5. AN EXAMPLE

The salary survey data of Chatterjee and Hadi (2006, p122), reproduced here in Table 1 for convenience, is used to illustrate our methods, in particular the CIs for percentile and survivor functions. The response variable is `Salary` and the predictors are years of experience (`Exp`), education (`Edu`) (1=high school diploma, 2=bachelor degree, 3=advanced degree), and management responsibility (`Man`) taking value 1 if a person bears management responsibility and 0 otherwise. The MLEs of the transformation parameter from the two transformations are, respectively, 0.183606 (Box-Cox) and 0.190988. The MLEs of the intercept, the coefficient of `Exp`, and the coefficients of three dummy variables (high school diploma, bachelor degree and `Man`) are (24.8645, 0.1913, -0.9647, 0.0367, 2.3575) from the Box-Cox power transformation, and (15.1719, 0.1053, -0.5308, 0.0202, 1.2974) from the dual power transformation. The MLEs of the error standard deviation are 0.3052 and 0.1679, respectively, from the Box-Cox power transformation and the dual power transformation.

Table 1: Salary Survey Data

Row	Salary	Exp	Edu	Man	Row	Salary	Exp	Edu	Man
1	13876	1	1	1	24	22884	6	2	1
2	11608	1	3	0	25	16978	7	1	1
3	18701	1	3	1	26	14803	8	2	0
4	11283	1	2	0	27	17404	8	1	1
5	11767	1	3	0	28	22184	8	3	1
6	20872	2	2	1	29	13548	8	1	0
7	11772	2	2	0	30	14467	10	1	0
8	10535	2	1	0	31	15942	10	2	0
9	12195	2	3	0	32	23174	10	3	1
10	12313	3	2	0	33	23780	10	2	1
11	14975	3	1	1	34	25410	11	2	1
12	21371	3	2	1	35	14861	11	1	0
13	19800	3	3	1	36	16882	12	2	0
14	11417	4	1	0	37	24170	12	3	1
15	20263	4	3	1	38	15990	13	1	0
16	13231	4	3	0	39	26330	13	2	1
17	12884	4	2	0	40	17949	14	2	0
18	13245	5	2	0	41	25685	15	3	1
19	13677	5	3	0	42	27837	16	2	1
20	15965	5	1	1	43	18838	16	2	0
21	12336	6	1	0	44	17483	16	1	0
22	21352	6	3	1	45	19207	17	2	0
23	13839	6	2	0	46	19346	20	1	0

Table 2 summarizes the CIs for percentile functions with $p = 0.05, 0.25, 0.5, 0.75$ and 0.95, and the CIs for survivor functions at y_0 chosen such that the values for the survivor function are estimated to be 0.95, 0.75, 0.5, 0.25, and 0.05, respectively. We choose $x_0' = (1, 10, 0, 0, 1)$, i.e. `Exp`=10, `Edu` = 3, and `Man`=1. The results show that the CIs based on Method 1 and the delta method are very similar. The CIs based on Method 2 are longest. These results are consistent with the Monte Carlo results – CIs based on Method 1 and the delta method often undercover the true quantity and as a result they are shorter. The results also show that the two transformations produce very similar sets of confidence intervals. Furthermore, a drawback of Method 1 and the delta method is clearly reflected in the CIs for the survivor function: the CIs based on these two methods can have an upper bound larger than 1 when y_0 is small, and a negative lower bound when y_0 is large. These problems do not occur with the corresponding CIs based on Method 2.

Table 2: CIs for percentile and survivor functions based on salary survey data

p or y_0	New Method 1		New Method 2		Delta Method	
Percentile Function, Box-Cox Power Transformation						
0.05	20981	22540	20705	22417	20970	22529
0.25	22081	23535	21929	23516	22072	23526
0.50	22819	24304	22755	24372	22810	24295
0.75	23524	25149	23548	25317	23513	25138
0.95	24484	26499	24634	26834	24468	26482
Percentile Function, Dual Power Transformation						
0.05	20981	22541	20705	22417	20969	22529
0.25	22081	23536	21929	23517	22072	23526
0.50	22819	24305	22755	24373	22810	24296
0.75	23525	25151	23549	25318	23514	25139
0.95	24486	26500	24636	26836	24469	26484
Survivor Function, Box-Cox Power Transformation						
21749	0.8742	1.0258	0.8186	0.9913	0.8742	1.0258
22799	0.5402	0.9595	0.5058	0.9088	0.5402	0.9595
23552	0.2383	0.7617	0.2559	0.7441	0.2383	0.7617
24325	0.0280	0.4723	0.0849	0.5099	0.0280	0.4723
25475	-0.0361	0.1360	0.0066	0.2088	-0.0361	0.1360
Survivor Function, Dual Power Transformation						
21749	0.8742	1.0258	0.8186	0.9913	0.8742	1.0258
22799	0.5402	0.9595	0.5058	0.9088	0.5402	0.9595
23553	0.2383	0.7617	0.2559	0.7441	0.2383	0.7617
24326	0.0280	0.4723	0.0849	0.5099	0.0280	0.4723
25477	-0.0360	0.1360	0.0066	0.2087	-0.0360	0.1360

6. CONCLUSIONS

We have provided general theories for conducting inferences, in a simple way, for the Box-Cox-type transformation model. More importantly, in the situations where the distribution of the λ-known statistic is completely known, we have introduced alternative procedures that lead to inference methods with better finite sample performance. Another important point as illustrated by the result (15) is that when the parametric function of interest is of limited range, a proper transformation of it may lead to a much better inference. Robustness of the methods are investigated by Monte Carlo simulation and the results show that our methods are quite robust against mild departures from normality of error distributions. No doubt, **simplicity**, **good finite sample performance**, and **robustness** are three important criteria for the applicability of a statistical inferential procedure. Our general results shed light in this direction and they may be extendable to the more complicated Box-Cox type of models.

APPENDIX

The detailed proofs of the theorems require the following set up and calculations. Let $\theta' = (\lambda, \beta', \sigma^2)$. Denote the partial derivatives of $h(Y_i, \lambda)$ with respect to Y_i and λ by adding relevant subscripts on h, i.e. $h_{y\lambda}(Y_i, \lambda) = \partial^2 h(Y_i, \lambda)/\partial Y_i \partial\lambda$. The score function $U(\theta) = \partial\ell(\theta)/\partial\theta$, partitioned according to λ, β, and σ^2, has elements,

$$
\begin{aligned}
U_1(\theta) &= -\frac{1}{\sigma^2}[h(\mathbf{Y},\lambda) - \mathbf{X}\beta]' h_\lambda(\mathbf{Y},\lambda) + \sum_{i=1}^{n} \frac{h_{y\lambda}(Y_i,\lambda)}{h_y(Y_i,\lambda)}, \\
U_2(\theta) &= \frac{1}{\sigma^2}\mathbf{X}'[h(\mathbf{Y},\lambda) - \mathbf{X}\beta], \\
U_3(\theta) &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}[h(\mathbf{Y},\lambda) - \mathbf{X}\beta]'[h(\mathbf{Y},\lambda) - \mathbf{X}\beta].
\end{aligned}
$$

The Hessian matrix $H = \partial\ell^2(\theta)/\partial\theta\partial\theta'$ has elements according to λ, β and σ^2,

$$
\begin{aligned}
H_{11} &= -\frac{1}{\sigma^2}\{[h(\mathbf{Y},\lambda) - \mathbf{X}\beta]' h_{\lambda\lambda}(\mathbf{Y},\lambda) + h_\lambda(\mathbf{Y},\lambda)' h_\lambda(\mathbf{Y},\lambda)\} \\
&\quad + \sum_{i=1}^{n} \frac{h_{y\lambda\lambda}(Y_i,\lambda) h_y(Y_i,\lambda) - h_{y\lambda}^2(Y_i,\lambda)}{h_y^2(Y_i,\lambda)}, \\
H_{22} &= -\frac{1}{\sigma^2}\mathbf{X}'\mathbf{X}, \\
H_{33} &= \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}[h(\mathbf{Y},\lambda) - \mathbf{X}\beta]'[h(\mathbf{Y},\lambda) - \mathbf{X}\beta], \\
H_{12} &= \frac{1}{\sigma^2} h_\lambda(\mathbf{Y},\lambda)'\mathbf{X}, \\
H_{13} &= \frac{1}{\sigma^4}[h(\mathbf{Y},\lambda) - \mathbf{X}\beta]' h_\lambda(\mathbf{Y},\lambda), \\
H_{23} &= -\frac{1}{\sigma^4}\mathbf{X}'[h(\mathbf{Y},\lambda) - \mathbf{X}\beta].
\end{aligned}
$$

The expected information matrix ($-\mathrm{E}[H(\theta)]$), denoted by $I(\theta)$ and partitioned as $I_{ij}, i, j = 1, 2, 3$, does not have explicit expressions for the elements involving subscript 1. For the other elements, we have $I_{22} = \sigma^{-2}\mathbf{X}'\mathbf{X}$, $I_{33} = \frac{n}{2\sigma^4}$, and $I_{23} = 0$. These expressions do not depend on which transformation function h the model employs. All the partial derivatives used in the theorems involving the Box-Cox power transformation, the dual power transformation, and the other transformations have analytical expressions and can all be derived easily.

Proof of Theorem 1. First, under regularity conditions for the ML estimation, we have

$$
\sqrt{n}\begin{pmatrix} \hat\beta(\lambda) - \beta \\ \hat\sigma^2(\lambda) - \sigma^2 \end{pmatrix} \xrightarrow{D} N\left(0, \begin{pmatrix} n\sigma^2(\mathbf{X}'\mathbf{X})^{-1}, & 0 \\ 0, & 2\sigma^4 \end{pmatrix}\right).
$$

This leads immediately to $\sqrt{n}[\hat\psi(\lambda) - \psi] \xrightarrow{D} N(0, v^2)$. Next, the condition ii) and the first-order Taylor series expansion give

$$
\begin{aligned}
\sqrt{n}[\hat\psi(\hat\lambda) - \psi] &= \sqrt{n}[\hat\psi(\lambda) - \psi] + \hat\psi_\lambda(\lambda)\sqrt{n}(\hat\lambda - \lambda) + o_p(1) \\
&= \sqrt{n}[\hat\psi(\lambda) - \psi] + \kappa\sqrt{n}(\hat\lambda - \lambda) + o_p(1).
\end{aligned}
$$

Thus, the second result of the theorem follows by showing that the asymptotic covariance of $\sqrt{n}(\hat\psi(\lambda) - \psi)$ and $\sqrt{n}(\hat\lambda - \lambda)$ is zero, which can be done by i) expressing $\hat\psi$ linearly in $\hat\beta(\lambda)$ and $\hat\sigma(\lambda)$; ii) expressing $\hat\beta(\lambda)$ and $\hat\sigma(\lambda)$ asymptotically in terms of U_2, U_3 and $I_{ij}, i, j = 2, 3$; and iii) expressing $\hat\lambda$ in terms of all elements of the score vector and the information matrix. Details of these calculations are tedious but available from the authors upon request.

Proof of Theorem 2. The first part is obvious. For the second part, condition iv) and a Taylor series expansion leads to

$$\begin{aligned}\sqrt{n}[c'\hat{\xi}(\hat{\lambda}) - f(\hat{\lambda}, \psi)] &= \sqrt{n}[c'\hat{\xi}(\lambda) - f(\lambda, \psi)] + [c'\hat{\xi}_\lambda(\lambda) - f_\lambda(\lambda, \psi)]\sqrt{n}(\hat{\lambda} - \lambda) + o_p(1) \\ &= \sqrt{n}[c'\hat{\xi}(\lambda) - f(\lambda, \psi)] + \kappa\sqrt{n}(\hat{\lambda} - \lambda) + o_p(1).\end{aligned}$$

The second result of the theorem follows from similar calculations as in the proof of Theorem 1, considering the fact that $c'\hat{\xi}(\lambda)$ is a special case of $\hat{\psi}(\lambda)$.

The Delta Method. The delta method is simply a consequence of the following well-known result of ML estimation: if the MLE $\hat{\theta}$ is asymptotically normal with mean θ and variance-covariance matrix $I^{-1}(\theta)$, then a smooth function of $\hat{\theta}$, $g(\hat{\theta})$ say, is the MLE of $g(\theta)$ and is asymptotically normal with mean $g(\theta)$ and variance $g'_\theta(\theta)I^{-1}(\theta)g_\theta(\theta)$. The variance can be consistently estimated by $g'_\theta(\hat{\theta})J^{-1}(\hat{\theta})g_\theta(\hat{\theta})$, where $J = -H$ is the observed information matrix. Based on this result, one quickly obtains the statistic for making inference about ψ

$$T(\hat{\lambda}, \psi) = \frac{\hat{\psi}(\hat{\lambda}) - \psi}{\sqrt{g'_\theta(\hat{\theta})J^{-1}(\hat{\theta})g_\theta(\hat{\theta})}}$$

with an asymptotic $100(1-\alpha)\%$ CI for ψ as

$$\hat{\psi}(\hat{\lambda}) \pm Z_{\alpha/2}\sqrt{g'_\theta(\hat{\theta})J^{-1}(\hat{\theta})g_\theta(\hat{\theta})}.$$

ACKNOWLEDGEMENTS

The authors are grateful to the Editor, Douglas Wiens, an Associate Editor and a referee for their constructive comments which have led to significant improvements in the paper. Z. L. Yang is grateful to the Department of Statistics, The Chinese University of Hong Kong and the Department of Mathematics and Statistics, University of Guelph for their hospitality during his visits. He gratefully acknowledges the research support from the Singapore Management University, and the excellent research assistance from Yihui Li. A. F. Desmond acknowledges with gratitude support via a Discovery Grant from the National Sciences and Engineering Research Council of Canada.

REFERENCES

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

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

G. E. P. Box & D. R. Cox (1982). An analysis of transformations revisited, rebutted. *Journal of the American Statistical Association*, 77, 209–210.

J. B. Burbidge, L. Magee & A. L. Robb (1988). Alternative transformations to handle extreme values of the dependent variable. *Journal of the American Statistical Association*, 83, 123–127.

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.

R. J. Carroll (1982). Tests for regression parameters in power transformation models. *Scandinavian Journal of Statistics*, 9, 217–222.

R. J. Carroll & D. Ruppert (1984). Power transformations when fitting theoretical models to data. *Journal of the American Statistical Association*, 79, 321–328

R. J. Carroll & D. Ruppert (1988). *Transformation and Weighting in Regression.* New York: Chapman & Hall

S. Chatterjee & A. S. Hadi (2006). *Regression Analysis by Example.* Hoboken, New Jersey: John Wiley & Sons. Inc.

A. M. Foster, L. Tian & 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.

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

F. Hernandez & R. A. Johnson (1980). The large-sample behavior of transformations to normality. *Journal of the American Statistical Association*, 75, 855–861.

D. V. Hinkley & G. Runger (1984). The analysis of transformed data. *Journal of the American Statistical Association*, 79, 302–320.

J. A. John & N. A. Draper (1980). An alternative family of transformations. *Applied Statistics*, 29, 190–197.

P. M. Hooper & Z. L. Yang (1997). Confidence intervals following Box-Cox transformation. *The Canadian Journal of Statistics*, 25, 401-416.

B. F. Manly (1976). Exponential data transformation. *The Statistician*, 25, 37–42.

J. G. MacKinnon & L. Magee (1990). Transforming the dependent variable in regression models. *International Economic Review*, 31, 315–339.

W. M. Patefield (1977). On the maximized likelihood function. *Sankhya Series B*, 39, 92–96.

H. White (1994). *Estimation, Inference and Specification Tests.* Cambridge: Cambridge University Press.

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

Z. L. Yang (2002a). Comment on " Box-Cox transformations in linear models: large sample theory and tests of normality ". *The Canadian Journal of Statistics*, 30, 222–225.

Z. L. Yang (2002b). Median estimation through a regression transformation. *The Canadian Journal of Statistics*, 30, 235–242.

Z. L. Yang (2006). A modified family of power transformations. *Economics Letters*, 92, 14–19.

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