the case of X4. It is quite remarkable how CIC is almost always able to improve on PCS of the best of the existing procedures, no matter whether the true model is one with the smallest, average or largest number of parameters. In this sense, it can be regarded as a powerful procedure.

5. Concluding Remarks

The aim of this paper was to investigate whether King *et al.*'s (1995) representative fixed points approach could be successfully applied to model selection problems in which parameters of interest are restricted to a closed interval. The answer to this question seems to be a very clear yes, at least for the particular problems considered above. It is difficult to think of another case in the literature where a new procedure has so clearly dominated other existing procedures in terms of PCS. Given this extraordinary result, it is perhaps worth asking why CIC works so well.

There are two things that are novel about CIC. The first is that nuisance parameters are handled via the bootstrap. The growing literature on modified likelihood functions (see for example Kalbfleisch and Sprott (1970), Cox and Reid (1987) and Laskar and King (1998)) does observe that the presence of nuisance parameters can affect the quality of likelihood based inference. The second is that the penalties are calculated to suit the circumstances - the particular sample size, design matrix and likely values of the nuisance parameters. This extra care in calculating appropriate penalties, which has a heavy price in terms of computation, seems to be well rewarded in the presence of interval restrictions on key parameters.

Acknowledgements

This research was supported in part by an ARC grant and in part by a Post Graduate Publication Award grant of Monash University where the research was conducted. We are grateful to Tony Hughes for helpful comments and assistance.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In: B.N. Petrov and F. Csaki, Eds., *Proceedings of the Second International Symposium on Information Theory*. Akademiai Kiado, Budapest, 267-281.

Box, G.E.P. and D.R. Cox (1964). An analysis of transformations. *Journal of the Royal Statistical Society* B, 26, 211-252.

Cox, D.R. and N. Reid (1987). Parameter orthogonality and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society* B, 49, 1-39.

Engle, R.F., D.F. Hendry and D. Trumble (1985). Small sample properties of ARCH estimators and tests. *Canadian Journal of Economics*, 18, 66-93.

Granger, C.W.J., M.L. King and H. White (1995). Comments on testing economic theories and the use of model selection criteria. *Journal of Econometrics* 67, 173-187.

Hannan, E.J. and B.G. Quinn (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society* B, 41, 190-195.

Hossain, M.Z. (1998). Model selection problems involving interval restricted parameters in econometrics, Ph.D. thesis, Monash University, unpublished.

Hossain, M.Z. and M.L. King (1998). Model selection using AIC in the context of inequality restrictions on key parameters. Mimeograph, Dept. of Econometrics and Business Statistics, Monash University.

Hughes, A.W. (1997). Improved model selection based on AIC-type criteria, Ph.D. thesis, Monash University, unpublished.

Hughes, A.W. and M.L. King (1994). One-sided model selection procedures. Paper presented at the 1994 Australasian Meeting of the Econometric Society, Armidale.

Hughes, A.W. and M.L. King (1998). Model selection using AIC in the presence of one-sided information. Mimeograph, Dept. of Econometrics and Business Statistics, Monash University.

Kalbfleisch, J.D. and D.A. Sprott (1970). Application of likelihood methods to models involving large numbers of parameters. *Journal of the Royal Statistical Society* B 32, 175-208.

Laskar, M.R. and M.L. King (1998). Estimation and testing of regression disturbances based on modified likelihood functions. *Journal of Statistical Planning and Inference*, forthcoming.

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

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

Spitzer, J.J. (1978). A Monte Carlo investigation of the Box-Cox transformation in small samples. *Journal of the American Statistical Association*, 73, 488-494.

Table 1 Comparison Between PAIC, AIC, PBIC, BIC and CIC Based on Estimated Probabilities of Correctly Choosing Model-1, Model-2 and Model-3.

			X1					X2				
Model	**λ**	***n***	PAIC	AIC	PBIC	BIC	CIC	PAIC	AIC	PBIC	BIC	CIC
1	**1**	**20**	.607	.611	.610	.611	.665	.549	.559	.556	.559	.600
2	**0**		.892	.957	.932	.978	1.00	.933	.973	.957	.984	1.00
3	**.1**		.233	.126	.172	.073	.010	.190	.098	.132	.079	.005
	.3		.221	.085	.127	.034	.320	.251	.104	.157	.045	.260
	.5		.089	.012	.037	.001	.600	.136	.029	.063	.004	.605
	.7		.021	.002	.006	.001	.545	.041	.006	.017	.000	.400
	.9		.005	.001	.002	.000	.260	.012	.001	.001	.000	.225
1	**1**	**50**	.755	.765	.765	.766	.790	.738	.758	.758	.758	.765
2	**0**		.865	.942	.940	.986	1.00	.890	.947	.945	.984	1.00
3	**.1**		.411	.263	.269	.117	.070	.404	.270	.273	.132	.075
	.3		.539	.365	.370	.114	.530	.567	.401	.406	.181	.545
	.5		.307	.078	.083	.003	.695	.423	.169	.176	.025	.710
	.7		.090	.005	.008	.000	.555	.162	.027	.029	.000	.425
	.9		.015	.000	.000	.000	.305	.042	.001	.002	.000	.175
1	**1**	**100**	.799	.826	.826	.826	.850	.778	.812	.814	.814	.825
2	**0**		.853	.937	.951	.990	1.00	.858	.940	.949	.991	1.00
3	**.1**		.482	.342	.312	.142	.135	.496	.370	.337	.181	.130
	.3		.692	.534	.483	.247	.690	.704	.570	.526	.309	.715
	.5		.562	.287	.217	.012	.785	.622	.374	.318	.053	.795
	.7		.238	.038	.017	.000	.670	.297	.093	.055	.000	.570
	.9		.053	.001	.001	.000	.420	.085	.007	.004	.000	.230
1	**1**	**500**	.842	.921	.969	.988	1.00	.833	.915	.960	.974	1.00
2	**0**		.838	.924	.966	.995	1.00	.847	.927	.962	.997	1.00
3	**.1**		.891	.790	.675	.388	.860	.882	.776	.660	.397	.850
	.3		.991	.970	.945	.810	.995	.977	.951	.911	.762	.950
	.5		.996	.984	.954	.788	1.00	.989	.963	.922	.705	1.00
	.7		.768	.619	.460	.122	.765	.739	.566	.407	.073	.850
	.9		.293	.157	.080	.000	.375	.283	.149	.053	.000	.295

Table 2 Comparison Between PAIC, AIC, PBIC, BIC and CIC Based on Estimated Probabilities of Correctly Choosing Model-1, Model-2 and Model-3.

			X3					X4				
Model	λ	*n*	PAIC	AIC	PBIC	BIC	CIC	PAIC	AIC	PBIC	BIC	CIC
1	**1**	**20**	.793	.885	.850	.924	.995	.860	.897	.861	.933	.990
2	**0**		.822	.906	.867	.953	1.00	.820	.910	.869	.941	1.00
3	**.1**		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	.3		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	.5		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	.7		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	.9		.922	.830	.873	.726	1.00	.997	.994	.994	.986	1.00
1	**1**	**50**	.847	.921	.920	.973	1.00	.838	.933	.931	.982	1.00
2	**0**		.820	.907	.902	.964	1.00	.827	.915	.914	.972	1.00
3	**.1**		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	.3		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	.5		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	.7		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	.9		1.00	1.00	1.00	.997	1.00	1.00	1.00	1.00	1.00	1.00
1	**1**	**100**	.831	.902	.929	.983	1.00	.816	.911	.924	.981	1.00
2	**0**		.843	.921	.936	.988	1.00	.849	.921	.933	.983	1.00
3	**.1**		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	.3		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	.5		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	.7		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	.9		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	**1**	**500**	.832	.912	.952	.992	1.00	.840	.923	.952	.989	1.00
2	**0**		.850	.939	.972	.993	1.00	.907	.969	.978	1.00	1.00
3	**.1**		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	.3		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	.5		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	.7		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	.9		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 3 Comparison Between PAIC, AIC, PBIC, BIC and CIC Based on Estimated Probabilities of Correctly Choosing Competing Models M1, M2, M3, M4, M5, M6, M7, M8 and M9 for Design Matrix **X2**.

X2			*n* = 20					*n* = 100				
Model	λ_1	λ_2	PAIC	AIC	PBIC	BIC	CIC	PAIC	AIC	PBIC	BIC	CIC
1	**1**	**1**	.175	.246	.186	.268	.360	.567	.656	.673	.738	.820
2	**1**	**0**	.181	.225	.155	.225	.370	.425	.551	.399	.551	.620
3	**0**	**1**	.832	.873	.857	.883	.980	.791	.895	.913	.978	1.00
4	**0**	**0**	.367	.593	.376	.600	.650	.431	.679	.469	.716	.790
5	**1**	**.25**	.031	.000	.024	.000	.120	.039	.000	.009	.000	.200
5	**1**	**.50**	.027	.000	.020	.000	.100	.105	.008	.006	.000	.240
5	**1**	**.75**	.008	.000	.009	.000	.070	.174	.050	.032	.000	.270
6	**0**	**.25**	.001	.000	.000	.000	.100	.027	.000	.000	.000	.120
6	**0**	**.50**	.003	.000	.000	.000	.120	.088	.010	.004	.000	.160
6	**0**	**.75**	.017	.001	.003	.000	.170	.207	.086	.062	.000	.210
7	**.25**	**1**	.043	.002	.024	.000	.080	.411	.318	.304	.178	.420
7	**.50**	**1**	.057	.001	.057	.000	.110	.361	.185	.158	.005	.280
7	**.75**	**1**	.076	.000	.074	.000	.130	.111	.004	.010	.000	.140
8	**.25**	**0**	.008	.001	.001	.000	.080	.246	.229	.139	.037	.270
8	**.50**	**0**	.001	.000	.001	.000	.070	.162	.027	.007	.001	.160
8	**.75**	**0**	.003	.000	.006	.000	.100	.018	.000	.000	.000	.110
9	**.25**	**.25**	.004	.000	.000	.000	.070	.000	.000	.000	.000	.050
9	**.25**	**.50**	.001	.000	.019	.000	.080	.001	.001	.000	.000	.060
9	**.25**	**.75**	.000	.000	.015	.000	.060	.016	.001	.000	.000	.100
9	**.50**	**.25**	.035	.000	.000	.000	.100	.000	.000	.000	.000	.060
9	**.50**	**.50**	.029	.000	.087	.000	.130	.000	.000	.000	.000	.070
9	**.50**	**.75**	.008	.000	.035	.000	.100	.007	.002	.000	.000	.090
9	**.75**	**.25**	.079	.000	.154	.000	.210	.003	.000	.037	.000	.110
9	**.75**	**.50**	.048	.000	.116	.000	.160	.000	.000	.013	.000	.080
9	**.75**	**.75**	.022	.000	.081	.000	.100	.000	.000	.001	.000	.070

Table 4 Comparison Between PAIC, AIC, PBIC, BIC and CIC Based on Estimated Probabilities of Correctly Choosing Competing Models M1, M2, M3, M4, M5, M6, M7, M8 and M9 for Design Matrix **X4**.

X4			*n*= 20					*n* = 100				
Model	λ_1	λ_2	PAIC	AIC	PBIC	BIC	CIC	PAIC	AIC	PBIC	BIC	CIC
1	**1**	**1**	.627	.748	.742	.872	.950	.721	.849	.911	.978	.990
2	**1**	**0**	.340	.508	.349	.520	.560	.633	.794	.737	.825	.850
3	**0**	**1**	.792	.882	.844	.925	1.00	.832	.910	.923	.978	1.00
4	**0**	**0**	.565	.843	.622	.888	.900	.701	.876	.846	.977	1.00
5	**1**	**.25**	.286	.233	.259	.186	.330	.574	.590	.651	.640	.660
5	**1**	**.50**	.416	.416	.441	.431	.460	.522	.526	.587	.592	.610
5	**1**	**.75**	.439	.440	.458	.460	.470	.473	.474	.502	.502	.570
6	**0**	**.25**	.679	.594	.652	.516	.650	.854	.861	.930	.936	.940
6	**0**	**.50**	.800	.803	.830	.837	.840	.709	.710	.784	.787	.800
6	**0**	**.75**	.695	.697	.738	.740	.750	.654	.661	.725	.730	.730
7	**.25**	**1**	.485	.485	.503	.503	.540	.441	.441	.453	.453	.470
7	**.50**	**1**	.487	.487	.507	.507	.550	.438	.438	.447	.447	.460
7	**.75**	**1**	.484	.484	.501	.502	.530	.451	.451	.453	.453	.460
8	**.25**	**0**	.108	.070	.067	.025	.200	.537	.507	.415	.307	.580
8	**.50**	**0**	.036	.008	.014	.000	.130	.516	.372	.269	.055	.520
8	**.75**	**0**	.002	.001	.000	.000	.090	.212	.058	.033	.000	.310
9	**.25**	**.25**	.120	.073	.156	.017	.400	.870	.856	.994	.675	.910
9	**.25**	**.50**	.556	.545	.791	.498	.630	.999	.999	.999	.999	1.00
9	**.25**	**.75**	.808	.807	.808	.791	.880	.999	.999	.999	.999	1.00
9	**.50**	**.25**	.023	.010	.504	.001	.540	.840	.820	.750	.615	.910
9	**.50**	**.50**	.442	.373	.313	.199	.510	1.00	1.00	.999	.998	1.00
9	**.50**	**.75**	.811	.811	.790	.789	.860	.999	.999	.999	.999	1.00
9	**.75**	**.25**	.004	.002	.000	.000	.110	.448	.421	.232	.162	.590
9	**.75**	**.50**	.334	.312	.274	.232	.460	.956	.956	.935	.934	1.00
9	**.75**	**.75**	.557	.456	.417	.282	.590	.995	.994	.970	.938	1.00