Appeared in: Economics Letters, 2006, 92, 14-19.

A Modified Family of Power Transformations

Zhenlin Yang*
School of Economics and Social Sciences, Singapore Management University
90 Stamford Road, Singapore 178903

December 12, 2005

Abstract

A modified family of power transformation, called the *dual power transformation*, is proposed. The new transformation is shown to possess properties similar to those of the well-known Box-Cox power transformation, but overcomes the long-standing truncation problem of the latter. It generates a rich family of distributions that is seen to be very useful in modeling and analysis of durations and event-times.

Keywords: Box-Cox transformation; Dual-power transformation; Duration; Trans-normal.

JEL Classification: C1; C5

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. Recent applications in economics include, among others, Machado and Mata (2000), Chen (2002), and Yang and Tsui (2004).

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 $-1/\lambda$

*Tel.: +65-6828-0852; fax: +65-6828-0833. *Email address:* zlyang@smu.edu.sg

Appeared in: Economics Letters, 2006, 92, 14-19.

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. See Yeo and Johnson (2000) and the references therein. 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. 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 $\lambda \neq 0$.

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. It generates a well-defined family of distributions, called the *trans-normal distribution*, that is shown to be very flexible in modeling and analysis of durations and event-times. As there is no longer a truncation problem, all the standard asymptotic results of the maximum likelihood theory apply to the transformed regression models induced by the new transformation. Section 2 introduces the new transformation. Section 3 introduces the trans-normal distribution. Section 4 presents two empirical illustrations and Section 5 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

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

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.*

Appeared in: Economics Letters, 2006, 92, 14-19.

Unlike the Box-Cox power transformation that leaves the data untransformed when $\lambda = 1$, the dual power transformation always transforms the data no matter what value λ takes. This may sound contradictory, but is consistent with y being nonnegative. Thus, the new transformation makes more sense and is technically more sound. Moreover, when the magnitude of y is large, the y^{-1} term becomes negligible and the dual power transformation becomes essentially linear. We now collect some properties of the new transformation.

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

(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 easy to derive. 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 (figures are available from the author). The plots show that the smaller the magnitude of λ, the closer is the two power transformation with the limits (at $\lambda = 0$) 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). Burbidge, et al. (1988) studied the IHS transformation and compared it with the Box-Cox power transformation. 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, like the Box-Cox power transformation, has a domain of positive half real line, but does not suffer from the truncation problem.

3 The Trans-Normal Distribution

One of the immediate econometric applications (from the modelling point of view) of the dual power transformation is that it generates a new family of distributions. If $h(Y, \lambda)$ follows $N(\mu, \sigma^2)$, then the probability density function (pdf) of Y is given as follows

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

where $y > 0$; $-\infty < \mu < \infty$, $\sigma > 0$, and $-\infty < \lambda < \infty$. Clearly, this is a well defined three-parameter distribution. As it is generated from a normalizing transformation, we call it in this paper the *Trans-Normal Distribution*.

In the analysis of economic durations, medical event-times, and engineering reliability data, one constantly faces the problem of choosing a suitable distribution from many such as lognormal, Weibull, gamma, inverse Gaussian, and Birnbaum-Saunders. Thus, it is of great interest to have a flexible distribution that either embeds or closely approximates those popular distributions to ease the model selection problem one faces in practice.

The trans-normal family overlaps with the ξ-normal family of Saunders (1974) as they both share the property that $h(y) = -h(y^{-1}), y > 0$, and they both contain the lognormal distribution as a special member. Some general theoretical properties are given below. Through out, we use the subscripted h function to denote the partial derivatives.

Proposition 2. Define

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

The pdf $f(y;\mu,\sigma,\lambda)$ defined in (4) satisfies:

(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.

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 at certain parameter settings.

To illustrate the versatility and usefulness of the trans-normal distribution, we pick a special dual power transformation corresponding to $\lambda = 0.5$, and plot the pdf and the hazard function for serval parameter configurations. The plots (available from the author) show that the pdf of the 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 μ of $\log Y$. When σ is small relative to μ, 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. The plots also exhibit 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). 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 very long.

4 Empirical Illustrations

From the discussions in the previous sections, it is clear that the use of dual power transformation is technically favored to the use of Box-Cox power transformation. We now present two empirical applications to compare the two transformations.

Example 4.1: Strike Durations. The strike duration data of Kennan (1985) has been analyzed by many authors using various models (see Greene, 2000). Fitting the newly proposed trans-normal distribution to the same data gives the MLEs of the model parameters (μ, σ^2, and λ) as ($3.7461, 1.8228$, and 0.2843). Correspondingly, fitting the Box-Cox transformation to the data gives the parameter estimates ($4.2876, 2.1009$, and 0.1663). 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, showing a slightly better fit from the dual power transformation. Also, both models fit the data better than the models considered in Greene (2000).

Example 4.2: Computer Execution Times. The data, representing the amount of time it took to execute a particular computer program (the response), on a multiuser computer system, as a function of system load (the regressor) at the time when execution was beginning, is taken from Meeker and Escobar (1998, p638). Fitting a dual-power transformation model results in an MLE of λ being zero, i.e., the loglinear model of Meeker and Escobar (1998) is warrant from the fitting of the dual power transformation. The MLEs of the intercept, the slope and the error standard deviation are, respectively, $4.4936, 0.2907$ and 0.3125. Fitting a Box-Cox transformation model results in 'MLEs' of λ and other parameters as -0.4340, $1.9829, 0.0291$ and 0.0338, very much different from the estimates from the dual power transformation. The multiple coefficient of regression 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.

5 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 general skewed distribution, possibly of two modes. 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. All the standard asymptotic results of the maximum likelihood theory apply to the trans-normal distribution as well as various regression models with transformation applied to the response and/or regressors.

Acknowledgements

The author thanks Lonnie Magee, Chuck Coleman and In-Kwon Yeo for helpful comments. Research support from the Wharton-SMU Research Center, Singapore Management University, is gratefully acknowledged.

References

1. 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.
2. 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.
3. Chen, S. (2002). Rank Estimation of Transformation Models. *Econometrica*, 70, 1683-1697.
4. Greene, W. H. (2000). *Econometric Analysis, 4th ed.* Prentice Hall.
5. Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika.* **36**, 149-176.
6. Kennan, J. (1985). The duration of contract strikes in U.S. manufacturing. *Journal of Econometrics.* **28**, 5-28
7. Kiefer, N. M. (1988). Economic duration data and hazard functions. *Journal of Economic Literature.* **26**, 646-679.
8. 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.
9. Meeker, W.Q. and Escobar, L. A. (1998). *Statistical Methods for Reliability Data.* John Wiley & Sons, Inc., New York.
10. Saunders, S. C. (1974). A family of random variables closed under reciprocation. *Journal of the American Statistical Association.* **69**, 533-539.

Appeared in: Economics Letters, 2006, 92, 14-19.

11. Yang, Z. L. and Tsui, A. K. (2004). Analytically calibrated Box-Cox percentile limits for duration and event-time models. *Insurance Mathematics and Economics.* **35**, 649-677.

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