COMM. STATIST.-THEOR. METH., 14(2), 363-379 (1985)

BOX-COX TRANSFORMATION AND THE PROBLEM OF HETEROSCEDASTICITY

Nityananda Sarkar

Indian Statistical Institute
203 B. T. Road, Calcutta 700 035

Key Words and Phrases : *Box-Cox model; functional form; heteroscedasticity; nonnormality; power transformation*

ABSTRACT

It is known that heteroscedasticity in the context of the well-known family of power transformations suggested by Box and Cox creates complications because it depends upon the heteroscedasticity of the original values of the dependent variable (in a regression model set-up) as well as upon the transformation parameter. In this paper we attempt at generalizing the original Box-Cox model in this direction, and suggest maximum likelihood method of estimation for the parameters of the model. Through illustrative examples we show the seriousness of the problem of heteroscedasticity in the context of this transformation, and indicate how one can separate out the problem of non-linearity from the influence of stabilization of error variance in an estimate of the transformation parameter in our generalized model.

0361-0926/85/1402-0363$3.50/0

1. INTRODUCTION

In a well-known paper Box and Cox (1964) proposed a family of power transformations given by $y^{(\lambda)} = (y^{\lambda} - 1)/\lambda$ if $\lambda \neq 0$, and $\ln y$ if $\lambda = 0$, of the dependent variable (in a regression model) so as to achieve (i) a linear relationship among the transformed dependent variable and the set of fixed regressors, and (ii) homoscedasticity and (iii) normality of the transformed dependent variable. This transformation, often referred to as Box-Cox (BC) transformation, has been extensively used; and currently there is a renewed interest in the theoretical problems relating to this transformation [see, for instance, Bickel and Doksum (1981), Box and Cox (1982), Spitzer (1982a, 1982b)]. In their paper Box and Cox also suggested maximum likelihood (ML) method of estimation for estimating the parameters of the model. Subsequently, Draper and Cox (1969) and Zarembka (1974) examined the extent to which all the three desirable properties hold simultaneously. While Draper and Cox found that Box-Cox precedure of estimation is robust to nonnormality so long as the disturbances in the regression model have reasonably symmetric distributions, Zarembka has shown that the method is not robust with respect to heteroscedasticity. More specifically, he has shown that there is a bias in estimating λ towards that transformation of the dependent variable which leads to stabilization of the error variance; λ and hence other parameters are consistently estimated only when the transformation that leads to linearity also leads to heteroscedastic error variance. Zarambka has also outlined a method of estimation for λ and other parameters of a (transformed) linear regression model under the assumption that $V(y_i) = \sigma^2 [E(y_i)]^h$, $E(y_i) > 0$ and h known. Recently Lahiri and Egy (1981) and Egy (1982) reconsidered the BC model under condition of heteroscedasticity given by $V(y_i^{(\lambda)}) = \sigma^2 m_i^{\delta}$ where m_i's are exogenously given and σ^2 and δ are unknown parameters.

It may be pointed out that Zarembka's model assumes h to be known, and that his estimation method yields only consistent but not efficient estimates of the parameters involved. For the generalized BC model (by Lahiri and Egy) m_i's are exogenously given, and the form of heteroscedasticity does not involve λ. Clearly, a straightforward assumption about $V(y_i^{(\lambda)})$ without taking into consideration the complications created by the transformation does not seem to be quite satisfactory. Zarembka has, in fact, noted this in ''... the problem is complicated by the fact that the heteroscedasticity in ε_i (the disturbance term) depends upon the heteroscedasticity in y_i as well as upon the unknown parameter λ'' (1974, p. 173).

In the light of the above observation a closer examination as to how the transformation affects the nature of heteroscedasticity of $y_i^{(\lambda)}$'s becomes quite important. Since in cross-sectional studies heteroscedasticity in the errors is often due to either heteroscedastic disturbances or estimation of incorrect functional form or both, the generalization of BC model incorporating proper heteroscedastic structure with varying degrees of heteroscedasticity is called for so that simultaneous testing for heteroscedasticity and functional form is possible. An attempt towards this direction is being made in this paper. The problem of heteroscedasticity in the context of the family of power transformations used by Box and Cox is discussed in the next section. The estimation method is described in section 3. Illustrative examples are given in section 4 to show that the use of standard BC model can produce misleading conclusions. This is followed by conclusions in section 5.

2. BOX-COX TRANSFORMATION AND THE PROBLEM OF HETEROSCEDASTICITY

The usual Box-Cox (BC) model is

$$y^{(\lambda)} = X\beta + \varepsilon \qquad \ldots (1)$$

where $y^{(\lambda)}$ is a $(n \times 1)$ vector of observations on the dependent variable with i-th element being $y_i^{(\lambda)} = (y_i^{\lambda} - 1)/\lambda$ if $\lambda \neq 0$, and $\ln y$ if $\lambda = 0$; X is a $(n \times k)$ matrix with rank k of observations on k regressors; β is the $(k \times 1)$ vector of associated regression coefficients and ε is a $(n \times 1)$ vector of disturbances. The elements of the first column of X are assumed to be all unity [cf. Schlesselman (1971)]. Box and Cox assumed that ε_i's are independently normally distributed with mean zero and variance σ^2 across observations. It may be stated here that all the dependent variables in X can as well be power transformed with the same value of λ or with different values of λ for different variables. This generalization would not, however, give rise to any additional problem than those arising in the specification in (1), nor would it change any of the essential conclusions.

Let us now look at the problem of heteroscedasticity more carefully. Using the well-known approximation

$$V(g(y_i)) \simeq V(y_i) \left[\frac{\partial g(y_i)}{\partial y_i}\right]^2_{y_i=E(y_i)}$$

for any function $g(y_i)$ of y_i, we have from the definition of $y^{(\lambda)}$,

$$V(y_i^{(\lambda)}) \simeq V(y_i)\ (E(y_i))^{2\lambda-2}, \quad i = 1, 2, \ldots, n \quad \ldots (2)$$

where $V(.)$ denotes the variance of the relevant variable. It follows from (2) that the value of λ that linearizes need not necessarily lead to homoscedasticity of $y_i^{(\lambda)}$'s even when the original observations are homoscedastic. The possible heteroscedasticity in ε_i is, therefore, expressed through $V(y_i)$, $E(y_i)$ and λ. We now make a specific assumption about the form of $V(y_i)$ viz.,

$$V(y_i) = \sigma^2\ (E(y_i))^h,\ E(y_i) > 0,\ i = 1, 2, \ldots, n \quad \ldots (3)$$

where h is unknown. This is quite a general form of heteroscedasticity, and is in line with those of Box and Hill (1974) who considered $|E(y_i)|$ rather than $E(y_i)$. In the context of multiple regression equations and random coefficient models such an assumption about the form of heteroscedasticity is quite appropriate. We thus have using (3) and assuming σ to be small i.e., $\sigma \to 0$ [cf. Bickel and Doksum (1981), Draper and Cox (1969) and Zarembka (1974)], from (2)

$$V(y_i^{(\lambda)}) \simeq \sigma^2 \; (E(y_i))^{2\lambda-2+h}. \qquad \ldots(4)$$

By using Taylor expansion around $E(y_i^{(\lambda)})$ and thereby approximating $E(y_i)$ in terms of $E(y_i^{(\lambda)})$, $V(y_i^{(\lambda)})$ in (4) easily reduces to

$$V(y_i^{(\lambda)}) = \sigma^2 \left[H(\lambda,\ x_i'\beta)\right]^{\zeta} \qquad \ldots(5)$$

where $H(\lambda,\ x_i'\beta) \simeq (1 + \lambda\ x_i'\beta)^{1/\lambda}$ for $\lambda \neq 0$, and $\exp(x_i'\beta)$ for $\lambda = 0$; $\zeta = (2\lambda - 2 + h)$, $x_i'\beta = E(y_i^{(\lambda)})$ and x_i' is the i-th row of X. The final form of $V(y_i^{(\lambda)})$ therefore becomes

$$V(y_i^{(\lambda)}) = \begin{cases} \sigma^2 \left[1 + \lambda\ x_i'\beta\right]^{\delta}, & \lambda \neq 0 \\ \sigma^2 \left[\exp(x_i'\beta)\right]^{\delta}, & \lambda = 0 \end{cases} \qquad \ldots(6)$$

where δ is assumed to be in a compact set and is equal to $(2\lambda - 2 + h)/\lambda$ for $\lambda \neq 0$ and $(h - 2)$ for $\lambda = 0$. The quantity within the third bracket is assumed to be positive and bounded away from zero. Thus, given that y_i's are heteroscedastic of the type considered, $y_i^{(\lambda)}$'s will have constant variance across observations only when $h = 2 - 2\lambda$ which, of course, is a very special case, although this might hold for more than one combination of values of λ and h. It may be noted here that the parameters of interst in our model, as given by (1) and (6), are, apart from β and σ^2, λ and δ.

We can, therefore, conclude from (6) that, in general, the power transformed dependent variable will have different variances across observations irrespective of whether original observations i.e., y_i's were homoscedastic or not, and that the structure of variance will, in addition, to the parameters involved in the assumed form of heteroscedasticity for y_i's, also involve λ.

3. ESTIMATION

In this section we briefly outline the estimation procedure for BC model with heteroscedastic disturbances, henceforth to be referred to as Box-Cox heteroscedastic (BCH) model, given by (1) and (6). While one can assume other forms of heteroscedasticity for $V(y_i)$ so that $V(y_i^{(\lambda)})$ would change, the estimation procedure, in principle, remains the same. Also, a power transformation, as originally pointed out by Poirier (1978), does not in general permit large negative values and hence, we assume, like others [e.g., Egy (1982), Lahiri and Egy (1981), and Spitzer (1978, 1982a, 1982b) to cite a few], that the probability for large negative values of the disturbance term is small so that the assumption of normality is not seriously affected.

Thus, under the assumption of normality of ε_i's, the log-likelihood function for BCH model becomes

$$L(\lambda, \beta, \sigma^2, \delta | y, X) = \text{const.} - \frac{n}{2} \ln\sigma^2 - \frac{1}{2} \ln |\Omega|$$
$$- \frac{1}{2\sigma^2} (y^{(\lambda)} - X\beta)' \, \Omega^{-1} \, (y^{(\lambda)} - X\beta) + \ln |J| \quad \ldots (7)$$

where Ω is a $(n \times n)$ diagonal matrix whose i-th diagonal element Ω_{ii} is equal to $(1 + \lambda x_i'\beta)^{\delta}$ for $\lambda \neq 0$ and $\exp(x_i'\beta)$ for $\lambda = 0$, and $|J|$ is the Jacobian of transformation.

Maximisation of (7) with respect to all the parameters is evidently very complicated because of the non-linearities involved in the log-likelihood function. While maximum likelihood (ML) estimates can be obtained by using any of the standard techniques for maximizing non-linear equations (see, Judge et al. (1980, Ch. 17) for detail discussion on these methods), one can as well use search techniques. For a given value of (λ, δ) pair, one first uses ordinary least squares (OLS) method to get an estimate of β, which together with the given value of (λ, δ) is used to obtain an estimate of Ω, say $\hat{\Omega}(\lambda, \delta)$, which in turn is used to obtain generalised least squares (GLS) estimates of β and σ^2, say $\hat{\beta}(\lambda, \delta)$ and $\hat{\sigma}^2(\lambda, \delta)$. The concentrated log-likelihood function therefore becomes

$$L(\lambda, \delta | y, X) = \text{const.} - \frac{n}{2} \ln \hat{\sigma}^2(\lambda, \delta) - \frac{1}{2} \ln|\hat{\Omega}(\lambda, \delta)| + (\lambda - 1) \Sigma \ln y_i \; . \qquad \ldots(8)$$

Further iterations with these new estimates are done till $L(\lambda, \delta | y, X)$ in (8) attains maximum for the fixed (λ, δ) value. The ML estimators of λ and δ and hence of β and σ^2 are then obtained by searching for values of (λ, δ) over a reasonable range until the maximum value of the log-likelihood function is attained. The essential logic in concluding that this procedure of estimation would result in ML estimates are derived from Dhrymes (1970, Ch. 3), Judge et al. (1980, Ch . 4), Maddala (1971, Ch. 12), Oberhofer and Kmenta (1974) and Spitzer (1982a).

Since the search procedure stated above is two-dimensional and hence the number of required regressions too large, we suggest a procedure [cf. Coondoo and Sarkar (1979)] where a systematic search over λ only is needed. Suppose for a given λ (without loss of generality $\lambda \neq 0$), $\hat{\beta}(\lambda) = (X'X)^{-1}X'y^{(\lambda)}$ is OLS estimator of β for the given λ. Now

$$\left.\frac{\partial L(.)}{\partial \sigma^2}\right|_{\beta = \hat{\beta}(\lambda)} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} e_i^2(\lambda) \tilde{\mu}_i(\lambda)^{-\delta} \qquad \ldots(9)$$

$$\left.\frac{\partial L(.)}{\partial \delta}\right|_{\beta = \hat{\beta}(\lambda)} = -\frac{1}{2} \sum_{i=1}^{n} \ln \tilde{\mu}_i(\lambda) + \frac{1}{2\sigma^2} \sum_{i=1}^{n} e_i^2(\lambda) \tilde{\mu}_i(\lambda)^{-\delta} \ln \tilde{\mu}_i(\lambda) \qquad \ldots(10)$$

where $e_i(\lambda) = y_i^{(\lambda)} - x_i'\hat{\beta}(\lambda)$, $\tilde{\mu}_i(\lambda) = \left[1 + \lambda\hat{\mu}_i(\lambda)\right]$ and $\hat{\mu}_i(\lambda) = x_i'\hat{\beta}(\lambda)$. Now setting (9) and (10) to zero and then substituting the solution of σ^2 thus obtained from (9) in (10), we have

$$\sum_{i=1}^{n} e_i^2(\lambda)\,\widetilde{\widetilde{\mu}}_i(\lambda)^{-\delta}\,\ln\widetilde{\widetilde{\mu}}_i(\lambda) = 0 \qquad \ldots(11)$$

where $\widetilde{\widetilde{\mu}}_i(\lambda) = \widetilde{\mu}_i(\lambda)\Big/\Big\{\prod_{i=1}^{n}\widetilde{\mu}_i(\lambda)\Big\}^{1/n}$ so that $\sum_{i=1}^{n}\ln\widetilde{\widetilde{\mu}}_i(\lambda) = 0.$

Denoting the left hand side of (11) by $S(\delta(\lambda))$, we can obviously check that $\lim_{\delta\to\infty} S(\delta(\lambda)) = -\infty$, $\lim_{\delta\to-\infty} S(\delta(\lambda)) = \infty$ and $\frac{\partial S(\delta(\lambda))}{\partial\delta} < 0$ for $-\infty < \delta < \infty$. Hence $S(\delta(\lambda)) = 0$ has a unique root for δ for a given λ, and this root can therefore be obtained by using the standard Newton-Raphson iterative procedure. The estimate of δ, say $\hat{\delta}(\lambda)$, thus obtained can now be used to calculate the log-likelihood function in (7). Improved estimates of β and σ^2 and hence of δ and $L(.)$ can then be obtained at each successive stage of iteration till the maximum value of $L(.)$ for the given λ is obtained. The unconditional ML estimates of all the parameters can be obtained by searching over reasonable range of λ till the maximum value of the log-likelihood function is attained.

Under standard conditions [see, Carroll et al. (1982), Magnus (1978), Theil (1971) etc. for the relevant conditions], both the methods would asymptotically yield ML estimates. In the numerical examples that follow we have found that the second method required much less computer processing time than the first one. As has been pointed out by Savin and White (1978) in the context of Box-Cox model with autocorrelated errors that it does not appear to be feasible to analytically derive the variance-covariance matrix of the limiting distribution of the ML estimates of the parameters. An approximation, however, to the estimated information matrix can be obtained numerically from the matrix of second derivatives. Incidentally, it is to be noted that the correct ML covariance estimates are not directly provided by these search procedures. Some additional computational work, though very insignificant compared to those needed for parameter estimates, is needed for this [see, Spitzer (1982a) for details].

While different conditional and unconditional hypotheses involving the principal parameters λ and δ may be considered for testing, the hypothesis that would be of main interest to us is whether or not $\delta = 0$. These tests are considered in the next section.

4. ILLUSTRATIVE EXAMPLES

In this section we report the results obtained by considering BCH model for two different sets of data. The first one is taken from O'Hara and McClelland (1964) and relates to radio-sales (y) and incomes (x), for 49 states in the United States for the year 1954. This was first analysed by Rutemiller and Bowers (1968) in terms of a linear regression model with heteroscedastic errors. The second example is based on the data available from Feigl and Zelen (1965) on survival time (y) of 17 leukemia patients and their white blood cell count (x). They used this data to estimate the linear regression of y on ln x with exponential errors. This data was also used by Amemiya (1973) and Cox and Snell (1968) for their studies.

We have used the above two examples to estimate the following regression equations

$$y_i^{(\lambda)} = \beta_1 + \beta_2 x_i + \varepsilon_i \ (i = 1, 2, \ldots, 49), \qquad \ldots(12)$$

for the first example and

$$y_i^{(\lambda)} = \beta_1 + \beta_2 x_i^{(\lambda)} + \varepsilon_i \ (i = 1, 2, \ldots, 17), \qquad \ldots(13)$$

for the second example. While different assumptions can be made about $V(y_i)$, we assume, as stated in section 2, that $V(y_i) = \sigma^2(E(y_i))^h$, $E(y_i) > 0$, h unknown, so that $V(y_i^{(\lambda)}) = \sigma^2(1 + \lambda x_i'\beta)^\delta$ for $\lambda \neq 0$ and $\sigma^2 \exp(x_i'\beta)$ for $\lambda = 0$ [cf. equation (6)].

TABLE I

Maximum log-likelihood values L(.) for different values of λ and δ

Model[a]	Radio-sales data			Leukemia data		
	λ	δ	L(.)	λ	δ	L(.)
LHOM	1*	0*	−159.524	1*	0*	−64.526
LHET	1*	1.501	−131.989	1*	0.582	−64.188
BCH	0.923[b]	1.326	−129.686	0.124[b]	c	−56.209
	(0.923)	(1.325)	(−129.732)	(0.124)	c	(−56.209)
BC	0.786	0*	−146.375	0.282	0*	−57.340
S(D)-LHOM	0*	0*	−187.682	0*	0*	−59.642
S(D)-LHET	0*	0.613	−184.316	0*	−1.500	−56.367

a LHOM : Linear Homoscedastic, LHET : Linear Heteroscedastic, S(D)LHOM : Semilog/Doublelog Homoscedastic and S(D)LHET : Semilog/Doublelog Heteroscedastic.

b Although both the methods of estimation have yielded identical ML estimates, we have, for the sake of illustration, given the results obtained by the second method in brackets for BCH model only.

c Since the estimate of δ for the Leukemia data lies in a range where L(.) is somewhat insensitive to changes in δ, we have kept the corresponding entry in the table blank.

* Indicates that the value of the parameter in question is given *a priori* from the model assumed.

In Table I, we present the maximum values of the log-likelihood function for each of the different assumed values of one or both of λ and δ. More specifically, in addition to the maximum log-likelihood values for BCH and Box-Cox (BC) models, we also report the maximum log-likelihood values for λ = 1 and 0 and/or δ = 0. For each of the two examples we have used both the methods stated in section 3 for obtaining ML estimates i.e., (i) search over λ and δ only and (ii) search over λ only. These two methods produce almost identical results (though the second one took much less computation time) and hence, for convenience, we have reported the results for the

TABLE II

Results of likelihood ratio test for different null hypotheses

Statistic	H_0	H_1	Ratio-sales data: Value of LR-test statistic	Ratio-sales data: Conclusions*	Leukemia data: Value of LR-test statistic	Leukemia data: Conclusions*
1(1)	$\lambda=1,\ \delta=0$	$\hat{\lambda}(0),\ \delta=0$	26.298	reject H_0	14.372	reject H_0
1(2)	$\lambda=1,\ \hat{\delta}(1)$	$\hat{\lambda},\ \hat{\delta}$	4.606	reject H_0 at 5°/.	15.956	reject H_0
1(3)	$\lambda=1,\ \delta=0$	$\lambda=1,\ \hat{\delta}(1)$	55.070	reject H_0	0.676	accept H_0
1(4)	$\hat{\lambda}(0),\ \delta=0$	$\hat{\lambda},\ \hat{\delta}$	33.378	reject H_0	2.260	accept H_0
1(5)	$\lambda=1,\ \delta=0$	$\hat{\lambda},\ \hat{\delta}$	59.676	reject H_0	16.632	reject H_0
1(6)	$\lambda=0,\ \delta=0$	$\lambda=0,\ \hat{\delta}(0)$	6.732	reject H_0	6.550	reject H_0 at 5°/.
1(7)	$\lambda=0,\ \delta=0$	$\hat{\lambda},\ \hat{\delta}$	116.244	reject H_0	6.864	reject H_0 at 5°/.
1(8)	$\lambda=0,\ \hat{\delta}(0)$	$\hat{\lambda},\ \hat{\delta}$	109.260	reject H_0	0.314	accept H_0
1(9)	$\lambda=0,\ \delta=0$	$\hat{\lambda}(0),\ \delta=0$	82.866	reject H_0	4.604	reject H_0 at 5°/.

* Unless mentioned, 'reject/accept H_0' means that the null hypothesis H_0 is rejected/accepted in favour of/against the alternative hypothesis H_1 at both 5°/. and 1°/. levels of significance.

first method only. Results of the likelihood-ratio (LR) test for different hypotheses involving λ and δ are given in Table II.

For the first example the ML method of estimation of BCH model has yielded estimates of λ and δ as $\hat{\lambda} = 0.923$ and $\hat{\delta} = 1.326$ and the maximum log-likelihood value is −129.686. The corresponding value for BC model is only −146.375. Thus, there is a considerable increase in the maximum value of the log-likelihood function in BCH model **compared** to that in BC model. Using LR test we find from 1(4) in Table II that $H_0 : \hat{\lambda}(0),\ \delta = 0$, where $\hat{\lambda}(0)$ denotes the ML estimate of λ when $\delta = 0$, is rejected in favour of $H_1 : \hat{\lambda}, \hat{\delta}$ i.e., BC model is clearly rejected in favour of BCH model. This, therefore, indicates that one would have chosen a wrong model by straightforwardly using Box-Cox procedure. If we now consider the test statistic l(2), we reject linear heteroscedastic model in favour of BCH model at 5 per cent level of significance, the critical value of χ_1^2 being 3.84. This test shows that the parameter indicating the degree of heteroscedasticity may be affected by a wrong choice of functional form. It is important to note that the advantage of using BCH model is that by rejecting the inappropriate null hypotheses against the unrestricted hypothesis, it helps us in choosing the proper model. This is shown by l(5), l(7) and l(8). It may also be seen from l(6), for example, how conditional hypotheses (where values of a parameter are a priori assumed to be known) may lead to wrong conclusion about the proper model. In this case λ is a priori fixed at $\lambda = 0$ and we find that $H_0 : \lambda = 0, \delta = 0$ is rejected in favour of $H_1 : \lambda = 0, \hat{\delta}(0)$ at 5 per cent level of significance and H_0 is almost accepted against H_1 at 1 per cent level of significance though the maximum values of the log-likelihood function are much less for both the hypotheses as compared to the maximum value for the BCH model. Thus we find that choice of functional form appears to be crucial in discriminating among different models and also, as in this example, that estimation of λ seems to be influenced by heteroscedasticity.

In the second example of survival time of leukemia patients our ML method estimation gives $\hat{\lambda} = 0.124$ for the BCH model. It was found here that the estimate of $[1 + \lambda x_i'\beta] \simeq 1$ for all i and hence σ_i^2's were insensitive to δ. In such cases of small values of λ, we suggest that one should test for $H_0 : \lambda = 0, \hat{\delta}(0)$ against $H_1 : \hat{\lambda}, \hat{\delta}$ and if the null hypothesis is accepted, one should proceed (for further studies) with the model where $\lambda = 0$ and for this case there is a separate expression for approximating $E(y_i)$ given in section 2. We find from Table I that the maximum log-likelihood values for the case ($\lambda = 0$, $\hat{\delta} = (0)$) and BCH model (i.e., $\hat{\lambda}$, $\hat{\delta}$) are close being −56.367 and −56.210 respectively and δ has a unique estimate at $\hat{\delta} = -1.500$ in the former case. The fact that $\hat{\delta}$ and hence $\hat{\zeta}$ (cf. equation (5)) have negative values indicates that the variance of $y_i^{(\lambda)}$'s decrease with increase in $E(y_i)$'s. We however, note that for $\lambda = 0$ and $\hat{\delta} = -1.5$, estimate of h comes out to be 0.5 which means that the original observations of the dependent variable i.e., y_i's have increasing variance with increase in $E(y_i)$ values. This clearly shows the extent to which transformation may affect the variance of the transformed observations compared to that of the original observations. This negative value of $\hat{\delta}$ also implies that heteroscedasticity is relatively less important for this data. This is also corroborated by l(4). How conditional hypotheses may lead to acceptance of wrong models are revealed here also by l(3), for example, where linear homoscedastic model is accepted against linear heteroscedastic model though both have much less maximum log-likelihood values as compared to BCH model and both are, as indicated by l(5) and l(2), rejected in favour of BCH model.

Thus we find that in a practical situation the departure in the estimate of λ from a model with heteroscedasticity to that of a model with homoscedasticity and hence the consequences of choosing an inappropriate model may depend upon the data. But since the actual situation cannot be known *a priori*, it is, in general, advisable to estimate λ within the framework of BCH model.

5. CONCLUSIONS

Box and Cox (1964) suggested a transformation of the dependent variable in a regression model in order to achieve linearity, homoscedasticity and normality of the transformed dependent variable and proposed maximum likelihood method of estimation of the parameters of such a model. Zarembka (1974) however showed that ML method of estimation suggested by Box and Cox is not robust to heteroscedasticity and that the estimate of λ will be biased towards the direction of stablizing the error variance.

In this paper we have asserted that the transformation that leads to linearity does not necessarily lead to homoscedasticity also. This is evident from the transformation itself. We have argued that since heteroscedasticity in transformed dependent variable is due to both transformation and heteroscedasticity in the original values of the dependent variable, one should estimate λ in the framework of heteroscedasticity as given by the transformation and the heteroscedasticity in the original dependent variable. We have advocated ML method of estimation of such a model and have also suggested search procedures for obtaining ML estimates of the parameters. We have also indicated the seriousness of the problem of heteroscedasticity in the context of the usual Box-Cox model. We have, through illustrative examples, shown how the problem can be circumvented and distinct improvement be brought about by considering Box-Cox heteroscedastic model where the form of heteroscedasticity should properly incorporate the complications created by the power transformation.

ACKNOWLEDGEMENTS

This paper is a revised version of a Chapter of the author's unpublished Ph.D. thesis at the Indian Statistical Institute. The author is grateful to Professor Robin Mukherjee for his guidance of the work. Thanks are also due to Professor Nikhilesh Bhattacharya and Dr. Dipankar Coondoo for their many useful suggestions at different stages of the work. Comments from Professor A. L. Nagar, and the two referees Professors Jan Magnus and Daniel J. Egy have helped in revising the paper. To them the author is thankful.

BIBLIOGRAPHY

Amemiya, T. (1973). Regression analysis when the variance of dependent variable is proportional to the square of its expectation, Journal of the American Statistical Association, 68, 928-934.

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

Box. G. E. P. and Cox, D. R. (1964). An analysis of transformations, Journal of the Royal Statistical Society, Series B, 26, 211-243.

Box, G. E. P. and Cox, D. R. (1982). An analysis of transformations revised, rebutted, Journal of the American Journal of Statistical Association, 77, 209-210.

Box, G. E. P. and Hill, W. J. (1974). Correcting inhomogeneity of variance with power transformation weighting, Technometrics, 16, 385-389.

Carroll, R. J., Ruppert, D. and Holt, R. N. Jr. (1982). Some aspects of estimation in heteroscedastic linear models, in S. S. Gupta and J. O. Berger, ed., Statistical Decision Theory and Related Topics III, vol. 1, Academic Press, 231-241.

Coondoo, D. and Sarkar, N. (1979). Regression analysis with heteroscedastic disturbances, Technical Report No. ERU/9/79, Indian Statistical Institute, Calcutta, India.

Cox, D. R. and Snell, E. J. (1968). A general definition of residuals, Journal of the Royal Statistical Society, Series B, 30, 248-265.

Dhrymes, P. J. (1970). Econometrics : Statistical Foundations and Applications, New York : Harper and Row.

Draper, N. R. and Cox, D. R. (1969). On distributions and their transformation to normality, Journal of the Royal Statistical Society, Series B, 31, 472-476.

Egy, D. J. (1982). A maximum-likelihood approach to treating error variance instability in Hougen-Watson rate models, Ind. Eng. Chem. Fundam., 21, 337-339.

Feigl, P. and Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information, Biometrics, 21, 826-838.

Judge, G. G., Griffiths, W. E., Hill, R. C. and Lee, T. C. (1980). The Theory and Practice of Econometrics, John Wiley and Sons, Inc.

Lahiri, K. and Egy, D. (1981). Joint estimation and testing for functional form and heteroscedasticity, Journal of Econometrics, 15, 299-308.

Maddala, G. S. (1977). Econometrics, McGraw-Hill Book Co.

Magnus, J. R. (1978). Maximum likelihood estimation of the GLS model with unknown parameters in the disturbance covariance matrix, Journal of Econometrics, 7, 281-312.

Oberhofer, W. and Kmenta, J. (1974). A general procedure for obtaining maximum likelihood estimates in generalised regression models, Econometrica, 42, 579-590.

O'Hara, J. B. and McClelland, R. C. (1964). Effective Use of Statistics in Accounting and Business, New York : Holt, Rinehart and Winston.

Poirier, D. J. (1978). Box-Cox transformation in limited dependent variable models, Journal of the American Statistical Association, 73, 284-289.

Rutemiller, H. and Bowers, D. A. (1968). Estimation in a heteroscedastic regression model, Journal of the American Statistical Association, 63, 552-557.

Savin, N. E. and White, K. J. (1978). Estimation and testing for functional form and autocorrelation : a simultaneous approach, Journal of Econometrics, 8, 1-12.

Schlesselman, J. (1971). Power families : a note on the Box and Cox transformation, Journal of the Royal Statistical Society, Series B, 33, 307-311.

Spitzer, J. J. (1978). A Monte Carlo investigation of the Box-Cox transformation in small samples, Journal of the American Statistical Association, 73, 488-495.

Spitzer, J. J. (1982a). A primer on Box-Cox transformation, The Review of Economics and Statistics, LXIV, 307-313.

Spitzer, J. J. (1982b). A fast and efficient algorithm for the estimation of parameters in models with Box-Cox transformation, Journal of the American Statistical Association, 77, 760-766.

BOX-COX TRANSFORMATION

Theil, H. (1971). Principles of Econometrics, John Wiley and Sons, Inc.

Zarembka, P. (1974). Transformation of variables in Econometrics, in P. Zarembka, ed., Frontiers of Econometrics, Academic Press, 81-104.

Received by Editorial Board member June, 1984; Revised October, 1984.

Recommended by J. P. C. Kleijnen, Tilburg University, Tilburh, The Netherlands.

Refereed by Jan Magnus, London School of Economics, London, United Kingdom, and Daniel J. Egy, Russell Sage College, Troy, NY