Scientific
Research
Publishing

Asymptotic Efficiency of the Maximum Likelihood Estimator for the Box-Cox Transformation Model with Heteroscedastic Disturbances

Kazumitsu Nawata

Graduate School of Engineering, University of Tokyo, Tokyo, Japan
Email: nawata@tmi.t.u-tokyo.ac.jp

Abstract

This paper considers the asymptotic efficiency of the maximum likelihood estimator (MLE) for the Box-Cox transformation model with heteroscedastic disturbances. The MLE under the normality assumption (BC MLE) is a consistent and asymptotically efficient estimator if the "small σ" condition is satisfied and the number of parameters is finite. However, the BC MLE cannot be asymptotically efficient and its rate of convergence is slower than ordinal order $1/\sqrt{n}$ when the number of parameters goes to infinity. Anew consistent estimator of order $1/\sqrt{n}$ is proposed. One important implication of this study is that estimation methods should be carefully chosen when the model contains many parameters in actual empirical studies.

1. Introduction

The Box-Cox transformation model (BC model) [1] is widely used in empirical studies. For details on this model, see Hossain [2] and Sakia [3]. The maximum likelihood estimator (MLE), which maximizes the likelihood function under the normality assumption (BC MLE), can be asymptotically efficient if the "small σ" condition described by Bickel and Doksum [4] is satisfied. On the other hand, the model with heteroscedastic disturbances, in which variances are different among groups, is also widely used in the analysis of various datasets such as panel data [5]. It is sometimes necessary to consider a model combining these two models. Nawata and Kawabuchi [6]-[11] analyzed length

of stay (LOS) in Japanese hospitals using the BC model. They found that the variances among hospitals were often very different among hospitals even after the transformation. Their studies are these cases.

It is well known that the MLE is usually an asymptotically efficient estimator when the number of parameters is finite. However, this may not be true when the number of parameters goes to infinity. It is often necessary for us to consider cases in which numbers of groups go to infinity. For example, the new medical payment system known as the Diagnostic Procedure Combination/Per Diem Payment System (DPC/PDPS) was introduced in 2003 in Japan, and as of April 2014, 1863 hospitals had either already joined or were preparing to join; this number has been increasing [12]. The hospitals joining the DPC/PDPS are required to computerize their medical information. This means that it has become possible to analyze a large scale dataset that contains information from many hospitals. In other words, it is necessary for us to consider the asymptotic properties of estimators when the number of groups (hospitals) goes to infinity.

This paper considers the estimation of the Box-Cox transformation model with heteroscedastic disturbances when the number of groups that increases to infinity. In such cases, the conventional maximum likelihood method yields only an estimator whose rate of convergence is slower than ordinal order of $1/\sqrt{n}$ even if the "small σ" condition is satisfied in all groups. Then a new estimation method that can handle these problems is proposed.

2. BC Model with Heteroscedastic Disturbances

Suppose that t_{ij} is the explanatory variable of observation j in group i (for example, LOS of patient j in hospital i in Nawata and Kawabuchi [6]-[11]). I consider the BC model:

$$y_{ij} = \left(t_{ij}^{\lambda} - 1\right)\big/\lambda \ \text{ if } \ \lambda \neq 0, \quad y_{ij} = \log\left(t_{ij}\right) \ \text{ if } \ \lambda = 0, \tag{1}$$

$$y_{ij} = x'_{ij}\beta + u_{ij}, \ \ i = 1, 2, \cdots, k, \ j = 1, 2, \cdots, n_i,$$

with heteroscedastic disturbances and variances given by

$$V\left(u_{ij}\right) = \sigma_i^2, \tag{2}$$

where λ is the transformation parameter, x_{ij} and β are the vectors of the explanatory variables and coefficients, k is the number of groups, and n_i is the number of people in group i. We assume that the "small σ" condition described by Bickel and Doksum [4] ($\lambda\sigma_i\big/\left(1+\lambda x'_{ij}\beta\right) \to 0$, in practice $P\left[t_{ij} < 0\right]$ under the normality assumption, is small enough, for all i and j) is satisfied.[1] Under this condition, we can assume that u_{ij} follows the normal distribution with mean 0 and variance σ_i^2. The

---

[1]If the "small σ" condition is not satisfied, we can use the estimator proposed by Nawata [15] instead of the BC MLE. Even in this case, however, we reach the same conclusion as that presented here; that is, a model that considers heteroscedasticity and a number of parameters that goes to infinity is simply a consistent estimator of order $1/\sqrt{m}$ and there exists a consistent estimator of order $1/\sqrt{n}$ by a modification of the homoscedastic case.

likelihood function is given by [13]

$$\log L(\lambda, \beta, \sigma_i) = \sum_{i,j} \left[\log \phi \left\{ \left(y_{ij} - x_{ij}'\beta \right) \middle/ \sigma_i \right\} - \log \sigma_i \right] + \sum_{i,j}^{n} (\lambda - 1) \log t_{ij} \tag{3}$$

where ϕ is the density function of the standard normal distribution. Although the likelihood function is a function of $\sigma_1, \sigma_2, \cdots, \sigma_k$, we simply write it as (3). Note that Showalter [14] reports a large bias of the BC MLE when heteroscedasticity is ignored.

3. Estimation of the Model When the Number of Groups Goes to Infinity

Let the numbers of observations be $n = \sum_i n_i$. We assume that $\sum_{i,j} x_{ij} x_{ij}' / n$ converges to a nonsingular matrix in probability, $k \to \infty$ and $n_i \to \infty$. Note that n_i is order of m and satisfies $n_i / m \to \gamma_i$ and $0 < c_1 < \gamma_i < c_2 < \infty$ for all i. Since $k \to \infty$, we get $m = o(n)$. k is assumed to increase at a slower rate than n_i; that means that $k = o(\sqrt{n})$. When k is finite, we can estimate $\lambda, \beta, \sigma_1, \sigma_2, \cdots, \sigma_k$ by maximizing (3). Let $\lambda_0, \beta_0, \sigma_{i0}$ be the true parameter values of λ, β, σ_i and let $\hat{\sigma}_i$ be the MLE of σ_i. We do not assume any specific forms of the variances, and simply assume $0 < c_3 < \sigma_{i0} < c_4 < \infty$ for all i. However, since n_i is $O(m)$, $\hat{\sigma}_i$ can only be a consistent estimator of order $1/\sqrt{m}$ by any estimation method (Baltagi and Griffin [5] considered different variance estimators). When we substitute $\hat{\sigma}_i$, the conditions that the estimators obtained by maximizing (3) become order $1/\sqrt{n}$; *i.e.*
$\sqrt{n}(\hat{\lambda} - \lambda) = O_p(1)$ and $\sqrt{n}(\hat{\beta} - \beta_0) = O_p(1)$, are given by

$$\left. \frac{\partial \log L}{\partial \lambda} \right|_{\lambda_0, \beta_{10}, \hat{\sigma}_i} = O_p(\sqrt{n}) \tag{4}$$

$$\left. \frac{\partial \log L}{\partial \beta} \right|_{\lambda_0, \beta_0, \hat{\sigma}_i} = O_p(\sqrt{n}). \tag{5}$$

As before, although the values of derivatives are at $\hat{\sigma}_1, \hat{\sigma}_2, \cdots, \hat{\sigma}_k$, we write them simply as (4). Then (5) becomes

$$\left. \frac{\partial \log L}{\partial \beta} \right|_{\lambda_0, \beta_0, \hat{\sigma}_i} = -\sum_i \frac{1}{\hat{\sigma}_i^2} \sum_j x_{1ij} \varepsilon_{ij} . \tag{6}$$

Here,

$$\hat{\sigma}_i - \sigma_i = O_p(1/\sqrt{m}) \tag{7}$$

$$\sum_j x_{ij}' \varepsilon_{ij} = O_p(\sqrt{m}). \tag{8}$$

We get

$$\frac{1}{\sqrt{n}} \left\{ \sum_i \left(\frac{1}{\hat{\sigma}_i} - \frac{1}{\sigma_{i0}} \right) \sum_j x_{ij} \varepsilon_{ij} \right\} = O_p(k/\sqrt{n}). \tag{9}$$

Therefore, if $k = o(\sqrt{n})$, (5) is satisfied. This means that we can use the standard method of dealing with heteroscedasticity if and only if $k = o(\sqrt{n})$ in the standard model.

However, for the transformation parameter λ, since

$$\frac{\partial \log L}{\partial \lambda} = -\sum_i \frac{1}{\sigma_i^2} \sum_j \left(y_{ij} - x_{ij}'\beta \right) \log\left(t_{ij} \right) y_{ij} + \sum_{i,j} \log t_{ij}, \tag{10}$$

we get [15] [16]

$$\begin{aligned}
\left.\frac{\partial \log L}{\partial \lambda}\right|_{\lambda_0, \beta_0, \hat{\sigma}_i} &= -\frac{1}{\hat{\sigma}_i^2 \lambda_0} \sum_{i,j} \left[\left\{ \frac{\left(\lambda_0 x_{ij}'\beta_0 + 1\right)\log\left(\lambda_0 x_{ij}'\beta_0 + 1\right)}{\lambda_0} - x_{ij}'\beta \right\} u_{ij} \right. \\
&\qquad \left. + \log\left(\lambda_0 x_{ij}'\beta_0 + 1\right) u_{ij}^2 + \frac{\lambda_0 u_{ij}^3}{\lambda_0 x_{ij}'\beta_0 + 1} \right] \\
&\qquad + \sum_{i,j} \left\{ \frac{1}{\lambda_0} \log\left(\lambda_0 x_{ij}\beta_0 + 1\right) + \frac{u_{ij}}{\lambda_0 x_{ij}'\beta_0 + 1} \right\} \\
&= \left.\frac{\partial \log L}{\partial \lambda}\right|_{\lambda_0, \beta_0, \sigma_{i0}} + \left(\left.\frac{\partial \log L}{\partial \lambda}\right|_{\lambda_0, \beta_0, \hat{\sigma}_i} - \left.\frac{\partial \log L}{\partial \lambda}\right|_{\lambda_0, \beta_0, \sigma_{i0}} \right) \\
&= \frac{1}{\lambda_0} \sum_i \left(\frac{1}{\sigma_{i0}^2} - \frac{1}{\hat{\sigma}_i^2} \right) \sum_j \left[\left\{ \frac{\left(\lambda_0 x_{ij}'\beta_0 + 1\right)\log\left(\lambda_0 x_{ij}'\beta_0 + 1\right)}{\lambda_0} - x_{ij}'\beta \right\} u_{ij} \right. \\
&\qquad \left. + \log\left(\lambda_0 x_{ij}'\beta_0 + 1\right) u_{ij}^2 + \frac{\lambda_0 u_{ij}^3}{\lambda_0 x_{ij}'\beta_0 + 1} \right] + O_p\left(1/\sqrt{n}\right)
\end{aligned} \tag{11}$$

under the "small σ" condition, where y_{ij}^* is the value of y_{ij} when $\lambda = \lambda_0$.

Since

$$E\left\{ \log\left(\lambda_0 x_{ij}'\beta_0 + 1\right) \varepsilon_{ij}^2 \right\} = E\left\{ \log\left(\lambda_0 x_{ij}'\beta_0 + 1\right) \right\} \sigma_{i0}^2 \neq 0, \tag{12}$$

we get

$$\left.\frac{\partial \log L^{**}}{\partial \lambda}\right|_{\lambda_0, \beta_{10}, \hat{\sigma}_i} = O_p\left(n/\sqrt{m}\right) \tag{13}$$

and (4) is not satisfied. This means that the MLE becomes a consistent estimator only of order $1/\sqrt{m}$; that is, the rate of convergence is slower than ordinal order $1/\sqrt{n}$ when $k \to \infty$. This means that the estimator of λ cannot be a consistent estimator of order $1/\sqrt{n}$. Here,

$$\frac{\partial^2 \log L}{\partial \lambda \partial \beta} = \sum_i \frac{1}{\sigma_i^2} \sum_j x_{ij} \log\left(t_{ij} \right) y_{ij} \tag{14}$$

Therefore, $\hat{\beta}$ cannot be an estimator of order $1/\sqrt{n}$ either.

4. A Consistent Estimator of Order $1/\sqrt{n}$

Here, an alternative estimator is proposed by an essential modification of the likelihood function. Suppose that disturbances are homoscedastic and that $\sigma_i^2 = \sigma^2$ for all i. Then the likelihood becomes

$$\log L^*\left(\lambda, \beta, \sigma\right) = \sum_{i,j} \log f_{ij}\left(\lambda, \beta, \sigma\right) \tag{15}$$

where $\log f_{ij}^{*}\left(\lambda,\beta,\sigma\right)=\log\phi\left\{\left(y_{ij}-x_{ij}'\beta\right)/\sigma\right\}-\log\sigma+\left(\lambda-1\right)\log\left(t_{ij}\right)$.

Instead of maximizing (15), we considered the roots of the equations,

$$\frac{\partial\log L^{*}}{\partial\lambda}=-\frac{1}{\sigma^{2}}\sum_{i,j}\left(y_{ij}-x_{ij}'\beta\right)\log\left(t_{ij}\right)y_{ij}+\sum_{i,j}\left(\lambda-1\right)\log t_{ij}=0,\tag{16}$$

$$\frac{\partial\log L^{*}}{\partial\beta}\sum x_{ij}\left(y_{ij}-x_{ij}'\beta\right)=0,\tag{17}$$

$$H\left(\theta\right)=\sigma^{2}-\frac{\sum_{i,j}\log\left(\lambda x_{ij}'\beta+1\right)\left(y_{ij}-x_{ij}'\beta\right)^{2}}{\sum_{i.j}\log\left(\left(\lambda x_{ij}'\beta+1\right)\right)}=0.\tag{18}$$

For the standard maximum likelihood method, the variance is estimated by the simple average. However, in this case, the variance is estimated by the weighted average of least squares residuals.

We assume

$$\lim_{n\to\infty}\bar{\sigma}_{n}^{2}=\bar{\sigma}^{2}>0\tag{19}$$

where $\bar{\sigma}_{n}^{2}=\dfrac{\sum_{i,j}\log\left(\lambda_{0}x_{ij}'\beta_{0}+1\right)\sigma_{i}^{2}}{\sum_{i.j}\log\left(\lambda_{0}x_{ij}'\beta_{0}+1\right)}$.

Here,

$$\begin{aligned}\left.\frac{\partial\log L^{*}}{\partial\lambda}\right|_{\lambda_{0},\beta_{0},\sigma^{2}}=&-\frac{1}{\sigma^{2}\lambda_{0}}\sum_{i,j}\left[\left\{\frac{\left(\lambda_{0}x_{ij}'\beta_{0}+1\right)\log\left(\lambda_{0}x_{ij}'\beta_{0}+1\right)}{\lambda_{0}}-x_{ij}'\beta\right\}u_{ij}\right.\\&\left.+\log\left(\lambda_{0}x_{ij}'\beta_{0}+1\right)u_{ij}^{2}+\log\left(\lambda_{0}x_{ij}'\beta_{0}+1\right)u_{ij}^{2}+\frac{\lambda_{0}u_{ij}^{3}}{\lambda_{0}x_{ij}'\beta_{0}+1}\right]\\&+\sum_{i,j}\left\{\frac{1}{\lambda_{0}}\log\left(\lambda_{0}x_{ij}'\beta_{0}+1\right)+\frac{u_{ij}}{\lambda_{0}x_{ij}'\beta_{0}+1}\right\}\end{aligned}\tag{20}$$

where $\vartheta_{0}'=\left(\lambda_{0},\beta_{0},\sigma^{2}\right)$. Therefore,

$$E\left[\left.\frac{\partial\log L^{*}}{\partial\lambda}\right|_{\vartheta_{0}}\right]=-\frac{1}{\lambda_{0}}\left\{\frac{1}{\sigma^{2}}\sum_{i.j}\log\left(\lambda_{0}x_{ij}'\beta_{0}+1\right)\sigma_{i0}+\sum_{i,j}\log\left(\lambda_{0}x_{ij}'\beta_{0}+1\right)\right\}.\tag{21}$$

From (21), we get

$$E\left[\left.\frac{\partial\log L^{*}}{\partial\lambda}\right|_{\bar{g}_{n}}\right]=0,\tag{22}$$

$$E\left[\left.\frac{\partial\log L^{*}}{\partial\beta}\right|_{\bar{g}_{n}}\right]=0,\tag{23}$$

$$E\left[H\left(\bar{\vartheta}_{n}\right)\right]=0.\tag{24}$$

where $\bar{\vartheta}_{n}=\left(\lambda_{0},\beta_{0},\bar{\sigma}_{n}^{2}\right)$. Based on the same argument presented by Nawata [15] [17],

there exist consistent estimators of λ and β among the roots of (16)-(18). Let $\hat{\theta}'_N = \left(\hat{\lambda}_N, \hat{\beta}_N, \hat{\sigma}^2_N\right)$ be the consistent root and let $\vartheta^*_0 = \left(\lambda_0, \beta_0, \bar{\sigma}^2\right)$. The asymptotic distribution of this estimator is given by

$$\sqrt{T}\left(\hat{\theta}_N - \vartheta^*_0\right) \to N\left[0, A^{-1} B \left(A'\right)^{-1}\right],\tag{25}$$

where $A = -\lim_{n\to\infty} \dfrac{1}{n}\sum_{i,j} E\left[\left.\dfrac{\partial \xi_{ij}}{\partial \theta'}\right|_{\vartheta^*_0}\right]$, $\xi_{ij}\left(\theta\right)' = \left[\dfrac{\partial \log L^*}{\partial \lambda}, x_{ij}\left(y_{ij} - x'_{ij}\beta\right), H\left(\theta\right)\right]$,

$B = -\lim_{n\to\infty} \dfrac{1}{n}\sum_{i,j} E\left[\left.\dfrac{\partial \varsigma_{ij}}{\partial \theta}\right|_{\vartheta^*_0} \left.\dfrac{\partial \varsigma_{ij}}{\partial \theta'}\right|_{\vartheta^*_0}\right]$, $\varsigma_{ij}\left(\theta_0\right)' = \left[\dfrac{\partial \log f^*_{ij}}{\partial \lambda}, x'_{ij}u_{ij}, w_i\left(u^2_{ij} - \sigma^2_i\right)\right]$, and

$w_{ij} = \log\left(\lambda_0 x'_{ij}\beta_0 + 1\right)\Big/\left(\sum_{i,j}\log\left(\lambda_0 x'_{ij}\beta_0 + 1\right)\Big/n\right).$

σ^2_i is estimated by $\hat{\sigma}^2_i = \sum_j\left(\hat{y}_{ij} - x'_{ij}\hat{\beta}\right)^2\Big/n_i$ where $\hat{y}_{ij} = \left(t^{\hat{\lambda}_N}_{ij} - 1\right)\Big/\hat{\lambda}_N$. This means that $\hat{\lambda}_N$ and $\hat{\beta}_N$ are consistent estimators of order $1\big/\sqrt{n}$ and are asymptotically more efficient than the BC MLE.

5. Conclusion

This paper considers the estimation of the BC model with heteroscedastic disturbances; that is, variances are different by groups. The BC MLE is a consistent and asymptotically efficient estimator if the "small σ" condition described by Bickel and Doksum [4] is satisfied and the number of parameters is finite. However, its rate of convergence is slower than ordinal order of $1\big/\sqrt{n}$ and the BC MLE cannot be efficient when the heteroscedasticity of disturbances is considered and the number of groups goes to infinity. An alternative consistent estimator based on a modification of the likelihood function is considered. It is a consistent estimator of order $1\big/\sqrt{n}$. One important result of this study is that the MLE might not be a good estimator and estimation methods should be carefully chosen when the model contains many parameters in the actual empirical studies.

Acknowledgements

References

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

[2] Hossain, M.Z. (2011) The Use of Box-Cox Transformation Technique in Economic and Statistical Analyses. *Journal of Emerging Trends in Economics and Management Sciences*, **2**, 32-39.

[3] Sakia, R.M. (1992) The Box-Cox Transformation Technique: A Review. *The Statistician*, **41**, 169-178. http://dx.doi.org/10.2307/2348250

[4] Bickel, P.J. and Doksum, K.A. (1981) An Analysis of Transformations Revisited. *Journal of American Statistical Association*, **76**, 296-311. http://dx.doi.org/10.1080/01621459.1981.10477649

[5] Baltagi, B.H. and Griffin, J.M. (1988) A Generalized Error Component Model with Heteroscedastic Disturbances. *International Economic Review*, **29**, 745-753. http://dx.doi.org/10.2307/2526831

[6] Nawata, K. and Kawabuchi, K. (2015) Evaluation of Length of Hospital Stay Joining Educational Programs for Type 2 Diabetes Mellitus Patients: Can We Control Medical Costs in Japan? *Health*, **7**, 256-269. http://dx.doi.org/10.4236/health.2015.72030

[7] Nawata. K. and Kawabuchi, K. (2015) Financial Sustainability of the Japanese Medical Payment System: Analysis of the Japanese Medical Expenditure for Educational Hospitalization of Patients with Type 2 Diabetes. *Health*, **7**, 1007-1021. http://dx.doi.org/10.4236/health.2015.78118

[8] Nawata, K. and Kawabuchi, K. (2015) An Econometric Analysis of Hospital Length of Stay for Cataract Operations in Japan by the Box-Cox Transformation Model and Hausman Tests: Evaluation of the 2010 Revision of the Medical Payment System. *Open Journal of Applied Science*, **5**, 559-570. http://dx.doi.org/10.4236/ojapps.2015.59054

[9] Nawata, K. and K. Kawabuchi (2016) Comparison of the Length of Stay and Medical Expenditures among Japanese Hospitals for Type 2 diabetes Treatments: The Box-Cox Transformation Model under Heteroscedasticity. *Health*, **8**, 49-63. http://dx.doi.org/10.4236/health.2016.81007

[10] Nawata, K. and Kawabuchi, K. (2016) Did the Revision of the Japanese Medical Payment System Work Properly?—An Analysis of Averages and Variances of Length of Hospital Stay for Type 2 Diabetes Patients by Individual Hospital. *Health*, **8**, 505-517. http://dx.doi.org/10.4236/health.2016.86054

[11] Nawata, K. and Kawabuchi, K. (2016) A Long Term Evaluation of the Japanese Medical Payment System for Cataract Surgeries: Did the Medical Policy Reduce the Long Hospital Stay in Japan? *Health*, **8**, 905-919. http://dx.doi.org/10.4236/health.2016.89094

[12] Central Social Insurance Medical Council (2014) DPC taishou hyouin Junbi hyouin no kibo (heisei 26 nen 4gatsu 1 nichi) mikomi (Estimated Sizes of DPC Hospitals and Preparing Hospitals as of April 1st, 2014). (In Japanese)

[13] Greene, W.H. (1997) Econometric Analysis. 3rd Edition, Prentice-Hall, Upper Saddle River, New Jersey.

[14] 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. http://dx.doi.org/10.2307/2109980

[15] Nawata, K. (2013) A New Estimator of the Box-Cox Transformation Model Using Moment Conditions. *Economics Bulletin*, **33**, 2287-2297.

[16] Nawata, K. and Kawabuchi, K. (2014) A New Test for the Box-Cox Transformation Model: An Analysis of Length of Hospital Stay for Diabetes Patients in Japan *Economics Bulletin*, **34**, 324-332.

[17] Nawata, K. (2015) Robust Estimation Based on the Third-Moment Restriction of the Error Terms for the Box-Cox Transformation Model: An Estimator Consistent Under Heteroscedasticity. *Economics Bulletin*, **3**, 1056-1064.

Scientific Research Publishing

Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: http://papersubmission.scirp.org/

Or contact ojs@scirp.org