The Canadian Journal of Statistics
Vol. 30, No. 2, 2002, Pages 235–242
La revue canadienne de statistique

Median estimation through a regression transformation

Zhenlin YANG

Key words and phrases: Box–Cox transformation; confidence interval; correction factor; linear model; median.

MSC 2000: Primary 62J05; secondary 62F25, 62E20.

Abstract: The author considers the problem of constructing confidence intervals for the median of a future observation at certain values of exogenous variables, following a normalizing transformation. He shows that when this transformation is estimated, the usual interval obtained through an inverse transformation needs to be corrected, even when the sample size is large. He then gives a simple analytical solution to this problem and provides simulation results confirming the good small-sample properties of the corrected interval. He also presents two concrete illustrations.

L'estimation de la médiane par l'intermédiaire d'une transformation de la régression

Résumé : L'auteur s'intéresse à la construction d'intervalles de confiance pour la médiane d'une observation future correspondant à certaines valeurs de variables exogènes, suite à une transformation normalisatrice. Il montre que lorsque cette transformation est estimée, l'intervalle usuel obtenu par transformation inverse doit être corrigé, même pour de grandes tailles d'échantillon. Il propose ensuite une solution analytique simple à ce problème et rapporte des résultats de simulation attestant des bonnes propriétés de l'intervalle corrigé dans de petits échantillons. Il présente en outre deux illustrations pratiques.

1. INTRODUCTION

Data arising from biological, environmental, medical and economic studies are typically skewed and an appropriate measure of the central tendency in these cases is the median. In analyzing these data, it is popular first to use the Box–Cox transformation technique to normalize the data and stabilize the variance and then to apply the usual regression technique to the transformed data. Attractive features of this approach include (i) standard normal theory inference methods can be used, and (ii) a simple inverse transformation gives point or interval estimation concerning a future observation. These are especially true for a model with only the response being transformed (Box & Cox 1964). The resulting intervals based on an inverse transformation are exact or approximate, depending on whether the data can be transformed to exact normality or only to near normality.

Often in practice, however, the transformation (or transformation parameter) is unknown and has to be estimated from the data. In this case, a common practice is to replace the unknown transformation by its estimate and treat the estimated transformation as the true one (Collins 1991; Hahn & Meeker 1991, p. 72). Such "plug-in" type intervals ignore the effect of transformation estimation, which should be studied or corrected to account for the transformation estimation.

In this paper, we study the confidence interval for the median of a future observation at certain values of concomitant variables, obtained through a simple inverse transformation, and show that when the transformation is unknown and is replaced by its estimate, the usual transformation-based interval needs to be corrected even when the sample size is large. We then give a simple analytical correction. Monte Carlo simulation shows that the corrected interval performs very well, having coverage probabilities very close to their nominal levels for small to moderate sample sizes. Real data examples indicate that the transformation approach with correction gives more reasonable confidence intervals than those without transformation and correction.

Section 2 outlines the Box–Cox transformation-based confidence interval for the median of an original future observation. Section 3 introduces the analytical correction. Section 4 contains some Monte Carlo simulation results. Section 5 presents two real data examples and Section 6 gives a general discussion.

2. THE BOX–COX CONFIDENCE INTERVAL

Let $\mathbf{Y}$ be an $n \times 1$ vector of original observations, let $h(\mathbf{Y}, \lambda)$ be a vector of transformed observations, and let $\mathbf{X}$ be an $n \times p$ matrix whose columns contain the values of the explanatory variables $X_1, \ldots, X_p$. The Box–Cox transformed linear model (Box & Cox 1964) has the form

$$h(\mathbf{Y}, \lambda) = \mathbf{X}\beta + \sigma \mathbf{e}, \tag{1}$$

where β is a $p \times 1$ vector of parameters, σ is the error standard deviation, $\mathbf{e}$ is an $n \times 1$ vector of independent and identically distributed (iid) errors with zero mean and median, and unit standard deviation, and $h(\cdot, \lambda)$ is a monotonically increasing function. Box and Cox assumed that there exists a λ such that the transformed linear model (1) is normal, which leads to the *Box–Cox estimators* for β, σ and λ as follows:

$$\hat\beta(\hat\lambda) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'h(\mathbf{Y}, \hat\lambda),\ \hat\sigma^2(\hat\lambda) = \frac{1}{n}\left\|Qh(\mathbf{Y}, \hat\lambda)\right\|^2,\ \hat\lambda = \arg\min_{\ell} \dot{Y}^{-\ell}\|Qh(\mathbf{Y}, \ell)\|, \tag{2}$$

where $Q = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and $\mathbf{I}_n$ is an $n \times n$ identity matrix. For a given λ, the Box–Cox estimators of β and σ are $\hat\beta(\lambda)$ and $\hat\sigma(\lambda)$. In general, the Box–Cox estimators are not maximum likelihood estimators (MLE) unless the e_i are exactly normal.

Consider the problem of predicting $\delta_0 = \text{median}(Y_0)$ at x_0. As the transformed observations are assumed to have the same mean and median and the transformation is monotonic, we have $\mathrm{E}\{h(Y_0, \lambda)\} = h(\delta_0, \lambda)$. A natural predictor for $h(\delta_0, \lambda)$ is $x_0'\hat\beta(\lambda)$ with variance $x_0'(\mathbf{X}'\mathbf{X})^{-1}x_0\sigma^2$. The assumptions of the model (1) suggest that

$$T(\lambda) = \frac{x_0'\hat\beta(\lambda) - h(\delta_0, \lambda)}{\hat\sigma(\lambda)\sqrt{x_0'(\mathbf{X}'\mathbf{X})^{-1}x_0}}\sqrt{\frac{n-p}{n}} \tag{3}$$

is distributed as a t_{n-p} if the model errors are normal and an approximate t_{n-p} if the model errors are approximately normal, where t_ν denotes Student's distribution with ν degrees of freedom. This immediately leads to an exact or approximate $100(1-\alpha)\%$ confidence interval (CI) for $h(\delta_0, \lambda)$: $x_0'\hat\beta(\lambda) \pm t_{n-p}(\alpha/2)k_n(x_0)\hat\sigma(\lambda)$, where

$$k_n(x_0) = \sqrt{\frac{nx_0'(\mathbf{X}'\mathbf{X})^{-1}x_0}{n-p}}\,.$$

If the function h in (1) stands for the Box–Cox power transformation, namely

$$h(Y, \lambda) = \begin{cases} (Y^\lambda - 1)/\lambda & \text{if } \lambda \neq 0, \\ \log Y & \text{if } \lambda = 0, \end{cases}$$

then a simple inverse transformation of this interval yields a $100(1-\alpha)\%$ CI for δ_0

$$\left[1 + \lambda\left\{x_0'\hat\beta(\lambda) \pm t_{n-p}(\alpha/2)k_n(x_0)\hat\sigma(\lambda)\right\}\right]^{1/\lambda}, \tag{4}$$

which becomes $\exp\{x_0'\hat\beta(0) \pm t_{n-p}(\alpha/2)k_n(x_0)\hat\sigma(0)\}$ when $\lambda = 0$. This interval is correct when exact normality is achieved by the transformation, and it is asymptotically correct when only a certain degree of symmetry is achieved by the transformation, i.e., the mean and median of the transformed observations are the same. Thus, when λ is known, the interval can be recommended for practical applications and should be very attractive to practitioners due to its simplicity.

Unfortunately, λ is often unknown and has to be estimated using the same set of data. In this case, a common practice is to replace λ in (4) by its estimator $\hat{\lambda}$ (Collins 1991; Hahn & Meeker 1991, p. 72). Thus, the confidence interval for δ_0 becomes

$$\left[1+\hat{\lambda}\left\{x_0'\hat{\beta}(\hat{\lambda}) \pm t_{n-p}(\alpha/2)k_n(x_0)\hat{\sigma}(\hat{\lambda})\right\}\right]^{1/\hat{\lambda}}, \tag{5}$$

which reduces to

$$\exp\left\{x_0'\hat{\beta}(0) \pm t_{n-p}(\alpha/2)k_n(x_0)\hat{\sigma}(0)\right\}$$

when $\hat{\lambda}=0$. Interval (5), referred to as the *Box–Cox Confidence Interval*, is then used without accounting for the extra variability introduced by $\hat{\lambda}$.

3. THE ADJUSTED BOX–COX CONFIDENCE INTERVAL

Clearly, for the confidence interval (5) to have a good performance, it is necessary that the pivotal quantity $T(\hat{\lambda}) = \{x_0'\hat{\beta}(\hat{\lambda}) - h(\delta_0,\hat{\lambda})\}/\{\hat{\sigma}(\hat{\lambda})k_n(x_0)\}$, obtained by replacing λ in (3) by $\hat{\lambda}$, have a distribution close to that of $T(\lambda)$. Unfortunately, this is not true, as shown below; hence a correction is necessary. Let h_λ and $h_{\lambda\lambda}$ be, respectively, the first- and second-order partial derivatives of h with respect to λ. Let $\psi = (\beta', \sigma, \lambda)'$ and let $\tau^2(\psi)$ be the asymptotic variance of $\sqrt{n}(\hat{\lambda}-\lambda)$. We have the following result.

THEOREM 1. *Assume that for some λ, the following are true*: (i) $\mathrm{E}\{h(Y_0,\lambda)\} = h(\delta_0,\lambda)$; (ii) $\hat{\lambda} \to^p \lambda$ *and* $\sqrt{n}(\hat{\lambda}-\lambda)/\tau(\psi) \to^D N(0,1)$; (iii) $\mathbf{X}'h_\lambda(\mathbf{Y},\lambda)/n$ *and* $\mathbf{X}'h_{\lambda\lambda}(\mathbf{Y},\lambda)/n$ *both converge in probability*; (iv) $\mathbf{X}'\mathbf{X}/n$ *converges to a positive definite matrix*; (v) $h'(\mathbf{Y},\lambda)Qh_\lambda(\mathbf{Y},\lambda)/n$ *converges in probability. Then*

$$T(\hat{\lambda}) \xrightarrow{d} N\left[0, 1+c^2(\psi,x_0)\right],$$

where

$$c(\psi,x_0) = \lim_{n\to\infty} \frac{\tau(\psi)\left\{x_0'(X'X)^{-1}X'\mathrm{E}h_\lambda(\mathbf{Y},\lambda) - h_\lambda(\delta_0,\lambda)\right\}}{\sqrt{n}k_n(x_0)\sigma}.$$

Theorem 1 is proved in Appendix A. The constant $c(\psi,x_0)$ quantifies the effect of estimating the transformation on the Box–Cox predictive pivot $T(\lambda)$, and hence on the Box–Cox confidence interval. As $c^2(\psi,x_0)$ is a nonnegative number, $T(\hat{\lambda})$ has the same limiting mean as that of $T(\lambda)$, but it has a limiting variance larger than that of $T(\lambda)$. This indicates that the Box–Cox confidence interval without accounting for the estimation of λ is liberal. The value of $c^2(\psi,x_0)$ generally depends on the values of ψ and x_0; its general expression is given in the next theorem.

THEOREM 2. *Assume that the first six moments of e_i are the same as those of a standard normal random variable. Assume that $\max|\theta_i|$ is small. Then we have, for large n*:

$$c(\psi,x_0) \approx \frac{x_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\{(1_n+\lambda\eta)\#\phi + \lambda\sigma\theta/2\} - (1+\lambda\eta_0)\phi_0}{\lambda\sigma k_n(x_0)\sqrt{\| Q(\theta^{-1}\#\phi+\theta/2)\|^2 + 2\|\phi-\bar{\phi}\|^2 + 3\|\theta\|^2/2}}, \tag{6}$$

when $\lambda \neq 0$, and

$$c(\psi,x_0) \approx \frac{x_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\eta^2+\sigma^2 1_n) - \eta_0^2}{k_n(x_0)\sqrt{\| Q\eta^2\|^2 + 8\sigma^2\|\eta-\bar{\eta}\|^2 + 6n\sigma^4}}, \tag{7}$$

when $\lambda = 0$, *where* $\eta = \mathbf{X}'\beta$, $\phi = \log(1+\lambda\eta)$, $\theta = \lambda\sigma(1_n + \lambda\eta)^{-1}$, $\eta_0 = x_0'\beta$, $\phi_0 = \log(1+\lambda\eta_0)$, $\bar{\eta} = 1_n'\eta/n$, $\bar{\phi} = 1_n'\phi/n$, *and* $1_n = (1)_{n\times 1}$.

In Theorem 2, we use the following notational conventions: "#" denotes the elementwise multiplication operator of two vectors, i.e., $a\#b = (a_i \times b_i)_{n\times 1}$ for two $n \times 1$ vectors a and b; common functions applied to a vector, b say, are carried out elementwise, e.g., $\log b = (\log b_i)_{n\times 1}$; and $\|\cdot\|$ is the Euclidean norm. Theorem 2 is proved in Appendix B.

Now, let $c^2(\hat{\psi}, x_0)$ be an estimator of $c^2(\psi, x_0)$. Based on the results of the above theorems, it is natural to define the *adjusted Box–Cox confidence interval* for δ_0 as

$$\left[1 + \hat{\lambda}\left\{x_0'\hat{\beta}(\hat{\lambda}) \pm t_{n-p}(\alpha/2)k_n(x_0)\hat{\sigma}(\hat{\lambda})\sqrt{1 + c^2\left(\hat{\psi}, x_0\right)}\right\}\right]^{1/\hat{\lambda}}, \tag{8}$$

which becomes

$$\exp\left\{x_0'\hat{\beta}(0) \pm t_{n-p}(\alpha/2)k_n(x_0)\hat{\sigma}(0)\sqrt{1 + c^2\left(\hat{\psi}, x_0\right)}\right\}$$

when $\hat{\lambda} = 0$. The confidence interval given in (8) is seen to be very simple. Once the $\hat{\lambda}$ value is obtained, it is very easy to calculate $c^2(\hat{\psi}, x_0)$ and the rest is just like calculating a usual regression confidence interval. The $\hat{\lambda}$ defined in (2) can be conveniently found by solving the following equation:

$$-n\frac{h'(\mathbf{Y},\lambda)Qh_\lambda(\mathbf{Y},\lambda)}{h'(\mathbf{Y},\lambda)Qh(\mathbf{Y},\lambda)} + \sum_{i=1}^{n}\log Y_i = 0.$$

A Fortran subroutine for doing so is available from the author. For more practical applications, a SAS/IML program is also available which calculates the quantity in the third equation of (2) over a grid values of λ, so that an approximate $\hat{\lambda}$ value can be given. This value can be further refined when necessary using the same program.

4. MONTE CARLO SIMULATIONS

The results given in the last section say that when the sample size is large, the adjusted Box–Cox confidence interval performs just as the confidence interval under the normal theory linear model. However, it is not clear how well the adjusted Box–Cox confidence interval performs when n is not large. We now study this issue using Monte Carlo simulation. Consider a simple linear transformation model:

$$h(Y_i, \lambda) = \beta_0 + \beta_1 X_{1i} + \sigma e_i, \quad i = 1, \ldots, n$$

where $\beta_0 = 5.0$, $\beta_1 = 0.5$, $X_{1i} = 0.5(i), i = 1, \ldots, n$ and the e_i are iid $N(0,1)$. For a given parameter configuration, $e_1, \ldots, e_n$ are first generated form $N(0,1)$ and then converted to $Y_1, \ldots, Y_n$ through the model relation. The quantities $\hat{\beta}$, $\hat{\lambda}$, $c(\hat{\psi}, x_0)$, $T(\hat{\lambda})$ and the adjusted confidence interval are calculated. This process is repeated 10,000 times. The average and the standard deviation of the 10,000 $\hat{\lambda}$ values give the Monte Carlo estimates of $\mathrm{E}(\hat{\lambda})$ and $\mathrm{sd}(\hat{\lambda})$. Similarly, Monte Carlo estimates of $\mathrm{E}\{T(\hat{\lambda})\}$ and $\mathrm{sd}\{T(\hat{\lambda})\}$ are obtained. The number of confidence intervals that cover δ_0 divided by 10,000 gives the simulated coverage probability of the adjusted Box–Cox confidence interval. Results for $n = 50$ are summarized in Table 1.

We see that the adjusted confidence interval has coverage very close to its nominal level. It is important to note that the correction is very much necessary, especially when x_0 is outside of the experimental region. For example, the value of c^2 can be as big as 8.2128 (the square of the 12th number in the fourth column of Table 1), which indicates that the confidence interval without correction can be very tight. The approximate asymptotic standard deviation formula $\sqrt{1+c^2(\psi,x_0)}$ for $T(\hat\lambda)$ is very accurate, as seen by comparing the true value with the simulated values. Changing the values of λ and σ does not have much effect on the coverage probabilities. This shows the good performance of this simple analytical correction factor.

TABLE 1. Simulation results. Monte Carlo estimates of the mean and standard deviation of $\hat\lambda$ (columns 5 and 6), and the mean and standard deviation of $T(\hat\lambda)$ (columns 7 and 8), and simulated coverage probabilities for the 95% adjusted Box–Cox confidence interval (last column).

(λ,σ)	x_{10}	$k_n(x_0)$	$c(\psi,x_0)$	$\hat{\mathrm{E}}(\hat\lambda)$	$\hat{\mathrm{sd}}(\hat\lambda)$	$\hat{\mathrm{E}}\left\{T(\hat\lambda)\right\}$	$\hat{\mathrm{sd}}\left\{T(\hat\lambda)\right\}$	95%CI
$(0.1, 1.0)$	0	0.2931	−0.8637	0.0989	0.0392	0.0899	1.4178	0.9434
	5	0.2118	−0.0093	0.0989	0.0393	0.0406	1.0456	0.9410
	15	0.1512	0.7290	0.0988	0.0391	−0.0159	1.3046	0.9411
	35	0.4679	−1.8385	0.0981	0.0390	−0.0524	2.3009	0.9403
$(0.0, 1.0)$	0	0.2931	−0.8200	0.0000	0.0185	0.0686	1.4005	0.9402
	5	0.2118	−0.0426	−0.0003	0.0187	0.0424	1.0425	0.9440
	15	0.1512	0.7612	0.0000	0.0185	0.0091	1.3125	0.9488
	35	0.4679	−2.1171	−0.0003	0.0183	−0.1597	2.5546	0.9389
$(0.0, 0.1)$	0	0.2931	−1.1410	0.0000	0.0024	0.0128	1.5733	0.9462
	5	0.2118	−0.1138	0.0000	0.0024	−0.0049	1.0538	0.9433
	15	0.1512	0.9418	0.0000	0.0024	−0.0222	1.4327	0.9452
	35	0.4679	−2.8658	0.0000	0.0024	−0.0310	3.1536	0.9439
$(0.5, 0.1)$	0	0.2931	−1.2737	0.4997	0.0156	0.0451	1.7001	0.9418
	5	0.2118	−0.0583	0.4998	0.0156	0.0133	1.0426	0.9449
	15	0.1512	0.8824	0.5000	0.0154	0.0075	1.3850	0.9455
	35	0.4679	−2.3368	0.5000	0.0157	−0.0148	2.6770	0.9407
$(2.0, 0.1)$	0	0.2931	−1.3076	1.9993	0.0550	0.0254	1.6928	0.9469
	5	0.2118	−0.0477	1.9991	0.0545	0.0055	1.0392	0.9456
	15	0.1512	0.8725	1.9991	0.0549	−0.0262	1.3760	0.9455
	35	0.4679	−2.2474	2.0002	0.0549	−0.0307	2.5915	0.9416
$(-1.0, 0.1)^*$	0	0.2931	1.2838	−0.9999	0.0287	0.0047	1.7272	0.9428
	5	0.2118	0.0515	−1.0007	0.0288	-0.0042	1.0397	0.9497
	15	0.1512	−0.8716	−0.9998	0.0290	0.0042	1.3768	0.9480
	35	0.4679	2.2631	−1.0000	0.0288	0.0396	2.6005	0.9475

*$\beta_0 = -5.0$ and $\beta_1 = -0.5$ are used for this case to ensure the positivity of the Y_i.

5. NUMERICAL EXAMPLES

Two real data examples in this section illustrate the adjusted Box–Cox confidence intervals. The original data are omitted for brevity.

Example 1 (Biological Data). A 3×4 factorial design with linear additive effects and four replicates is fitted to the biological data of Box & Cox (1964). The responses are survival times (in 10 hr) of animals; the two factors are poison and treatment. Confidence intervals for the median survival time at each of the 12 design points are calculated; results are summarized in Table 2. Notice that $k_n(x_0)$ is constant, due to the symmetric structure of the design. In this special factorial design case, the estimated correction is not large, but it cannot be ignored in many cases. It is interesting to note that the adjusted Box–Cox confidence intervals are often shorter than the ones obtained from fitting a linear model to the original data, i.e., $\lambda = 1$ and $c(\hat{\psi}, x_0) = 0$. Compared with the original data, it seems that the transformation approach with correction provides more reasonable confidence intervals.

TABLE 2. A summary of confidence intervals for the biological data, $\hat{\lambda} = -0.7502$.

x_0'						$k_n(x_0)$	$c(\hat{\psi}, x_0)$	90% CI		95% CI		99% CI	
1	1	0	1	0	0	.3780	−.3310	0.34	0.43	0.33	0.44	0.32	0.47
1	1	0	0	1	0	.3780	.2018	0.74	1.15	0.71	1.22	0.66	1.37
1	1	0	0	0	1	.3780	−.1827	0.42	0.55	0.41	0.57	0.39	0.61
1	1	0	−1	−1	−1	.3780	.0705	0.61	0.88	0.59	0.92	0.55	1.00
1	0	1	1	0	0	.3780	−.1283	0.29	0.36	0.29	0.36	0.28	0.38
1	0	1	0	1	0	.3780	.0215	0.57	0.80	0.55	0.83	0.52	0.91
1	0	1	0	0	1	.3780	−.0826	0.35	0.44	0.34	0.45	0.33	0.48
1	0	1	−1	−1	−1	.3780	−.0113	0.48	0.65	0.47	0.67	0.44	0.72
1	−1	−1	1	0	0	.3780	.3529	0.19	0.23	0.19	0.23	0.19	0.24
1	−1	−1	0	1	0	.3780	−.4073	0.31	0.38	0.30	0.39	0.29	0.41
1	−1	−1	0	0	1	.3780	.1420	0.22	0.26	0.22	0.27	0.21	0.28
1	−1	−1	−1	−1	−1	.3780	−.2190	0.28	0.34	0.27	0.34	0.26	0.36

Example 2 (Salinity Data). In forecasting the shrimp harvest in Pamlico Sound, North Carolina, USA, & Carroll (1980) give a set of 28 observations on the salinity of water during the spring. The three predictors are the salinity lagged two weeks, the trend dummy variable for the time period, and the water flow or river discharge. These data were extensively analyzed by Atkinson (1985) to illustrate techniques of model checking, data transformation, etc. As in Atkinson's paper, the water flow value for observation 16 is modified and the third observation is deleted. The resulting confidence intervals are summarized in Table 3. Notice that the correction can be quite large if x_0 is outside the experimental region. The confidence intervals without transformation and correction can have negative lower limits, which is clearly unreasonable.

TABLE 3. A summary of confidence intervals for the salinity data, $\hat{\lambda} = -0.1489$.

x_{01}	x_{02}	x_{03}	$k_n(x_0)$	$c(\hat{\psi}, x_0)$	90% CI		95% CI		99% CI	
2.0	1.0	20.0	1.0233	0.0947	6.52	8.64	6.34	8.90	5.97	9.48
6.0	2.0	32.0	1.0094	0.7075	3.61	4.92	3.50	5.09	3.28	5.45
8.0	5.0	20.0	0.5076	−0.1401	9.59	11.11	9.45	11.29	9.16	11.66
10.0	1.0	25.0	0.3060	−0.6406	8.63	9.57	8.54	9.67	8.36	9.89
16.0	6.0	18.0	0.6949	1.5654	17.53	26.53	16.82	27.73	15.45	30.44

6. DISCUSSION

The Box–Cox transformation is a popular technique in analyzing the skewed data commonly occurring in biological, medical and environmental sciences, as well as in economics and engineering reliability. When the transformation is known, standard normal inference theories can be applied and simple inverse transformations lead to inferences corresponding to the original response. However, the transformation parameter is often unknown and has to be estimated from the data. A common practice in this case is to use the so-called “plug-in” method, i.e., plugging $\hat{\lambda}$ into the λ-assumed-known intervals. Such a practice ignores the effect of estimating the transformation, and a correction seems to be necessary in most parametric inference problems.

A simple analytical correction on the “plug-in” method, if it exists, extends the standard theory to more complicated modelling situations and at the same time preserves the simplicity of the standard normal inference theories. This was the original driving force behind the present work. The results given in this paper shed light in this direction. The problems that may follow include the confidence limits for regression quantiles, tolerance intervals for a future observation, bounds on reliability function, etc. The results given in this paper are consistent with those of Carroll & Ruppert (1981), where it is shown that there is a cost associated with the estimation of the transformation when estimating the median of the original future observation. Luckily, this cost is generally not severe.

The results of this paper may be modified to suit other transformation functions; see recent work of Yeo & Johnson (2000). The transformability of a data set to normality can be tested using a recent result of Chen, Lockhart & Stephens (2002).

APPENDIX A: PROOF OF THEOREM 1

Second-order Taylor expansions of $\hat{\beta}(\hat{\lambda})$ and $h(\delta_0, \hat{\lambda})$ give

$$\begin{aligned} x_0'\hat{\beta}(\hat{\lambda}) &= x_0'\hat{\beta}(\lambda) + (\hat{\lambda} - \lambda)\left\{x_0'(X'X)^{-1}X'h_\lambda(\mathbf{Y}, \lambda)\right\} \\ &\quad + \frac{1}{2}(\hat{\lambda} - \lambda)^2\left\{x_0'(X'X)^{-1}X'h_{\lambda\lambda}(\mathbf{Y}, \lambda) + R_n\right\}, \\ h(\delta_0, \hat{\lambda}) &= h(\delta_0, \lambda) + (\hat{\lambda} - \lambda)h_\lambda(\delta_0, \lambda) + \frac{1}{2}(\hat{\lambda} - \lambda)^2\left\{h_{\lambda\lambda}(\delta_0, \lambda) + R_n'\right\}, \end{aligned}$$

where R_n and R_n' converge to zero as $\hat{\lambda} \to \lambda$. Taking the difference yields

$$x_0'\hat{\beta}(\hat{\lambda}) - h(\delta_0, \hat{\lambda}) = x_0'\hat{\beta}(\lambda) - h(\delta_0, \lambda) + (\hat{\lambda} - \lambda)\left\{x_0'(X'X)^{-1}X'h_\lambda(\mathbf{Y}, \lambda) - h_\lambda(\delta_0, \lambda)\right\} + O_p(n^{-1}).$$

It is easy to see that $\hat{\sigma}^{-1}(\hat{\lambda}) = \hat{\sigma}^{-1}(\lambda) + O_p(n^{-1/2}) = \sigma^{-1} + O_p(n^{-1/2})$. This and the above result lead to

$$T(\hat{\lambda}) = T(\lambda) + \frac{(\hat{\lambda} - \lambda)\left\{x_0'(X'X)^{-1}X'\mathrm{E}h_\lambda(\mathbf{Y}, \lambda) - h_\lambda(\delta_0, \lambda)\right\}}{\sigma k_n(x_0)} + O_p(n^{-1/2}).$$

The result of Theorem 1 follows since $T(\lambda)$ and $\hat{\lambda}$ are asymptotically independent, which can be easily shown using standard asymptotic theory.

APPENDIX B: PROOF OF THEOREM 2

Yang (1999) obtained an accurate approximation to $\tau^2(\psi)$, namely

$$\begin{aligned} \tau^2(\psi) &\approx \frac{n\lambda^2}{\| Q(\theta^{-1}\#\phi + \theta/2) \|^2 + 2 \| \phi - \bar{\phi} \|^2 + 3 \| \theta \|^2 /2}, && \text{if } \lambda \neq 0, \\ &\approx \frac{4n\sigma^2}{\| Q\eta^2 \|^2 + 8\sigma^2 \| \eta - \bar{\eta} \|^2 + 6n\sigma^2}, && \text{if } \lambda = 0. \end{aligned}$$

For the power transformation, we have $h_\lambda(y, \lambda) = \{y^\lambda \log y - h(y, \lambda)\}/\lambda$ for $\lambda \neq 0$, and $0.5 \log^2(y)$ for $\lambda = 0$. The latter leads directly to the result (7) of Theorem 2 for the case of $\lambda = 0$. When $\lambda \neq 0$, using the following approximation

$$\lambda \log(Y_i) \approx \phi_i + \theta_i e_i - \theta_i^2 e_i^2/2,$$

one obtains approximations to $\mathrm{E}h_\lambda(Y_i, \lambda)$ and $h_\lambda(\delta_0, \lambda)$, which lead to the result (6) of Theorem 2 after some algebra.

ACKNOWLEDGEMENTS

This research was supported in part by Grant RP3982709 from the National University of Singapore. The author is grateful to the Editor, Professor Richard A. Lockhart, and the two referees for helpful comments.

REFERENCES

A. C. Atkinson (1985). *Plots, Transformation and Regression.* Oxford University Press.

G. E. P. Box & D. R. Cox (1964). An analysis of transformation (with discussion). *Journal of the Royal Statistical Society Series B*, 26, 211–252.

R. J. Carroll & D. Ruppert (1981). On prediction and power transformation family. *Biometrika*, 68, 609–615.

G. Chen, R. A. Lockhart & M. A. Stephens (2002). Box–Cox transformations in linear models: large sample theory and tests of normality (with discussion). *The Canadian Journal of Statistics*, 30, 177–234.

S. Collins (1991). Prediction techniques for Box–Cox regression models. *Journal of Business and Economic Statistics*, 9, 267–277.

G. J. Hahn & W. Q. Meeker (1991). *Statistical Intervals, A Guide for Practitioners.* Wiley, New York.

D. Ruppert & R. J. Carroll (1980). Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association*, 75, 828–838.

Z. Yang (1999). Estimating a transformation and its effect on Box–Cox *T*-ratio. *Test*, 8, 167–190.

I. K. Yeo & R. A. Johnson (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, 87, 954–959.

Received 11 January 2001
Accepted 19 February 2002

Zhenlin YANG: zlyang@smu.edu.sg
School of Economics and Social Sciences
Singapore Management University
469 Bukit Timah Road, Singapore 259756