20th International Congress on Modelling and Simulation, Adelaide, Australia, 1–6 December 2013
www.mssanz.org.au/modsim2013

Robust estimation based on the first- and third-moment restrictions of the power transformation model

Nawata, K. [a]

[a] *Graduate School of Engineering, University of Tokyo, Bunkyo-ku, Tokyo 113-8656, Japan*
Email: nawata@tmi.t.u-tokyo.ac.jp

Abstract: The Box-Cox (1964) transformation model (hereafter called the BC model) is widely used in various fields of econometrics and statistics. However, since the error terms cannot be normal except in cases in which 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. However, these alternative estimators are not inconsistent if the distributions of the error terms are misspecified. Foster, Tain, and Wei (2001) and Nawata and Kawabuchi (2013) proposed semiparametric estimators. However, their estimators are not consistent under heteroscedasticity.

Powell (1996) proposed a semiparametric estimator based on the moment restriction. Although Powell's estimator is consistent under heteroscedasticity, the problems of the estimator are: (i) to identify the transformation parameter, λ, we need to introduce one or more instrumental variables, w_t, which satisfy $E(w_t \cdot u_t) = 0$ and are not included in x_t, and the result of the estimation changes depending on the selection of instrumental variables, (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 a rather arbitrary rescaling of y_t is necessary, and (iii) its finite-sample properties are not good and it often performs poorly, as shown in the Monte Carlo experiments.

Here I propose a new robust estimator of the power transformation model (the Box-Cox transformation model excluding the cases in which the transformation parameter is zero) given by

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

The estimator is based on only the first- and third-moment restrictions of the error terms and does not require the assumption of a specific distribution. The estimator is a root of the equations;

$$\sum_t (z_t - x_t\beta)^3 = 0, \ \text{ and } \quad \sum_t x_t(z_t - x_t'\beta) = 0.$$

The estimator is consistent if the first- and third-moments of the error terms are zero; that is, it is consistent even under heteroscedasticity. Moreover, it can be 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 and Powell's estimator.

Keywords: *Box-Cox transformation, power transformation, heteroscedasticity, robust estimator, moment restriction*

1. INTRODUCTION

The Box-Cox (1964) transformation model (hereafter called BC model) is widely used in various fields of econometrics and statistics. However, since the error terms cannot be normal except in cases in which 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. However, these alternative estimators are not inconsistent if the distributions of the error terms are misspecified. Foster, Tain, and Wei (2001) and Nawata and Kawabuchi (2013) proposed semiparametric estimators. However, these alternative estimators are not consistent under heteroscedasticity.

Powell (1996) proposed a semiparametric estimator based on the moment restriction. Although Powell's estimator is consistent under heteroscedasticity, the problems of the estimator are: (i) to identify the transformation parameter, λ, we need to introduce one or more instrumental variables, w_t, which satisfy $E(w_t \cdot u_t) = 0$ and are not included in x_t, and the result of the estimation changes depending on the selection of instrumental variables, (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 a rather arbitrary rescaling of y_t is necessary, and (iii) its finite-sample properties are not good and it often performs poorly as shown in the Monte Carlo experiments.

Here I propose a new robust estimator of the power transformation model: the Box-Cox transformation model excluding the cases in which the transformation parameter is zero. The estimator is based on only the first- and third-moment restrictions of the error terms and does not require the assumption of a specific distribution. The estimator is consistent even under heteroscedasticity. Its asymptotic distribution is obtained, and the results of Monte Carlo experiments are also presented

2. MODEL

We consider the simple power transformation model

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

where x_t and β are k-th dimensional vectors of explanatory variables and the coefficients, respectively, and λ is the transformation parameter. Let $(y_t^{\lambda} - 1)/\lambda = x_t^{*}'\beta^{*} + v_t$ and $v_t = u_t/\lambda$, in which case we obtain the BC model. However, to ensure the asymptotic distribution of the estimator, we only considered the $\lambda \neq 0$ case and did not consider the $\lambda = 0$ case. Therefore, we call this model a power transformation model rather than a BC model. $\{x_t\}$ and $\{u_t\}$ do not have to be independent and identically distributed (i.i.d.) random variables, and heteroscedasticity can be assumed. The following assumptions are made:

Assumption 1. $\{(x_t, u_t)\}$ are independent but not necessary identically distributed. The distribution of u_t may depend on x_t.

Assumption 2. u_t follows distributions whereby the supports are bounded from below (i.e., $f_t(u) = 0$ if $u \leq -a$ for some $a > 0$ where $f_t(u)$ is the probability (density) function.) For any t, the following moment conditions are satisfied: (i) $E(u_t \mid x_t) = 0$, (ii) $E(u_t^3 \mid x_t) = 0$, and (iii) $\delta_1 < E(u_t^6 \mid x_t) < \delta_2$ for some $0 < \delta_1 < \delta_2 < \infty$.

Assumption 3. $\{x_t\}$ are independent and its fourth moments are finite. The distributions of $\{x_t\}$ and the parameter space of β are restricted so that $\inf(x_t'\beta_0) > a$ and $\inf(x_t'\beta) > c$ for some $c > 0$ in the neighborhood of β_0 where β_0 is the true parameter value of β.

Here, instead of the BC likelihood function, we use the third-moment restriction and the roots of the equations;

$$G_T(\theta) = \sum_t (z_t - x_t\beta)^3 \equiv \sum_t g_t(\theta) = 0, \text{ and } \sum_t x_t(z_t - x_t'\beta) = 0, \tag{2}$$

where $\theta' = (\lambda, \beta)$, are considered. Note that the second equation of (2) gives the least-squares estimator when the value of λ is given. Let $\theta_0' = (\lambda_0, \beta_0)$ be the true parameter value of θ. Since $E[G(\theta_0)] = 0$, we obtain the following proposition:

Proposition 1

Among the roots of (2), there exists a consistent root.

Let $\hat{\theta}'=(\hat{\lambda},\hat{\beta}')$ be the consistent root. The asymptotic distribution of $\hat{\theta}$ is obtained by the following proposition.

Proposition 2

Let $\xi_t(\theta)=x_t(z_t-x_t'\beta)$ and $\ell_t(\theta)'=[g_t(\theta),\xi_t(\theta)']$. Suppose that $\frac{1}{T}\sum_t \frac{\partial \ell_t(\theta)}{\partial \theta'}|_{\theta_0}$ converges to a nonsingular matrix A in probability and that $\frac{1}{T}\sum_t E[\ell_t(\theta_0)\ell_t(\theta_o)']$ converges to a nonsingular matrix B. Then the asymptotic distribution of $\hat{\theta}$ is given by

$$\sqrt{T}(\hat{\theta}-\theta_0)\rightarrow N[0,A^{-1}B(A')^{-1}], \tag{3}$$

where $A=p\lim_{T\rightarrow\infty}\sum_t \frac{\partial \ell_t(\theta)}{\partial \theta'}|_{\theta_0}$, and $B=\lim_{T\rightarrow\infty}\frac{1}{T}\sum_t E[\ell_t(\theta_0)\ell_t(\theta_o)']$.

[Proof]

Let

$$\ell(\theta)=\sum_t \ell_t(\theta)=\begin{bmatrix} G_T(\theta) \\ \sum_t \xi_t(\theta)\end{bmatrix}. \tag{4}$$

Then

$$\sqrt{T}(\hat{\theta}-\theta_0)=-[\frac{1}{T}\frac{\partial \ell}{\partial \theta'}|_{\theta^*}]^{-1}\frac{1}{\sqrt{T}}\ell(\theta_0), \tag{5}$$

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

$$\ell_t(\theta_0)=\begin{bmatrix} u_t^3 \\ \frac{1}{\sigma_0^2}x_t u_t \end{bmatrix}. \tag{6}$$

Therefore, $E[\ell_t(\theta_0)]=0$. Since the variables $\{\ell_t(\theta_0)\}$ are independent and the Lindberg condition is satisfied under Assumptions 2 and 3, we obtain

$$\frac{1}{\sqrt{T}}\ell(\theta_0)\rightarrow N(0,B), \tag{7}$$

from Theorem 3.1.6 in Amemiya (1985, p. 92).

Since $\partial \ell/\partial \theta=\begin{bmatrix} 3\sum_t (z_t-x_t'\beta)^2 z_t\log(y_t) & -3\sum_t (z_t-x_t'\beta)^2 x_t \\ \sum_t z_t x_t'\beta\log(y_t) & -\sum_t x_t x_t' \end{bmatrix}$,

$$\frac{1}{T}\frac{\partial \ell(\theta)}{\partial \theta'}|_{\theta^*}\xrightarrow{P}A, \tag{8}$$

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 Equation (3).

3. MONTE CARLO STUDY

In this section some Monte Carlo results are presented for the BC MLE, the newly proposed estimator, and Powell's estimator. The behaviors of the estimators under both homoscedasticity and heteroscedasticity are studied. The basic model is

$$z_t = \beta_1 + \beta_2 x_t + u_t\,, \qquad z_t = y_t^{\lambda_0}, \quad y_t \geq 0, \qquad t = 1,2,..., T. \tag{9}$$

Note that when λ is given, β_1 and β_2 are obtained by the least-squares method. The BC MLE and the proposed estimator are calculated by the following scanning method (Nawata 1994; Nawata and Nagase 1996).

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

ii) Calculate $\hat{\beta}_1(\lambda)$ and $\hat{\beta}_2(\lambda)$ 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 proposed estimator, choose $\hat{\lambda}_{N_1}$, which satisfies $G_T(\theta_i) \cdot G_T(\theta_{i+1}) < 0$ where $\theta_i' = (\lambda_i, \hat{\beta}_1(\lambda_i), \hat{\beta}_2(\lambda_i))$.

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 steps (ii) and (iii).

v) Determine the final estimator.

For the proposed estimator, there are two possible problems. They are: i) Equation (2) has multiple solutions, and Equation (2) does not have a solution. However, all trials have just one solution and the above problems do not occur in the Monte Carlo study.

Since Powell (1996) suggested a function of x_t as the instrument variable w_t, we use $w_t = x_t^2$ and consider the moment restrictions, $E(x_t u_t) = 0$ and $E(x_t^2 u_t) = 0$. Since heteroscedasticity is also considered, the generalized method of moment (GMM) type estimator is not used and Powell's estimator is obtained by minimizing.

$$S = \{\sum_t x_t (z_t - \beta_0 - \beta_1 x_t)\}^2 + \{\sum_t x_t^2 (z_t - \beta_0 - \beta_1 x_t)\}^2. \tag{10}$$

Powell's estimator is also calculated by the scanning method over $\lambda \in [0, 2.0]$. As the proposed estimator, there are two possible problems for Powell's estimator. They are: (i) S is not minimized in $\lambda \in (0.01, 2.0)$ and S is minimized on the boundary (i.e., $\lambda = 0.01$ or $\lambda = 2.0$) , and (ii) S becomes 0 by multiple values of θ. Unlike the proposed estimator, these problems happen in many trials. Since we cannot get accurate values of the estimator in these trials, the results of Powell's estimator are calculated for trials without these problems.

3.1. Under homoscedasticity

In this section, the behavior of the estimators under homoscedasticity is analyzed. $\{x_t\}$ are i.i.d. random variables distributed uniformly on (0, 10). $\{u_t\}$ are i.i.d. random variables distributed uniformly on (-5, 5). The true parameter values are λ_0 = 0.4, $\beta_1 = 5.0$ and $\beta_2 = 0.1$. Values of 50, 100 and 200 are considered for the sample size *T*. The number of trials was 1,000 for all cases.

The results are presented in Table 1. Note that the following notation is used in the tables: STD, standard deviation; Ql, first quartile; and Q3, third quartile. For Powell's estimator, the following notation is also used: N1, number of trials where S is minimized at $\lambda = 0.01$; N2, number of trials where S is minimized at $\lambda = 2.0$; and N3, number of trials where $S = 0$ becomes 0 at multiple values of θ . The BC MLE underestimates λ and has a fairly large bias for all cases. Although the standard deviations of the proposed estimator are about 1.7 times larger than those of the BC MLE, the biases of the proposed estimator are much smaller. The bias almost disappears, even when T = 50. In terms of the mean squared error (MSE), the proposed estimator is better than the BC MLE if $T \geq 100$. (When T= 50, the MSEs of the two estimators are similar values.) The BC MLE also underestimates β_1 and β_2 . Although the standard deviations of the proposed estimator for β_1 and β_2 are about 1.5 and 2.0 times larger than those of the BC MLE, they are mainly caused by the scaling effect of the transformation. Since the BC MLE underestimates λ, the equation $z_t = y_t^{\lambda}$ holds and its variation become smaller than the true value. This effect makes the standard deviations smaller. Therefore, the smaller standard deviation does not directly indicate the superiority of the estimators for β_1 and β_2. Powell's estimator performs poorly. In many trials, we cannot get accurate values of the estimator because of the problems mentioned earlier. Moreover, although the biases are smaller than those of the BC MLE, the standard deviations are much larger than those of the newly proposed estimator even for trials without the problems.

3.2. Under heteroscedasticity

In this section, the effect of heteroscedasticity is analyzed. The values of x_t are chosen in the same way as in the previous section. The true parameter values are λ_0 = 0.4, $\beta_1 = 2.5$ and $\beta_2 = 0.25$. The error terms are given by

$$u_t = (1 + x_t / 10) \times \varepsilon_t \tag{12}$$

where $\{\varepsilon_t\}$ represent i.i.d. random variables distributed uniformly on (-2.5, 2.5). As before, Values of 50, 100 and 200 are considered for the sample size *T.,* are considered in the Monte Carlo study. The results are presented in Tables 2.

The BC MLE underestimates λ and the biases of the BC MLE are larger than those under homoscedasticity for all cases. This coincides with a previous report (Showalter, 1994) in which large biases of the BC MLE under heteroscedasticity were described. The standard deviations of the proposed estimator are about 2.5 times larger than those of the BC MLE. However, the biases of the proposed estimator are much smaller than those of the BC MLE. As a result, in terms of the MSE, the proposed estimator is better than the BC MLE in all cases. As before, Powell's estimator performs poorly. In many trials, we cannot get accurate values of the estimator. The standard deviations are much larger than those of the newly proposed estimator even for the trials without the problems.

4. CONCLUSION

Although the BC model is widely used in various fields, the BC MLE is not consistent. In this paper, a new robust estimator of the power transformation model is proposed. The estimator is based on the first- and third-moment restrictions of the error terms. The estimator is consistent even under heteroscedasticity 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 and Powell's estimator. However, the performance of the estimators may depend on the model; the findings may differ in other models. Further investigation is thus necessary to determine the conditions under which the proposed estimator shows superiority

REFERENCES

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

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.

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: 37-379.

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.

Nawata K. and K. Kawabuchi (2013). Evaluation of the 2006 revision of the medical payment system in Japan by a new estimator of the power transformation model -An analysis of the length of the hospital stay for cataract operations-. *mimeo*, Graduate School of Engineering, University of Tokyo.

Nawata, K., and Nagase, N. (1996), Estimation of sample selection biases models. *Econometric Reviews*, 15: 387-400.

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.

Table 1. BC MLE, proposed estimator, and Powell's estimator under homoscedasticity (λ_0 =0.4)

		Mean	STD	Q1	Median	Q3
BC MLE						
$T = 50$	λ	0.2956	0.0689	0.2490	0.2916	0.3397
	β_1	3.4299	1.4403	2.4383	3.0770	4.0417
	β_2	0.0534	0.0868	0.0021	0.0464	0.0935
$T = 100$	λ	0.2971	0.0457	0.2660	0.2956	0.3269
	β_1	3.2930	0.8513	2.6735	3.1507	3.7539
	β_2	0.0518	0.0568	0.0174	0.0485	0.0836
$T = 200$	λ	0.2962	0.0325	0.2741	0.2960	0.3162
	β_1	3.2358	0.5995	2.8214	3.1439	3.5691
	β_2	0.0500	0.0377	0.0261	0.0478	0.0728
Proposed Estimator						
$T = 50$	λ	0.4105	0.1200	0.3267	0.3985	0.4863
	β_1	6.7042	6.1429	3.3644	4.9468	7.6518
	β_2	0.1375	0.3435	0.0035	0.0871	0.2059
$T = 100$	λ	0.4071	0.0776	0.3562	0.3994	0.4540
	β_1	5.6734	2.6783	3.9680	5.0203	6.6415
	β_2	0.1164	0.1599	0.0304	0.0899	0.1814
$T = 200$	λ	0.4040	0.0556	0.3663	0.4008	0.4379
	β_1	5.3683	1.7211	4.1652	5.0267	6.1420
	β_2	0.1045	0.0906	0.0488	0.0939	0.1520
Powell's Estimator						
$T = 50$	λ	0.5050	0.3834	0.1754	0.3985	0.8302
	β_1	89.0886	889.8028	1.8422	4.8711	33.6328
	β_2	1.1844	3.7146	0.0013	0.0450	0.6169
	N1=56, N2=0, N3=242					
$T = 100$	λ	0.4933	0.3591	0.1809	0.3792	0.7761
	β_1	40.9710	144.2587	1.9585	4.3445	27.7987
	β_2	1.2856	3.0677	0.0082	0.0538	0.7242
	N1=23, N2=0, N3=119					
$T = 200$	λ	0.4713	0.3257	0.1879	0.3519	0.7190
	β_1	27.5082	69.8728	2.0090	4.1275	22.4976
	β_2	1.0232	2.4328	0.0190	0.0608	0.4831
	N1=3, N2=0, N3=165					

Table 2. BC MLE, proposed estimator, and Powell's estimator under heteroscedasticity (λ_0 =0.4)

		Mean	STD	Q1	Median	Q3
BC MLE						
$T = 50$	λ	0.2561	0.0460	0.2249	0.2544	0.2848
	β_1	1.7630	0.3765	1.4876	1.6988	1.9863
	β_2	0.1017	0.0645	0.0594	0.0936	0.1390
$T = 100$	λ	0.2569	0.0320	0.2347	0.2553	0.2777
	β_1	1.7486	0.2770	1.5564	1.7143	1.8933
	β_2	0.0994	0.0406	0.0711	0.0952	0.1256
$T = 200$	λ	0.2564	0.0217	0.2428	0.2562	0.2705
	β_1	1.7370	0.1832	1.6060	1.7312	1.8408
	β_2	0.0975	0.0287	0.0778	0.0957	0.1163
Proposed Estimator						
$T = 50$	λ	0.4234	0.1293	0.3356	0.4053	0.4836
	β_1	2.6058	1.4509	1.9266	2.4520	3.1006
	β_2	0.5666	1.7228	0.1248	0.2483	0.5043
$T = 100$	λ	0.4135	0.0776	0.3598	0.4066	0.4570
	β_1	2.6158	0.7699	2.1544	2.4843	2.9169
	β_2	0.3324	0.2783	0.1617	0.2624	0.4228
$T = 200$	λ	0.4066	0.0562	0.3708	0.4021	0.4391
	β_1	2.5570	0.4757	2.2451	2.4935	2.8151
	β_2	0.2896	0.1597	0.1843	0.2504	0.3520
Powell's Estimator						
$T = 50$	λ	0.4532	0.3360	0.1433	0.3869	0.7521
	β_1	3.3420	4.0932	1.2228	1.7813	3.6622
	β_2	1.8866	3.1677	0.0271	0.1809	2.2608
	N1=79, N2=0, N3=127					
$T = 100$	λ	0.4085	0.3174	0.1362	0.2987	0.6841
	β_1	2.9038	3.0840	1.2667	1.7109	3.3203
	β_2	1.4593	2.7281	0.0286	0.1210	1.1863
	N1=34, N2=0, N3=164					
$T = 200$	λ	0.4080	0.2990	0.0744	0.1492	0.3170
	β_1	2.9569	2.2461	1.3233	1.7966	3.7187
	β_2	1.2648	2.4834	0.0343	0.1159	1.0618
	N1=15, N2=0, N3=206					