7-2002

A Modified Family of Power Transformations

Zhenlin YANG
Singapore Management University, zlyang@smu.edu.sg

SMU ECONOMICS & STATISTICS
WORKING PAPER SERIES

A Modified Family of Power Transformations

Yang Zhenlin

July 2002

Paper No. 12-2002

A Modified Family of Power Transformations

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

July 23, 2002

Abstract

A modified family of power transformation, called the *Dual Power Transformation*, is proposed, which overcomes the truncation problem of the Box-Cox power transformation. The new transformation possesses properties similar to those of the Box-Cox power transformation. It generates a rich family of distributions that is seen to be very useful in modeling and analysis of economic durations and medical/engineering event-times. Further, it gives rise to transformed (regression) models such that all the standard asymptotic results of the maximum likelihood theory apply. Empirical results presented are more favorable to the new transformation than to the Box-Cox power transformation in terms of model fit.

Keywords: *Box-Cox power transformation; Dual power transformation; Transformed regression model; Trans-normal distribution.*

1 Introduction

Box and Cox (1964) proposed to transform the response variable to achieve a model with simple structure, normal errors and constant error variance. They used a power transformation:

$$h(y,\lambda) = \begin{cases} (y^\lambda - 1)/\lambda, & \lambda \neq 0, \\ \log y, & \lambda = 0, \end{cases} \quad y > 0. \tag{1}$$

to demonstrate their methodology. Soon after, this method became very popular and influential among the applied scientists and researchers, in particular among the economists. Noticeable applications of the Box-Cox method in economics include, among others, Zarembka (1968), White (1972), Leech (1975), Granger and Newbold (1976), Collins (1991), Higgins and Bera (1992), Buchinsky (1995), Hentschel (1995), Kim and Hill (1995), Machado and Mata (2000), Chen (2002), Yang and Abeysinghe (2002).

Besides all the successes of this method, there is a truncation problem associated with the use of the Box-Cox power transformation, i.e., $h(y, \lambda)$ is either bounded below or above at $-1/\lambda$ depending on whether λ is positive or negative. Hence, exact normality is incompatible with the distribution of the transformed variable unless $\lambda = 0$. To overcome this problem, many alternative transformations have been proposed, including the most recent one by Yeo and Johnson (2000). Most of the alternative transformations are constructed along the line of extending the domain of $h(y, \cdot)$ from half real line to the whole real line so that unbounded range of h can be achieved (e.g., Manly, 1976; John and Draper, 1980; and Bickel and Doksum, 1981). In most of the economic applications, however, the data are nonnegative. To normalize nonnegative data, it may still be best to use the Box-Cox power transformation although it is impossible to achieve exact normality when $\lambda \neq 0$. Practically, this truncation effect may be negligible as claimed by many researchers, but technically it causes difficulties in deriving the asymptotic properties of the estimated model, etc..

In this paper, we propose a modified family of power transformation, called the *dual power transformation*, that overcomes the shortcoming of the Box-Cox power transformation. It is shown that this new transformation has properties similar to those of the Box-Cox power transformation. This transformation also generates a well-defined family of distributions, called *trans-normal distribution*, that is shown to be very useful and flexible in modeling and analysis of economic durations and medical/engineering event-times. Once the truncation problem is removed, all the standard asymptotic results of the maximum likelihood theory apply. Further, empirical results given are more favorable to the dual power transformation than to the Box-Cox power transformation in terms of model fit.

Section 2 introduces the new transformation and its properties. Section 3 introduces the trans-normal distribution and presents some properties of this new distribution. Section 4 considers the common inference problems such as estimation, testing and confidence intervals in the framework of a transformed regression model. Section 5 presents some numerical examples and Section 6 concludes.

2 The Dual Power Transformation

We have noticed that $h(y, \lambda)$ is bounded below at $-1/\lambda$ when $\lambda > 0$, bounded above at $-1/\lambda$ when $\lambda < 0$, and unbounded when $\lambda = 0$. Removing the bound in the Box-Cox power transformation while at the same time preserving the nonnegativeness of y is the key motivation of the new transformation. For example, when $\lambda > 0$ in the Box-Cox power transformation, the h is bounded below at $1/\lambda$. If we replace $'1'$ in the numerator of h by $y^{-\lambda}$, then this bound is extended to $-\infty$, confirmable to the domain of a normal distribution. To make the limit of h when λ approaches zero the same as that of the Box-Cox power transformation, a $'2'$ is added to the denominator. The modified power transformation thus takes the form

$$h(y,\lambda) = \begin{cases} (y^\lambda - y^{-\lambda})/2\lambda, & \lambda > 0, \\ \log y, & \lambda = 0, \end{cases} \quad y > 0, \tag{2}$$

As this modified power transformation consists of two power functions, one with positive power and the other with negative power, we call this transformation the *Dual Power Transformation*. Unlike the Box-Cox power transformation that leaves the data untransformed when $\lambda = 1$, meaning there is no need to transform the data and the data itself is already normal, the dual power transformation always transform the data no matter what value λ takes. This may sound contradictory to our usual understanding of the Box-Cox transformation method, but as it has been claimed at beginning that y is nonnegative, this means that y has to be transformed to make it confirmable with a normal random variable, at least to have a desirable range. Thus, the new transformation makes more sense and is technically more sound. We now collect some properties of the new transformation in the following proposition.

Proposition 1. The dual power transformation defined in (2) satisfies the following:

(i) as a function of y, $h(y,\lambda)$ is increasing, concave when $|\lambda| \leq 1$, and concave and then convex as y increases when $|\lambda| > 1$, with the turning point $y_0 = [(\lambda+1)/(\lambda-1)]^{1/2\lambda}$;

(ii) as a function of λ, $h(y,\lambda)$ is symmetric around $\lambda = 0$, concave when $y \leq 1$ and convex when $y > 1$;

(iii) letting $z = h(y,\lambda)$, the inverse transformation is

$$y = g(z,\lambda) = \begin{cases} \left(\lambda z + \sqrt{1+\lambda^2 z^2}\right)^{1/\lambda}, & \lambda \neq 0, \\ \exp(z), & \lambda = 0, \end{cases} \tag{3}$$

(iv) $h(y,\lambda) = -h(y^{-1},\lambda)$.

Proof. The proofs of (i) and (ii) are based on the relevant partial derivatives of h which are given in Appendix. To prove (iii), solve $2\lambda z = y^\lambda - y^{-\lambda}$ for y^λ. There are two roots. One of them is obliviously inadmissible due to the fact that y is positive and the other is that given in (iii). Property (iv) is obvious.

Note that similar to the Box-Cox power transformation, the dual power transformation is also monotonic increasing, covers lognormal as a special case, and possesses partial derivatives of any order. As h is symmetric in λ around 0, it is sufficient to consider the positive values of λ. To give a visual comparison of the dual power transformation with the Box-Cox power transformation, we plot the two functions in Figure 1. The plots show that the smaller the λ value, the closer is the two power transformation with the limits (at

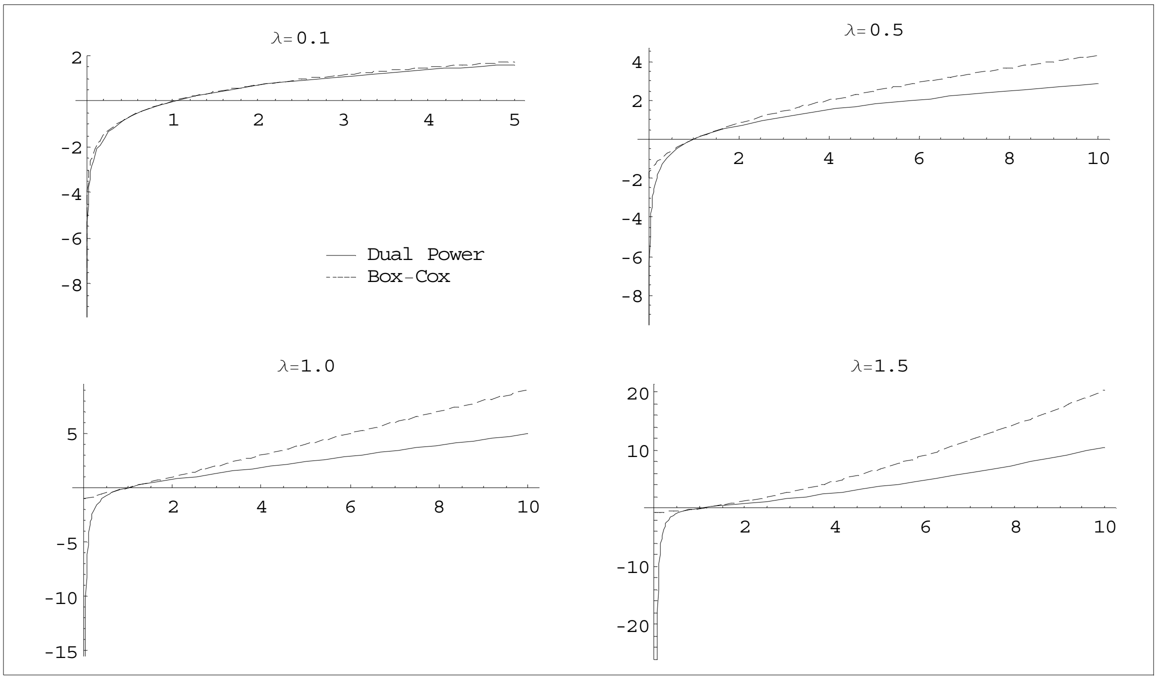

Figure 1: Plots of the dual power transformation and Box-Cox power transformation

$\lambda = 0$) being the identical log transformation. When λ moves away from 0, the difference between the two transformation becomes more and more substantial. When $\lambda > 0$, the main difference between the two transformations happens at the part where y takes values from 0 to 1: the Box-Cox power transformation maps $[0, 1]$ to $[-1/\lambda, 0]$, whereas the dual power transformation translates it to $(-\infty, 0]$. Both functions translate the $[1, \infty)$ into $[0, \infty)$ with the curve of the dual power transformation lying below that of the Box-Cox power transformation. When $\lambda < 0$, the Box-Cox transformation is able to translate the $[0, 1]$ part into $[-\infty, 0)$, but maps the $[1, \infty)$ part into $[0, -1/\lambda)$. The dual power transformation is symmetric in λ, and hence a negative λ gives the same function as a positive one.

The dual power transformation is related to the inverse hyperbolic sine (IHS) transformation of Johnson (1949), which was studied and compared with the Box-Cox power transformation by Burbidge, at al. (1988). If we let $y = exp(x)$ in the dual power transformation, then h becomes, as a function of x, a hyperbolic sine function. The IHS transformation, however, works with the inverse hyperbolic sine, thus has a domain of whole real line. The dual power transformation has a domain of positive half real line with an intention of modifiing the Box-Cox power transformation so that the truncation problem is removed.

3 The Trans-Normal Distribution

The dual power transformation generates a new family of distributions. In other words, if $h(Y, \lambda)$ follows $N(\mu, \sigma^2)$, then the probability density function (pdf) of Y is given as follows

$$f(y; \mu, \sigma, \lambda) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2\sigma^2}[h(y, \lambda) - \mu]^2 \right\} (y^{\lambda-1} + y^{-\lambda-1}), \tag{4}$$

where $y > 0$; $-\infty < \mu < \infty$, $\sigma > 0$, and $-\infty < \lambda < \infty$. Note that the distributional family generated by the dual power transformation generalizes the ξ-normal family of Saunders (1974). As this family of distributions is generated from a normalizing transformation, it is called the *Trans-normal Distribution.* We now give some general theoretical properties of the trans-normal family. Through out, we use the subscripted h function to denote the partial derivatives.

Proposition 2. The pdf $f(y; \mu, \sigma, \lambda)$ defined in (4) satisfies the following:

(i) it is a monotonic function of y if $m(y) = 0$ does not have a real root;

(ii) it is a unimodal pdf if $m(y) = 0$ has a unique real root in the interior of $[0, \infty)$;

(iii) it has two stationary points if $m(y) = 0$ has two real roots;

(iv) it is bimodal if $m(y) = 0$ has three real roots, etc., where

$$m(y) = \frac{(\lambda - 1)y^{\lambda} - (\lambda + 1)y^{-\lambda}}{(y^{\lambda} + y^{-\lambda})^2} - \frac{y^{\lambda} - y^{-\lambda} - 2\lambda\mu}{2\lambda\sigma^2}.$$

Proof. Let $k(y) = \exp\{-[h(y, \lambda) - \mu]^2/(2\sigma^2)\}$. Then, $f(y; \mu, \sigma, \lambda) \propto k(y)\ h_y(y, \lambda)$, and $\partial f(y; \mu, \sigma, \lambda)/\partial y = k(y)h_y^2(y)[h_{yy}(y, \lambda)/h_y^2(y, \lambda) - (h(y, \lambda) - \mu)/\sigma^2] = k(y)h_y^2(y, \lambda)m(y)$. Since the function $k(y)h_y^2(y, \lambda)$ is a positive function of y, how many times that $\partial f/\partial y$ changes its sign as y changes depends on how many real roots that $m(y) = 0$ has, which determines the behavior of f. The results of Proposition 2 thus follow.

Note that the case (i) in Proposition 2 rarely happens, case (ii) is the most typical case and it happens as long as f vanishes at both ends and $h_{yy}(y, \lambda)/h_y^2(y, \lambda)$ is monotonic in y. The cases (iii) is also not common and (iv) can happen for certain special functions at certain parameter settings. All the necessary partial derivatives of $h(y, \lambda)$ are given in Appendix.

To illustrate the versatility and usefulness of the trans-normal distribution, we pick a special modified power transformation corresponding to $\lambda = 0.5$, i.e., $h(y, \lambda) = y^{.5} - y^{-.5}$, and plot the pdf, the survivor function (sf) and the hazard function (hf) for serval parameter configurations. From the plots summarized in Figure 2, we see that the pdf of this trans-normal distribution has all kinds of shapes: it can be nearly symmetric, bimodal, or very skewed depending whether σ is small, medium, or large relative to the mean of

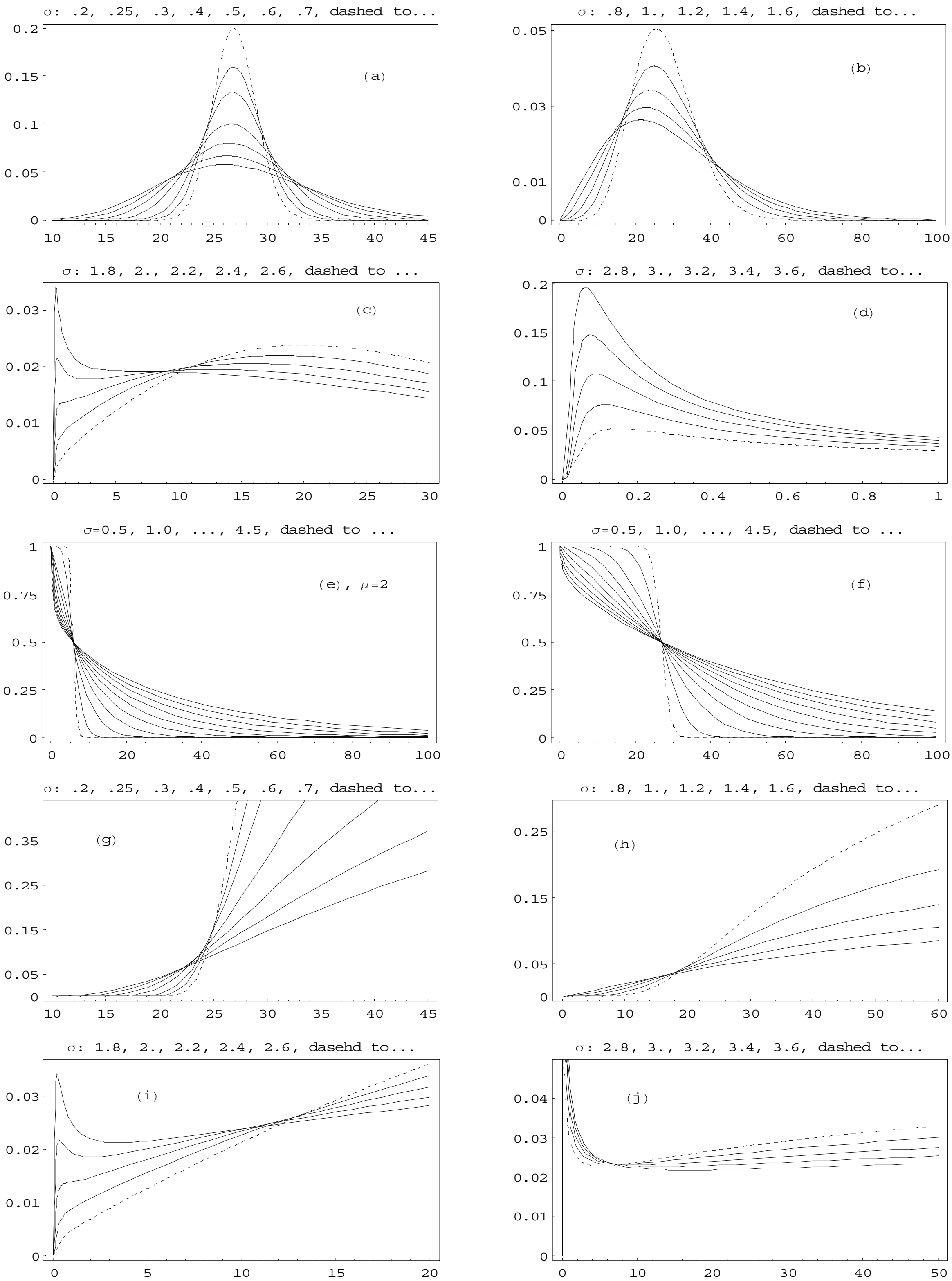

Figure 2: Plots of the trans-normal pdfs (a)-(d), sfs (e)-(f), and hfs (g)-(j), $\lambda = 0.5$ through out, and $\mu = 5$ except in Plot (e)

Y. When σ is small relative to the mean, the pdf has one bump at the center part; as σ increases, another bump shows up at the left of the center and as σ further increases, the first bump disappeared and the distribution becomes unimodal again. Figure 1 also exhibits serval shapes of hazard function, including the interesting 'bath-tub' shape, which has a popular engineering interpretation: first bump represents the 'burn-in' period, the center flat part represents the 'stable period' and the second bump represents the 'wear-out' period. Econometricians call this the U-shaped hazard (Kiefer, 1988) and some evidence for its existence is provided by Kennan (1985) from the analysis of the strike duration data. It is interesting to note that when σ is large, the hf has a sharp increase at the very beginning and then quickly becomes flat for a long period of 'time'. This exactly reflects the failure mechanisms of certain engineering systems and electronic components which are very fragile at the very beginning, but once stabilized, can last for a very long period of time.

4 The Transformed Regression Model

Let $\mathbf{Y}$ be an $n \times 1$ vector of original observations, $h(\mathbf{Y}, \lambda)$ be a vector of transformed observations, and $\mathbf{X}$ be an $n \times p$ matrix whose columns contain the values of the explanatory variables $X_1, X_2, \ldots, X_k$. The Box-Cox transformed linear model (Box and Cox, 1964) has the form

$$h(\mathbf{Y}, \lambda) = \mathbf{X}\beta + \sigma\mathbf{e} \tag{5}$$

where β is a $k \times 1$ vector of parameters, σ is the error standard deviation, $\mathbf{e}$ is an $n \times 1$ vector of independent and identically distributed (iid) normal errors, and $h(\cdot, \lambda)$ is the dual power transformation. The log likelihood function in terms of original variables is

$$\ell(\beta, \sigma^2, \lambda) \propto -\frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}[h(\mathbf{Y}, \lambda) - \mathbf{X}\beta]'[h(\mathbf{Y}, \lambda) - \mathbf{X}\beta] + \log J(\lambda), \tag{6}$$

where $J(\lambda) = |\prod_{i=1}^n \partial h(Y_i, \lambda)/\partial Y_i| = \prod_{i=1}^n (Y_i^\lambda + Y_i^{-\lambda})$.

Parameter Estimation and Asymptotic Properties. For a given λ, the model (5) is just an ordinary regression model and thus the maximum likelihood estimators (MLE) for β and σ are

$$\hat\beta(\lambda) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'h(\mathbf{Y}, \lambda) \text{ and } \hat\sigma^2(\lambda) = \frac{1}{n}\|\mathbf{M}h(\mathbf{Y}, \lambda)\|^2, \tag{7}$$

where $\|\cdot\|$ is the Euclidian norm, $\mathbf{M} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, and $\mathbf{I}_n$ is an $n \times n$ identity matrix. Substituting $\hat\beta(\lambda)$ and $\hat\sigma^2(\lambda)$ into (6) for β and σ^2 gives the profile likelihood of λ

$$\ell_p(\lambda) \propto -\frac{n}{2}\log\|\mathbf{M}h(\mathbf{Y}, \lambda)\|^2 + \log J(\lambda) \tag{8}$$

Maximizing $\ell_p(\lambda)$ gives the MLE $\hat{\lambda}$ of λ. Equivalently, $\hat{\lambda}$ can be obtained by solving the equation $S_p(\lambda) = 0$, where $S_p(\lambda) = d\ell_p(\lambda)/d\lambda$ is the profile score with the expression

$$S_p(\lambda) = -\frac{nh(\mathbf{Y},\lambda)'\mathbf{M}h_\lambda(\mathbf{Y},\lambda)}{h(\mathbf{Y},\lambda)'\mathbf{M}h(\mathbf{Y},\lambda)} + \sum_{i=1}^{n} \frac{(Y_i^\lambda - Y_i^{-\lambda})\log Y_i}{Y_i^\lambda + Y_i^{-\lambda}}. \tag{9}$$

Finally, substituting $\hat{\lambda}$ back into $\hat{\beta}(\lambda)$ and $\hat{\sigma}^2(\lambda)$ gives the unrestricted MLEs of β and σ^2.

It is easy to check that the model (5) satisfies the usual regularity conditions (see, for example, Rao, 1973) and hence the MLEs possess the standard asymptotic properties. In particular, $\hat{\psi}$ is consistent and $\hat{\psi}$ is asymptotic normal with mean ψ and variance-covariance matrix $I^{-1}(\psi)$, where $\psi = \{\beta', \sigma^2, \lambda\}'$ and $I(\psi)$ is the expected information matrix.

Tests of Transformation. Once the parameter estimates are obtained, a problem that is of primary interest is to test the transformation parameter to decide in which scale the subsequent inferences be carried out. In other words, we want to perform a test of the hypothesis $H_0 : \lambda = \lambda_0$, where λ_0 often represents convenient values such as $0, 1/2, 1/3$, etc, corresponding to the log, square-root, cubic-root transformation, etc. The traditional tests such as sign-square-root likelihood ratio test, score test (econometricians call it the Lagrange multipliers test) and Wald test can be easily implemented in this context. We now outline the three tests below.

The *sign-square-root likelihood ratio test* has the form

$$T_L = 2\,\mathrm{sign}(\hat{\lambda} - \lambda_0)[\ell_p(\hat{\lambda}) - \ell_p(\lambda_0)]^{1/2}$$

The *Wald test* is given by

$$T_W = \frac{\hat{\lambda} - \lambda_0}{se(\hat{\lambda})}$$

where $se(\hat{\lambda})$ is the estimated standard error of $\hat{\lambda}$ and can be obtained as the square root of the last diagonal element of $J^{-1}(\hat{\psi})$, where $J(\hat{\psi})$ is the observed information matrix evaluated at the full MLEs of the parameters (see Appendix for the full expression of $J(\psi)$).

The *score test* has the form

$$T_S = S_p(\lambda_0)se(\lambda_0)$$

where $se(\lambda_0)$ is the same quantity used in the Wald test but evaluated at the restricted MLEs $\hat{\beta}(\lambda_0)$ and $\hat{\sigma}(\lambda_0)$.

The three tests have the same limiting null distribution (standard normal), but their finite sample distributions are not known and are not necessarily the same. The score test is the easiest one to implement as it requires only the restrictive MLEs of β and σ, both of which have explicite expressions. Note that the score and Wald tests given here are based on the observed information.

In the context of the Box-Cox power transformation, Yang (1999b, 2000) constructed the score and Wald tests based on the expected information and have shown that, within the parameter regions such that the truncation effect is small, the two tests behave better than those based on the observed information. Similar work could be carried out for the dual power transformation, but apparently it is more difficult to approximate the expected information in this case.

Inferences Concerning Regression Parameters. There have been some debates on which parameter should the inference focus on after a transformation has been decided on the response (Bickel and Doksum, 1981, Box and Cox, 1982, Hinkley and Runger, 1984). Hooper and Yang (1997) and Yang (1999c) showed that if inference concern a regression parameter defined on the selected transformation scale, the usual inference methods remain asymptotically valid. Recently, Chen, et, al. (2002) argued that the inference should be carried out on the scaled regression coefficient β/σ, and showed that estimation on this parameter is much more stable with respect to the transformation estimation than the estimation of β itself. See also the comment by Yang (2002a).

All the theoretical results mentioned above (except Bickel and Doksum, 1981) are based on the Box-Cox power transformation. These results can be extended to the case of dual power transformation in a straightforward manner, hence the details will not be given here.

Prediction in Original Scales. One of the attractive feature of the transformation method is it allows the use of the standard normal-theory methods to give prediction or confidence interval concerning a future transformed observation, and then inverse-transform the end points of the intervals to give prediction and confidence intervals concerning the original observation. When the transformation parameter is unknown, it is then replaced by its estimate. This method is usually termed as the *plug-in* method.

It has been shown that the plug-in method works well for prediction intervals for a future response Y_0 at a future predictor value x_0 (Carroll and Ruppert, 1991; Yang, 1999a, and Chu at al., 2001), but needs a correction for confidence intervals of quantiles of Y_0 (Carroll and Ruppert, 1981, 1991; Yang 2001, 2002b). A result of Yang and Tse (2002) can be used to give an analytically adjusted confidence interval for the quantile of Y_0.

5 Numerical Examples

We now use two real data sets, the strike duration data of (Keenan, 1985) and the computer execution time data of Meeker and Escobar (1998), to demonstrate the application of the new transformation. Also, comparisons are made with the applications of the Box-Cox power transformation on the same data sets.

The Strike Duration Data. The data set (Kennan, 1985): 7 9 13 14 26 29 52 130 9 37 41 49 52 119 3 17 19 28 72 99 104 114 152 153 216 15 61 98 2 25 85 3 10 1 2 3 3 3 4 8 11 22 23 27 32 33 35 43 43 44 100 5 49 2 12 12 21 21 27 38 42 117, represent the durations in days of 62 strikes in the period from 1968 to 1976. It has been subsequently analyzed by many authors including Keifer (1988) and Greene (2000), using models such as exponential, lognormal, log logistic and Weibull. The data is positively skewed. We fit the trans-normal model with the dual power transformation and with the Box-Cox power transformation to the data. The MLE of (μ, σ^2, λ) is $(4.2876, 2.1009, 0.166253)$ if the Box-Cox power transformation is used, and $(3.7461, 1.8228, 0.284335)$ if the dual power transformation is usaed. The two models give quite different values of the transformation parameter.

Figure 3 gives histograms of the original data (Duration), the dual power transformed data (DPTrans), and the Box-Cox power transformed data (BCTrans). It also gives the probability plots of the two sets of transformed data, as well as plots of the fitted densities using the two transformations. The goodness of fit statistic (Anderson-Darling) is 0.499 when dual power transformation is used, and 0.531 when Box-Cox power transformation is used. The trans-normal model with the dual power transformation gives a better fit to the data as compared with the model with Box-Cox power transformation. Also, both models fit the data better than the models used in Greene (2000).

Computer Execution Time Data. The data given in Table 1 is taken from Meeker and Escobar (1998, p638). It represents the amount of time it took to execute a particular computer program, on a multiuser computer system, as a function of system load (obtained with the Unix `uptime` command) at the time when execution was beginning. The data was analyzed by Meeker and Escobar using a simple log-linear regression model.

Table 1: Computer Program Execution Time Versus Load

Seconds	load	Seconds	load	Seconds	load
123	2.74	78	0.51	317	5.86
704	5.47	98	0.29	142	1.18
184	2.13	240	0.96	127	0.57
113	1.00	110	0.60	96	1.10
94	0.32	213	2.10	111	1.89
76	0.31	284	3.10		

Fitting a transformed regression model with the dual power transformation results in an MLE of the transformation parameter being zero (significant up to 8th decimal point). In other words, the loglinear model Meeker and Escobar (1998) is warrant from the fitting of

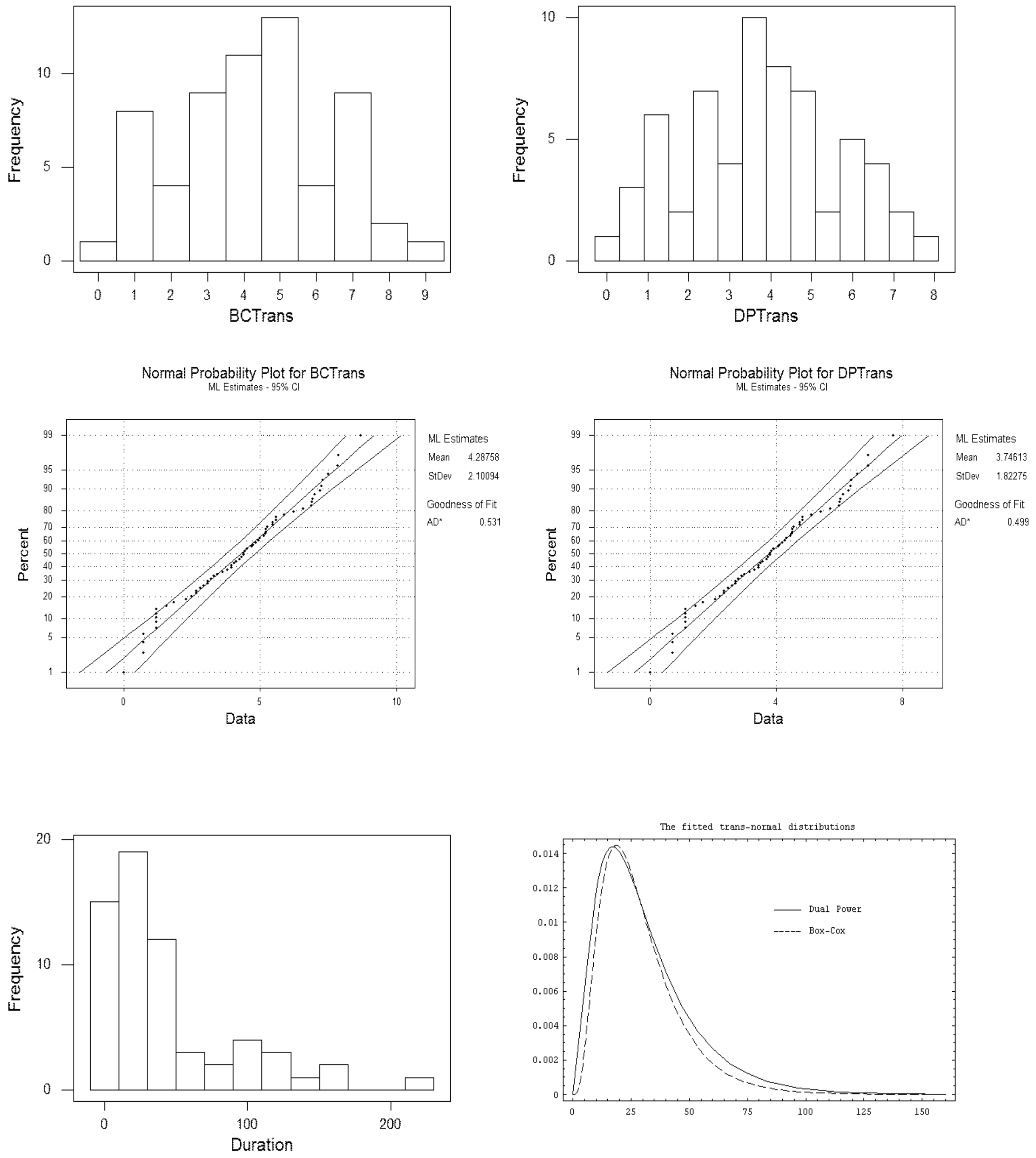

Figure 3: Histograms, Normal Probability Plots and Fitted Distributions for Strike Duration Data

the dual power transformation. The MLEs of the regression coefficient and the error standard deviation are, respectively, $4.4936, 0.2907$ and 0.3125. Fitting a transformed regression model with the Box-Cox power transformation results in a Box-Cox estimate of the transformation parameter being -0.434015, which is very much different from the estimate of the dual power transformation. The MLEs of β_0, β_1 and σ are, respectively, $1.9829, 0.0291$ and 0.0338. The multiple coefficient of regression (R^2) is 70.2% for the model with dual power transformation, and 67.0% for the model with the Box-Cox power transformation, showing that the model with dual power transformation gives a better fit to the data.

6 Conclusions

A new power transformation is proposed. It overcomes the truncation problem of the Box-Cox power transformation. It generates a rich family of distributions that can be applied to economics, engineering, medicine, and other fields to model the nonnegative data with a skewed distribution, with two modes, etc.. As the transformation is a monotonic smooth function that has a domain half real line and a range of whole real line, it turns out that the normality assumption is technically valid for the transformed observations. It follows that all the standard asymptotic results of the maximum likelihood theory apply. The empirical results favor the dual power transformation.

Appendix A: Derivatives, Score and Observed Information

Denote the partial derivatives of $h(y,\lambda)$ by adding the relevant subscripts on h. Then,

$$
\begin{aligned}
h_\lambda(y,\lambda) &= \frac{1}{2\lambda}(y^\lambda + y^{-\lambda})\log y - \frac{1}{\lambda}h(y,\lambda) \\
h_{\lambda\lambda}(y,\lambda) &= h(y,\lambda)(\frac{2}{\lambda^2} + \log^2 y) - \frac{1}{\lambda^2}(y^\lambda + y^{-\lambda})\log y \\
h_y(y,\lambda) &= \frac{1}{2}[y^{\lambda-1} + y^{-\lambda-1}], \\
h_{yy}(y,\lambda) &= \frac{1}{2}[(\lambda-1)y^{\lambda-2} - (\lambda+1)y^{-\lambda-2}], \\
h_{y\lambda}(y,\lambda) &= \frac{1}{2}(y^{\lambda-1} - y^{-\lambda-1})\log y \\
h_{y\lambda\lambda}(y,\lambda) &= \frac{1}{2}(y^{\lambda-1} + y^{-\lambda-1})\log^2 y
\end{aligned}
$$

The score functions are

$$
\begin{aligned}
S_\beta(\psi) &= \frac{1}{\sigma^2}\mathbf{X}'[h(\mathbf{Y},\lambda) - \mathbf{X}\beta] \\
S_{\sigma^2}(\psi) &= -\frac{n}{2\sigma^2} + \frac{1}{\sigma^4}[h(\mathbf{Y},\lambda) - \mathbf{X}\beta]'[h(\mathbf{Y},\lambda) - \mathbf{X}\beta] \\
S_\lambda(\psi) &= -\frac{1}{\sigma^2}[h(\mathbf{Y},\lambda) - \mathbf{X}\beta]'[h(\mathbf{Y},\lambda) - \mathbf{X}\beta] + \sum_{i=1}^n \frac{(Y_i^\lambda - Y_i^{-\lambda})\log Y_i}{Y_i^\lambda + Y_i^{-\lambda}}.
\end{aligned}
$$

The elements of the observed information matrix are

$$
\begin{aligned}
J_{\beta\beta} &= \frac{1}{\sigma^2}\mathbf{X}'\mathbf{X} \\
J_{\sigma^2\sigma^2} &= -\frac{n}{2\sigma^4} + \frac{1}{\sigma^6}[h(\mathbf{Y},\lambda) - \mathbf{X}\beta]'[h(\mathbf{Y},\lambda) - \mathbf{X}\beta] \\
J_{\lambda\lambda} &= \frac{1}{\sigma^2}[h'_\lambda(\mathbf{Y},\lambda)h_\lambda(\mathbf{Y},\lambda) + (h(\mathbf{Y},\lambda) - \mathbf{X}\beta)'h_{\lambda\lambda}(\mathbf{Y},\lambda)] \\
&\quad + \sum_{i=1}^n \frac{[Y_i^\lambda + Y_i^{-\lambda} - (Y_i^\lambda - Y_i^{-\lambda})^2]\log Y_i}{(Y_i^\lambda + Y_i^{-\lambda})^2} \\
J_{\beta\sigma^2} &= \frac{1}{\sigma^4}\mathbf{X}'[h(\mathbf{Y},\lambda) - \mathbf{X}\beta] \\
J_{\beta\lambda} &= -\frac{1}{\sigma^2}\mathbf{X}'h_\lambda(\mathbf{Y},\lambda) \\
J_{\sigma^2\lambda} &= -\frac{1}{\sigma^4}[h(\mathbf{Y},\lambda) - \mathbf{X}\beta]'h_\lambda(\mathbf{Y},\lambda)
\end{aligned}
$$

REFERENCES

1. Bickel, P. J. and Doksum, K. A. (1981). An analysis of transformations revisited. *Journal of the American Statistical Association.* **76**, 296-311.
2. Box, G.E.P. and Cox,D.R. (1964). An analysis of transformation (with discussion). *Journal of the Royal Statistical Society B.* **26**, 211-252.
3. Box, G.E.P. and Cox,D.R. (1982). An analysis of transformation revisited, rebutted. *Journal of the American Statistical Association.* **77**, 209-210.
4. Burbidge, J. B., Magee, L. and Robb, A. L. (1988). Alternative transformations to handle extreme values of the dependent variable. *Journal of the American Statistical Association.* **83**, 123-127.
5. Buchinsky, M. (1995). Quantile regression, Box-Cox transformation model, and the U.S. wage structure, 1963-1987. *Journal of Econometrics.* **65**, 109-154.
6. Carroll, R. J. and Ruppert, D. (1981). On prediction and power transformation family. *Biometrika.* **68**, 609-615.
7. Carroll, R. J. and Ruppert, D. (1991). Prediction and tolerance intervals with transformations and/or weighting. *Technometrics*, **33**, 197-210.
8. Collins, S. (1991). Prediction Techniques for Box-Cox Regression Models. *Journal of Business and Economic Statistics.* **9**, 267-277.
9. Chen, S. (2002). Rank Estimation of Transformation Models. *Econometrica.* To appear.
10. Chen, G., Lockhart, R. A. and Stephens, M. A. (2002). Box-Cox transformations in linear models: large sample theory and tests of normality. *Canadian Journal of Statistics.* Vol. 30, No. 2. To appear.
11. Chu, K., Yeo, I. K., Johnson, R. A. and Loh, W. Y. (2001). Prediction interval estimation in transformed linear models. *Statistics and Probability Letters.* **51**, 345-350.
12. Granger, C. W. J. and Newbold, P. (1976). Forecasting transformed time series. *Journal of the Royal Statistical Society B.* **38**, 189-203.
13. Greene, W. H. (2000). *Econometric Analysis, 4th ed.* Prentice Hall.
14. Hentschel, L. (1995). All in the family: Nesting symmetric and asymmetric GARCH models. *Journal of Financial Economics.* **39**. 71-104.
15. Higgins, M. L. and Bera, A. (1992). A class of nonlinear ARCH model. International Economic Review. **33**, 137-158.
16. Hinkley, D. V. and Runger, G. (1984). The analysis of transformed data. *Journal of the American Statistical Association.* **79**, 302-320.
17. Hooper and Yang (1997). Confidence intervals following Box-Cox transformation. *Canadian Journal of Statistics.* **25**, 401-416.
18. John, J. A. and Draper, N. A. (1980). An alternative family of transformations. *Applied Statistics.* **29**, 190-197.

19. Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika.* **36**, 149-176.
20. Kennan, J. (1985). The duration of contract strikes in U.S. manufacturing. *Journal of Econometrics.* **28**, 5-28
21. Kiefer, N. M. (1988). Economic duration data and hazard functions. *Journal of Economic Literature.* **26**, 646-679.
22. Kim, M. and Hill, R. C. (1995). Shrinkage estimation in nonlinear regression The Box-Cox transformation. *Journal of Econometrics.* **66**, 1-33.
23. Leech, D. (1975). Testing the error specification in nonlinear regression. *Econometrica.* **43**, 719-725.
24. Machado, J. A. F. and Mata, J. (2000). Box-Cox quantile regression and the distribution of firm sizes. *Journal of Applied Econometrics* **15**, 253-274.
25. Manly, B. F. (1976). Exponential data transformation. *The Statistician.* **25**, 37-42.
26. Meeker, W.Q. and Escobar, L. A. (1998). *Statistical Methods for Reliability Data.* John Wiley & Sons, Inc., New York.
27. Rao, C. R. (1973). *Linear Statistical Inference and its Applications, 2nd ed.* John Wiley & Sons, New York.
28. Saunders, S. C. (1974). A family of random variables closed under reciprocation. *Journal of the American Statistical Association* 69, 533-539.
29. White, K. J. (1972). Estimation of the liquidity trap with a generalized functional form. *Econometrica.* **40**, 193-199.
30. Yang, Z. (1999a). Predicting a future lifetime through Box-Cox transformation. *lifetime Data Analysis.* **5**, 265-279.
31. Yang, Z. (1999b). An alternative approximation to the variance of transformation Score. *Journal of Statistical Computation and Simulation.* **62**, 181-188.
32. Yang, Z. (1999c). Estimating transformation and its effects on Box-Cox T-ratio. *Test.* **8**, 167-190.
33. Yang, Z. (2000). A new statistics for regression transformation. *Test.* **9**, 123-132.
34. Yang, Z. (2001). Predicting a future median life through a power transformation. *Lifetime Data Analysis.* **7**, 305-317.
35. Yang, Z. (2002a). Comment on "Box-Cox transformations in linear models: large sample theory and tests of normality". *Canadian Journal of Statistics.* Vol. 30, No. 2. To appear.
36. Yang, Z. (2002b). Median estimation through a regression transformation. *Canadian Journal of Statistics.* Vol. 30, No. 2. To appear.
37. Yang, Z. and Abeysinghe, T (2002). An explicit variance formula for the Box-Cox functional form estimator. *Economics Letters.* **76**, 259-265.
38. Yang, Z. and Tse, Y. K. (2002). Quantile estimation through a regression transformation. *Unpublished Manuscript.* Singapore Management University, School of

Economics and Social Sciences.

39. Yeo, I. K. and Johnson, R. A. (2000). A new family of power transformation to improve normality or symmetry. *Biometrika*, **87**, 954-959.
40. Zarembka, P. (1968). Functional for in the demand for money. *Journal of the American Statistical Association*. **63**, 502-511.