11-2001

A Note on Visualizing Response Transformations in Regression

R. Dennis Cook
University of Minnesota - Twin Cities

David J. Olive
Southern Illinois University Carbondale, dolive@math.siu.edu

Published in *Technometrics*, 43, 443-449.

A Note on Visualizing Response Transformations in Regression

R. Dennis Cook* and David J. Olive†

March 29, 2008

Abstract

A new graphical method for assessing parametric transformations of the response in linear regression is given. Simply regress the response variable Y on the predictors and find the fitted values. Then dynamically plot the transformed response $Y^{(\lambda)}$ against those fitted values by varying the transformation parameter λ until the plot is linear. The method can also be used to assess the success of numerical response transformation methods and to discover influential observations. Modifications using robust estimators can be used as well.

KEY WORDS: Box-Cox Transformation; Graphics.

---

*R. Dennis Cook is Professor, School of Statistics, University of Minnesota, St. Paul, MN 55108, USA. Research for this article was supported in part by NSF grant DMS-9703777.

†David J. Olive is Assistant Professor, Department of Mathematics, Southern Illinois University, Carbondale, IL 62901, USA.

The authors are grateful to the editors and referees for a number of helpful suggestions.

1 Introduction

It has long been recognized that the applicability of the linear regression model can be expanded by allowing response transformations, and issues related to determining a suitable transformation have generated an enormous literature.

Incorporating the possibility of transforming the positive response Y, the usual linear regression model can be represented in terms of an unknown transformation parameter λ_o,

$$Y^{(\lambda_o)} = \alpha_0 + \boldsymbol{\beta}_0^T \boldsymbol{x} + \epsilon \tag{1}$$

Here, $\boldsymbol{x}$ is a $p \times 1$ vector of predictors that are assumed to be measured with negligible error, the errors ϵ are assumed to be iid and symmetric about 0, and the transformed response is restricted to be a member of the power transformation family

$$Y^{(\lambda)} = \frac{Y^{\lambda} - 1}{\lambda} \tag{2}$$

for $\lambda \neq 0$ and $Y^{(0)} = \log(Y)$. Generally $\lambda \in \Lambda$ where Λ is some interval such as $[-1, 1]$ or a coarse subset such as $\Lambda_c = \{\pm 1, \pm 2/3, \pm 1/2, \pm 1/3, 0\}$. This family is a special case of the response transformations considered by Tukey (1957). In a classic paper, Box and Cox (1964) developed numerical methods for estimating λ_o. It is well known that the Box-Cox normal likelihood method for estimating λ_o can be sensitive to remote or outlying observations. Cook and Wang (1983) suggested diagnostics for detecting cases that influence the estimator, as did Tsai and Wu (1992), Atkinson (1986) and Hinkley and Wang (1988).

According to Tierney (1990, p. 297), one of the earliest uses of dynamic graphics was to examine the effect of power transformations. McCulloch (1993) gave a graphical method for finding response transformations, and Cook and Weisberg (1994) described how to use an inverse response plot of fitted values vs Y to visualize the needed transformation.

In this article we propose a new graphical method for assessing response transformations under model (1). The appeal of the proposed method rests with its simplicity and its ability to show the transformation against the background of the data. We introduce the method by example in Section 2. Details underlying the method are presented in Section 3, an example is presented in Section 4 and a concluding discussion is given in Section 5.

2 Introductory Illustrations

2.1 Textile Data

In their pioneering paper on response transformations, Box and Cox (1964) analyze data from a 3^3 experiment on the behavior of worsted yarn under cycles of repeated loadings. The response Y is the number of cycles to failure and the three predictors are the length, amplitude and load. Using the normal profile log likelihood for λ_o, Box and Cox determine $\hat{\lambda}_o = -0.06$ with approximate 95 percent confidence interval -0.18 to 0.06. These results give a strong indication that the log transformation may result in a relatively simple model, as argued by Box and Cox. Nevertheless, this basic method provides no direct way of judging the transformation against the data. This remark applies also to many of the diagnostic methods for response transformations in the literature. For example, the influence diagnostics studied by Cook and Wang (1983) and others are largely numerical. It would seem useful to have a simple method for judging transformations against the data.

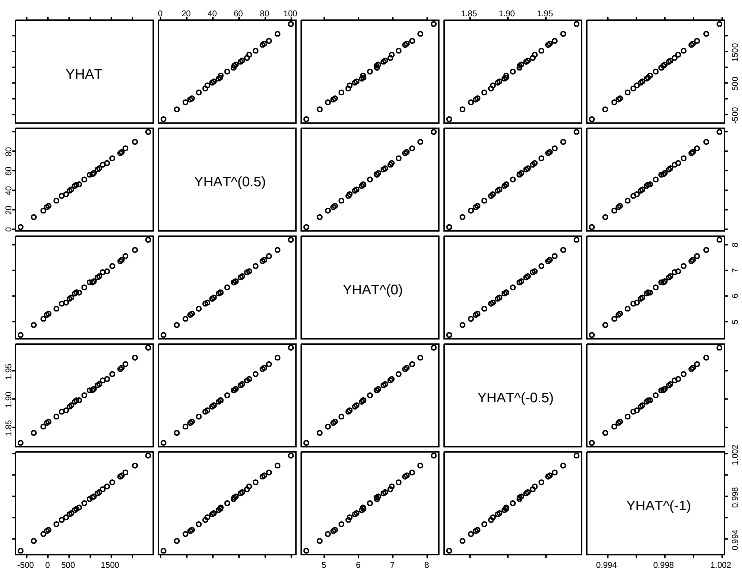

Figure 1: FFλ Plot for the Textile Data

Suppose we use OLS to fit the linear model (1) for two selected values λ_1 and λ_2 of λ in the usual interval $[-1, 1]$, and then plot one set of fitted values versus the other, $\hat{Y}^{(\lambda_1)}$ versus $\hat{Y}^{(\lambda_2)}$. We call this a "fit-fit" or an FFλ plot. How will the FFλ plot look? It turns out that in many regressions $\hat{Y}^{(\lambda_1)}$ and $\hat{Y}^{(\lambda_2)}$ have a strong linear relationship.For the textile data this linearity is evident in Figure 1 which shows a scatterplot matrix of FFλ plots for five selected values of λ. The smallest sample correlation among the pairs in the scatterplot matrix is about 0.9995.

The strong linearity in the FFλ plots of Figure 1 means that, if $\lambda_o \in [-1, 1]$, then

$$\hat{Y}^{(\lambda)} \approx c_\lambda + d_\lambda \hat{Y}^{(\lambda_o)} \tag{3}$$

for all $\lambda \in [-1, 1]$. Consequently, for any value of $\lambda \in [-1, 1]$, $\hat{Y}^{(\lambda)}$ is essentially a linear function of the fitted values $\hat{Y}^{(\lambda_o)}$ for the true λ_o, although we do not know λ_o itself. However, to estimate λ_o graphically, we could select any fixed value $\lambda^* \in [-1, 1]$ and then plot $Y^{(\lambda)}$ versus $\hat{Y}^{(\lambda^*)}$ for several values of λ and find the one for which the plot is linear with constant variance. This simple graphical procedure will work in this example because a plot of $Y^{(\lambda)}$ versus $\hat{Y}^{(\lambda^*)}$ is equivalent to a plot of $Y^{(\lambda)}$ versus $c_{\lambda^*} + d_{\lambda^*}\hat{Y}^{(\lambda_o)}$ by (3). Dynamic plotting using λ as a control seems quite effective.

Shown in Figure 2 are plots of $Y^{(\lambda)}$ versus $\hat{Y}^{(1)}$ (so that $\lambda^* = 1$) for four values of λ. The plots show how the transformations bend the data to achieve a homoscedastic linear trend. Perhaps more importantly, they indicate that the information on the transformation is spread throughout the data in the plot since changing λ causes all points along the curvilinear scatter in Figure 2a to form along a linear scatter in Figure 2c.

The next example illustrates that the transformation plots of Figure 2 can show characteristics of data that might influence the choice of a transformations by the usual Box-Cox procedure.

2.2 Mussel Data

Cook and Weisberg (1994) gave a data set on 82 mussels sampled off the coast of New Zealand. The response is muscle mass M in grams, and the four predictors are the length L, and height H of the shell in mm, the logarithm $\log W$ of the shell width W and the logarithm $\log S$ of the shell mass S. With this starting point, we might expect a log transformation of M to be needed because M and S are both mass measurements and

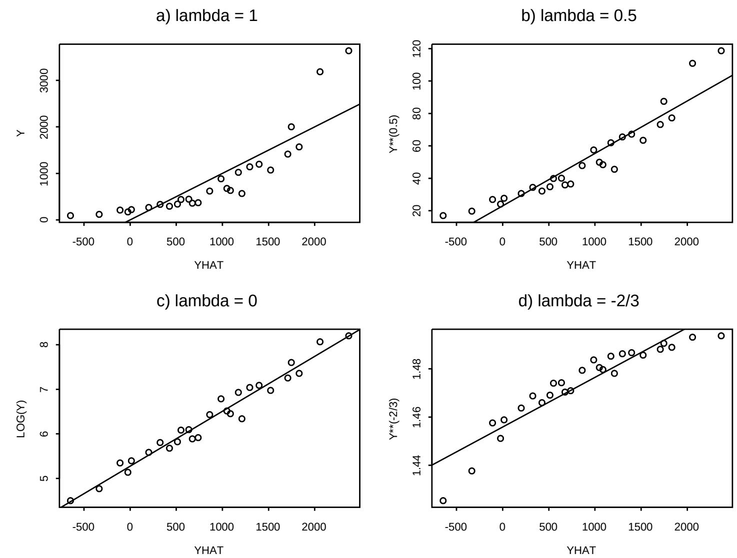

Figure 2: Four Transformation Plots for the Textile Data

$\log S$ is being used as a predictor. Using $\log M$ would essentially reduce all measurements to the scale of length. The Box-Cox likelihood method gave $\hat{\lambda}_0 = 0.28$ with approximate 95 percent confidence interval 0.15 to 0.4. The log transformation is excluded under this inference leading to the possibility of using different transformations of the two mass measurements.

The FFλ plots for $\lambda \in \Lambda_c$ (not shown) exhibit strong linear relations, the correlations ranging from 0.9716 to 0.9999. Shown in Figure 3 are transformation plots of $M^{(\lambda)}$ versus $\hat{M}$ (so $\lambda^* = 1$) for four values of λ. A striking feature of these plots is the two highlighted points that stand out in three of the four plots. The Box-Cox estimate $\hat{\lambda} = 0.28$ is evidently influenced by the two outlying points and, judging deviations from the OLS line in Figure 3c, the mean function for the remaining points is curved. In other words, the Box-Cox estimate is allowing some visually evident curvature in the bulk of the data so it can accommodate the two outlying points. Recomputing the estimate of λ_o

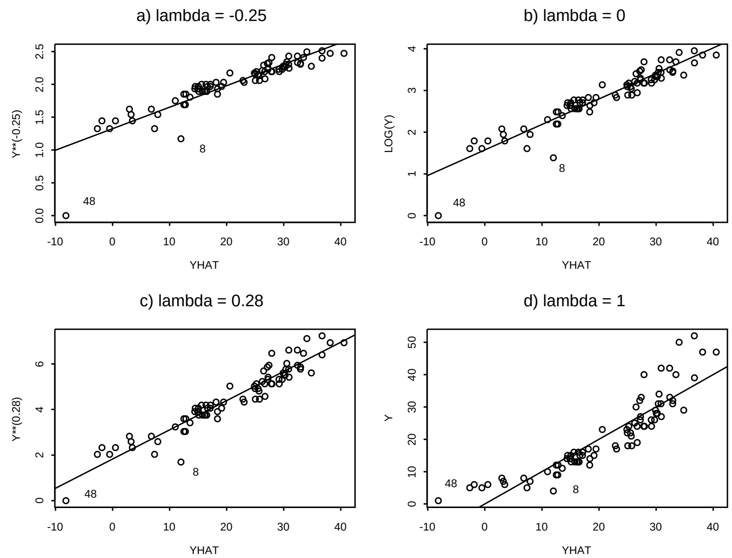

Figure 3: Transformation Plots for the Mussel Data

without the highlighted points gives $\hat{\lambda}_o = -0.02$, which is in good agreement with the log transformation anticipated at the outset. Reconstruction of the plots of $M^{(\lambda)}$ versus $\hat{M}$ indicated that now the information for the transformation is consistent throughout the data on the horizontal axis of the plot.

The essential point of this illustration is that observations that influence the choice of power transformation are often easily identified in a dynamic transformation plot of $Y^{(\lambda)}$ versus $\hat{Y}$ when the FFλ plots are strongly linear.

3 Explanations

The proposed graphical method is very simple. Use ordinary least squares (OLS) to regress the positive response variable on the predictors. Then dynamically plot $Y^{(\lambda)}$ against the fitted values by varying λ until the plot is linear. We call this the "dynamic transformation plot." This plot helps visualize $\hat{\lambda}_o$ against the data, and also the curvature and heteroscedasticity in competing models with different values of λ.

The starting point the dynamic transformation plot is that for many data sets, the FFλ plots are strongly linear. Hence any choice of λ^* can be used for the fitted values on the horizontal axis of the plot; in particular, $\lambda^* = 1$. An alternative choice is $\lambda^* = 0$. Because $\lambda^* = 0$ is in the middle of the usual range $[-1, 1]$, the correlation between $\hat{Y}^{(0)}$ and other fitted values will tend to be larger than the correlations using $\hat{Y}^{(1)}$. This may be of little consequence when the FFλ correlations are quite large, but may help in cases when the FFλ correlations begin to weaken. By using a single set of fitted values, say $\hat{Y}^{(1)}$ or $\hat{Y}^{(0)}$, on the horizontal axis, influential points or outliers that might be masked in plots of $Y^{(\lambda)}$ vs $\hat{Y}^{(\lambda)}$ for $\lambda \in \Lambda$ will show up unless they conform on *all* scales.

The fact that the FFλ correlations are frequently all quite high can be justified in at least two ways.

3.1 Local Approximation

The success of the dynamic transformation plot rests on finding fitted values $\hat{Y}^{(\lambda^*)}$ that are highly correlated with the fitted values $\hat{Y}^{(\lambda_o)}$ for the true λ_o. If λ_o and λ^* are both in Λ and Λ is sufficiently narrow then, by a simple Taylor series argument, $\hat{Y}^{(\lambda^*)}$ will be highly correlated with $\hat{Y}^{(\lambda_o)}$. The FFλ plots are used to assess the strength of the linear relationships between the different sets of fitted values $\hat{Y}^{(\lambda)}$ obtained by varying λ in Λ. This line of reasoning suggests that Λ should be a narrow interval containing λ_o. This is one reason why we used the interval $[-1, 1]$ in the previous examples; another is that useful transformations are often found in this interval. For example, recomputing the FFλ correlations after adding the four points ± 2 and ± 3 to the coarse set Λ_c defined following (2) resulted in reducing the smallest sample correlation from 0.9995 to 0.9940 in the textile regression and from 0.9716 to 0.8474 in the mussel regression. This reduction seems inconsequential for the textile regression but may not be so for the mussel regression.

The strength of the correlation between $\hat{Y}^{(\lambda^*)}$ and $\hat{Y}^{(\lambda_o)}$ can decrease as the error variance in (1) increases. For illustration, we replaced the response Y in the textile regression with a simulated response $\tilde{Y}$ constructed as $\log(\tilde{Y}) = \log(Y) + 0.38\varepsilon$ where ε is a standard normal random variable and 0.38 is about twice the estimated standard deviation of the model error ϵ from the OLS fit of $\log(Y)$ on the three predictors. Reconstructing the FFλ plots of Figure 1 using $\tilde{Y}$ reduced the minimum correlation to 0.9919 from 0.9995. Consequently, the correlations in FFλ plots will need to be assessed in each regression since they depend on more than just the choice of Λ.

3.2 Predictor Distribution

The second reason to expect that the FFλ plots may exhibit linear relationships rests with the distribution of the predictors. The following discussion of this reason incorporates a comparison of the dynamic transformation plot with other related plots in the literature.

Consider a model of the form

$$Y = h(\boldsymbol{\beta}^T \boldsymbol{x}, \epsilon). \tag{4}$$

where h is an unknown function. Let $\hat{\boldsymbol{\beta}}$ denote the coefficients of $\boldsymbol{x}$ obtained by minimizing an objective function L:

$$(\hat{a}, \hat{\boldsymbol{\beta}}) = \arg\min \sum_{i=1}^{n} L(a + \boldsymbol{b}^T \boldsymbol{x}_i, Y_i)$$

where the function $L(u, v)$ is a strictly convex function of u when v is held fixed. For example, the OLS objective function is obtained by setting $L(u, v) = (v - u)^2$. Li and Duan (1989) showed that if the conditional predictor expectation $\mathrm{E}(\boldsymbol{x}|\boldsymbol{\beta}^T\boldsymbol{x})$ is linear then $\hat{\boldsymbol{\beta}}$ is a consistent estimator for $\tau\boldsymbol{\beta}$ where τ is a scalar. Thus least squares, M-estimators with monotone ψ functions, and generalized linear model (GLM) estimators can be used to produce $\hat{\boldsymbol{\beta}}$, but high breakdown regression estimators may produce a $\hat{\boldsymbol{\beta}}$ that is not consistent for $\tau\boldsymbol{\beta}$. Further background on this result was given by Cook (1998, Ch. 8).

The linearity condition is the key new ingredient for the discussion of this section. It applies to the marginal distribution of the predictors and does not involve the response. It is implied when the predictors follow an elliptically contoured distribution, including the normal. Hall and Li (1993) show that the linearity condition will hold to a reasonable approximation in many problems. The intuition here is that conditional expectations of the form $\mathrm{E}(\boldsymbol{x}|\boldsymbol{\beta}^T\boldsymbol{x})$ become more linear as p increases. This is related to the work of Diaconis and Freedman (1984) who argue that almost all low-dimensional projections of high-dimensional data sets are nearly normal. In addition, the linearity condition holds to a good approximation in balanced designed experiments where the design points are "uniformly spaced" on the surface of a p-dimensional hypersphere (Ibrahimy and Cook 1995). Examples of standard designs that are included are 3^3 designs and central composite designs. In addition, the linearity condition might be induced by using predictor transformations and predictor weighting (Cook and Nachtsheim 1994).

When using a strictly convex objective function L and the linearity condition holds to a reasonable approximation, the *forward response plot* of Y versus the fitted values $\hat{Y}$ constructed using L provides a visual estimate of h. See Cook (1998, Ch. 8) and Cook and Weisberg (1999, section 18.5.3) for further discussion of this situation.

For a transformation of the response to be most useful, h must be a strictly monotone function t^{-1} :

$$Y = t^{-1}(\alpha + \boldsymbol{\beta}^T \boldsymbol{x} + \epsilon). \tag{5}$$

Hence

$$t(Y) = \alpha + \boldsymbol{\beta}^T \boldsymbol{x} + \epsilon. \tag{6}$$

Cook and Weisberg (1994) consider the forward (Y versus $\hat{Y}$) and inverse ($\hat{Y}$ versus Y) response plots in this setting. With $\hat{\boldsymbol{\beta}}$ still constructed by minimizing a strictly convex objective function, the inverse response plot shows t if the joint distribution of $\boldsymbol{\beta}^T \boldsymbol{x}$ and ϵ is elliptically contoured. It should also provide a useful display of the transformation t if the signal to noise ratio $\text{Var}(\boldsymbol{\beta}^T \boldsymbol{x})/\text{Var}(\epsilon)$ is large. It may be useful to emphasize that t is not parameterized in this approach. As long as t is strictly monotonic, nearly any function is possible as a choice for t.

The dynamic transformation plot proposed in this article adds another constraint by restricting t to a parsimoniously parameterized family. In particular, $t(Y) = Y^{(\lambda)}$, which brings us back to model (1). Now the key condition for the transformation plot is that we find a λ^* so that

$$\hat{Y}^{(\lambda^*)} \approx c_{\lambda^*} + d_{\lambda^*} \hat{Y}^{(\lambda_o)} \tag{7}$$

The linearity condition guarantees that for *any* λ (not necessarily confined to a selected

Λ), the *population* fitted values $\hat{Y}^{(\lambda)}_{\text{pop}}$ are of the form

$$\hat{Y}^{(\lambda)}_{\text{pop}} = \alpha_\lambda + \tau_\lambda \boldsymbol{\beta}^T \boldsymbol{x} \tag{8}$$

so that any one set of population fitted values is an exact linear function of any other set provided the τ_λ's are nonzero. This result indicates that sample FFλ plots will be linear, although it does not by itself guarantee high correlations. However, the strength of the relationship (7) can be checked easily by inspecting a few FFλ plots. The conditions necessary for the plots associated with (4) and (6) cannot be checked easily.

The idea of the transformation plot is similar to a graphical check suggested by Cook and Weisberg (1994, p. 734): suppose that t^* is the estimate of t. Then find $\hat{\boldsymbol{\beta}}$ from the regression of $t^*(y)$ on the predictors and verify that the plot of $t^*(y)$ versus $\hat{\boldsymbol{\beta}}^T \boldsymbol{x}$ is linear and homoscedastic. However, the dynamic transformation plots suggested here are not available in the context of (6) since its not clear how to choose a collection of t's that may be "near" the true t. That is, while we can plot the data at t^*, a series of ordered nearby t's would be required to use the idea underlying the dynamic transformation plots.

The dynamic transformation plot seems to work well in the textile data for at least three reasons. First, the design is a 3^3 so that (8) holds to a good approximation (Ibrahimy and Cook 1995). Second, the signal to noise ratio is large so that trends should be visually evident in a transformation plot. Third, $\Lambda = [-1, 1]$ seems narrow enough to gain benefit of the local approximation. In many regressions these reasons evidently work together to produce strongly linear FFλ plots and therefore useful dynamic transformation plots.

If the FFλ plots are not strongly linear then it is often useful to consider two possibilities: Either the transformation model (1) does not provide a good representation of the data, including the possibilities of outliers and influential cases, or there are strong *nonlinear* relationships among the predictors. How to proceed in the former case depends on the applications context. In the latter case, linearizing the predictor relationships by using marginal power transformations prior to application of model (1) is often a useful procedure.

4 Mussel Data Again

In this section we return to the mussel data, this time considering the regression of M on the four untransformed predictors L, H, W and S. Assuming model (1) in the four untransformed predictors, the Box-Cox likelihood estimate of λ_o is 0.49, suggesting the square root transformation. FFλ plots for this regression are shown in Figure 4; There is notable curvature in some of the plots and the FFλ correlations may be too small for the transformation plots to work well.

Nevertheless, application of the transformation plot with $\hat{Y}^{(0)}$ on the horizontal axis clearly points to the two highlighted cases in Figure 3 as influential. Deleting those cases noticeably improved the linearity in the FFλ plots, the sample correlations now ranging between 0.89 and 0.94. Even so, there is still substantial curvature in the FFλ scatterplot matrix. The Box-Cox likelihood estimate of the transformation parameter is now 0.23, suggesting the fourth root transformation. However, the usual diagnostic plots do not sustain model (1) with the original predictors and $\lambda_o = 0.25$. For example, the plot of the residuals versus S is clearly curved, the quadratic S^2 adding significantly to the model, and there is evidence of heteroscedasticity as well. Both of these conclusions are suggested by the transformation plots.

One way to attempt improvement of the model for the mussel data is to transform the predictors to remove gross nonlinearities. This might be done, for example, by using simultaneous power transformations $\boldsymbol{\lambda} = (\lambda_L \ldots \lambda_S)^T$ of the predictors so that the vector of transformed predictors $\boldsymbol{x}^{(\boldsymbol{\lambda})} = (L^{(\lambda_L)}, \ldots, S^{(\lambda_S)})^T$ is approximately multivariate normal. A method for doing this was developed by Velilla (1993). The basic idea is the same as that underlying the likelihood approach of Box and Cox for estimating a power transformation of response in regression, but the likelihood comes from the assumed multivariate normal distribution for $\boldsymbol{x}^{(\boldsymbol{\lambda})}$. We used the procedure suggested by Velilla (1993), resulting in the transformed predictors used in the analysis described in Section 2.2 and a better model. Of course, there is no guarantee that the transformation model (1) is "correct", and the usual diagnostic tests should still be performed. However, if it is correct, then the methodology discussed in this paper should be helpful when estimating a transformation and judging its effects against the data.

5 Discussion

Construction. We think simplicity is one of the advantages of the proposed method for visualizing response transformations in regression. Apart from the usual assumption that transformation model (1) is correct, the only requirement behind the transformation plot is that the FFλ plots be strongly linear. This requirement can be checked easily in practice, so the various conditions discussed in Section 3 need not be worrisome, and

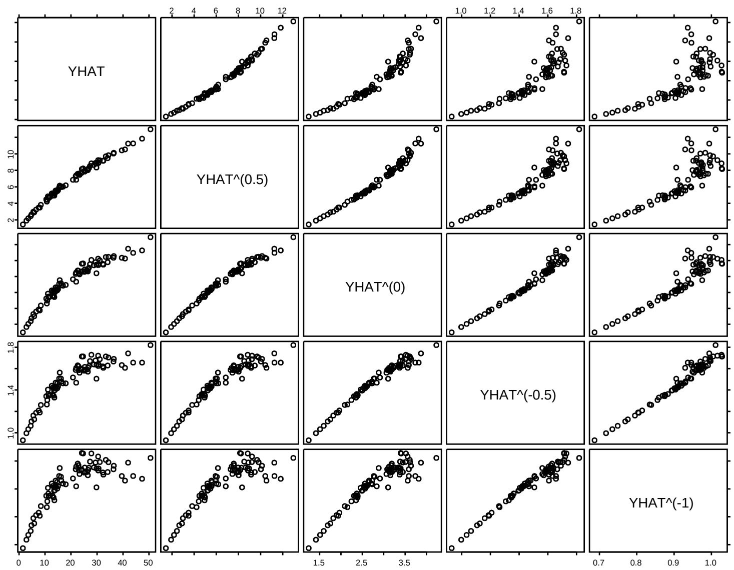

Figure 4: FFλ Plot for Mussel Data with Original Predictors

it might be induced by predictor transformations as necessary. Indeed, the proposed transformation plots seem to work quite well for the poison data (Box and Cox 1964) which has two qualitative predictors incorporated into the model (1) as factors, the smallest correlation in the FFλ plots over Λ_c being 0.9833.

The construction of a dynamic transformation plot requires that the user select a coarse subset Λ_c for constructing the separate views, and for assessing linearity in the corresponding FFλ plots. We have found it useful to insure that the Box-Cox estimate $\hat{\lambda}_o$ is near the center of Λ_c.

While we restricted attention to power transformations, the basic ideas here can be adapted straightforwardly to other parameterized transformation families.

Influence. The dynamic transformation plot shows the effect of a transformation against the data, and seems quite effective for identifying influential cases, but we make no claim that it will find influential cases that cannot be found by numerical methods or that it will find all influential cases. Indeed, we have found that dynamic transformation plots can play a useful role in studies that involve assessing influence by numerical methods. Highlighting in the frames of a dynamic transformation plot potentially influential points found by numerical methods often provides useful visual information about their role in determining the transformation. This visual information may sustain the results of numerical methods or raise questions about them, for example.

Robustness. Consider the possibility that there are outliers so model (1) may not hold for an unknown subset of cases. To address this issue, almost any good robust method can replace OLS in the construction of the transformation plots. If the objective function L is convex, we will still gain the benefits of the linearity condition discussed in Section 3.2. If the linearity condition holds and there are no outliers, then the coefficient estimate will be converging to a population value that is proportional to $\boldsymbol{\beta}$. If there are outliers then the robust estimate may be still converge to a vector proportional to $\boldsymbol{\beta}$, even if a transformed response $Y^{(\lambda)}$ is used, while the OLS estimate converges to a population vector that is not proportional to $\boldsymbol{\beta}$. In other words, we can address the issue of outliers by using a robust objective function for L and proceeding as before.

Strong Linearity. The requirement that the FFλ plots be "strongly linear" is quantitative and so giving a definitive cutoff is problematic. Nevertheless, our experience indicates that the dynamic transformation plots usually give informative results when the FFλ plots are linear with correlations larger than about 0.85. As the correlations drop substantially below this level, the value of transformation plots seems to decrease, although useful results can still be obtained. For example, analysis of the untransformed mussel data in Section 4 revealed the influential cases, although the FFλ plots were curved and the smallest FFλ correlation was about 0.76.

Inverse Response Plots. Perhaps the closest graphical method in the literature is based on inverse response plots by Cook and Weisberg (1994). They require assumptions that, while plausible in many applications, cannot be checked directly. Additionally, use of inverse response plots requires visualizing deviations from curves superimposed on an inverse response plot. We find it easier to visualize deviations from lines rather than curves. On the other hand, the primary context for the Cook-Weisberg plots is quite different from that for the transformation plots proposed here: They emphasize non-parametric transformations that linearize the mean function, while the approach here is for parsimoniously parameterized transformation families that linearize the mean function and stabilize the variance function. Either method may be appropriate depending on the applications context.

Access. All of the methods discussed in this article, including simultaneous power transformations to multivariate normality, are available in *Arc* (Cook and Weisberg 1999), which is available at the internet site http://www.stat.umn.edu/arc.

6 References

Atkinson, A.C. (1986) "Diagnostic Tests for Transformations," *Technometrics,* 28, 29-37.

Box, G.E.P., and Cox, D.R. (1964), "An Analysis of Transformations," *Journal of the Royal Statistical Society, B,* 26, 211-246.

Cook, R.D. (1998), *Regression Graphics: Ideas for Studying Regression through Graphics,* New York: John Wiley and Sons, Inc.

Cook, R.D., and Nachtsheim, C.J. (1994), "Reweighting to Achieve Elliptically Contoured Covariates in Regression," *Journal of the American Statistical Association,* 89, 592-599.

Cook, R.D., and Wang, P.C. (1983), "Transformations and Influential Cases in Regression," *Technometrics,* 25, 337-343.

Cook, R.D., and Weisberg, S. (1994), "Transforming a Response Variable for Linearity," *Biometrika,* 81, 731-737.

Cook, R.D., and Weisberg, S. (1999), *Applied Regression Including Computing and Graphics,* John Wiley and Sons, Inc., NY.

Diaconis, P., and Freedman, D. (1984), "Asymptotics of Graphical Projection Pursuit," *The Annals of Statistics,* 12, 793-815.

Hall, P., and Li, K.C. (1993), "On Almost Linearity of Low Dimensional Projections from High Dimensional Data, *The Annals of Statistics,* 21, 867-889.

Hinkley, D.V., and Wang, S. (1988), "More about Transformations and Influential Cases in Regression," *Technometrics,* 30, 435-440.

Ibrahimy, A., and Cook, R. D. (1995), "Regression Design for One-dimensional Subspaces," In Kitsos, C. P. and Müller, W. G. (Eds.), *Model Oriented Data Analysis*, New-York: Springer-Verlag, 125–134.

Li, K.C., and Duan, N. (1989), "Regression Analysis Under Link Violation," *The Annals of Statistics,* 17, 1009-1052.

McCulloch, R.E. (1993), "Fitting Regression Models with Unknown Transformations Using Dynamic Graphics," *The Statistician,* 42, 153-160.

Tierney, L. (1990), *Lisp-Stat,* John Wiley and Sons, Inc, NY.

Tsai, C.L., and Wu, X. (1992), "Transformation-Model Diagnostics," *Technometrics,* 34, 197-202.

Tukey, J.W. (1957), "Comparative Anatomy of Transformations," *Annals of Mathematical Statistics,* 28, 602-632.

Velilla, S. (1993). "A Note on Multivariate Box-Cox Transformations to Normality." *Statistics and Probability Letters,* 17, 259-263.