The Korean Communications
in Statistics Vol. 5, No. 2, 1998
pp. 297-306

A Recursive Method of Transforming a Response Variable for Linearity

Han Son Seo[1)]

Abstract

We consider a graphical method for visualizing the strictly monotonic transformation of $t(y)$ so that the regression function $E(t(y) \mid x)$ is linear in the predictor vector x. Cook and Weisberg (1994) proposed an inverse response plot which relies on the results of Li & Duan (1989) to obtain consistent estimates. Based on the recursive addition of the results from the two dimensional plots, we propose a new procedure which can be used when the consistency result is in doubt.

1. Introduction

The linear regression model is considered as a standard method to analyze statistical data. The assumptions behind the linear model include homogeneity of variance, additivity and normality. Transformation of the response which would result in satisfying the above assumptions was firstly considered by Box and Cox (1964). They formulate this problem by restricting transformation of the response within the parametric family of power transformations indexed by an unknown power. For the estimation of unknown power they suggest a maximum likelihood type estimation by maximizing an objective function. This method gives transformation toward normality because unknown power is estimated to make errors as nearly like a normal sample as possible. Cook and Weisberg (1989) used a dynamic graphical method for the determination of unknown power. Atkinson (1973, 1982) suggested a score statistic for transformation and proposed partial residual plot for the estimation of transformation parameter.

Here we concentrate on the linearity issue. Suppose that there is a strictly monotonic transformation of the response $t(y)$ with response y and the predictor vector x :

$$t(y) \mid x = \beta_0 + \beta^T x + \varepsilon \tag{1.1}$$

where $E(\varepsilon) = 0$, $Var(\varepsilon) = \sigma^2$ and ε is independent of x. In this formulation any linear function of $t(y)$ which satisfies (1.1) will also satisfy (1.1) with corresponding value of β_0 and β. Hence $t(y)$ is not unique and any strictly monotonic transformation that satisfies

1) Assistant Professor, Dept. of Applied Statistics, Kon-Kuk University, Seoul, Korea

(1.1) will be denoted by $t(y)$. Since $t(y)$ is defined only up to linear transformations, we assume without loss of generality that $\beta_0=0$. Cook and Weisberg (1994) suggested a graphical method for visualizing the form of the transformation of $t(y)$, which will be reviewed shortly later. Under the certain conditions, Cook and Weisberg's inverse response plot seems to perform well. However inverse response plot relies on the results of Li & Duan(1989) to obtain a consistent estimate of $k\beta$ in (1.1). In this paper we propose a new procedure, based on the recursive addition of the results from two dimensional plots, which can be used when the consistency result is in doubt. Section 2 reviews the inverse response plot suggested by Cook and Weisberg (1994). A new procedure and a couple of examples are given in section 3. Section 4 contains some remarks.

2. Inverse response plot

We denote two dimensional plot as $\{h\ v\}$ with the understanding that h is assigned to the horizontal axis and v is assigned to the vertical axis. Assume (1.1) and consider the problem of finding a response transformation $t(y)$. If $k\beta$ were known the plot $\{ y,\ k\beta^T x \}$ will provide a visualization of an appropriate transformation. According to the different value of k, the transformations displayed may not be same, but they are related linearly and are all satisfy (1.1). Cook and Weisberg (1994) suggested an inverse response plot taking the following approach. Since exact value of $k\beta$ is not so sure in most cases maximum likelihood-type regression based on a linear model $y=\alpha_0+\alpha^T x+e$ is considered as a practically useful estimators of $k\beta$. The maximum likelihood estimate of (α_0, α) is obtained by minimizing an objective function $n^{-1}\sum L(\alpha_0+\alpha^T x_i, y_i)$, where $L(m,y)$ is convex in m for each y. Ordinary least squares is an example of possible estimation method. Following Li & Duan (1989), $\hat{\alpha}$ is a consistent estimator of $k\beta$ if $E(x \mid \beta^T x)$ is linear in $\beta^T x$. This condition hold for all β if and only if x has an elliptically contoured distribution (Eaton, 1986). Since we can get a consistent estimate of $k\beta$ using Li & Duan's results, we now assume that β is known in the argument of population case.

We consider the plot $\{ y,\ \beta^T x\}$ for obtaining a good impression of an appropriate transformation. The plot $\{ y,\ \beta^T x\}$ is useful when, at least approximately, $E(\beta^T x \mid y)=t(y)$. Since $t(y)$ is assumed to be monotonic, the following holds :

$$\begin{aligned} E(\beta^T x \mid y) &= E(\beta^T x \mid t(y)) \\ &= E(\beta^T x + \varepsilon - \varepsilon \mid t(y)) \\ &= E(t(y) - \varepsilon \mid t(y)) \\ &= t(y) - E(\varepsilon \mid \beta^T x + \varepsilon) \end{aligned}$$

Thus to satisfy the condition $E(\beta^T x \mid y) = t(y)$, it requires that $E(\varepsilon \mid \beta^T x + \varepsilon)$ should be linear in $t(y)$. It follows from Cambanis, Huang and Simons (1981) that this condition will hold if $(\beta^T x, \varepsilon)$ follows an elliptically contoured distribution. As another example that the condition holds, if ε and $\beta^T x + \varepsilon$ have same marginal distribution f then

$$\begin{aligned} E(\varepsilon \mid \beta^T x + \varepsilon = l) &= \frac{\int \varepsilon f(\varepsilon) f(l - \varepsilon) d\varepsilon}{\int f(k) f(l - k) dk} \\ &= l - E(\varepsilon \mid \beta^T x + \varepsilon = l) \end{aligned}$$

and thus $E(\varepsilon \mid \beta^T x + \varepsilon)$ is linear in $\beta^T x + \varepsilon = t(y)$.

To measure the degree of linearity, the population correlation coefficient between $E(\beta^T x \mid t(y))$ and $t(y)$ can be used.

$$\rho = \frac{Cov[E(\beta^T x \mid t(y)), t(y)]}{Var(E(\beta^T x \mid t(y)))^{1/2} \cdot Var(t(y))^{1/2}} \tag{2.1}$$

Let

$$E(\beta^T x \mid t) = E(\beta^T x) + \delta(t - E(t)) + \gamma \tag{2.2}$$

where

$$\delta = \frac{Cov(t, E(\beta^T x \mid t))}{Var(t)} = \frac{Var(\beta^T x)}{Var(\beta^T x) + Var(\varepsilon)} \quad \text{and} \quad \gamma = E(\beta^T x \mid t) - E(\beta^T x) - \delta(t - E(t)).$$

Since δ is defined to be the population OLS regression of $E(\beta^T x \mid t(y))$ on t, $Cov\{\delta(t - E(t)), \gamma\} = 0$ and (2.1) becomes

$$\rho = \left\{1 + \frac{Var(\gamma)}{Var(\beta^T x)} \left(1 + \frac{Var(\varepsilon)}{Var(\beta^T x)}\right)\right\}^{-1/2} \tag{2.3}$$

The applicability of the inverse plot depends only on the noise-to-signal ratio $Var(\varepsilon) / Var(\beta^T x)$ and on a nonlinearity ratio $Var(\gamma) / Var(\beta^T x)$. Once $t(y)$ is estimated from the inverse plot applicability of the method can be checked by calculating a sample correlation between $\widehat{t(y)}$ and an estimate of $E(\beta^T x \mid y)$ which can be obtained by smoothing the plot

$\{\widehat{t(y)}, \widehat{\beta^T x}\}$. A graphical check is also available. Fit the regression of $\widehat{t(y)}$ on the predictors, and draw the plot $\{\widehat{\beta^T x}, \widehat{t(y)}\}$. The linear and homoscedastic trend implies that the deviations γ from (2.2) are all negligible and an appropriate transformation has been found. For examples, see Cook & Weisberg (1994).

3. A recursive procedure

We suggest a new procedure which can be applied when the consistency result is in doubt. Cook & Weisberg's inverse response plot becomes free of consistency restriction in simple linear regression case. Suppose that there is a strictly monotonic transformation of the response $t(y)$ with response y and a predictor x :

$$t(y) \mid = \beta_0 + \beta_1 x + \varepsilon \tag{3.1}$$

where $E(\varepsilon) = 0$, $Var(\varepsilon) = \sigma^2$ and ε is independent of x. We have $x = \{t(y) - \beta_0 - \varepsilon\}/\beta_1$ by just solving the equation (3.1) with respect to x. Thus we consider the plot $\{y, x\}$ rather than $\{y, \widehat{\beta_1} x\}$ to obtain a graphical display for a function $t(y)$. Similar arguments in section 2 are applied here. The plot $\{y, x\}$ is useful when, at least approximately, $E(x \mid y) = t(y)$ which is equivalent to $E(\beta_0 + \beta_1 x \mid y) = t(y)$. As a measurement of linearity, the population correlation coefficient $E(\beta_0 + \beta_1 x \mid t(y))$ and $t(y)$ can be written as

$$\rho = \left\{ 1 + \frac{Var(\gamma)}{Var(\beta_0 + \beta_1 x)} \left(1 + \frac{Var(\varepsilon)}{Var(\beta_0 + \beta_1 x)} \right) \right\}^{-1/2} \tag{3.2}$$

with similarly defined γ in (2.2).

Now we go back to the multiple linear regression case. For the simplicity of the problem, we consider linear regression with two predictors. Suppose that there are strictly monotonic transformation of the response such that $E(t(y) \mid x_1, x_2)$ is linear in x_1, x_2 and that $E(h(y) \mid x_1)$, $E(g(y) \mid x_2)$ are approximately linear in x_1 and x_2:

$$\left.\begin{aligned} t(y) \mid x_1, x_2 &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \\ h(y) \mid x_1 &\approx \alpha_0 + \alpha_1 x_1 + \varepsilon_1 \\ g(y) \mid x_2 &\approx \gamma_0 + \gamma_1 x_2 + \varepsilon_2 \end{aligned}\right\} \tag{3.3}$$

then $t(y)$ is approximately same as $ah(y) + bg(y)$ for suitable values of a and b. $h(y)$ and $g(y)$ can be estimated from the plots $\{y, x_1\}$ and $\{y, x_2\}$ respectively. Though a condition (3.3) seems well guaranteed in most cases, when assumption (3.3) is in doubt we may take a transformation on independent covariates. For the determination of $t(y)$ with estimates of

$h(y)$ and $g(y)$, we can use animation. Animation is accomplished by varing λ over -1 to 1 and redrawing the plot:

$$\{y,\ \lambda h(y)+(1-|\lambda|)g(y)\} \tag{3.4}$$

A transformation $t(y)$ can be estimated by fitting a curve the plot. As λ varies corresponding plot of (3.4) and residual plot from the regression of $t(y)$ on x_1 and x_2 are redrawn and R^2 is also calculated. We stop at the plot with the biggest R^2 among the moderate residual plots and choose the corresponding $t(y)$ as an appropriate transformation. The applicability of this procedure depends on the applicability of the method for finding $h(y)$ and $g(y)$. This depends on the linearity between $E(\alpha_0+\alpha_1 x_1 \mid y)$ and $t(y)$ and the linearity between $E(\gamma_0+\gamma_1 x_2 \mid y)$ and $t(y)$, which are enhanced by letting $Var(x_1)$ and $Var(x_2)$ as large as possible from (3.2). Sample version of this condition can be roughly accomplished by orthogonalizing x_1 and x_2 (Cook & Weisberg, 1990). For the estimation of $h(y)$ and $g(y)$, although we may use any set of two predictors which span the same subspace as spanned by the columns of x_1 and x_2 we use orthonormal predictors to get good impression of $h(y)$ and $g(y)$. When the dimension of predictors are more than two we use the proposed method recursively. We summarize the procedure as follows:

1. Orthonormalize predictors and let them x_i^*, $i=1,,,p$.
2. Make a plot of $\{y,\ x_1^*\}$.

   Fit a curve the plot and let the fitted values be $\hat{h}(y)$.
3. Do the step 2 with x_2^* and let the fitted values be $\hat{g}(y)$.
4. As λ moves from 0 to 1 :

   Draw a plot of $\{y,\ \lambda\hat{h}(y)+(1-|\lambda|)\hat{g}(y)\}$.

   Fit a curve the plot and let's denote the fitted values as $\hat{t}_\lambda(y)$.

   Do the regression of $\hat{t}_\lambda(y)$ on x_1^* and x_2^*.

   Draw a residual plot and Calculate R^2.
5. Stop at the plot with the biggest R^2 among the suitable residual plots and let the corresponding $\hat{t}_\lambda(y)$ be $t^*(y)$.
6. Treat $t^*(y)$ as new $\hat{h}(y)$.

   Go back to the step 3 with next predictor x_3^* and find new $\hat{g}(y)$.

7. Repeat steps 4 and 5 and 6 until the last predictor is applied.
   The final $t^*(y)$ is considered as an appropriate response transformation.

If the number of predictors is small, for example p=3, then we may let

$$\hat{t}_{\lambda_1,\lambda_2}(y) = \text{ fitted curve of } \{y,\ \lambda_2\{\lambda_1\hat{h}(y)+(1-|\lambda_1|)\hat{g}(y)\}+(1-|\lambda_2|)\hat{s}(y)\} \quad (3.5)$$

where $s(y) \mid x_3 \approx \delta_0 + \delta_1 x_1 + \varepsilon_3$. We consider λ_1 and λ_2 together at step 4 and proceed the next steps. Examples are now given to illustrate the usefulness of the new procedure. Programs for examples are coded by Xlisp-stat (Tierney, 1990).

example 1: (Artificial data)

We generate a small sample size of 10 observations with predictors which does not follow elliptically contoured distribution so that consistency result is in doubt. 10 observations were generated according to the following models:

$$\left.\begin{array}{l} \log y = 18x_1 + 6x_2 + \varepsilon \\ y^3 = x_1 \cdot \varepsilon_1 \\ \log y - \dfrac{18}{\varepsilon_1} y^3 = 6x_2 + \varepsilon \end{array}\right\} \quad (3.6)$$

where ε is normal random variable with mean 0 and variance 0.05^2, and $\varepsilon_1 = \exp(3\varepsilon)$. x_1 is linearly defined by a beta random variable as $x_1 = 0.2 + 0.8 \times w$, where w follows beta distribution with parameters 0.5 and 3. y and x_2 were generated according to the models, $y = (x_1 \cdot \varepsilon_1)^{1/3}$ and $x_2 = (1/6) \times \{1/3 \log x_1 - 18x_1\}$ respectively. An artificial data is given in Table 1. Since $\log x_1$ is approximately linear in $x_1 \in (0.2,\ 1)$, the following holds:

$$\begin{array}{l} t(y) = \log y = 18x_1 + 6x_2 + \varepsilon, \quad \varepsilon \sim N(0, 0.3^2) \\ h(y) = \log y \approx (1/3)x_1 + 3\varepsilon \\ g(y) = \log y - 18y^3 \approx 6x_2 + \varepsilon \end{array}$$

Note that $t(y)$, $h(y)$ and $g(y)$ are all strictly monotonic functions. Figure 1 shows the inverse response plot of $\{y,\ \hat{y}\}$. A superimposed line 1 is a fitted curve to the points which is obtained from the ordinary least regression of $\hat{y}$ on y and y^2. Line 2 on the figure 1 is ordinary regression of $\hat{y}$ on $\log y$, representing transformation $t(y)$. Two lines on the figure 1 show that response transformation indicated by the plot $\{y, \hat{y}\}$ is far different from the correct transformation $t(y)$. We now apply the new procedure. $h(y)$ and $g(y)$ are estimated parametrically by the model, $E(x_i) = \beta_{0i} + \beta_{1i}y + \beta_{2i}y^2 + \beta_{3i}y^3$, $i = 1,2$. $\hat{t}_\lambda(y)$ is derived from the linear regression $\lambda\hat{h}(y) + (1 - |\lambda|)\hat{g}(y)$ on $y, y^2, \ldots, y^6$. Four frames of an animated plot in figure 2 correspond to $\{y, \hat{t}_\lambda(y)\}$ with indicated value of λ. Figure 3 is $\{y, \hat{t}_\lambda(y)\}$ and corresponding residual plot of $\lambda = 0.9$ with the biggest $R^2 = 0.831$. A superimposed line on the figure 3 is obtained from the ordinary least squares regression of $\widehat{t_{0.9}}(y)$ on $\log y$. The curve gives an excellent visual fit to the points in the figure.

Table 1. Data generated according to (3.6)

lbs	1 / 6	2 / 7	3 / 8	4 / 9	5 / 10
y	0.616831	0.572603	0.557519	0.66435	0.635319
	0.706679	0.647205	0.59759	0.691073	0.696159
x_1	0.264858	0.206599	0.215221	0.289896	0.229528
	0.408801	0.282782	0.212863	0.328726	0.286997
x_2	-0.868383	-0.707407	-0.731001	-0.938479	-0.770347
	-1.2761	-0.918517	-0.724539	-1.047985	-0.93034

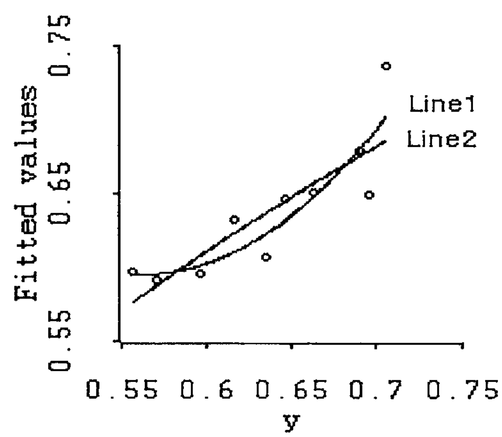

Figure 1. Plot of {y, $\hat{y}$} for Table 1.
Line 1 is fitted line to the point.
Line 2 is OLS line of $\hat{y}$ on *log y*.

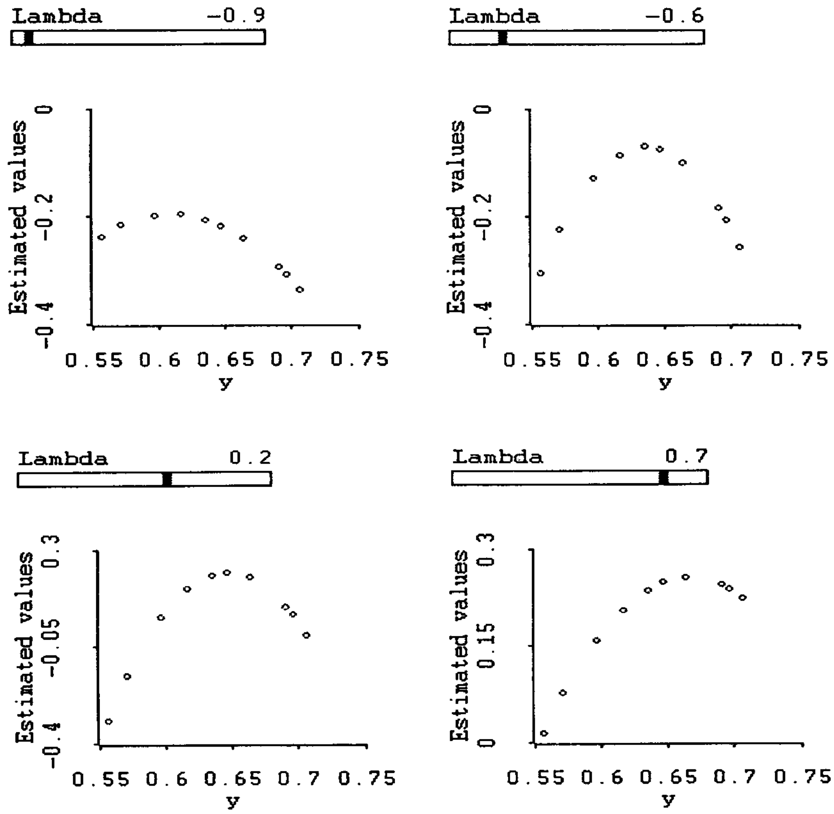

Figure 2. Plot of $\{y, \hat{t}_\lambda(y)\}$

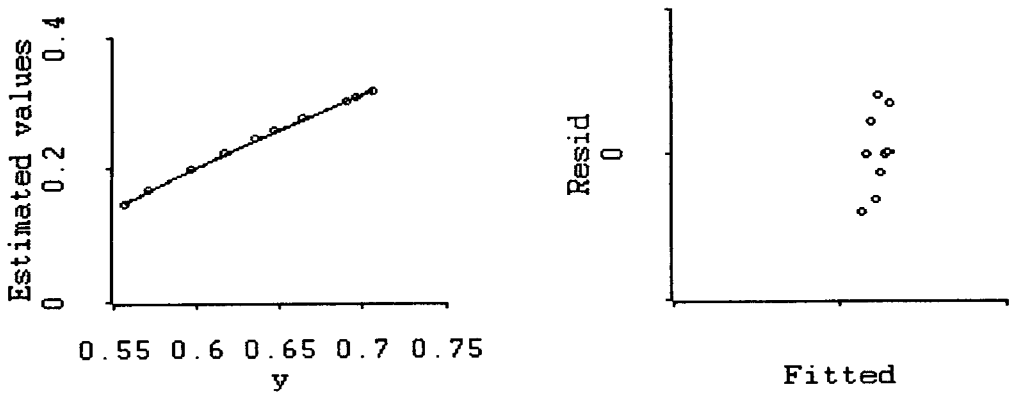

Figure 3. Plot $\{y, \widehat{t_{0.9}}(y)\}$ and residual plot from regression of $\hat{t}_{0.9}(y)$ on x_1 and x_2
Line on the plot $\{y, \widehat{t_{0.9}}(y)\}$ is from OLS regression of $\hat{t}_{0.9}(y)$ on *log y*.

example 2: (Wool data)

We consider wool data which was discussed by Box & Cox(1964) and Atkinson (1985, p.81). Cook and Weisberg (1994) also used this data to show the usefulness of the inverse plot. Since wool data has three predictors we consider $\widehat{t_{\lambda_1,\lambda_2}}(y)$ in (3.5). Figure 4 shows $\{y, \hat{t}_\lambda(y)\}$ and residual plot of $\lambda_1 = -0.6$ and $\lambda_2 = -1$ which has the biggest R^2=0.808. A superimposed line on the figure 4 is obtained from the ordinary least squares regression of $\hat{t}_{-0.6,0.1}(y)$ on $\log y$. The curve indicates that the log transformation is a strong candidate for achieving linearity in (1.1) agreeing with the conclusions of Box & Cox, of Atkinson and of Cook & Weisberg.

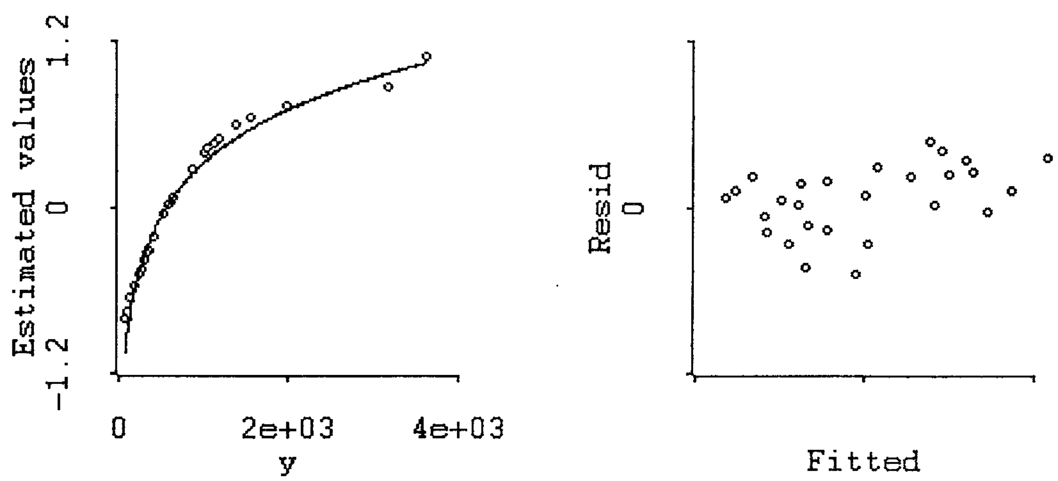

Figure 4. Plot $\{y, \hat{t}_{-0.6,0.1}(y)\}$ and residual plot from regression of $\hat{t}_{-0.6,0.1}(y)$ on x_1 and x_2 for wool data. Line on the plot $\{y, \hat{t}_{-0.6,0.1}(y)\}$ is from OLS regression of $\hat{t}_{-0.6,0.1}(y)$ on *log y*.

4. Concluding remarks

The inverse response plot suggested by Cook and Weisberg relies on the results of Li & Duan(1989) to obtain a consistent estimate of $k\beta$ in (1.1). The methodology proposed in this paper is designed to be useful when the consistency result is in doubt. Example 1 shows that this method suggests proper response transformation while Cook & Weinberg's method fails.

Furthermore, as we can see in the example 2, it also works with a usual data set as well as other methods do. New method can also be used to find a transformation which compromises the linearity and the normality assumption by displaying $\{y,\ \hat{t}_\lambda(y)\}$ and Q-Q plot, which change as λ varies, simultaneously.

References

[1] Atkinson, A. C. (1973). Testing transformations to normality, *Journal of Royal Statistical Society*, B, 35, 473-479.

[2] Atkinson, A. C. (1982). Regression Diagnostics, transformations and Constructed Variables (with discussion), *Journal of Royal Statistical Society*, B, 44, 1-35.

[3] Atkinson, A. C. (1985), *Plots, Transformations and Regression*, Oxford University Press, Oxford.

[4] Box, G. E. P. & Cox, D. R. (1964). An Analysis of Transformations (with discussion), *Journal of Royal Statistical Society*, B. 26, 211-246.

[5] Cambanis, S., Huang, S., and Simons, G. (1981). On the Theory of Elliptically Contoured Distributions, *Journal of Multiplicate Analysis*, 7, 368-385.

[6] Cook, R. D. and Weisberg, S. (1989), Regression Diagnostics With Dynamic Graphics *Technometrics*, 31, 277-311.

[7] Cook, R. D. and Weisberg, S. (1990), Three Dimensional Residual Plots *Proceedings of the 21st Symposium on the Interface: Computing Science and Statistics*, Washington: American Statistical Association, 162-166.

[8] Cook, R. D. and Weisberg, S. (1994). Transforming a response variable for linearity, *Biometrika*, 81, 731-737.

[9] Eaton, M. L. (1986). A Characterization of Spherical Distributions, *Journal of Multivariate Analysis*, 20, 272-276.

[10] Li, K. C. and Duan, N. (1989). Regression Analysis under Link Violation, *Annals of Statistics,* 17, 1009-1052.

[11] Tierney, L. (1990), *LISP-STAT*, Wiley, New York.