Kagawa University Economic Review
Vol. 71, No. 3, December 1998, 217-256

Investigation of Nested x^2 Test and AIC in the Box-Cox Transformation Model*

Feng YAO**

Abstract

This paper presents some of the simulation results of the nested x^2 test in the simultaneous identification of the Box-Cox transformation (BCT) model. It focus on the investigation of fitting a BCT polynomial regression model to data generated by nonlinear model which does not belong to the family of the BCT regression models. Three nonlinear models are introduced for the simulation. Except for a second order polynomial regression model, the rest two models are the exponential regression model and the logistic regression model. The performances of applied the nested x^2 test and the Akaike's information criterion to simultaneous identification of the BCT model are compared.

1. Introduction

In the analysis of complex economic activities with general linear regression models, it was often shown that some economic data do not fit models with Gaussian noise. The Box-Cox transformation (BCT) model can offer

---

* A contributed paper to the festshrift in honour of Professor Hiroaki Seto who is retiring from Faculty of Economics, Kagawa University, in March 1998.

** The simulation was conducted on ACOS3900 (Computer Center of Tohoku University). I express my thanks to Prof. Y. Hosoya for his kindly help and advices.

the choice between linear and linear-in-logrithm models with Gaussian noise [Box and Cox (1964)]. Chang (1977, 1980) gave a successful use of the BCT in the analysis of demand for meat in the United States. He pointed out that the linear or logarithm function is not suitable for the analysis of demand for meat in the United States. James and David (1982) investigated the income and food expenditure distribution by use of the BCT. Poirier and Melino (1978) gave a discussion on the interpretation of estimated coefficients in the BCT model. The other details of the theoretical works and applications can be mainly seen in: Poirier (1978), Huang and Grawe (1980), Bickel and Doksum (1981), Box and Cox (1982), Seaks and Layson (1983), Tse (1984), Zarembka (1990) etc. But all of the analyses are within the confines of the fixed order BCT regression model. The estimate of the BCT parameter is mainly based on the traditional maximum likelihood (ML) method.

Yao (1992, 1994) [see also Yao and Hosoya (1994)] investigated the simultaneous identification of the BCT regression model by the general information criterion (GIC) and Akaike's information criterion (AIC) [Akaike (1973)]. The GIC discussed there is a developed result of an asymptotic approximation of the cross-entropy risk for the purpose of estimating the parameters of the BCT and the clan of regressions [the original idea is given by Takeuchi (1976) and Hosoya (1983)]. The Monte Carlo simulation showed that the estimate of the BCT parameter determined by the GIC is a little precise than that of determined by the AIC; but it is on the contrary in identifying the order of the BCT polynomial regression model or of the BCT autoregressive model. Being a successful application to the empirical analysis, a nonlinear model of Tokyo stock price index was presented. Yao (1995) discussed the simultaneous identification of the BCT model in view of the information criteria and the nested x^2 test [see Hosoya (1986)]. The emphases there are on the simultaneous estimation of the BCT parameter and on the order of the regression part. The Monte Carlo simulation results showed that the AIC and the GIC are very similar in identification performance, the nested x^2 method, when used for the point estimation, has a somewhat different feature. The latter has the ability to control the probability of the identified orders those exceeding the true order. The nested x^2 test tends to underestimate the order with comparably larger probability, especially in the case of the disturbance variance is large compared with the magnitude of variation of the regressor part. But on the other hand, the AIC (and the GIC also) tends to overestimate the order in general.

The investigations so far have not touched the problem of fitting the BCT polynomial regression model to data generated by the model which does not belong to the family of the BCT polynomial regression models. It is clear that this is important for understanding the properties of the information criteria methods and the nested x^2 method, and also very necessary for the application of these methods to the investigation of complex economic phenomena. Yao (1996b) gives some simulation results about this kind of investigation in view of the information criteria methods. The simulation results of Monte Carlo experiment showed that there is almost no significant difference between the GIC and the AIC with the application to the BCT model identification. It also showed that, in fitting a BCT polynomial regression model to data set generated by logistic regression model, the frequency distribution of the identified order depends on the sample variance.

This paper investigates the properties of applying the nested x^2 test to fit a p-th order BCT polynomial regression model to data generated by the three models as discussed in Yao (1996b), [see model (3-1), (3-2) and (3-3) below], to each of the model with three levels of disturbance term variance. Furthermore, the critical value α used in the nested x^2 test are chosen in five levels from 0.10 to 0.30 by a fixed step of 0.05. For each of the cases we studied, the Monte Carlo experiment is conducted for 5000 times replication. The simulation purpose here is to find a BCT polynomial regression model that can best fit the data set generated by a model, even the model itself does not belong to the family of the BCT polynomial regression models. The usually used 0.05 critical value is not discussed in this paper because Yao (1995) has pointed out that the comparatively small critical value (for example $\alpha \leq 0.05$) makes the nested x^2 test underestimating the true model order.

Our simulation results show that, both the nested x^2 test and the AIC plays good performance in fitting the BCT polynomial regression model to data generated by the three models which do not belong to the family of the BCT polynomial regression models. It is reconfirmed the fact that the nested x^2 test has good ability to control the probability of the identified orders those exceeding the true order. It also shows that the nested x^2 test tends to underestimate the order especially for the case with large sample variance. The AIC method in general tends to overestimate the true order. In the identification of the true model order, for comparatively small critical values, the nested x^2 test is seen superior to the AIC. As for the cases of underestimating the true order of the BCT polynomial regression model, the properly estimated BCT parameter might make compensate for the information loss by the underestimated order. This is true for both of the two methods.

This paper proceeds as follows: In section 2, we first give an overview of the BCT, then summarize the nested x^2 test and the AIC used in the simultaneous identification of the BCT regression model. At the last part of this section, we present three nonlinear models including two models that do not belong to the family of the BCT polynomial regression models. In section 3 we discuss the Monte Carlo simulation of fitting a BCT polynomial regression model to data generated by the three specified models introduced in section 2. The simulation results are listed in tables and are plotted in graphs for the three models, respectively. For each of the models, we discuss three levels of the disturbance term variance and five levels of critical value. The discussion of the simulation results is summarized in section 4. We give the conclusions and remarks in section 5.

2. Models and Methods

2.1 The Box-Cox transformation Model

As a special power transformation, for any positive variable y, the Box-Cox transformation (BCT) [Box-Cox (1964)] is defined as

$$y^{[\lambda]} = \begin{cases} [y^{\lambda}-1]/\lambda & \lambda \neq 0 \\ \log y & \lambda = 0, \end{cases} \tag{2-1}$$

or for the case $y < 0$ but $y > -a \quad (a > 0)$,

$$y^{[\lambda]} = \begin{cases} [(y+a)^{\lambda}-1]/\lambda & \lambda \neq 0 \\ \log (y+a) & \lambda = 0, \end{cases} \tag{2-1$'$}$$

where λ is an unknown parameter called the BCT parameter. In general, it is assumed that for each λ, $y^{[\lambda]}$ is a monotonous function of y over the admissible range. Because of (2-1) is continuous at $\lambda = 0$ [see Yao (1994)], so it is preferable for theoretical analysis. The following investigations are only based on the transformation defined by (2-1). It is clear that all the results will be hold for (2-1)$'$ if only we change y with $(y+a)$ in the corresponding definitions.

As for the BCT regression model, since both the dependent and independent variables can be transformed, the general BCT model has the form

$$y^{[\lambda_1]} = \beta_1 + \beta_2 x_2^{[\lambda_2]} +, \cdots, + \beta_p x_p^{[\lambda_p]} + \varepsilon, \tag{2-2}$$

where ε is random disturbance term generated by $i.i.d.\ N(0, \sigma^2)$. The BCT regression model expressed in (2-2) can be specified and estimated. For $\lambda_1 = 1$, 0, and -1, $y^{[\lambda_1]}$ enters into model (2-2) linearly, as $\log y$ and as the reciprocal of y. Thus the estimation procedure itself can choose the transformation which best fits the data.

For different λ_i $(i = 1, 2, \cdots, p)$, model (2.2) can be specified into mainly three types of BCT regression models [see Spitzer (1982)] or six types of BCT regression models [see Yao and Hosoya (1994), there the model classification cover all possible main versions of the BCT regression models]. We consider in this paper the specified model that only the independent variable is transformed

$$y^{[\lambda]} = \beta_0 + \beta_1 x + \beta_2 x^2 +, \cdots, + \beta_p x^p + \varepsilon. \tag{2-3}$$

We will fit this BCT polynomial regression model to the data sets generated by models defined by (3-1) and (3-2) as well as (3-3), respectively.

2.2 The Nested x^2 Test for BCT Regression Model

We summarize the nested x^2 test [which is given by Hosoya (1986)] in this subsection. The application to hierarchical statistical models is studied by Hosoya (1989). It applies the generalized likelihood ratio (GLR) test of equal marginal error rate to the model selection problems. The process of applied this method to the identification of the BCT model can be seen in Yao (1995).

Suppose that the parameter $\theta = (\theta_1, \theta_2, \cdots, \theta_p)$ specifying a density of observations, where θ_j is r_j-dimensional parameter vector. The hypothesis H_j implies that $\theta_{j+1} = \theta_{j+2} = \cdots = \theta_p = 0$ (H_p implies that no such specification is imposed). For $i < j$ $(1 \leq i,\ j \leq p)$ denote by L_{ij} the log-likelihood ratio for testing H_i against H_j. The test for H_i in the presence of such nested alternative hypotheses would use L_{ij}, by using a test with critical region

$$R = \{L_{ij} \leq c_j, \text{ for some } j \in (i+1, \cdots, p)\},$$

where the c_j's are determined so that $\Pr\{R \mid H_i\}$ is equal to the required size. The P-value which corresponds to this test is evaluated as $P(q^*) = \Pr\{Q \leq q^* \mid H_i\}$, where $Q = \min(P_j \mid i+1 \leq j \leq p)$ is the P-value based on L_{ij} and q^* is the observed value of Q. The test for the critical region R will here be termed a GLR test [see Hosoya (1989)]. For the case all degrees of freedom are 1 and $p \leq 13$, the algorithm for the P-value is available in Hosoya and Katayama (1987).

Now we consider the P-th order BCT polynomial regression model (2-2) with the expression of density function

$$f_n(y_1, y_2, \cdots, y_n \mid \beta_0, \beta_1, \cdots, \beta_p, \sigma^2, \lambda) = \tag{2-4}$$
$$\frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{n}\left(y_t^{[\lambda]} - \beta_0 - \sum_{t=1}^{n}(\beta_k x_t^k)\right)^2\right\} \prod_{t=1}^{n} y_t^{\lambda-1}.$$

For a set of n independent observations $\{y_i,\ x_i\}_{i=1}^{n}$, the simultaneous estimation of the BCT parameter and the order of the regression model by the nested x^2 test proceeds by the following steps:

1. Given p, we consider the family of polynomial regression models that each of the orders is less than or equals to p. We first calculate the $n_\lambda \times p$ matrix of the maximum log-likelihood for given n_λ and different fixed $\lambda_i = 1, 2, \cdots, n_\lambda$. Then by the maximum log-likelihood estimation, the BCT parameter $\hat{\lambda}_j$ for j-th order polynomial regression model can be determined. We denote

the maximum log-likelihood by $L(j, \hat{\lambda}_j)$, $j = 1, 2, \cdots, p$.

2. For $k = 1, 2, \cdots, p-1$, we calculate the difference of the two maximum log-likelihood ratio $LR(k, j)$ defined as follows:

$$LR(k, j) = 2[L(j, \hat{\lambda}_j) - L(k, \hat{\lambda}_k)], \; j = k+1, \cdots, p, \tag{2-5}$$

where k is the order of the polynomial regression model. We treat this as the input to the subroutine program given by Hosoya and Katayama (1987). By this way, the P-values $\{P_k\}_{k=1}^{p-1}$ can be obtained.

3. For the given significance level α, if there exists some index k that satisfied $P_k > \alpha$, we choose the first k to be the estimate of the order of the polynomial regression model. That is to say $\hat{p} = k$. If $P_k \leq \alpha$ for all the k, $k = 1, 2, \cdots, p-1$, then we have $\hat{p} = p$. The corresponding $\hat{\lambda}_{\hat{p}}$, which we want to estimate, is the estimator of the BCT parameter that best fits the model.

2.3 The AIC for BCT Regression Model

The AIC was first introduced for the purpose of comparison and selection among several models [Akaike (1973)]. The introduction of objective criterion enables the objective comparison of models that are usually selected subjectively by the analysts. The details of the AIC theory can be seen in Sakamoto, Ishiguro, Kitagawa (1986). Yao (1994) shows the process of using the AIC in the simultaneous identification of the BCT regression model.

Suppose $y_1, y_2, \cdots, y_n$ be independent, positive random variables and consider the p-th order BCT polynomial regression model, say model (2-3) with an expression of probability density function (2-4) indexed by the BCT parameter λ. For a set of n independent observations $\{y_i, x_i\}_{i=1}^{n}$, the simultaneous estimation of the BCT parameter and the order of the regression model can be determined by the AIC. Because the number of free parameters in model (2-4) for given (p, λ) is $(p+1)$, therefore the AIC for the BCT model (2-4) is:

$$\mathrm{AIC}(p, \lambda) = C + n\log(\hat{\sigma}^2(p, \lambda)) - 2(\lambda-1)\sum_{t=1}^{n}\log(y_t) + 2(p+1), \tag{2-6}$$

where $C = n(1+\log(2\pi))$. The identification problem between the two given probability density function $f(p_1, \lambda_1)$ and $f(p_2, \lambda_2)$ is dealt by the minimum principle of the AIC. Choose the model $f(p_1, \lambda_1)$ if

$$\mathrm{AIC}(p_1, \lambda_1) < \mathrm{AIC}(p_2, \lambda_2), \tag{2-7}$$

and the model $f(p_2, \lambda_2)$ otherwise.

2.4 Three Nonlinear Models

For the purpose of evaluating the performances of the nested x^2 test and the AIC in model identification of the BCT regression model by Monte Carlo simulation, data generation is considered by three types of nonlinear regression models. Except for a second order polynomial regression model, the others are the so called exponential regression model and logistic regression model. The three models that will be used in the Monte Carlo experiment are:

$$\text{Model A}: y = a_1 + b_1 x + c_1 x^2 + \varepsilon_1, \tag{2-8}$$

$$\text{Model B}: y = a_2 + b_2^x + \varepsilon_2, \tag{2-9}$$

$$\text{Model C}: y = a_3 + 1/(1+\exp(-x)) + \varepsilon_3, \tag{2-10}$$

where ε_i $(i = 1, 2, 3)$ is random disturbance term generated by $i.i.d.$ $N(0, \sigma^2)$, a_i and b_j $(i = 1, 2, 3, j = 1, 2)$ are constant. We call model B an exponential regression model, and Model C a logistic regression model. It is clear that Model B and Model C do not belong to the family of the BCT polynomial regression models.

3. Data Generation and Numerical Evaluation

In this section we conduct Monte Carlo experiment based on model (2-8) and model (2-9) and also model (2-10). The purpose is to investigate the properties of the nested x^2 method and the AIC method in fitting the BCT polynomial regression model (2-3) to data generated by the above mentioned nonlinear regression models. The numerical calculations are conducted by the FORTRAN programs presented by Yao (1992).

In view of model (2-8), (2-9) and (2-10), let us consider the following three specified nonlinear regression models:

$$y_{t1} = 5+0.8x_t+0.64x_t^2+\varepsilon_{t1}, \qquad (3\text{-}1)$$

$$y_{t2} = 5+2^{x_t}+\varepsilon_{t2}, \qquad (3\text{-}2)$$

$$y_{t3} = 5+1/(1+\exp(-x_t))+\varepsilon_{t3}. \qquad (3\text{-}3)$$

For the sample size n, we need first to generate random data set $\{\varepsilon_{ti}\}_{t=1}^{n}$, the value of the disturbance term ε_{ti} $(i=1, 2, 3)$ is random number and obtained from $i.i.d.$ $N(0, \sigma_0^2)$. We choose the sample size $n=100$ in this paper. The independent variable x_t is defined as $x_t = t/40$ $(t=1, 2, \cdots, n)$. By this procedure, for a given data set $\{x_t\}_{t=1}^{n}$, we can generate data set $\{y_{ti}, x_t\}_{t=1}^{n}$, $i=1, 2, 3$.

We fit the BCT polynomial regression model (2-3) to data generated by the above three models, respectively. The largest order of the BCT polynomial regression model used in the simulation is chosen to be 7. Then, for all the estimated orders those exceeding 7 are put to be 7. For both the nested x^2 method and the AIC method, the experiments are performed for three levels of the variance of the disturbance term, $\sigma_0^2 = 0.5$, 1.0, 1.5. Corresponding to these selected models, the critical values used in the nested x^2 test are chosen from 0.10 to 0.30 with a fixed step of 0.05. The Monte Carlo experiments are performed for all the three models with

5000 replications.

In the Monte Carlo simulation, for each of generated data sets and an initial given BCT parameter λ_0, we need to estimate the P-value (and the AIC) for order $p = 1, 2, \cdots, 7$ and the BCT parameter $\lambda_j = \lambda_0 + \left(\frac{n_\lambda + 1}{2} - j\right)\Delta\lambda$, $j = 1, 2, \cdots, n_\lambda$, where n_λ is a given (odd) number and $\Delta\lambda$ is a given real value. The initial given λ_0 and n_λ as well as $\Delta\lambda$ should be determined by a pre-test or based on some prior information about the BCT parameter and the pattern of the frequency distribution. In the following experiments, we choose $\lambda_0 = 1.0$, $n_\lambda = 17$ and $\Delta\lambda = 0.25$. In view of the nested x^2 test and according to the minimum principle of the AIC, we can simultaneously get the estimates of $\hat{p}$ and $\hat{\lambda}$ for both of the two methods. The estimator $(\hat{p}, \hat{\lambda})$ can determine the best fitted BCT regression model for the generated data set. We can choose an enough large n_λ to satisfy any required precise of the estimated BCT parameter.

Remarks: For the space of this paper, in the following Tables 3.2 to 3.5, we only listed the simulation results for $n_\lambda = 15$. It gives no effects on the BCT regression model identification because the estimated frequencies are zero at the BCT parameter of -1 or 3.

To show the simultaneous identification processes, for the space of this paper, we only give a distribution of the estimated ML and the estimated P-values for one experiment (in 5000 times) of fitting the BCT polynomial regression model (2-3) to data generated by the nonlinear regression model (3-1). In table 3.1 the estimated minimum ML for the seven orders in levels of the BCT parameter are marked by underline. Based on these estimated ML, the P-value can be estimated for the corresponding order. We listed them at the last row. For the 0.1 critical value, for example, it can be seen that the P-value is first over 0.1 at the 2nd order and the corresponding estimated BCT parameter is λ_8 $(= 0.75)$. That is to say in this experiment, the best fitted BCT regression model to the data generated by model (3-1) is the 2nd order BCT polynomial regression model with the estimated BCT parameter 0.75. This experiment gives a contribution of '1' to the upper block of table 3.2 at (0.75, 2). The upper block in Table 3.2 is the results of repeating the experiment by 5000 times.

Table 3.1 Distribution of the Estimated ML Values and the P-values by the Nested x^2 Method

λ_j	Order 1	2	3	4	5	6	7
-1.00	-26.13	-26.13	-24.78	-23.74	-22.30	-18.73	-17.08
-0.75	-20.96	-20.90	-19.62	-18.48	-17.06	-13.81	-12.47
-0.50	-16.58	-16.34	-15.17	-13.93	-12.51	-9.60	-8.54
-0.25	-13.07	-12.48	-11.45	-10.11	-8.70	-6.13	-5.31
0.00	-10.47	-9.35	-8.50	-7.06	-5.65	-3.42	-2.81
0.25	-8.82	-7.00	-6.34	-4.79	-3.38	-1.47	-1.04
0.50	-8.15	-5.43	-4.97	-3.31	-1.90	-0.30	-0.01
0.75	-8.47	-4.68	-4.40	-2.63	-1.20	0.11	0.29
1.00	-9.74	-4.74	-4.61	-2.73	-1.28	-0.23	-0.13
1.25	-11.93	-5.61	-5.57	-3.58	-2.11	-1.30	-1.24
1.50	-14.99	-7.28	-7.28	-5.16	-3.67	-3.05	-3.03
1.75	-18.84	-9.72	-9.68	-7.44	-5.91	-5.47	-5.46
2.00	-23.42	-12.91	-12.75	-10.38	-8.81	-8.50	-8.50
2.25	-28.65	-16.80	-16.45	-13.93	-12.31	-12.12	-12.12
2.50	-34.46	-21.36	-20.74	-18.07	-16.39	-16.28	-16.27
2.75	-40.78	-26.54	-25.57	-22.75	-21.00	-20.94	-20.94
3.00	-47.55	-32.31	-30.93	-27.94	-26.10	-26.09	-26.08
P-value	0.0162	0.1122	0.0638	0.1177	0.1538	0.5444	***

Fit BCT model (2-3) to data generated by model (3-1), $\sigma_0^2 = 1.0$

Table 3.2 shows the distribution of the estimated frequencies in 5000 times replicated Monte Carlo experiments by the nested x^2 test and the AIC. The simulation result by use of the nested x^2 test is only one of the results of fitting the BCT polynomial regression model to data generated by model (3-1) in the case of the variance of $\sigma_0^2 = 1.0$ and the given critical value $\alpha = 0.15$. The estimated BCT parameter $\hat{\lambda}(p)$ for order $p\,(p = 1,$

2, ⋯, 7) is the weighted mean of $\{\lambda_j\}_{j=1}^{n_\lambda}$, namely

$$\hat{\lambda}(p) = \sum_{j=1}^{n_\lambda} \lambda_j \times \frac{N(\lambda_j,\ p)}{N(p)} \tag{3-4}$$

where $N(\lambda_j,\ p)$ which being listed in the main block respectively for the nested x^2 method and the AIC method, is the number of the estimated frequency at the BCT parameter λ_j and the order p,

$$N(p) = \sum_{j=1}^{n_\lambda} N(\lambda_j,\ p). \tag{3-5}$$

The estimated BCT parameter $\hat{\lambda}$ is the weighted mean of $\{\lambda_j\}_{j=1}^{n_\lambda}$, namely

$$\hat{\lambda} = \sum_{j=1}^{n_\lambda} \lambda_j \times \frac{N(\lambda_j)}{N}, \tag{3-6}$$

where $N(\lambda_j)$ is the number of the estimated frequency at the BCT parameter λ_j, $N(\lambda_j) = \sum_{p=1}^{7} N(\lambda_j,\ p)$ for $j = 1,\ 2,\ \cdots,\ n_\lambda$, and $N = \sum_{j=1}^{n_\lambda} N(\lambda_j)$. The estimate of order p should be determined by

$$\hat{p} = \max_p N(p). \tag{3-7}$$

If there exists one more estimated orders, i. e. for example $\max_{p_1} N(p_1) = \max_{p_2} N(p_2)$, we usually choose the larger one, $\hat{p} = \max(p_1,\ p_2)$.

The results of the simulation experiment for fitting the BCT polynomial regression model to data generated by the exponential regression model (3-2) and the logistic regression model (3-3) in the same situation are summarized and given in the following table 3.3 and table 3.4, respectively.

Table 3.2 The Frequencies Distribution in 5000 Times Replications by the Nested x^2 Method and the AIC

λ_j	Order 1	2	3	4	5	6	7	Total
	Nested χ^2 Method							$N(\lambda_j)$
-0.75	0	0	0	0	0	0	0	0
-0.50	0	0	0	0	0	0	0	0
-0.25	2	0	0	0	0	0	0	2
0.00	10	0	0	1	0	0	1	12
0.25	77	16	4	1	2	1	2	103
0.50	280	165	29	13	9	8	16	520
0.75	332	691	67	33	36	24	25	1208
1.00	136	1237	108	63	28	40	46	1658
1.25	25	901	66	49	32	22	32	1127
1.50	2	250	33	9	11	7	9	321
1.75	0	31	4	5	3	1	3	47
2.00	0	1	1	0	0	0	0	2
2.25	0	0	0	0	0	0	0	0
2.50	0	0	0	0	0	0	0	0
2.75	0	0	0	0	0	0	0	0
$N(p)$	864	3292	312	174	121	103	134	5000
$\lambda(p)$	0.6690	1.0326	1.0088	1.023	1.0062	0.9903	0.9851	0.9651
	AIC							$N(\lambda_j)$
-0.75	0	0	0	0	0	0	0	0
-0.50	0	0	0	0	0	0	0	0
-0.25	1	0	0	0	0	0	0	1
0.00	6	1	0	1	0	0	1	9
0.25	41	20	7	1	2	1	2	74
0.50	145	185	49	21	16	9	19	444
0.75	152	740	125	59	50	36	33	1195
1.00	70	1183	198	118	56	63	53	1741
1.25	10	803	133	82	56	39	38	1161
1.50	1	211	43	25	14	16	15	325
1.75	0	24	6	5	6	2	4	47
2.00	0	1	2	0	0	0	0	3
2.25	0	0	0	0	0	0	0	0
2.50	0	0	0	0	0	0	0	0
2.75	0	0	0	0	0	0	0	0
$N(p)$	426	3168	563	312	200	166	165	5000
$\lambda(p)$	0.6585	1.0100	1.0004	1.0313	1.0175	1.0301	0.9985	0.9809

Fit BCT model (2-3) to data generated by model (3-1), $\sigma_0^2 = 1.0$.
The nested method used critical value $\alpha = 0.15$.

Table 3.3 The Frequencies Distribution in 5000 Times Replications by the Nested x^2 Method and the AIC

λ_j	Order 1	2	3	4	5	6	7	Total
				Nested χ^2 Method				$N(\lambda_j)$
-0.75	2	0	0	0	0	0	0	2
-0.50	0	0	0	1	0	0	0	1
-0.25	7	2	0	0	0	1	0	10
0.00	47	12	2	0	0	0	3	64
0.25	132	82	16	6	5	3	1	245
0.50	225	316	33	14	17	9	22	636
0.75	226	695	61	40	28	20	27	1097
1.00	144	912	94	53	19	29	33	1284
1.25	54	741	93	31	32	24	27	1002
1.50	14	333	52	25	18	14	15	471
1.75	1	100	26	12	5	3	8	155
2.00	0	16	4	5	2	1	1	29
2.25	0	2	2	0	0	0	0	4
2.50	0	0	0	0	0	0	0	0
2.75	0	0	0	0	0	0	0	0
$N(p)$	852	3211	383	187	126	104	137	5000
$\lambda(p)$	0.6408	1.0117	1.077	1.0602	1.0278	1.0313	0.9982	0.9557
				AIC				$N(\lambda_j)$
-0.75	0	0	0	0	0	0	0	0
-0.50	1	1	0	1	0	0	0	3
-0.25	2	2	0	0	0	2	0	6
0.00	25	14	4	0	1	0	3	47
0.25	60	90	27	10	4	5	3	199
0.50	109	315	65	24	24	15	24	576
0.75	101	708	133	56	48	27	32	1105
1.00	71	850	185	98	37	50	39	1330
1.25	27	673	152	73	52	34	32	1043
1.50	10	284	78	42	24	25	23	486
1.75	1	91	35	16	5	7	12	167
2.00	0	16	4	5	6	1	2	34
2.25	0	2	2	0	0	0	0	4
2.50	0	0	0	0	0	0	0	0
2.75	0	0	0	0	0	0	0	0
2.75	0	0	0	0	0	0	0	0
$N(p)$	407	3046	685	325	201	166	170	5000
$\lambda(p)$	0.6529	0.9924	1.0288	1.0654	1.0336	1.0407	1.0309	0.9791

Fit BCT model (2-3) to data generated by model (3-2), $\sigma_0^2 = 1.0$.
The nested method used critical value $\alpha = 0.15$.

Table 3.4 The Frequencies Distribution in 5000 Times Replications by the Nested χ^2 Method and the AIC

λ_j	Order 1	2	3	4	5	6	7	Total
			Nested χ^2 Method					$N(\lambda_j)$
-0.75	1	1	1	0	0	0	0	3
-0.50	11	1	0	0	0	0	0	12
-0.25	33	4	2	3	1	0	2	45
0.00	117	19	5	2	3	3	4	153
0.25	248	20	10	10	3	3	9	303
0.50	503	29	21	11	16	9	13	602
0.75	826	56	17	15	14	13	20	961
1.00	936	62	31	26	14	22	23	1114
1.25	731	60	24	23	15	17	19	889
1.50	450	39	19	15	7	9	15	554
1.75	208	15	11	8	8	3	6	259
2.00	58	6	4	4	0	4	2	78
2.25	20	1	0	1	0	1	1	24
2.50	1	0	0	0	1	0	0	2
2.75	0	0	0	0	1	0	0	1
$N(p)$	4143	313	145	118	83	84	114	5000
$\lambda(p)$	0.9584	0.9433	0.9741	1.0169	0.9819	1.0387	0.9583	0.961
				AIC				$N(\lambda_j)$
-0.75	1	1	1	0	0	0	0	3
-0.50	10	2	0	0	0	0	0	12
-0.25	28	6	5	3	1	2	2	47
0.00	102	18	9	4	6	5	5	149
0.25	209	41	22	13	7	6	9	307
0.50	432	56	41	21	23	16	16	605
0.75	698	108	57	28	27	20	24	962
1.00	798	123	64	45	34	29	26	1119
1.25	624	104	51	34	18	22	20	873
1.50	391	70	35	15	12	15	19	557
1.75	179	32	14	12	10	5	6	258
2.00	48	10	6	7	1	5	5	82
2.25	16	4	0	0	1	1	1	23
2.50	1	0	0	0	1	0	0	2
2.75	0	0	0	0	1	0	0	1
$N(p)$	3537	575	305	182	142	126	133	5000
$\lambda(p)$	0.9588	0.9722	0.9295	0.9835	0.9489	0.9841	0.9774	0.9603

Fit BCI model (2-3) to data generated by model (3-3), $\sigma_0^2 = 1.0$
The nested method used critical value $\alpha = 0.15$.

To investigate the performance of applied the nested x^2 test to model identification in case of the disturbance term variance and the given critical values changed, we summarize 45 three-dimension graphs in nine figures from figure 3.1.1 to figure 3.3.3. They are the simulation results of fitting the BCT polynomial regression model (2-3) to the data sets generated by the three nonlinear models mentioned above. Figure 3.1.1. to figure 3.1.3 show the results of fitting the BCT polynomial regression model (2-3) to the data sets generated by model (3-1) in three levels of the disturbance term variance of $\sigma_0^2 = 0.5,\ 1.0,\ 1.5$, respectively. In each of the figures, we give five plots corresponding to the critical values α from 0.10 to 0.30 with a fixed step of 0.05. Each of the bar graphs is based on the distribution of the estimated frequency $N(\lambda_j,\ p)$ in 5000 times replicated Monte Carlo experiments. Table 3.2 is concentrated into one of the plot in figure 3.1.2 for $\alpha = 0.15$. All of the simulations for data sets generated by the exponential regression model (3-2) are conducted by the same way, the results are plotted in figures 3.2.1, 3.2.2, 3.2.3 for three levels of variance and five levels of critical values, respectively. The same Monte Carlo experiment results for fitting the data generated by the logistic regression model (3-3) are plotted in figure 3.3.1 and figure 3.3.2 as well as figure 3.3.3. In each of the figures, the simulation result by the AIC method is also listed there for the model and the disturbance term variance indicated. Then the difference of the performances between the nested x^2 test and the AIC can be observed visually. A number of properties of the Monte Carlo simulations can be obtained by a detailed observation of those graphs. This will be discussed in the next section in detail.

For the further investigation of the nested x^2 method, we then summarize the distribution of the estimated frequencies and the BCT parameter in 5000 times replicated Monte Carlo experiments for nonlinear model (3-1) in the three cases of the variance of $\sigma_0^2 = 0.5,\ 1.0,\ 1.5$, with five levels of critical values $\alpha = 0.10,\ 0.15,\ 0.20,\ 0.25,\ 0.30$. The results are listed in table 3.5 respectively for the two methods. As for table 3.5, the upper block shows the frequency distribution of the estimated $N(p)$ [see (3-5)] in percentage form for different order p $(p = 1, 2, \cdots, 7)$. The simulation results in view of the AIC are listed at the last three lines. The last column that just on the right of this block lists the estimated percentage values of those exceeding the true order. We denote it as $d(\alpha, \sigma^2)\%$ and in the tables as $d\ \%$. As the true model is the 2nd order BCT regression model (3-1), $d(0.1, 0.5) = 12.08$ is the sum of the percentage from the 3rd order to the 7th order. The lower block shows the estimated BCT parameters $\hat{\lambda}(p)$ $(p = 1, 2, \cdots, 7)$ [see (3-4)] for models distinguished by three levels of variance and five critical values used in the nested x^2 test. The last three lines are the simulation results estimated by the AIC method. The last column just on the right of this block lists the estimated BCT parameters $\hat{\lambda}$ which is defined by (3-6).

The distribution of the estimated order and the BCT parameters for fitting model (2-3) to data generated by exponential regression model (3-2) are listed in table 3.6. The same Monte Carlo simulation results for logistic regression model (3-3) are summarized in table 3.7. The $d\%$ in these two tables shows the percentage of the estimated frequencies those exceeding the estimated order which is determined by (3-7).

Table 3.5 Distribution of the Estimated Order and the BCT Paramater for Data Generated by Model (3-1)

α	σ_0^2	Order 1	2	3	4	5	6	7	$d\%$ k>2
		Nested χ^2 Method							
0.10	0.5	7.56	80.36	4.72	2.40	1.62	1.44	1.90	12.08
	1.0	22.98	65.22	4.56	2.38	1.58	1.42	1.86	11.80
	1.5	35.26	53.34	4.26	2.28	1.56	1.42	1.88	11.40
0.15	0.5	4.90	77.82	6.52	3.48	2.42	2.12	2.74	17.28
	1.0	17.28	65.84	6.24	3.48	2.42	2.06	2.68	16.88
	1.5	27.86	55.80	5.92	3.44	2.26	2.00	2.72	16.34
0.20	0.5	3.54	73.82	8.18	4.08	3.26	3.02	4.10	22.64
	1.0	13.52	64.20	8.02	4.00	3.28	2.96	4.02	22.28
	1.5	23.00	55.26	7.78	3.86	3.18	2.90	4.02	21.74
0.25	0.5	2.74	68.76	9.28	5.56	3.92	4.08	5.66	28.50
	1.0	10.74	61.66	8.96	5.38	3.76	4.04	5.46	27.60
	1.5	19.46	53.96	8.58	5.06	3.74	3.86	5.34	26.58
0.30	0.5	2.04	64.70	10.18	6.38	4.56	4.84	7.30	33.26
	1.0	8.50	58.98	10.00	6.22	4.42	4.70	7.18	32.52
	1.5	15.84	52.74	9.46	5.96	4.28	4.58	7.14	31.42
		AIC							
	0.5	1.84	69.28	11.68	6.34	4.08	3.42	3.36	28.88
	1.0	8.52	63.36	11.26	6.24	4.00	3.32	3.30	28.12
	1.5	15.82	57.40	10.52	5.88	3.90	3.24	3.24	26.78
		Nested χ^2 Method							$\hat{\lambda}$
0.10	0.5	0.411	1.023	0.995	1.021	0.963	0.993	1.016	0.974
	1.0	0.672	1.042	0.991	1.011	0.975	1.014	1.005	0.952
	1.5	0.767	1.048	0.994	1.024	0.962	1.004	0.995	0.943
0.15	0.5	0.403	1.014	1.018	1.030	1.015	1.000	1.006	0.984
	1.0	0.669	1.033	1.009	1.023	1.006	0.990	0.985	0.965
	1.5	0.770	1.037	0.998	1.026	0.998	0.990	0.982	0.956
0.20	0.5	0.400	1.007	1.010	1.038	1.032	1.007	1.013	0.988
	1.0	0.666	1.024	0.996	1.035	1.023	1.020	0.998	0.973
	1.5	0.772	1.027	0.990	1.039	1.006	1.002	0.991	0.963
0.25	0.5	0.394	1.003	1.009	1.039	1.017	1.017	1.006	0.990
	1.0	0.664	1.015	0.996	1.042	1.015	1.019	1.014	0.977
	1.5	0.767	1.022	0.983	1.041	0.996	1.019	1.009	0.968
0.30	0.5	0.392	1.000	1.012	1.021	1.011	1.013	1.013	0.992
	1.0	0.659	1.011	0.999	1.023	1.015	1.017	1.001	0.980
	1.5	0.766	1.016	0.996	1.020	0.994	1.014	1.002	0.973
		AIC							
	0.5	0.400	1.000	1.003	1.030	1.017	1.023	1.009	0.993
	1.0	0.659	1.010	1.000	1.031	1.018	1.030	0.999	0.981
	1.5	0.768	1.016	0.989	1.037	1.001	1.014	0.994	0.974

Replicated Monte Carlo Experiments by 5,000 Times.

Table 3.6 Distribution of the Estimated Order and the BCT Paramater for Data Generated by Model (3-2)

α	σ_0^2	Order 1	2	3	4	5	6	7	$d\%$ k>2
					Nested χ^2 Method				
0.10	0.5	6.28	78.26	7.76	2.70	1.62	1.56	1.82	15.46
	1.0	23.42	63.24	6.08	2.30	1.56	1.52	1.88	13.34
	1.5	36.40	51.30	5.22	2.26	1.46	1.44	1.92	12.30
0.15	0.5	4.00	74.40	9.72	4.18	2.60	2.16	2.94	21.60
	1.0	17.04	64.22	7.66	3.74	2.52	2.08	2.74	18.74
	1.5	28.90	53.84	6.84	3.52	2.36	1.94	2.60	17.26
0.20	0.5	2.76	70.08	11.28	4.92	3.46	3.04	4.46	27.16
	1.0	13.04	62.46	9.38	4.80	3.26	2.90	4.16	24.50
	1.5	24.06	53.22	8.04	4.46	3.24	2.88	4.10	22.72
0.25	0.5	1.98	65.24	13.36	5.86	4.04	3.78	5.74	32.78
	1.0	10.58	59.88	10.76	5.58	3.88	3.82	5.50	29.54
	1.5	20.06	52.16	9.56	5.38	3.84	3.80	5.20	27.78
0.30	0.5	1.50	59.82	14.78	6.88	4.50	4.88	7.64	38.68
	1.0	8.46	56.48	12.32	6.42	4.40	4.60	7.32	35.06
	1.5	16.62	50.62	10.92	6.08	4.28	4.40	7.08	32.76
					AIC				
	0.5	1.24	59.38	16.24	7.64	5.06	4.82	5.62	39.38
	1.0	7.26	57.06	13.92	7.14	4.68	4.68	5.26	35.68
	1.5	14.76	51.36	12.76	6.80	4.62	4.46	5.24	33.88
					Nested χ^2 Method				$\hat{\lambda}$
0.10	0.5	0.312	0.974	1.160	1.135	1.006	1.010	1.047	0.954
	1.0	0.651	1.021	1.083	1.109	0.987	1.020	1.021	0.940
	1.5	0.779	1.026	1.050	1.095	1.007	1.021	1.016	0.938
0.15	0.5	0.314	0.965	1.131	1.078	1.081	1.056	1.026	0.967
	1.0	0.641	1.012	1.077	1.060	1.028	1.031	0.998	0.956
	1.5	0.771	1.022	1.039	1.047	1.015	1.031	1.010	0.951
0.20	0.5	0.288	0.960	1.102	1.089	1.085	1.038	1.046	0.975
	1.0	0.641	0.999	1.065	1.079	1.045	1.028	1.043	0.967
	1.5	0.767	1.015	1.029	1.061	1.026	1.009	1.044	0.960
0.25	0.5	0.268	0.959	1.066	1.090	1.068	1.078	1.051	0.981
	1.0	0.642	0.994	1.045	1.069	1.040	1.060	1.035	0.973
	1.5	0.769	1.006	1.025	1.067	1.017	1.041	1.034	0.967
0.30	0.5	0.293	0.956	1.072	1.068	1.050	1.036	1.041	0.986
	1.0	0.644	0.993	1.030	1.053	1.035	1.037	1.029	0.979
	1.5	0.773	1.002	1.014	1.049	1.025	1.023	1.025	0.971
					AIC				
	0.5	0.274	0.959	1.065	1.073	1.078	1.048	1.035	0.984
	1.0	0.653	0.992	1.029	1.065	1.034	1.041	1.031	0.979
	1.5	0.769	1.002	1.023	1.066	1.023	1.029	1.025	0.972

Replicated Monte Carlo Experiments by 5,000 Times.

Table 3.7 Distribution of the Estimated Order and the BCT Paramater for Data Generated by Model (3-3)

α	σ_0^2	Order 1	2	3	4	5	6	7	$d\%$ k>1
					Nested χ^2 Method				
0.10	0.5	87.94	4.86	1.94	1.54	1.00	1.16	1.56	12.06
	1.0	88.36	4.50	1.96	1.54	0.96	1.16	1.52	11.64
	1.5	88.46	4.44	1.98	1.50	0.94	1.12	1.56	11.54
0.15	0.5	82.48	6.66	2.90	2.48	1.58	1.66	2.24	17.52
	1.0	82.86	6.26	2.90	2.36	1.66	1.68	2.28	17.14
	1.5	82.88	6.20	2.96	2.28	1.64	1.76	2.28	17.12
0.20	0.5	77.10	8.12	4.06	2.62	2.42	2.30	3.38	22.90
	1.0	77.52	7.72	4.10	2.62	2.36	2.32	3.36	22.48
	1.5	77.66	7.58	4.06	2.62	2.38	2.32	3.38	22.34
0.25	0.5	72.04	9.06	4.94	3.46	3.02	2.92	4.56	27.96
	1.0	72.40	8.76	4.88	3.46	3.04	2.98	4.48	27.60
	1.5	72.48	8.60	4.76	3.56	3.06	3.00	4.54	27.52
0.30	0.5	66.94	10.18	5.82	4.14	3.44	3.62	5.86	33.06
	1.0	67.34	9.70	5.82	4.06	3.52	3.70	5.86	32.66
	1.5	67.46	9.54	5.74	4.12	3.58	3.70	5.86	32.54
					AIC				
	0.5	69.92	12.22	6.22	3.62	2.78	2.58	2.66	30.08
	1.0	70.74	11.50	6.10	3.64	2.84	2.52	2.66	29.26
	1.5	70.68	11.56	6.08	3.68	2.86	2.52	2.62	29.32
					Nested χ^2 Method				$\hat{\lambda}$
0.10	0.5	0.955	0.977	1.034	1.104	0.910	1.047	1.010	0.961
	1.0	0.955	0.983	1.015	1.046	0.943	1.056	1.020	0.961
	1.5	0.958	0.988	0.995	1.030	0.931	1.040	1.016	0.963
0.15	0.5	0.959	0.937	0.990	1.061	0.930	1.051	0.958	0.962
	1.0	0.958	0.943	0.974	1.017	0.982	1.039	0.958	0.961
	1.5	0.961	0.943	0.980	1.000	0.979	1.026	0.963	0.963
0.20	0.5	0.960	0.950	0.953	0.998	0.942	0.967	0.988	0.961
	1.0	0.959	0.964	0.916	1.013	0.960	0.974	0.993	0.961
	1.5	0.962	0.962	0.932	1.000	0.943	0.963	0.991	0.962
0.25	0.5	0.956	0.975	0.929	1.006	0.916	0.964	1.007	0.960
	1.0	0.959	0.968	0.932	0.984	0.942	0.966	0.992	0.960
	1.5	0.961	0.962	0.956	0.979	0.933	0.958	0.990	0.962
0.30	0.5	0.957	0.972	0.936	1.004	0.935	0.956	0.969	0.959
	1.0	0.958	0.976	0.947	0.991	0.947	0.955	0.968	0.960
	1.5	0.959	0.972	0.957	0.992	0.950	0.945	0.968	0.961
					AIC				
	0.5	0.961	0.964	0.931	0.996	0.923	0.973	0.993	0.961
	1.0	0.959	0.972	0.930	0.984	0.949	0.984	0.977	0.960
	1.5	0.961	0.971	0.942	0.974	0.948	0.980	0.975	0.962

Replicated Monte Carlo Experiments by 5,000 Times.

4. Discussion

The simulation results of fitting the BCT model (2-3) to data generated by the 2nd order polynomial regression model (3-1) for the case of $\sigma_0^2 = 1.0$ are listed in table 3.2. The critical value $\alpha = 0.15$ is used in the nested x^2 test. The distributions of the estimated frequencies in 5000 times experiments by the nested x^2 method and the AIC method are listed in two blocks, respectively. It shows that both of the two methods have good performance of fitting the BCT polynomial regression model to the data generated by polynomial regression model. In the 5000 times replicated experiments by the nested x^2 test, there are 1237 times that just fitting the true model (3-1), and there appears 1183 times by the AIC method. As far as the identification of the true 2nd order model, it shows 3292 times in 5000 replicated experiments by the nested x^2 method and 3168 times in 5000 replicated experiments by the AIC method. Each of the above four frequencies takes the maximum in the case it is discussed. The frequencies of fitting the true model or identifying the true model order by the nested x^2 test are larger than that of by the AIC. In view of the nested x^2 test, the percentage of the overestimated frequency those exceeding the true model order is 16.88% (approximately equals to the critical value $\alpha = 0.15$). But the percentage of the overestimated frequency by the AIC is 28.12%, which is significantly higher than that of determined by the nested x^2 test. As for the estimation of the BCT parameter, the AIC seems better than the nested x^2 test but the difference is not too large. In view of the AIC method, the estimate of the BCT parameter is 1.01, this is the weighted mean determined by the frequencies against the identified true model order 2, $N(\lambda_i, 2)$. It is better than that of determined by the nested x^2 method, which is 1.03. If we choose the BCT parameter by the weighted mean determined by $N(\lambda_i)$, it shows the result of 0.98 by the AIC method, and 0.97 by the nested x^2 method. In the whole, for comparatively small critical values ($\alpha < 0.25$), the nested x^2 method is superior to the AIC method in the identification of the true model order, but it is the reverse in the estimate of the BCT parameter. This conclusion is coincidence with the result obtained by Yao (1995).

Table 3.3 shows the simulation result of fitting the BCT model (2-3) to data generated by the exponential regression model (3-2) for the case of $\sigma_0^2 = 1.0$ and the critical value $\alpha = 0.15$ is also used in the nested x^2 test. An observation of the two blocks tells us that the performances of the two methods in model identification are very similar to the case where fitting the data generated by model (3-1) as discussed above. The nested x^2 test suggests us to choose the 2nd order BCT polynomial regression model with the BCT parameter 1.01 to fit model (3-2). The AIC suggests us to choose the 2nd order BCT polynomial regression model with the BCT parameter 0.99 to fit the same model. Table 3.4 shows the distributions of the estimated frequencies by the nested x^2 test and the AIC, for fitting the BCT model (2-3) to data generated by the specified logistic regression model (3-3). For the case of variance of the disturbance term $\sigma_0^2 = 1.0$, both the two methods suggest us to use the order 1 BCT polynomial regression model with the BCT parameter 0.96 to fit the data set generated by model (3-3). In view of the nested x^2 method, the estimated frequency 4143 suggests us to choose order 1 BCT model. This estimated frequency is highly larger than that of estimated by the AIC method (3537). In this meaning we can say the power of the nested x^2 test is higher than that of the AIC in this case.

The plots in figure 3.1.1, for five levels of critical α respectively, give details of the estimated frequencies in fitting the BCT polynomial regression model (2-3) to data generated by nonlinear model (3-1) with disturbance term variance of $\sigma_0^2 = 0.5$. The last graph is the result given by the AIC method. The simulation performances for different critical values can be observed by the peak of the column and the symmetry against the BCT parameter $\lambda\,(\lambda = -1, 0, \cdots, 3)$. In the identification of the BCT regression model order in view of the nested x^2 test, it is most powerful when choosing the lower critical value 0.10. The power will be degenerate as the critical value increased. The symmetry of the frequency distribution against the BCT parameter tells us that the high level of critical value gives good estimate of the BCT parameter. The performance of the AIC method is almost the same as the nested x^2 method when choosing critical value 0.25. Figure 3.1.2 and figure 3.1.3 give the results for the cases the disturbance term variance of $\sigma_0^2 = 1.0$, 1.5, respectively. The performances of the two methods in the identification of the BCT model can be compared by the plots with the same critical value in the three figures. An observation of these figures shows that both of the two methods tends to underestimate the order of the BCT regression model as the disturbance term variance increased. The nested x^2 test is more sensitive to the changes of the variance.

The plots in figure 3.2.1, for five levels of critical α respectively, give the details of the estimated frequencies in fitting the BCT polynomial regression model to data generated by exponential model (3-2). It is the case of the disturbance term variance of $\sigma_0^2 = 0.5$. Figure 3.2.2 and figure 3.2.3 give the same simulation results for the case the disturbance term variance of $\sigma_0^2 = 1.0$, 1.5, respectively. The properties are very similar to the above discussions on figure 3.1.1 to figure 3.1.3. For both of the data sets generated by nonlinear model (3-1) and exponential model (3-2), in the case of the disturbance term variance is large, for example $\sigma_0^2 = 1.5$, the performance of the AIC in the estimate of the true order seems better than that of the nested x^2 test. This can observed by figure 3.1.3 and figure 3.2.3, as the peak of the column at the 2nd order with BCT parameter $\lambda = 1$ determined by the AIC is higher than that of determined by the nested x^2 test for all the given different given critical values.

The plots in figure 3.3.1, for five levels of critical value α respectively, give the details of the estimated frequencies in fitting the BCT polynomial regression model (2-3) to data generated by the logistic model (3-3) for the case of the disturbance term variance of $\sigma_0^2 = 0.5$. The simulation result in view of the AIC is listed at the last of this figure. Figure 3.3.2 and figure 3.3.3 give the simulation results for the case of the disturbance term variance of $\sigma_0^2 = 1.0$, 1.5, respectively. The observation of the three figures show that there is almost no difference between the two methods in estimate of the BCT parameter. All the cases suggest choosing the 1st order BCT polynomial regression model to fit the logistic model (3-3). The power of estimating the simulation model order by the nested x^2 test with lower critical values is higher than that of the AIC. The performance of the AIC method is almost the same as the nested x^2 method with critical value of $\alpha = 0.25$ for all the three cases of model (3-3) specified by $\sigma_0^2 =$ 0.5, 1.0, 1.5.

The upper block in table 3.5 shows that the nested x^2 test can give a good identification of the true 2nd order in fitting the data generated by model (3-1). The changes of the disturbance term variance only give effects on the distribution of the estimated frequencies that the identified order less than the true order. The larger the disturbance term variance is, the smaller the estimated percentage at the true order is. It is the reverse for the percentage of the underestimated order. Furthermore, the percentage of the estimated orders those exceeding the true order is almost fixed no matter how the disturbance term variance moved. From table 3.5 we can see that as the critical value α increases, the estimated frequency of the identified order less than the true order decreases, the frequency of the identified order exceeding the true order increases. From this table we can also see that the percentage of the frequencies exceeding the true order increases with the significance level increases. The value of $d(\alpha, \sigma^2)$ is very near to the given critical value α. This result is just in accordance with what the P-value implies. The AIC method tends to overestimate the model order. By choosing the critical value, the nested x^2 test holds the good property of controlling the levels of overestimate the model order. We can also see that for comparatively small critical value (for example, $\alpha \leq 0.2$), the nested x^2 test is better than the AIC in identifying the order of the BCT polynomial regression model. The above result is coincidence with the conclusion that we have got in the earlier investigations [see Yao (1996*b*)]. The estimated results of the BCT parameter listed in the lower block show that except for the case of order 1, the estimated $\hat{\lambda}_p$ for order p $(p = 2, \cdots, 7)$ and $\hat{\lambda}$, which being estimated by the total 5000 times experiments, are very near to 1. The disturbance term variance has not significant effects on the estimate of the BCT parameter. There is no significant difference between the nested x^2 method and the AIC method.

As for the case of underestimating the true order of the BCT model, the estimated BCT parameter seems to make good compensate for the information loss by the lower estimated order. The virtues of simultaneous identification of the BCT regression model by the use of the nested x^2 test or the AIC can be seen here. In case of underestimate the true order, the estimate of the BCT parameter is very sensitive to the disturbance term variance. This conclusion is true for both the nested x^2 method and the AIC method.

Table 3.6 show the estimates of the BCT parameter in fitting BCT regression model (2-3) to data generated by the exponential regression model (3-2). The pattern of the percentage distribution of the estimated order and the BCT parameter are very similar to the results showed by table 3.5. The simulation result suggest us to use the 2nd order BCT polynomial regression model with the BCT parameter $\hat{\lambda}_2 \approx 0.98$ to fit model (3-2). The estimated BCT parameters we get here for the three levels of the variance are in the interval of (0.96, 1.0). Furthermore, it can be seen that for each of the given critical value α, for all the three levels of the disturbance term variance especially for the lower variance, the estimate of $d(\alpha, \sigma^2)$ is much higher than α. It seems, to some extent, in conflict with the property of the nested x^2 test. It may be explained by the fact that the divergence between the BCT polynomial regression model and the exponential regression model is too large. The properly explanation need a further investigation and we left it as an open question. Table 3.7 lists the experiment results in fitting the BCT polynomial regression model (2-3) to data generated by the logistic regression model (3-3). In view of the distribution of the estimated order in 5000 replication experiments for three levels of the variance, we can see that the best fitted model is the 1st order BCT polynomial regression model. The variance of the disturbance term almost has no effect on the estimate of the BCT parameter. Both of the two methods suggest to choose the BCT parameter $\hat{\lambda}_1 \approx 0.96$.

5. Conclusions and Remarks

To investigate the basic characteristics of the nested x^2 test and the AIC in the application of fitting the BCT polynomial regression model to data generated by nonlinear models, which do not belong to the family of the BCT polynomial regression models, we conduct 5000 replicated simulation experiments for three nonlinear models, a 2nd order polynomial regression model and an exponential regression model as well as a logistic regression model, respectively. For each of the models, we discuss three levels of the variance of disturbance term, and also five levels of the critical values used in the nested x^2 test. The simulation results capture quite well characteristics of the nested x^2 test and the AIC.

In fitting the BCT polynomial regression model (2-3) to data generated by models that do not belong to the family of the BCT polynomial regression models, simulation results show that both of the two methods play good performances in model identification. The nested x^2 test has the ability to control the probability of the identified orders those exceeding the true order, but it tends to underestimate the true order, especially for the case with a large variance of disturbance term. To avoid overestimating the order of the BCT regression model, we suggest to choose comparatively small critical value, or to choose larger critical value on the reverse. The AIC method in general tends to overestimate the true order of the model. As for the case of underestimating the true order of the BCT regression model, the properly estimated BCT parameter seems to make compensate for the information loss by the underestimated lower order. This is true for both of the two methods. For comparatively small critical values, the nested x^2 test plays good performance than the AIC in the estimation of the true model order. Furthermore, simulation results show that both of the two methods suggest using the 2nd order BCT polynomial regression model with the BCT parameter $\hat{\lambda}_2 \approx 0.98$ to fit the exponential regression model (3-2), and using the 1st order BCT polynomial regression model with the BCT parameter $\hat{\lambda}_1 \approx 0.96$ to fit the logistic regression model (3-3).

In the cases of fitting data set generated by exponential regression model, it is seen that for all the given critical value α and for the three levels of the variance of disturbance term, especially for the case with lower variance, the percentage of the estimated frequencies those exceeding the true order is much higher than 100 α. This result seems in conflict with the property of the nested x^2 test. We will left it as an open question. The conclusions reached in this paper should be tempered for the stochastic specifications which have been made and the general nature of Monte Carlo experimentation. The robustness of our conclusions, seems difficult to prove but very important, should also be discussed.

It is so regret to say that, for some constraints on the computation, in this paper the concrete simulation models are not presented. The discussion based on the concrete simulation models may give good contributions to the comparison of the nested x^2 test and the AIC. Also the mixed method based on the nested x^2 test and the AIC [see Yao (1995)] may play an important role in the model identification. Our forthcoming paper will discuss these issues.

Figure 3.1.1 Distribution of the Estimated Frequencies for Fitting Model (2-3) to Data Generated by Model (3-1) in 5000 Times Replicated Experiments

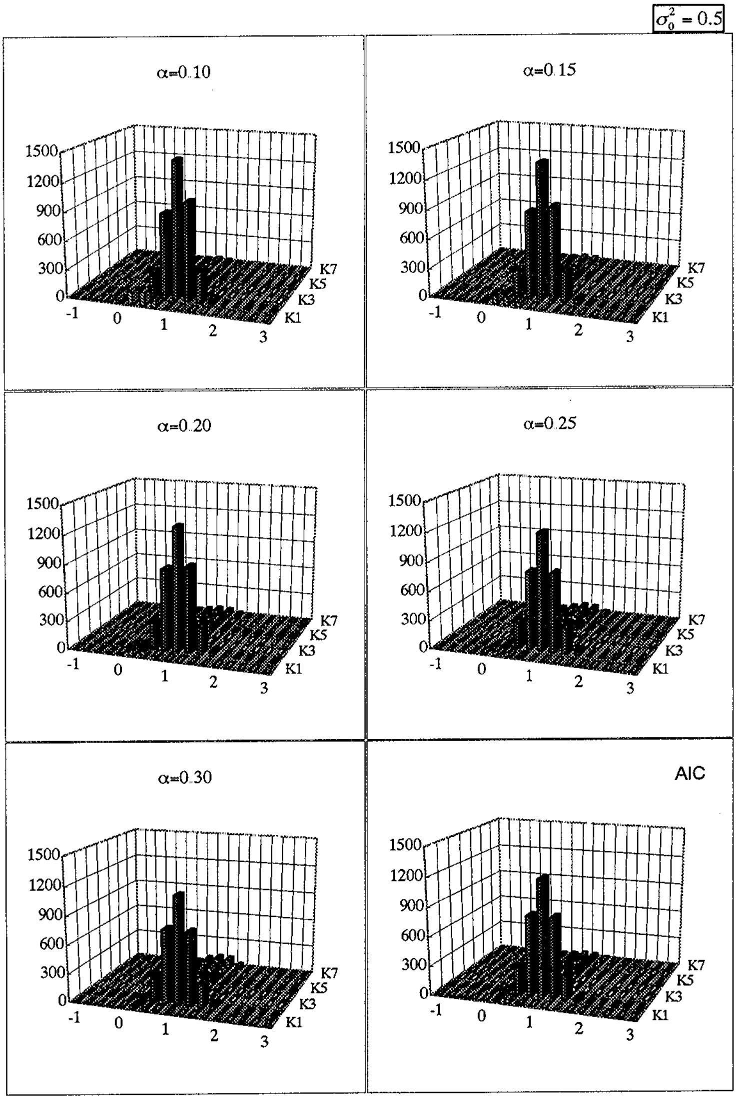


Figure 3.1.2 Distribution of the Estimated Frequencies for Fitting Model (2-3) to Data Generated by Model (3-1) in 5000 Times Replicated Experiments

$\sigma_0^2 = 1.0$

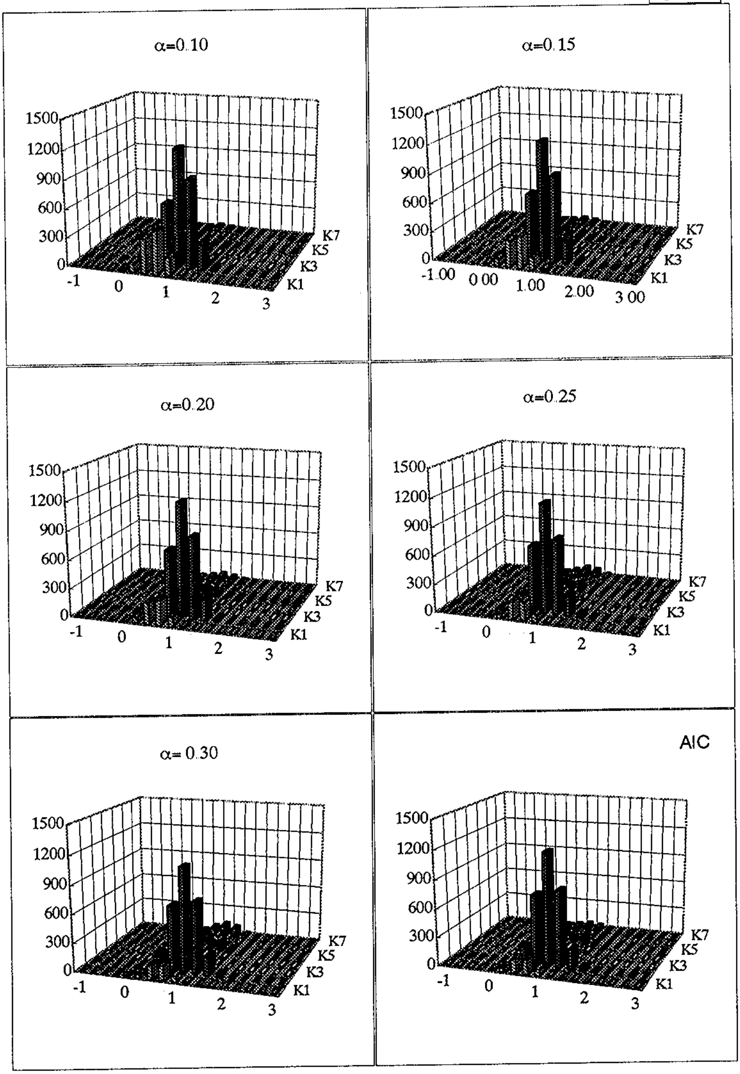


Figure 3.1.3 Distribution of the Estimated Frequencies for Fitting Model (2-3) to Data Generated by Model (3-1) in 5000 Times Replicated Experiments

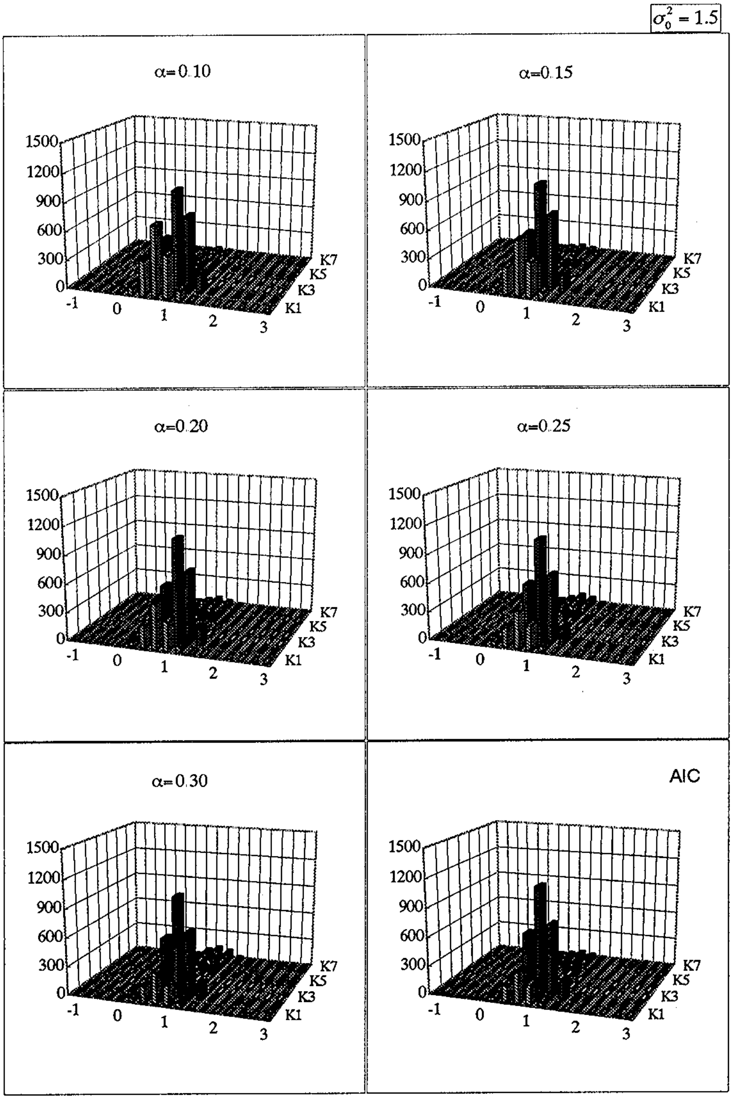


Figure 3.2.1 Distribution of the Estimated Frequencies for Fitting Model (2-3) to Data Generated by Model (3-2) in 5000 Times Replicated Experiments

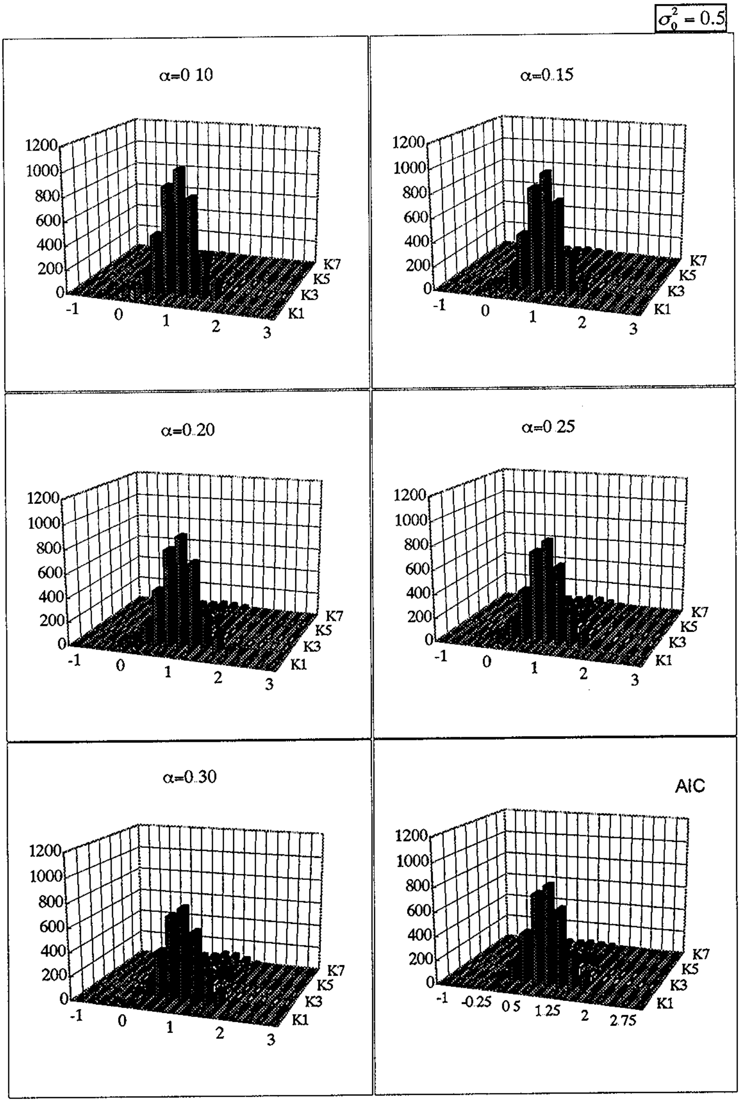


Figure 3.2.2 Distribution of the Estimated Frequencies for Fitting Model (2-3) to Data Generated by Model (3-2) in 5000 Times Replicated Experiments

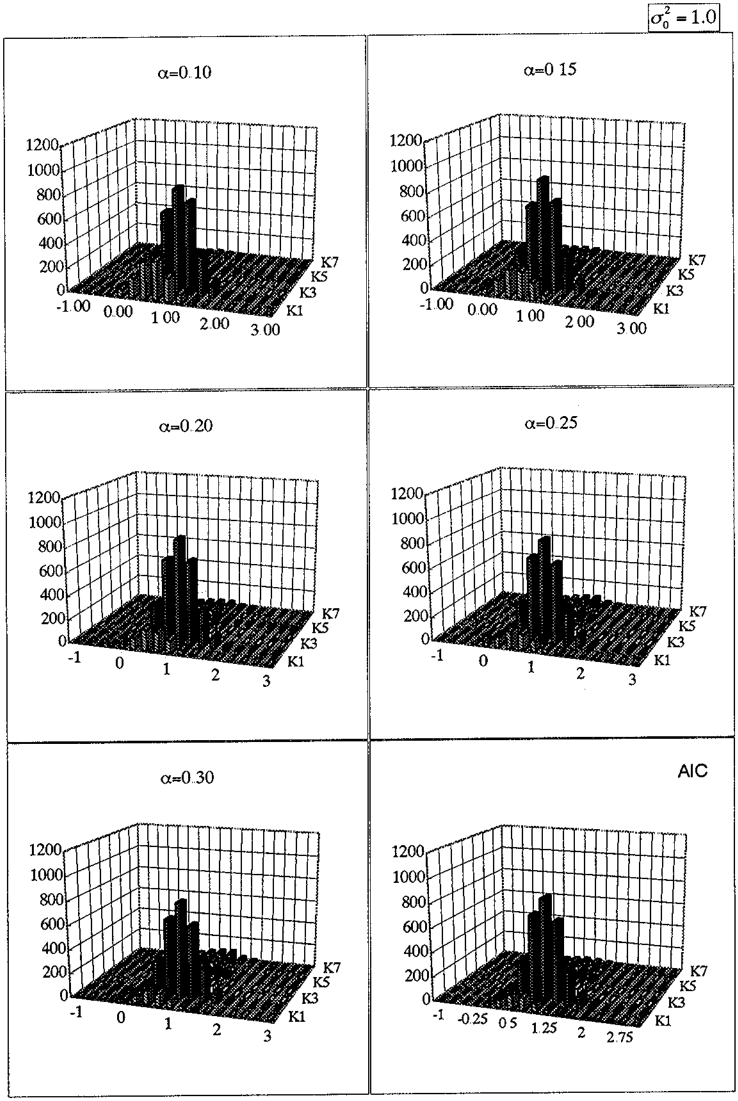


Figure 3.2.3 Distribution of the Estimated Frequencies for Fitting Model (2-3) to Data Generated by Model (3-2) in 5000 Times Replicated Experiments

$\sigma_0^2 = 1.5$

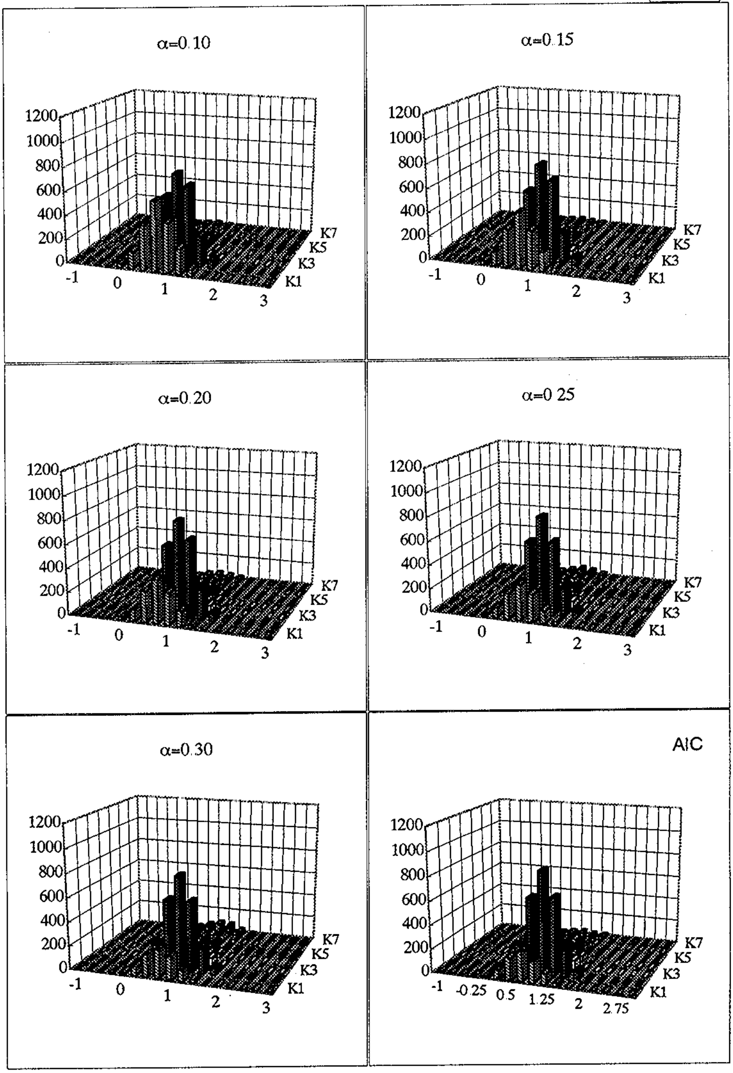


Figure 3.3.1 Distribution of the Estimated Frequencies for Fitting Model (2-3) to Data Generated by Model (3-3) in 5000 Times Replicated Experiments

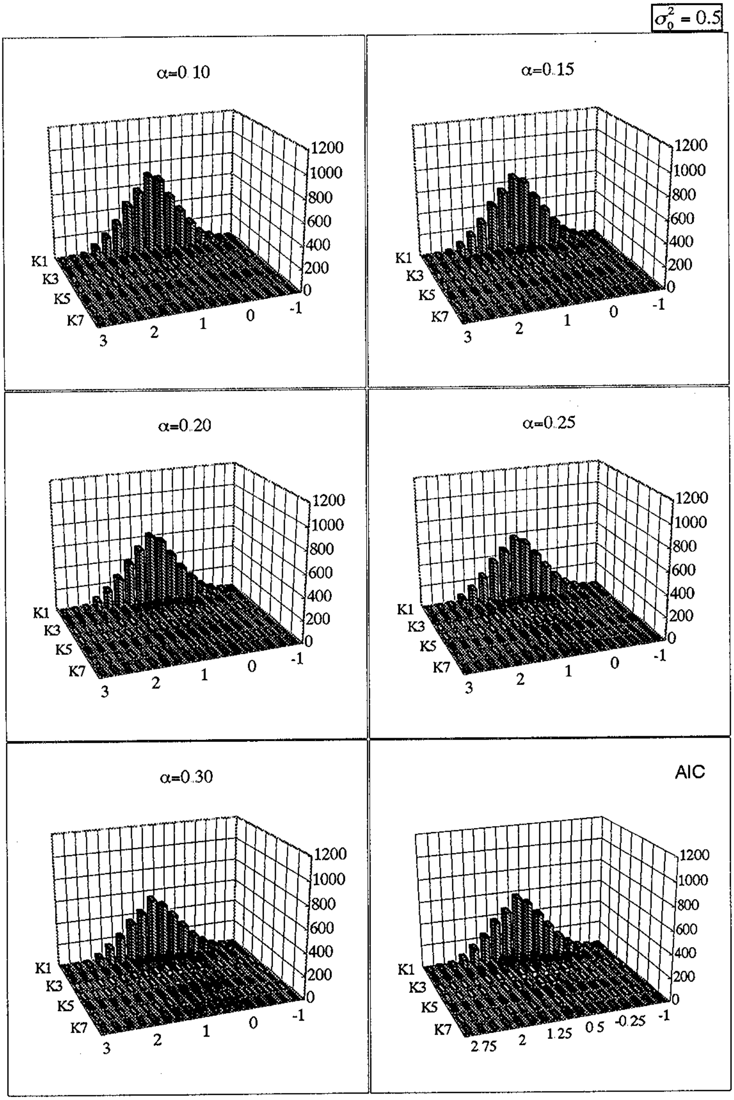


Figure 3.3.2 Distribution of the Estimated Frequencies for Fitting Model (2-3) to Data Generated by Model (3-3) in 5000 Times Replicated Experiments

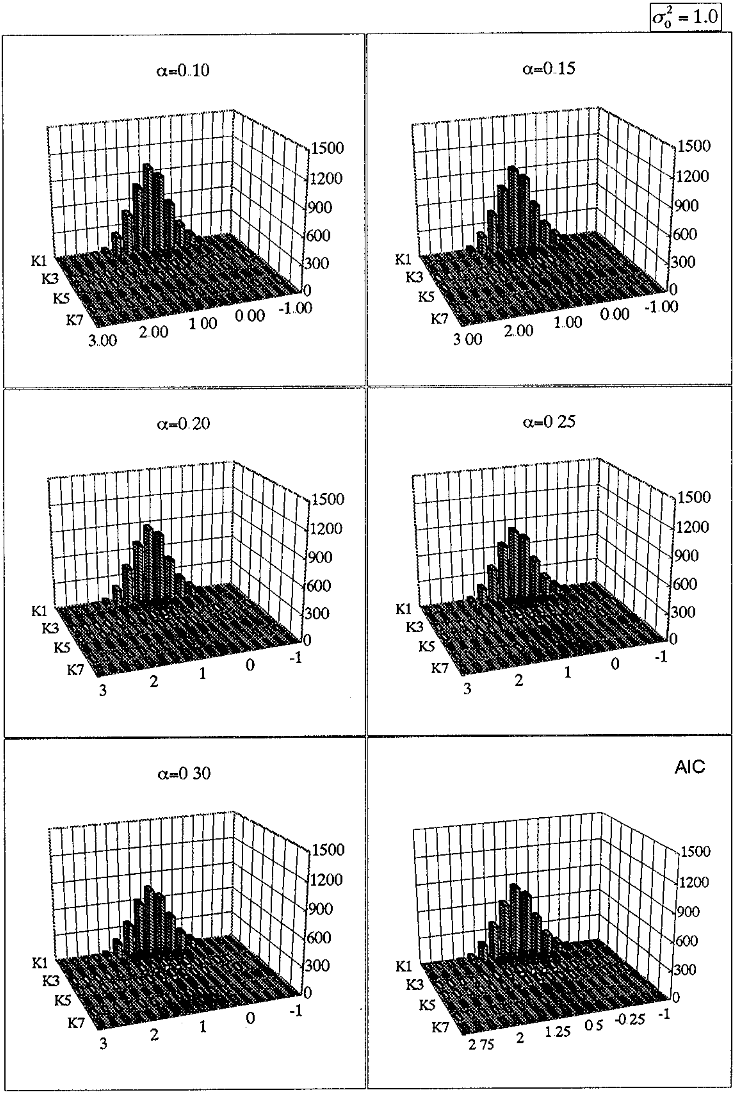
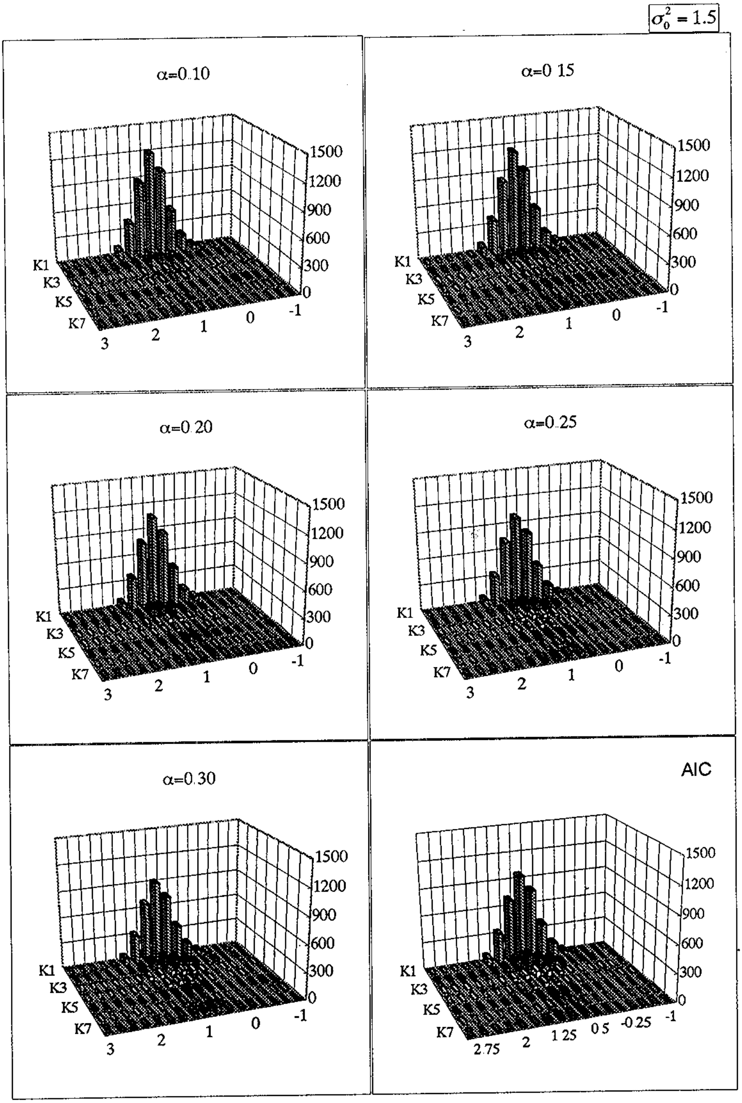


Figure 3.3.3 Distribution of the Estimated Frequencies for Fitting Model (2-3) to Data Generated by Model (3-3) in 5000 Times Replicated Experiments



References

Akaike, H., (1973). "Information theory and an extension of the maximum likelihood principle", *2nd International Symposium on Information Theory*, Eds. B. P. Petrov and F. Csaki, 267-281, Budapest: Akademia Kiado.

Bickel, P. J and Doksum, K. A., (1981), "An analysis of transformations revisited", *Journal of the American Statistical Association,* Vol. 76, 296-311.

Box, G. E. P., and Cox, D. R., (1964). "An analysis of transformations", *Journal of the Royal Statistical Society* B26, 211-43.

——(1982). "An analysis of transformations revisited, rebutted", *Journal of the American Statistical Association,* Vol. 77, 209-10.

Chang, H. S., (1977). "Functional forms and the demand for meat in the United States", *The Review of the Economics and Statistics* Vol. 59, 355-9.

Chang, H. S., (1980). "Functional forms and the demand for meat in the United States: A Reply", *The Review of the Economics and Statistics* Vol. 62, 148-50.

Hosoya, Y., (1983), "Some results related to model selection based on estimated risk", *Discussion Paper* No. 44, Faculty of Economics, Tohoku University.

——(1984). "Information criteria and tests for time-series models", *Time Series Analysis: Theory and Practice 5,* 39-50, Ed. Anderson O. D., Elsevier Science Publishers, Amsterdam.

——(1986). "A simultaneous test in the presence of nested alternative hypotheses", *Journal of Applied Probability,* 23A, 187-200.

——(1989). "Hierarchical statistical models and generalized likelihood ratio test", *Journal of Royal Statistical Society,* B. 51, 435-47.

Hosoya, Y. and Katayama, S. (1987). "A p-value algorithm for a nested x^2 test", *Annual Report of the Economic Society,* Tohoku University, Vol. 49, No. 1, 83-90.

Huang, G. J. and Grawe, O. R., (1980). "Functional forms and the demand for meat in the United States: A Comment", *The Review of the Economics and Statistics* Vol. 62, 144-6.

James, B. R. and Davied, S. M., (1982). "Analysis of income and food expenditure distributions: A Flexible Approach", *The Review of the Economics and Statistics* Vol. 64, 104-9.

Poirier, D. J., (1978). "The use of the Box-Cox transformation in limited dependent variable models", *Journal of the American Statistical Association,* Vol. 73, 284-7.

Poirier, D. J. and Melino, A., (1978). "A note on the interpretation of regression coefficients within a calss of truncated distributions", *Econometrica* 46, 1207-9.

Sakamoto, Y. and Isiguro, M. and Kitagawa, G., (1986). *Akaike information criterion statistics,* D. Reidel Publishing Company, Tokyo.

Seaks, T. G. and Layson, S. K., (1983). "Box-Cox estimation with standard econometric problems", *The Review of the Economics and Statistics* Vol. 65, 160-4.

Spitzer, J. J., (1982). "A fast and efficient algorithm for the estimation of parameters in models with the Box-Cox transformation", *Journal of the American Statistical Association,* Vol. 77, 760-6.

Takeuchi, K., (1976), "The distribution of information statistic and the criterion of fitness of model", *Suri Kagaku,* (in Japanese), Vol. 153, 12-18.

Tse, Y. K., (1984). "Testing for linear and log-linear regressions with heteroscedasticity", *Economics Letters* Vol. 16, 63-9.

Yao, F., (1992). "Analysis of the Box-Cox transformation with general information criterion and Akaike information criterion", *Master Degree Thesis,* Graduate School of Economics, Tohoku University.

——(1994). "Analysis of the Box-Cox transformation with general information criterion and Akaike information criterion", *Annual Report of the Economic Society,* Tohoku University Vol. 56, No. 1, 123-39.

——(1995). "Information criteria and the nested x^2 method for the Box-Cox data transformation", *Annual Report of the Economic Society,* Tohoku University Vol. 57, No. 2, 85-99.

——(1996a). "Econometric analysis of nonlinear and nonstationary relationships: Inference and Computational Methods", *Working Paper* No. 17, the Institute of Economic Research, Kagawa University, April 1996.

——(1996b). "Investigation of information criteria in the Box-Cox transformation model", *Kagawa University Economic Review* Vol. 69, No. 2-3, 267-95.

Yao, F. and Hosoya, Y. (1994). "Statistical model identification in the Box-Cox data transformation", *Discussion Paper* No. 109, Faculty of Economics, Tohoku University.

Zarembka, P. (1990). "Transformation of variables in econometrics", *Econometrics,* 261-64, Eds. Eatwell, J., Milgate, M., Newman, P., the Macmillan Press Ltd.