Economics Bulletin

Volume 33, Issue 3

A new estimator of the Box-Cox transformation model using moment conditions

Kazumitsu Nawata
University of Tokyo

Abstract

The maximum likelihood estimator (MLE) under the normality assumption of error terms is widely used to estimate the Box-Cox transformation model. However, since the error terms cannot be normally distributed, it is not a proper estimator. In other words, the estimator is inconsistent. In this paper, I propose a new estimator of the Box-Cox transformation model that modifies the MLE in key ways. I demonstrate that the estimator is consistent, and that an asymptotic distribution is obtained. The results of Monte Carlo experiments are also presented.

This study was supported by a Grant-in-Aid for Scientific Research, namely "Analyses of the Large Scale Medical Survey Data and the Policy Evaluations in Japan (Grant Number: 24330067)", from the Japan Society of Science. I would like to thank the anonymous referee, Professor Takeshi Amemiya at Stanford University and Yukari Ueno at the University of Tokyo for their helpful comments.
Citation: Kazumitsu Nawata, (2013) "A new estimator of the Box-Cox transformation model using moment conditions", *Economics Bulletin*, Vol. 33 No. 3 pp. 2287-2297.
Contact: Kazumitsu Nawata - nawata@tmi.t.u-tokyo.ac.jp.
Submitted: November 13, 2012. **Published:** September 10, 2013.

1. Introduction

The Box-Cox (1964) transformation model (hereafter, the BC model) is widely used in various fields of econometrics and statistics. However, since the error terms cannot have a normal distribution except in the case where the transformation parameter is zero, the likelihood function under the normality assumption (hereafter, the BC likelihood function) is misspecified and the maximum likelihood estimator (hereafter, the BC MLE) cannot be consistent. Alternative distributions of the error terms and transformations for the BC model have been proposed by various authors (for details, see Amemiya and Powell (1981), Yeo and Johnson (2000) and Yang (2006)). Because the simplicity of the model is lost (Showalter, 1994), these alternatives have not been widely used. In this paper, I propose a new estimator of the Box-Cox transformation model. The estimator is a modification of the BC MLE and proved to be consistent. An asymptotic distribution was obtained for it. The results of Monte Carlo experiments are also presented.

2. Model and Estimator

I consider the Box-Cox transformation model

$$z_t = x_t'\beta + u_t, \quad y_t \geq 0, \qquad t = 1,2,...,T, \tag{1}$$

$$z_t = \begin{cases} \dfrac{y_t^{\lambda}-1}{\lambda}, & \text{if } \lambda \neq 0, \\ \log(y_t), & \text{if } \lambda = 0, \end{cases}$$

where x_t and β are k-th dimensional vectors of explanatory variables and the coefficients, respectively, and λ is the transformation parameter. Random variables $\{u_t\}$ are independent and identically distributed (i.i.d.) and follow a distribution whereby the support is bounded from below, the first and third moments are zero, and the sixth moment exists and finite (i.e., $f(u) = 0$ if $u \leq -a$ for some $a > 0$ where $f(u)$ is the probability density function, $E(u_t) = E(u_t^3) = 0$, and $E(u_t^6) = M_6 < \infty$). We do not have to assume a specific distribution and the model is semiparametric in this sense. $\{x_t\}$ are i.i.d. random variables with the finite third moment. $\{u_t\}$ and $\{x_t\}$ are independently distributed. For the identification of the model, the distribution of x_t and the parameter space of β are restricted so that $\inf_x(\lambda_0 x'\beta_0 + 1) - \lambda_0 a > 0$ where β_0 and λ_0 is the true parameter values of β and λ, and $\inf_{x,\beta}(\lambda_0 x'\beta + 1) > c$ for some $c > 0$ in the neighborhood of β_0. Unlike the case under the normality assumption, $y_t > 0$ under this assumption, and we can obtain a consistent model.

Powell (1996) proposed a semiparametric estimator based on the generalized method of moment (GMM). However, the problems of the estimator are: i) to identify λ, we need to introduce one or more instrumental variables, w_t, which satisfy $E(u_t \mid w_t) = 0$ and are not included in x_t, and the result of estimation changes depending on the selection of instrumental variables, and ii) as pointed out by Khazzoom (1989), when all observations are $y_t < 1$, the objective function is always minimized at $\lambda = \infty$ (or at $\lambda = -\infty$ if $y_t > 1$ for all observations), so that rather arbitrary rescaling of y_t is necessary. Foster, Tain, and Wei (2001) also proposed a semiparametric estimator. However, their estimator has a problem similar to Powell's second problem.

Let $\theta' = (\lambda, \beta', \sigma^2)$. The BC likelihood function is given by

$$\log L(\theta) = \sum_t \left[\log\phi\{(z_t - x_t'\beta)/\sigma\} - \log\sigma\right] + (\lambda - 1)\sum_t \log y_t, \tag{2}$$

where ϕ is the probability density function of the standard normal distribution and σ^2 is the variance of u_t. The BC MLE is obtained by

$$\frac{\partial \log L}{\partial \lambda} = -\frac{1}{\sigma^2 \lambda} \sum_t \{\log(y_t) y_t^{\lambda} - z_t\}(z_t - x_t{}'\beta) + \sum_t \log(y_t) = 0, \qquad (3)$$

$$\frac{\partial \log L}{\partial \beta} = \frac{1}{\sigma^2} \sum_t x_t (z_t - x_t{}'\beta) = 0, \text{ and } \frac{\partial \log L}{\partial \sigma^2} = \sum_t \frac{(z_t - x_t{}'\beta)^2 - \sigma^2}{2\sigma^4} = 0.$$

Let $\theta_0{}' = (\lambda_0, \beta_0{}', \sigma_0^2)$ be the true parameter value of θ. Since $E[\frac{\partial \log L}{\partial \lambda}|_{\theta_0}] \neq 0$, the BC MLE cannot be consistent.

Instead of $\partial \log L / \partial \lambda$, I use

$$G_T(\theta) = \sum_t g_t(\theta), \qquad (4)$$

$$g_t(\theta) \equiv g(\theta, x_t, y_t) = -\frac{1}{\sigma^2 \lambda}[\{\frac{\log(\lambda x_t{}'\beta + 1)}{\lambda} + \frac{z_t - x_t{}'\beta}{\lambda x_t{}'\beta + 1}\} y_t^{\lambda} - z_t\}](z_t - x_t{}'\beta)$$

$$+ \frac{1}{\lambda}\log(\lambda x_t{}'\beta + 1) + \frac{z - x_t{}'\beta}{\lambda x_t{}'\beta + 1}, \qquad \text{if } \lambda \neq 0, \text{ and}$$

$$g_t(\theta) = \lim_{\lambda \to 0} g_t(\theta) = -\frac{1}{\sigma^2}[-\frac{1}{2}(x_t \beta)^2 - (x_t{}'\beta)\{\log(y_t) - x_t{}'\beta\} + \{\log(y_t)\}^2]\{\log(y_t) - x_t{}'\beta\}$$

$$+ \log(y_t), \qquad \text{if } \lambda = 0.$$

$G_T(\theta)$ is obtained by the approximation of $\partial \log L / \partial \lambda$ as shown in Appendix A.

I consider the roots of the equations,

$$G_T(\theta) = 0, \quad \frac{\partial \log L}{\partial \beta} = 0, \text{ and } \frac{\partial \log L}{\partial \sigma^2} = 0. \qquad (5)$$

Then, unlike the previous case, the estimator obtained by Equation (5) is consistent and we get the following proposition. Note that $G_T(\theta) = 0$ is equivalent to $\lambda G_T(\theta) = 0$ when $\lambda \neq 0$.

<u>Proposition 1</u>

Among the roots of Equation (5), there exists a consistent root.

The proof is given in Appendix B. Let $\hat{\theta}' = (\hat{\lambda}, \hat{\beta}', \hat{\sigma}^2)$ be the consistent root. The asymptotic distribution of $\hat{\theta}$ is obtained by the following proposition.

<u>Proposition 2</u>

The asymptotic distribution of $\hat{\theta}$ is given by

$$\sqrt{T}(\hat{\theta} - \theta_0) \to N[0, A^{-1}B(A')^{-1}], \qquad (6)$$

where $A = -E[\frac{\partial \ell_t(\theta)}{\partial \theta'}|_{\theta_0}]$, $B = E[\ell_t(\theta_0)\ell_t(\theta_o)']$, $\ell_t(\theta)' = [g_t(\theta), \xi_t(\theta)', \varsigma_t(\theta)]$,

$$\xi_t(\theta) = \frac{1}{\sigma^2} x_t (z_t - x_t'\beta), \text{ and } \zeta_t(\theta) = \frac{(z_t - x_t'\beta)^2 - \sigma^2}{2\sigma^4}$$

[Proof]

Let

$$\ell(\theta)' = \sum_t \ell_t(\theta)' = [G_T(\theta), \frac{\partial \log L}{\partial \beta'}, \frac{\partial \log L}{\partial \sigma^2}]. \tag{7}$$

Then

$$\sqrt{T}(\hat{\theta} - \theta_0) = -[\frac{1}{T}\frac{\partial \ell}{\partial \theta'}|_{\theta^*}]^{-1} \frac{1}{\sqrt{T}} \ell(\theta_0), \tag{8}$$

where θ^* is some value between $\hat{\theta}$ and θ_0. Here, when $\lambda_0 \neq 0$,

$$\ell_t(\theta_0) = \begin{bmatrix} -\frac{1}{\sigma_0^2 \lambda_0}[\{\frac{(\lambda_0 x_t'\beta_0 + 1)\log(\lambda_0 x_t'\beta_0 + 1)}{\lambda_0} - x_t\beta\}u_t + \log(\lambda_0 x_t'\beta_0 + 1)u_t^2 + \frac{\lambda_0 u_t^3}{\lambda_0 x_t'\beta_0 + 1}] \\ + \frac{\log(\lambda_0 x_t'\beta_0 + 1)}{\lambda_0} + \frac{u_t}{\lambda_0 x_t'\beta_0 + 1}\} \\ \frac{1}{\sigma_0^2} x_t u_t \\ \frac{u_t^2 - \sigma_0^2}{2\sigma_0^4} \end{bmatrix}. \tag{9}$$

Therefore, $E[\ell_t(\theta_0)] = 0$ and $E[\ell_t(\theta_0)\ell_t(\theta_0)']$ exists. Since $\{\ell_t(\theta_0)\}$ are i.i.d. random variables with finite second moments, we get

$$\frac{1}{\sqrt{T}} \ell(\theta_0) \to N(0, B), \tag{10}$$

from the multivariate central limit theorem.

Since all elements of $\partial \ell / \partial \theta$ are differentiable,

$$-\frac{1}{T}\frac{\partial \ell(\theta)}{\partial \theta'}|_{\theta^*} \xrightarrow{P} A, \tag{11}$$

from Theorem 4.1.4 in Amemiya (1985, pp.112-113). From Theorem 4.1.3 in Amemiya (1985, p.111) , the asymptotic distribution of $\hat{\theta}$ is given by (6). When $\lambda_0 = 0$, we use $\lim_{\lambda_0 \to 0}\{\ell_t(\theta_0)\}$, $\lim_{\lambda_0 \to 0} A$ and $\lim_{\lambda_0 \to 0} B$ for $\ell_t(\theta_0)$, A and B. Since they are continuous even when $\lambda_0 = 0$ by the same argument of Appendix B, we can get the asymptotic distribution given by the same formula. Note that matrices A and B can be estimated by $\hat{A} = -\frac{1}{T}\frac{\partial \ell(\theta)}{\partial \theta'}|_{\hat{\theta}}$ and $B = \frac{1}{T}\sum [\ell_t(\hat{\theta})\ell_t(\hat{\theta})']$.

3. Monte Carlo Experiments

In this section some Monte Carlo results are presented for the BC MLE and the newly proposed estimator (hereafter the N-estimator). The model is given by

$$z_t = \beta_0 + \beta_1 x_t + u_t, \quad z_t = (y_t^{\lambda_0} - 1)/\lambda_0 \qquad (12)$$

$\{x_t\}$ and $\{u_t\}$ are i.i.d. random variables. The following items are considered in the Monte Carlo study:
i) The effect of transformation parameter λ_0; values of 0.8, 0.5 and 0.2 are considered.
ii) The effect of different distributions of error terms.

As the distributions of error terms, the three cases are considered. First, the case where $\lambda_0 u_t$ is distributed uniformly on (-1, 1) is considered. Secondly, I consider the case where $\lambda_0 u_t$ follows the doubly truncated normal distribution such that $N(0, 0.5^2)$ is truncated between -1 and 1 so that the support becomes (-1,1). Note that the degree of truncation is 4.55% in this case. In the two previous cases, the supports are bounded. The behavior of the estimator may be different if this condition is not satisfied. Therefore, for the third case, I consider the power transformation of the gamma distribution following the suggestion of Hinkley (1975). Let ξ be a random variable which follows the chi-squared distribution with degrees of freedom one, $\chi^2(1)$. $\lambda_0 u_t$ is obtained by $\lambda_0 u_t = (\xi^a - b)/b, a = 0.830, b = 0.229$. The values of a and b are chosen so that the first and third moments of $\lambda_0 u_t$ become zero and its support becomes $(-1, \infty)$.

x_t is distributed uniformly on (0, 2). The true parameter values are:

$$\beta_0 = 0.0 \text{ and } \beta_1 = 0.1. \qquad (13)$$

The sample size is 200 and the number of trials is 1,000 for all cases. Note that when λ is given, β and σ^2 are obtained by the least squares method. The BC MLE and the N-estimator were calculated by the following scanning method (Nawata, 1994).

i) Choose $\lambda_1 < \lambda_2 < \lambda_3 < ... < \lambda_n$ from -1.0 to 2.0 with an interval of 0.01

ii) Calculate $\hat{\beta}_0(\lambda), \hat{\beta}_1(\lambda)$ and $\hat{\sigma}(\lambda)^2$ for each λ by the least squares method.

iii) For BC MLE, choose $\hat{\lambda}_{BC_1}$, which maximizes the BC likelihood function. For the N-estimator, choose $\hat{\lambda}_{N_1}$, which satisfies $G_T(\theta_i) \cdot G_T(\theta_{i+1}) < 0$ where $\theta_i' = (\lambda_i, \hat{\beta}_0(\lambda), \hat{\beta}_1(\lambda), \hat{\sigma}(\lambda)^2)$.

iv) Choose λ_i in the neighborhood of $\hat{\lambda}_{BC_1}$ and $\hat{\lambda}_{N_1}$ with an interval of 0.0001 and repeat the steps (ii) and (iii).

v) Determine the final estimators.

For the N-estimator, there are two possible problems. They are: i) Equation (5) has multiple solutions, and ii) Equation (5) does not have a solution. Figure 1 is the graph of $G_T(\theta_i)$ in the trial where the distribution is uniform and $\lambda_0 = 0.5$.The graph is monotonically decreasing, and Equation (5) has just one solution. Although it cannot be proved that Equation (5) has a unique solution, all trials have just one solution as Figure 1 and the above problems do not occur in the Monte Carlo experiment.

The results are presented in Tables I-III. Note that the following notations are used in the tables: STD, standard deviation; Ql, first quartile; and Q3, third quartile. The BC MLE underestimates λ and has large biases for the uniform case. The biases of the BC MLE are fairly large for the doubly truncated normal case even though the degree of the truncation is less than 5%. Although the biases are smallest among three distributions, some degree of biases are still observed for the power transformation case. The standard deviations of the N-estimator are slightly larger than those of the BC MLE, however, the biases of the N-estimator are much smaller. The bias almost disappears for all cases.

4. Conclusion

Although the BC model is widely used in various fields, the BC MLE cannot be consistent. In this paper, I propose a new estimator of the BC model. The estimator is a modification of the BC MLE. The estimator is consistent and its asymptotic distribution is also obtained. Moreover, the estimator is easily calculated by the least squares and scanning methods. The results of the Monte Carlo experiments show the superiority of the proposed estimator over the BC MLE for all cases. However, the performance of the estimators may depend on the model. The findings of the study may be different in other models. Therefore, further investigation is necessary to determine the superiority of the estimators.

Appendix A: Approximation of $\partial \log L / \partial \lambda$

Here,

$$\frac{\partial \log L}{\partial \lambda}|_{\theta_0} = -\frac{1}{\sigma_0^2 \lambda_0} \sum_t \{y^{\lambda_0} \log(y_t) - z_t^*\} u_t \quad + \sum_t \log(y_t) \qquad (14)$$

where $z_t^* = \{y_t^{\lambda_0} - 1\} / \lambda_0$ if $\lambda_0 \neq 0$ and $z_t^* = \log(y_t)$ if $\lambda_0 = 0$. If $|\lambda_0 u_t / (\lambda_0 x_t ' \beta_0 + 1)|$ is small and $\lambda_0 \neq 0$, we get

$$\log(y_t) = \frac{1}{\lambda_0} \log(\lambda_0 x_t ' \beta_0 + 1 + \lambda_0 u_t) = \frac{1}{\lambda_0} \{\log(\lambda_0 x_t ' \beta_0 + 1) + \log(1 + \frac{\lambda_0 u_t}{\lambda_0 x_t ' \beta_0 + 1})\}$$

$$\approx \frac{1}{\lambda_0} \log(\lambda_0 x_t ' \beta_0 + 1) + \frac{u_t}{\lambda_0 x_t ' \beta_0 + 1}. \qquad (15)$$

Therefore, if $\lambda_0 \sigma_0 / (\lambda_0 x_t ' \beta_0 + 1) \approx 0$ for all observations (following Bickel and Doksum (1981), I call "small σ" cases), we get

$$\frac{\partial \log L}{\partial \lambda}|_{\theta_0} \approx -\frac{1}{\sigma_0^2 \lambda_0} \sum_t [\{\frac{(\lambda_0 x_t ' \beta_0 + 1) \log(\lambda_0 x_t ' \beta_0 + 1)}{\lambda_0} - x_t ' \beta_0\} u_t + \log(\lambda_0 x_t ' \beta_0 + 1) u_t^2 + \frac{\lambda_0 u_t^3}{\lambda_0 x_t ' \beta_0 + 1}] \qquad (16)$$

$$+ \sum_t \{\frac{1}{\lambda_0} \log(\lambda_0 x_t ' \beta_0 + 1) + \frac{u_t}{\lambda_0 x_t ' \beta_0 + 1}\} = G_T(\theta_0).$$

The BC MLE is considered to be a method to make the distribution of the error terms close to the normal distribution by the transformation for the small σ cases. However, normality approximations are impossible for non-small σ cases. The proposed estimator makes the distributions close to symmetric rather than the normal distribution for the non-small σ cases and we can obtain a consistent estimator even under those conditions. (Hinkley (1975) considered the power transformation to make the distributions close to symmetric based on quantiles. However, the method cannot be directory applied to a regression type model and a result of the estimation changes depending on the selection of quantiles.) For the small σ cases, where the BC MLE performs well (for the details of "small σ asymptotics," see Bickel and Doksum (1981)), the properties of the proposed estimator are similar to those of the BC MLE due to Equation (16). These factors are considered to constitute the superiority of the proposed estimator.

Appendix B: Proof of Proposition 1

Here, I prove the consistency of the estimator using the mean value and intermediate value theorems. Since

$$y^{\lambda} = 1 + \log(y)\lambda + \frac{\log(y)\lambda^2}{2} + o_p(\lambda^2), \quad z = \frac{y^{\alpha} - 1}{\lambda} = \log(y) + \frac{\{\log(y)\}^2 \lambda}{2} + \frac{\{\log(y)\}^3 \lambda^2}{6} + o_p(\lambda^2) \quad (17)$$

$$\frac{1}{\lambda}\log(1+\lambda x'\beta) = \lambda x'\beta - \frac{1}{2}x'\beta + (x'\beta)^2\lambda^2 + o_p(\lambda^2), \quad \frac{1}{1+\lambda x'\beta} = 1 - (x'\beta)\lambda + (x'\beta)\lambda^2 + o_p(\lambda^2)$$

in the neighborhood of $\lambda = 0$, we get

$$g(\theta, x, y) = -\frac{1}{\sigma^2}[-\frac{1}{2}(x'\beta)^2 + \{\log(y)\}^2 - x'\beta\{\log(y) - x'\beta\}]\{\log(y) - x'\beta\} + \log(y) \quad (18)$$

$$-\frac{\lambda}{\sigma^2}\Big[\ [-\frac{1}{2}(x'\beta)^2 + \{\log(y)\}^2 - x'\beta\{\log(y) - x'\beta\}]\frac{1}{2}\{\log(y)\}^2 +$$

$$+\frac{1}{\sigma^2}[-\frac{5}{6}\{\log(y)\}^3 + x'\beta\log(y)\{\log(y) - x_t'\beta\}]\{\log(y_t) - x_t'\beta\}]$$

$$+\{\frac{1}{2}(\log y)^2 - x'\beta\log(y) - (x'\beta)^2\}\ \Big] + o_p(\lambda^2).$$

Let $g^*(\theta) = \partial g/\partial\lambda$ if $\lambda \neq 0$ and $g^*(\theta) = \lim_{\lambda\to 0} g^*(\theta)$ if $\lambda = 0$. Since $g(\theta)$ and $g^*(\theta)$ are continuous functions of θ at $\lambda = 0$ from (18), we can treat the $\lambda = 0$ case same as the $\lambda \neq 0$ case. When λ is given, β and σ^2 are uniquely estimated by the least squares method. Let $\hat\beta(\lambda)$ and $\hat\sigma^2(\lambda)$ be the estimators. Define

$$h_T(\lambda) = \frac{1}{T}G_T\{\lambda, \hat\beta(\lambda), \hat\sigma(\lambda)^2\} \quad \text{and} \quad h(\lambda) = \operatorname*{plim}_{T\to\infty} h_T(\lambda). \quad (19)$$

If $\lambda = \lambda_0$, the model becomes an ordinary regression model and $\hat\beta(\lambda_0)$ and $\hat\sigma(\lambda_0)^2$ are consistent. Therefore,

$$h(\lambda_0) = \operatorname*{plim}_{T\to\infty}\frac{1}{T}G_T(\theta_0) = E[g_t(\theta_0)]. \quad (20)$$

When $\lambda_0 = 0$,

$$g_t(\theta_0) = -\frac{1}{\sigma_0^2}[\frac{1}{2}(x_t'\beta_0)^2 u_t + x_t'\beta_0 u_t^2 + u_t^3] + x_t'\beta_0 + u_t \quad (21)$$

From (16) and (21), we get

$$E[g_t(\theta_0)] = 0, \quad (22)$$

for any value of λ_0. From (20) and (22),

$$h(\lambda_0) = 0. \quad (23)$$

Let $h'(\lambda) = dh/d\lambda$ if $\lambda \neq 0$ and $h'(0) = \lim_{\lambda\to 0} dh/d\lambda$. Since $g^*(\theta)$ is a continuous functions of θ, $h'(\lambda)$ is continuous in the neighborhood of λ_0. Therefore, there exists $\delta > 0$ such that $sign\{h'(\lambda)\} = sign\{h'(\lambda_0)\}$ and $|h'(\lambda)| \geq \gamma \equiv |h'(\lambda_0)|/2$ if $\lambda \in [\lambda_0 - \delta, \lambda_0 + \delta]$. From (23) and the mean value theorem, for any $\varepsilon \in (0, \delta)$,

$$h(\lambda_0 + \varepsilon) = h(\lambda_0 + \varepsilon) - h(\lambda_0) = h'(\lambda^*)\varepsilon \quad \text{and} \quad h(\lambda_0 - \varepsilon) = h(\lambda_0 - \varepsilon) - h(\lambda_0) = -h'(\lambda^{**})\varepsilon, \quad (24)$$

where λ^* and λ^{**} are some values in $[\lambda_0 - \varepsilon, \lambda_0 + \varepsilon]$. Therefore,

$$sign\{h(\lambda_0 - \varepsilon)\} \neq sign\{h(\lambda_0 + \varepsilon)\}, \quad |h(\lambda_0 - \varepsilon)| > \gamma\varepsilon, \quad \text{and} \quad |h(\lambda_0 + \varepsilon)| > \gamma\varepsilon. \quad (25)$$

Since $h_T(\lambda_0 - \varepsilon) \xrightarrow{P} h(\lambda_0 - \varepsilon)$ and $h_T(\lambda_0 + \varepsilon) \xrightarrow{P} h(\lambda_0 + \varepsilon)$,

$$P\Big[\ sign\{h_T(\lambda_0 - \varepsilon)\} \neq sign\{h_T(\lambda_0 + \varepsilon)\}, |h_T(\lambda_0 - \varepsilon)| > 0, \ \text{and} |h_T(\lambda_0 + \varepsilon)| > 0\ \Big] \to 1. \quad (26)$$

From the intermediate value theorem, $h_T(\lambda) = 0$ for some $\lambda \in [\lambda_0 - \varepsilon, \lambda_0 + \varepsilon]$ if $sign\{h_T(\lambda_0 - \varepsilon)\} \neq sign\{h_T(\lambda_0 + \varepsilon)\}$, $|h_T(\lambda_0 - \varepsilon)| > 0$, and $|h_T(\lambda_0 + \varepsilon)| > 0$. Therefore,

$$P[\text{ There exists } \hat{\lambda} \text{ such that } h_T(\hat{\lambda}) = 0 \text{ and } \hat{\lambda} \in [\lambda_0 - \varepsilon, \lambda_0 + \varepsilon]\ ] \to 1. \quad (27)$$

Since (27) holds for any $\varepsilon \in (0, \delta)$, $h_T(\lambda) = 0$ has a consistent root of λ_0. $\hat{\beta}(\hat{\lambda})$ and $\hat{\sigma}(\hat{\lambda})^2$ are obtained by the least squares method, they are consistent estimators when $\hat{\lambda} \xrightarrow{P} \lambda_0$. Hence there exists a consistent root among the roots of Equation (5).

References

Amemiya, T . (1985) *Advanced Econometrics*, Harvard University Press: Cambridge, MA.

Amemiya, T ., and J. L. Powell (1981) "A Comparison of the Box-Cox Maximum Likelihood Estimator and the Non-Linear Two Stage Least Squares Estimator," *Journal of Econometrics* **17**, 351-381.

Bickel, P. J. and K. A. Doksum (1981) "An Analysis of Transformations Revisited," *Journal of the American Statistical Association* **76**, 296-311.

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

Foster, A. M., L. Tain, and L. J. Wei (2001) "Estimation for the Box-Cox Transformation Model without Assuming Parametric Error Distribution," *Journal of the American Statistical Association* **96**, 1097-1101.

Hinkley, D.V. (1975) "On power transformations to symmetry," *Biometrika* **62**, 101-111.

Khazzoom, D. J. (1989) "A Note on the Application of the Non-linear two-stage least squares estimator to a Box-Cox Transformed Model," *Journal of Econometrics* **42** 377-379

Nawata, K. (1994) "Estimation of sample selection bias models by the maximum likelihood estimator and Heckman's two-step estimator," *Economic Letters* **45**, 33-40.

Powell, J. L. (1996) "Rescaled Method-of-Moments Estimation for the Box–Cox Regression Model," *Economics Letters* **51**, 259–265.

Showalter, M. H. (1994) "A Monte Carlo Investigation of the Box-Cox Model and a Nonlinear Least Squares Alternative," *Review of Economics and Statistics* **76**, 560-570.

Yang, Z. (2006) "A Modified Family of Power Transformations," *Economics Letters* **92**, 14–19.

Yeo, I. K., and R. A. Johnson (2000) "A New Family of Power Transformation to Improve Normality or Symmetry," *Biometrika* **87**, 954–959.

Table I. Uniform distribution, sample size =200.

		Mean	STD	Q1	Median	Q3
λ_0 =0.2						
BC MLE	λ	0.145	0.014	0.135	0.144	0.154
	β_0	-0.267	0.473	-0.608	-0.263	0.083
	β_1	0.089	0.377	-0.170	0.080	0.351
N−Estimator	λ	0.200	0.026	0.183	0.197	0.218
	β_0	0.021	0.465	-0.306	0.022	0.370
	β_1	0.078	0.362	-0.168	0.072	0.329
λ_0 =0.5						
BC MLE	λ	0.362	0.037	0.338	0.361	0.387
	β_0	-0.116	0.186	-0.232	-0.112	0.003
	β_1	0.105	0.150	0.006	0.110	0.203
N−Estimator	λ	0.497	0.063	0.452	0.495	0.539
	β_0	-0.004	0.187	-0.122	-0.001	0.112
	β_1	0.098	0.145	0.001	0.106	0.195
λ_0 =0.8						
BC MLE	λ	0.587	0.064	0.544	0.586	0.627
	β_0	-0.067	0.116	-0.146	-0.068	0.014
	β_1	0.104	0.090	0.042	0.104	0.164
N−Estimator	λ	0.801	0.108	0.729	0.792	0.868
	β_0	0.001	0.119	-0.078	0.000	0.083
	β_1	0.098	0.088	0.039	0.096	0.158

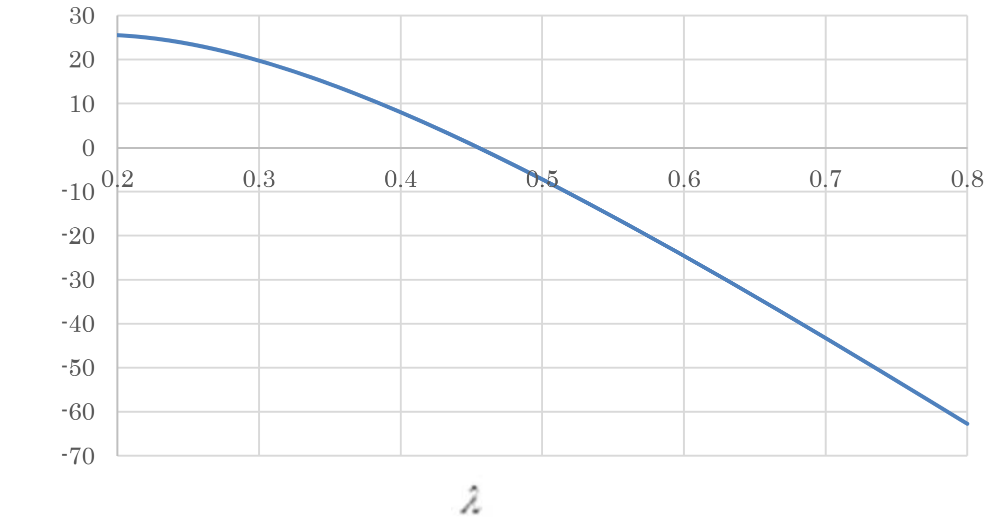
Table II. Doubly truncated normal distribution, sample size =200.

		Mean	STD	Q1	Median	Q3
λ_0 =0.2						
BC MLE	λ	0.175	0.015	0.165	0.176	0.186
	β_0	-0.048	0.323	-0.267	-0.050	0.168
	β_1	0.088	0.272	-0.091	0.098	0.275
N−Estimator	λ	0.201	0.020	0.187	0.201	0.214
	β_0	0.022	0.323	-0.197	0.016	0.242
	β_1	0.087	0.269	-0.086	0.096	0.270
λ_0 =0.5						
BC MLE	λ	0.440	0.042	0.414	0.440	0.468
	β_0	-0.020	0.135	-0.108	-0.018	0.068
	β_1	0.094	0.113	0.017	0.095	0.167
N−Estimator	λ	0.501	0.056	0.465	0.501	0.536
	β_0	0.007	0.135	-0.084	0.008	0.095
	β_1	0.093	0.112	0.014	0.093	0.164
λ_0 =0.8						
BC MLE	λ	0.710	0.068	0.666	0.711	0.753
	β_0	-0.013	0.081	-0.070	-0.009	0.043
	β_1	0.100	0.069	0.053	0.097	0.148
N−Estimator	λ	0.805	0.087	0.748	0.806	0.862
	β_0	0.003	0.081	-0.054	0.006	0.058
	β_1	0.100	0.069	0.053	0.098	0.148

Table III. Power transformation of $\chi^2(1)$, sample size = 200.

		Mean	STD	Q1	Median	Q3
λ_0 =0.2						
BC MLE	λ	0.182	0.021	0.167	0.183	0.195
	β_0	-0.055	0.307	-0.264	-0.060	0.164
	β_1	0.104	0.249	-0.075	0.102	0.284
N-Estimator	λ	0.203	0.027	0.184	0.202	0.219
	β_0	-0.009	0.306	-0.215	-0.015	0.205
	β_1	0.103	0.248	-0.073	0.104	0.278
λ_0 =0.5						
BC MLE	λ	0.457	0.055	0.424	0.459	0.491
	β_0	-0.014	0.115	-0.088	-0.012	0.059
	β_1	0.097	0.098	0.030	0.095	0.164
N-Estimator	λ	0.509	0.069	0.463	0.510	0.553
	β_0	0.004	0.115	-0.068	0.004	0.077
	β_1	0.097	0.098	0.031	0.095	0.164
λ_0 =0.8						
BC MLE	λ	0.732	0.091	0.672	0.734	0.792
	β_0	-0.010	0.074	-0.061	-0.008	0.038
	β_1	0.100	0.062	0.056	0.099	0.142
N-Estimator	λ	0.811	0.115	0.735	0.811	0.882
	β_0	0.000	0.074	-0.051	0.003	0.048
	β_1	0.100	0.062	0.056	0.099	0.143

Figure 1. Values of $G_T(\theta)$ (uniform distribution, $\lambda_0 = 0.5$)



Economics Bulletin, 2013, Vol. 33 No. 3 pp. 2287-2297