The Relationship Between Transformation and Weighting in Regression, With Application to Biological and Physical Science

Marie Davidian and Perry Haaland*

* Marie Davidian is Assistant Professor, Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203; and Perry Haaland is Senior Scientist-Statistics, Becton Dickinson and Company Research Center, Research Triangle Park, NC 27709. The authors thank Jim Hubbell for providing the data set used in the example.

ABSTRACT

Regression analysis is a commonly used tool in the biological and physical sciences. In addition to fitting a regression model, further goals may include prediction, calibration, and estimation of detection limits. The data are often systematically heteroscedastic, with variance small relative to the range of the mean response, and possibly asymmetrically distributed. Transformation is frequently used to induce both constant variance and normality of the response, so that subsequent inference presumably can be undertaken in a standard manner. Recent approaches to data transformation and accomodation of heteroscedasticity with variance models are briefly reviewed in a discussion based on Carroll and Ruppert (1988). There are situations in which transformation can not remedy all of the nonstandard features of the data simultaneously. As an illustration, an assay data set is used to point out the relationship between transformation and weighted regression as it might occur in many applications, especially in biological and physical science.

KEY WORDS: Assays; Asymmetry; Heteroscedasticity; Small Sigma; Subsampling; Variance Functions.

1. INTRODUCTION

The purpose of this article is to describe and illustrate with an example how transformation and weighted regression are related in a situation that occurs commonly in many applications, especially in the biological and physical sciences. It is also meant to be a resource for readers not familiar with some of the current methods for dealing with violation of standard assumptions in regression analysis, giving a brief indication of useful approaches and of descriptions and examples in the literature. Approaches we discuss are commonly known in the statistical literature, but many references we cite may be new to readers in applied fields. It is hoped that the discussion will encourage readers unfamiliar with the references and techniques to investigate their usefulness.

Most of the literature survey in Section 2 is paraphrased from the excellent monograph of Carroll and Ruppert (1988), abbreviated CR, and should serve as a starting point for those interested in but unfamiliar with the recent literature. Data transformation is commonly used to induce the "usual" regression assumptions of constant variance and normality. In Sections 3 we show how and why transformation sometimes cannot function in these capacities simultaneously, as described in CR (p. 123), and illustrate this with a data set in Section 4. The intention is not to analyze the data, but simply to informally illustrate why it is not uncommon that one cannot "transform away" all of one's problems, as is commonly perceived. We also illustrate how the assumption of independence can sometimes be violated for assay data.

2. A BRIEF SURVEY OF THE LITERATURE

Regression analysis is a useful tool in a variety of applications such as the biological and physical sciences, engineering, and economics. In biological, biopharmaceutical, and other applications, a standard analysis is to investigate the relationship between an independent variable x, such as concentration of a substance or dose of a drug, and a response y by regression techniques. Our discussion is in this context, but the

implications apply to more general situations for which x is a vector of covariates, as in econometrics and physical applications. Often, a subsequent goal is to base inference such as prediction, calibration, and estimation of quantities such as detection limits on the fitted regression. Generally, one assumes that observations on y are independent and that the x's are fixed and measured without error, and this is assumed throughout the following discussion. See Section 3 for more information.

Often, such data exhibit systematic heterogeneity of variance, such as variability increasing with level of mean response, and possibly asymmetric distribution, often right-skewed. Ignoring the first problem may yield estimated regressions that do not differ greatly from those based on accomodating heteroscedasticity somehow, as is commonly observed for assay data, see Oppenheimer, et al. (1983), but subsequent inference can be severely affected. CR (p. 53) show that ignoring nonconstant variance can result in grossly conservative prediction intervals. Davidian, Carroll, and Smith (1988) show how estimation of minimum detectable concentration for assays can be similarly affected. Inference such as prediction, meant to be performed on approximately normal data, may be rendered erroneous by serious asymmetry.

A common feature is apparent nonlinearity of the relationship on the original scales, denoted in the absence of error as

$$y = f(x, \beta), \tag{2.1}$$

where β is the $p \times 1$ regression parameter. The form of f, such as sigmoidal models for assay data, may be suggested by theoretical considerations or be based on empirical evidence so β may have meaningful interpretation on the original scale. The analyst's task is complex: find a meaningful model that accomodates the features that render a standard "textbook" regression analysis inappropriate. Much work in the statistical literature deals with these issues; a good summary of philosophy and methods, detailed

examples, and a comprehensive bibliography is given in CR, while a good recent example of creative application of such methods to a bioassay problem is described in Rudemo, Ruppert, and Streibig (1988).

A common method of dealing with these characteristics is to transform y. One may also transform x, see below, although under our assumptions this will have no effect on distributional characteristics. The transformation is often chosen by the analyst and regarded as fixed and known or it may be estimated. Originally, transformation was viewed as a tool to induce simultaneously "ideal" properties of normality, constant variance, and linearity of the regression in the parameters, as in the seminal paper of Box and Cox (1964). They defined the power transformation family indexed by λ as

$$y^{(\lambda)} = (y^{\lambda} - 1)/\lambda;\ \lambda \neq 0;\ = \log y\ ;\ \lambda = 0. \tag{2.2}$$

Other transformations are possible, but our discussion is in the context of the common family (2.2).

Subsequent investigations allow for current sophisticated computing technology that has made fitting of many nonlinear models a routine prospect, so that the major function of transformation is now perceived to be that of simultaneously inducing homoscedasticity and normality; linearity is not an issue. We express the idea of transforming the response only as

$$y^{(\lambda)} = f(x, \beta) + \sigma\, \epsilon, \tag{2.3}$$

where ϵ is assumed be standard normal, so that f is a possibly nonlinear model for the mean response on the transformed scale. A contested issue for (2.3) and (2.5) below is whether to view inference when λ is estimated as conditional on the estimated scale, which makes practical sense, or unconditional, which is mathematically appealing. This controversy is nicely explained in CR (Ch. 4); see also Hinkely and Runger (1981).

The perception that a single transformation induces both homoscedasticity and normality is not always appropriate; as CR (p. 118) state, "Although transformations can remove (both) heteroscedasticity and nonnormality, there is no guarantee that one transformation can do both." The latter part of this article is a case in which transformation corrects for heteroscedasticity but has no effect on the shape of the distribution of the data.

Transformations of type (2.3) and (2.5) below may make interpretation of β difficult or be undesirable if there is a theoretical model for y on the original scale. If a relationship like (2.1) is thought to exist, a useful approach to avoid this, see CR (p. 121), is the "transform-both-sides" method formally investigated by Carroll and Ruppert (1984). Here, one applies the same transformation to both response and regression function to preserve the meaningful relationship on the orginal scale. Write this as

$$y^{(\lambda)} = \{f(x,\beta)\}^{(\lambda)} + \sigma\epsilon, \tag{2.4}$$

where ϵ is standard normal. An example of application in chemical kinetics given in Bates, Wolf, and Watts (1985).

Another approach that does not preserve parameter meaning is to transform both y and x by the same transformation or two possibly different transformations,

$$y^{(\lambda)} = f(x^{(\gamma)},\beta) + \sigma\epsilon, \tag{2.5}$$

where x is also transformed by (2.2) indexed by γ. The case $\gamma = \lambda$ is useful when y and x are measurements on the same phenomenon. Possible inequality of γ and λ is useful for developing empirical models for which the data satisfy the usual regression assumptions when no theoretical or empirical model exists on the original scale.

In (2.3) – (2.5), λ and γ may be fixed in advance or estimated; when estimated,

the usual method is joint normal-theory maximum likelihood for $(\beta, \sigma, \lambda, \gamma)$, see CR (p. 124-126, 162-164) for implementation with standard software. Technically, if y is positive, transformation of y can induce strict normality only if $\lambda = 0$, see Ruppert and Aldershof (1988), so the assumption is really that of approximate normality.

Another approach is to assume a relationship like (2.1) and do something related to weighted regression. One might use replication at each x to form weights given by the inverse of the sample variance at x. This can be quite inefficient and yield inconsistent estimates for β for asymmetric data, as shown analytically by Carroll and Cline (1988). A similar statement can be made for assuming, for example, Poisson-like variability and using the reciprocal of the actual response at x as a weight. Both approaches are advocated in some standard regression texts, but should be avoided. More sophisticated, flexible, and reliable methods are available.

Since heteroscedasticity is often systematic, another approach is to postulate a mean model and a variance model on the original scale and use fitting techniques based on weighted regression. A benefit is that replication is not necessary to assess variance structure. One approach is to entertain a distributional model and assume the data arise from an exponential family, thus including variance models suggested by Poisson, gamma, and inverse Gaussian distributions, as with generalized (non)linear models, see McCullagh and Nelder (1983). This framework strictly allows normal data with homoscedastic variances only. Maximum likelihood leads to estimation by iteratively reweighted least squares (IRLS) to convergence.

One more generally might consider models for the first two moments of the data without reference to a distribution. Quasi-likelihood refers to considering variance to be proportional to a known function of the mean and using IRLS, see McCullagh and Nelder (1983, Ch. 8). For example, for pharmacokinetic data, a common variance model is that of constant coefficient of variation (CV), $\mathrm{Var}(y) = \sigma^2 f(x, \beta)^2$, see Beal

and Sheiner (1988), also suggested by the gamma distribution, but the data are often thought to be normal.

An even more flexible approach is to consider a general model

$$\mathrm{Var}(y) = \sigma^2 g^2\{f(x,\beta),\theta,z\} \tag{2.6}$$

for known vector of constants z and variance function g, where the vector of variance parameters θ may be unknown. Common models are the power model $g\{f(x,\beta),\theta,z\} = f(x,\beta)^\theta$ or the related model $g\{f(x,\beta),\theta,z\} = x^\theta$. Such models may not fit any known distributional framework. The power model is used for assay data, see Rodbard and Frazier (1975), Raab (1981), and Davidian, Carroll, and Smith (1988). Here, y is often a count with variability that exhibits increase with the mean more profound than that of Poisson; typically, $\theta = 0.6 - 0.9$. In addition, variability is quite small relative to the range of the mean response. Both features are accomodated by the power model, which allows for the possibility $\theta \neq 0.5$ and small relative variability through the value of σ. One can often assume approximate normality on the original scale since the counts are large, and empirical models for the original scale are common and well-interpreted, so one would not want to transform.

A phenomenon in biopharmaceutical data is that of a separate component of variance at low concentrations due to inability of equipment to measure the response precisely. A likely model is

$$\mathrm{Var}(y) = \theta_1 + \theta_2 f(x,\beta)^{\theta_3},\ f(x,\beta) > 0,$$

where θ_1 expresses the component at small x overwhelming the usual variance structure because of measurement error. See CR (p. 61) for an example of detection of such a characteristic; see also Rudemo, et al. (1988).

We can write the general form of y under a heteroscedastic regression

model as

$$y = f(x,\beta) + \sigma g\{f(x,\beta),\theta, z\}\epsilon, \tag{2.7}$$

where ϵ is a random variable with mean 0, variance 1 whose character describes the true distribution. A general account of variance modeling and estimation methods for θ is given in CR (Chs. 2 and 3); theoretical properties of methods are compared in Davidian and Carroll (1987). Some common methods are by a "regression" of a transformation of absolute residuals from a preliminary fit on the same transformation of σg, where $f(x,\beta)$ is replaced by its predicted value and (σ,θ) is the "regression" parameter. When the "regression" is based on squares, the method is called "pseudo-likelihood." Estimation of β is by generalized least squares, i.e., weighted regression with weights based on a previous $\hat{\beta}$ and some $\hat{\theta}$, and the process may be iterated. See Giltinan and Ruppert (1989) and CR (p. 72) for implementation with standard software. Beal and Sheiner (1988) suggest maximum likelihood based on a normal distribution with the given mean and variance functions to jointly estimate (β,σ,θ), which can yield a fitting method different from weighted regression. See CR (p. 20) for discussion.

Assuming a model like (2.7) accomodates nonconstant variance only and can have no effect on the shape of the data distribution. CR (Ch. 5) describe how one can often combine transformations and weighting to allow the former to address inducing a normal (or at least symmetric) distribution and the variance model to describe the variability, since, often, one transformation cannot accomplish both. Since the mean model may be based on deterministic considerations, they suggest a combination of (2.4) and (2.7) written as

$$y^{(\lambda)} = \{f(x,\beta)\}^{(\lambda)} + \sigma g\{f(x,\beta),\theta, z\}\epsilon, \tag{2.8}$$

where the transformation is assumed to induce approximate normality, so we assume standard normal ϵ, and the variance model accomodates the heteroscedasticity.

Transformation and weighting approaches can be adapted to provide robustness against outliers and high leverage; see Giltinan, Carroll, and Ruppert (1986) and Carroll and Ruppert (1987). The representation we have used implies the assumption of independence. For cases of, for example, subsampling at each x, nested models with possibly several heteroscedastic components as in Morton (1987) would be appropriate.

3. THE CASE OF "SMALL σ"

The parameter σ in a situation where a model like (2.7) is appropriate reflects the size of variance relative to the range of the means; it is the coefficient of variation for constant CV models. Thus, variability for biological and physical data is often well-described by thinking of σ as "small," where "small" in many applications may be around $\sigma = 0.10$ or less. Theoretical research applicable to regression for these data has capitalized on the fact that σ "small" is a reasonable approximation to reality, see Carroll and Ruppert (1984), Davidian and Carroll (1987), and Davidian, et al. (1988). Technical arguments allow sample size to become large and $\sigma \to 0$ simultaneously. For many problems, implications of the theory are valid for $\sigma < 0.30$. Indeed, the implications of "σ small" can be valid for regression situations in other fields.

Thinking of transformations in the context of σ "small" provides insight into why they sometimes do not seem to induce both constant variance and approximate normality. Suppose that true variance structure is as in (2.6), with heteroscedasticity a systematic function of mean response or level of x. As in CR (p. 123), solve (2.4) for y, and expand about $\sigma = 0$. This yields

$$y \approx f(x,\beta) + \sigma f(x,\beta)^{1-\lambda}\epsilon. \tag{3.1}$$

If σ is sufficiently small, then, (2.4) will behave like (2.7) with $E(y) = f(x,\beta)$ and $Var(y) = \sigma^2 f(x,\beta)^{2-2\lambda}$. Thus, if we remained on the original scale with variance model

$\mathrm{Var}(y) = \sigma^2 f(x,\beta)^{2\theta}$, thinking of (3.1) with θ replacing $1-\lambda$, we would accomodate heteroscedasticity only, where ϵ expresses the nature of the distribution of y. If σ is small, then, transformation does not change the character of the distribution from the way it was on the original scale.

Similarly, (2.3) and (2.5) yield, respectively,

$$\begin{aligned} y &\approx \{1 + \lambda f(x,\beta)\}^{1/\lambda} + \sigma[\{1 + \lambda f(x,\beta)\}^{1/\lambda}]^{1-\lambda}\epsilon; \\ y &\approx \{1 + \lambda f(x^{(\gamma)},\beta)\}^{1/\lambda} + \sigma[\{1 + \lambda f(x^{(\gamma)},\beta)\}^{1/\lambda}]^{1-\lambda}\epsilon, \end{aligned} \tag{3.2}$$

showing that for small σ, transformation yields implicit models on the original scale with variance modeled as a power 2θ of the "mean function," $\theta = 1-\lambda$; the shape of the distribution is unaffected. (Take limits to obtain the exponential "mean function" for $\lambda = 0$.) Thus, if σ is small and data exhibiting apparent skewness along with heterogeneity of variance are transformed in the hope of "correcting" both, the transformed data will retain the asymmetry. This is not to say that transformations can not be quite effective at achieving both aims simultaneously in many problems, see CR (Chs. 4 and 5) for numerous examples for which σ is not small.

Note that if we expand (2.8),

$$y \approx f(x,\beta) + \sigma f(x,\beta)^{1-\lambda} g\{f(x,\beta),\theta,z\}\epsilon, \tag{3.4}$$

so again transformation does not affect the distribution. The variance model in the small σ case functions to permit approximate models for variance besides that of the power model, which could be similarly undertaken by postulating an appropriate variance model in (2.7) on the original scale.

Even if the independence assumption is not valid, transformation will not affect the shape of the distribution or the correlation structure. In biopharmaceutical research, it

is common for assays to be conducted by subsampling a "batch" of substance at concentration x_i, so that replicate observations y_{ij}, say, at x_i ought to be correlated. Let $\sigma\omega_{ij}$ represent the "error" associated with such observations in a transformation model such as $y_{ij}^{(\lambda)} = \{f(x_i,\beta)\}^{(\lambda)} + \sigma\omega_{ij}$. The ω_{ij} can be thought of as composed of two homosecedastic components on the transformed scale, $\omega_{ij} = \delta_i^* + \epsilon_{ij}^*$, say, so the assumption is that the transformed data follow a nested homoscedastic error structure. The small σ approximation yields

$$y_{ij} \approx f(x_i,\beta) + \sigma f(x_i,\beta)^{1-\lambda}\omega_{ij}. \tag{3.5}$$

The implication is a heteroscedastic model on the original scale which could be written as $y_{ij} = f(x_i,\beta) + \sigma_\delta f(x_i,\beta)^{\theta_1}\delta_i + \sigma_\epsilon f(x_i,\beta)^{\theta_2}\epsilon_{ij}$ for independent random variables δ and ϵ with mean 0 and variance 1 and $\theta_1 = \theta_2$. Thus, when variance is small relative to the range of the means, the result of transformation is a heteroscedastic components-of-variance model; the representation with $\theta_1 \neq \theta_2$, is a generalization of (3.5). The transformation only affects the heterogeneity of the covariance struture. If correlation is present on the original scale of y as a result of such additive effects, it will obtain on the transformed scale in the same additive manner. The usual approach to fitting regression relationships for assays is to assume independence of the observations, see Rodbard (1978) and Raab (1981), thus ignoring the error δ_i due to "batch." This is often reasonable because variability among batches can be small relative to variability in the measurements. We discuss the implications for a set of data in the next section.

4. AN EXAMPLE

Understanding of how transformations work when σ is small can allow the analyst to be more effective at assessing unusual features of data. We consider a set of assay data given in Table 1 which was provided by Jim Hubbell of Burroughs Wellcome &

Co. (Research Triangle Park, North Carolina) and for which we do not discuss details. Our intention is not analysis but to show that the implications of Section 3 are valid.

The material assayed was obtained from two sources, with every other concentration starting with 0.476 from the first source and the remainder from the second. Replicate observations were obtained by subsampling at each concentration. The actual assay was performed by design, so that groups consisting of one sample from each concentration were run together, with run orders determined by the rows of a Latin Square. A plot of the data is given in Figure 1, from which systematic increase in variance with level of mean response, small variability relative to the range of the means, and reasonably straight-line relationship are evident. For purposes of illustration and to adhere to what is often done in practice, we proceed as if all observations were independent, noting the possibility for violation of this assumption in light of the potential sources of variability.

We fit a naive simple linear model to the data by least squares, $y = \beta_1 + \beta_2 x + \sigma\epsilon$, ignoring heteroscedasticity, obtaining $\hat{\beta}_1$= -0.021, $\hat{\beta}_2 = 0.144$, $\hat{\sigma} = 0.095$. We do not report standard errors since our goal is not analysis but simple illustration; here, the usual standard error estimates would be biased. The plot of studentized residuals in Figure 2 shows clear heteroscedasticity and evidence of some right-skewness. The skewness coefficient is 3.51, but these residuals do not have the same distribution. As a rough measure we averaged skewness values for studentized residuals at each x level, obtaining "skewness" of 0.90, suggesting some asymmetry. For small x, an ill-fitting mean model seems possible. There is also a slight sort of systematic "curvy" pattern to the groups of residuals, indicating possible source and/or subsampling effects.

A usual approach would be to consider $\log y = \beta_1^* + \beta_2^* \log x + \sigma\epsilon$, for which $\hat{\beta}_1^*$ = -2.05, $\hat{\beta}_2^* = 1.036$, $\hat{\sigma} = 0.037$. This is (2.5) with $\lambda = \gamma = 0$; $\hat{\sigma}$ suggests that σ is "small." A plot of studentized residuals based on the transformed data is given in Figure 3. The

horizontal axis is predicted values that have been transformed back to the original scale for comparison with Figure 2. The heteroscedasticity seems accomodated, but the spread of residuals at each x is different and is particularly large at the lowest x, which could suggest a component of variability due to imprecision as in Section 2 or an inappropriate mean model at small x. Note that what appears to be asymmetry remains, and that a systematic trend to the groups of residuals is evident. We invoked our "rough" skewness measure, which gave 0.82, on the order of that above.

For comparison, we estimated transformations by maximum likelihood, fitting (2.5) with $y^{(\lambda)} = \beta_1^* + \beta_2^* x^{(\gamma)} + \sigma\epsilon$. Even though the transformed data may not be normal, these estimates are consistent for σ small. For $\lambda = \gamma$ ($\neq \gamma$), $\hat{\beta}_1^*$ = -1.822 (-1.890), $\hat{\beta}_2^*$ = 0.814 (0.888), $\hat{\sigma}$ = 0.034 (0.033), $\hat{\lambda}$ = 0.120 (0.085), $\hat{\gamma}$ = 0.076 for $\lambda \neq \gamma$. Both sets of values suggest the log-log transformation. A residual plot for $\lambda \neq \gamma$ is given in Figure 4; that for $\lambda = \gamma$ is similar. Note the similarity to Figure 3, although the residuals at each x seem to exhibit a more uniform amount of spread. Possibly the log transformation "overcompensates" for nonconstant variance; a scatterplot on the log-log scale shows observations at the smallest x to have greatest spread. As for distribution, no transformation seems to have much effect. For $\lambda \neq \gamma$, for example, the "rough" skewness measure was 0.83. The apparent correlation structure is also evident.

Since the relationship appears linear on the original scale, we fit (2.4) $y^{(\lambda)} = (\beta_1 + \beta_2 x)^{(\lambda)} + \sigma\epsilon$, with $\hat{\beta}_1$ = -0.011, $\hat{\beta}_2$ = 0.142, $\hat{\sigma}$ = 0.039, $\hat{\lambda}$ = 0.065. For comparison, we tried weighting, fitting $y = \beta_1 + \beta_2 x + \sigma(\beta_1 + \beta_2 x)^{\theta}\epsilon$ by generalized least squares with pseudo-likelihood estimation of θ, obtaining $\hat{\beta}_1$ = -0.011, $\hat{\beta}_2$ = 0.142, $\hat{\sigma}$ = 0.039, and $\hat{\theta}$ = 0.931. Note the identical regression parameter estimates and that $\hat{\theta} \approx 1 - \hat{\lambda}$. This is exactly the behavior expected for σ small. The two approaches gave residual plots (not shown) that were virtually identical to each other and to Figures 3 and 4, suggesting again asymmetry in the original data.

We also fit models of form (2.8), taking transformation models already considered and adding the power variance model. $y^{(\lambda)} = \beta_1^* + \beta_2^* x^{(\gamma)} + \sigma(\beta_1^* + \beta_2^* x^{(\gamma)})^\theta \epsilon$ yielded $\hat{\beta}_1^* = -1.874$, $\hat{\beta}_2^* = 0.873$, $\hat{\lambda} = 0.093$, $\hat{\gamma} = 0.084$, $\hat{\theta} = -0.050$, suggesting that the implicit variance model in the unweighted case above has already compensated for nonconstant variance. For $y^{(\lambda)} = (\beta_1 + \beta_2 x)^{(\lambda)} + \sigma(\beta_1 + \beta_2 x)^\theta \epsilon$, the regression parameter estimates were identical to those for the unweighted model, $\hat{\lambda} = -4.886$, $\hat{\theta} = -4.951$, so that from (3.4), $1 - \hat{\lambda} + \hat{\theta} \approx 0.936$, suggesting the result obtained for the unweighted transform-both-sides case. Note the identifiability problem for small σ, so the $\hat{\lambda}$ and $\hat{\theta}$ values are questionable. Fitting, for example, $y^{(\lambda)} = \beta_1^* + \beta_2^* x^{(\gamma)} + \sigma x^\theta \epsilon$ gave $\hat{\beta}_1^* = -1.537$, $\hat{\beta}_2^* = 0.572$, $\hat{\lambda} = 0.299$, $\hat{\gamma} = 0.298$, $\hat{\theta} = 0.226$, $\hat{\sigma} = 0.040$; rough consideration of these values in the "back-transformed" model shows that the resulting fit is similar in spirit. All residual plots were very similar to Figure 4.

Note that for the log-log and fits based on (2.5), the "back-transformed" model for implies $y \approx e^{\beta_1^*} x^{\beta_2^*} + \sigma(e^{\beta_1^*} x^{\beta_2^*})\epsilon$ for small σ, and that from the transform-both-sides and weighted fits, $\hat{\beta}_1 \approx 0$, $\hat{\beta}_2 \approx e^{\hat{\beta}_1^*}$, and $\hat{\beta}_2^* \approx 1$. From such rough analogies and the comparisons above, apparently all attempts fit approximately the same mean model on the original scale with variance proportional to a power of the mean of around 2 under the assumption of independent observations. Because σ is small, all attempts act as models of form (2.7). No transformation can alter the apparent distribution of the data; the only effect is to model variance as a power of the implied mean model for the original scale. The possibly more complicated error structure is evident for both transformation fits and weighted fits, so is not affected by transformation of the response, as in (3.5).

Understanding the function of transformations when σ is small puts us in a better position to evaluate the pattern of apparent correlation or source effect common to all residual plots. From Section 3 and the discussion so far, we know that since σ is small,

the unusual pattern is the result of a phenomenon that we can think about on the original scale. We considered the fact that the pattern of residuals could in fact be due to an inappropriate mean model by fitting more general sigmoidal, polynomial, and exponential mean models with variance models such as the power model or power of x, with little improvement in the "curviness" exhibited in Figures 3 and 4. We tried different variance models to capture the possibility of a component of variance for small x, with some success in that the wide spread of residuals at $x = 0.476$ was reduced. We concluded that the pattern is indeed the result of the implied error structure. Our discussion shows that evaluation of this feature could be undertaken on the original scale, where one might investigate whether the nature of the heteroscedasticity is derived from the "batches" or subsampling or if in fact there is a heteroscedastic component from each. We do not pursue this issue here.

Jim Hubbell intentionally sets up these experiments by alternating sources at each concentration to obtain evidence as to whether the sources currently being used are uniform in composition. This allows for the effects of "batch" and "source" to be confounded, but he feels that the effect of batch is minor relative to the source effect. He fits a modification of the log-log model assuming independence and examines the pattern of residuals. Our discussion explains the mechanics of this fit. The models we have used do not contain a component for "source," which could be included and estimated for a more formal investigation.

Another feature of these data is apparent from further examination of residuals for the original data. Recalling that the assay was run in "groups," we identified each residual in a typical plot by "group" and found by informal inspection that the largest and second largest residuals for almost every x came from the same "group," as do the largest and second largest observations at each x. The implication is some nonuniformity to the assay procedure, although no ready explanation is available.

Again, transformation will not alter the character of this feature from its appearance for analyses on the original scale.

5. DISCUSSION

The illustration in Section 4 shows that the implications of the "small σ" arguments of Section 3 can be quite reasonable and important in practice. When variability, which in fact may be from several sources, is small relative to the range of the means, one may expect data transformation to have virtually no effect on the distribution or correlation structure of the data. Understanding of this phenomenon can aid in interpretation of analysis. For seriously skewed data which are also heteroscedastic, no transformation can be expected to induce normality as required to validate the usual normal-theory prediction and calibration inference. In this case, alternate methods of determining appropriate critical values may be required. For data with nested error structure that can not be ignored, methods to determine how to routinely model and fit several heteroscedastic components are still under development. Penalties for ignoring "batch" effects for assay data in terms of subsequent inference such as estimation of minimum detectable concentration are unexplored.

When σ is not small, transformation will affect the distribution of the data, and the discussion of Section 3 may not be relevant. Whether or not σ is small, functional modeling and estimation of the variance structure can be a meaningful way to accomodate systematic heteroscedasticity. The most effective fitting techniques, as in CR (Ch. 3), are often not routinely applied in practice. When σ is small, transformation may be able to correct some, but not all of the "nonstandard" features of the data, so cannot be regarded as a "panacea" in this case.

Functional modeling of variance and formal estimation of variance and transformation parameters can be useful tools for understanding features of data.

REFERENCES

Bates, D.M., Wolf, D.A., and Watts, D.G. (1985), "Nonlinear Least Squares and First Order Kinetics," *Proc. Computer Science and Statistics: Seventeenth Symposium on the Interface* ed. D. Allen. North-Holland: New York.

Beal, S.L. and Sheiner, L.B. (1988), "Heteroscedastic Nonlinear Regression," *Technometrics*, 30, 327-338.

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

Carroll, R.J. and Cline, D.B.H. (1988), "An Asymptotic Theory for Weighted Least Squares With Weights Estimated by Replication," *Biometrika,* 75, p. 35-43.

Carroll, R.J. and Ruppert, D. (1984), "Power Transformations When Fitting Theoretical Models to Data," *Journal of the American Statistical Association,* 79, 321-328.

——————— (1987), "Diagnostics and Robust Estimation When Transforming the Regression Model and the Response." *Technometrics,* 29, 287-299.

——————— (1988), *Transformation and Weighting in Regression,* New York: Chapman Hall.

Davidian, M. and Carroll, R.J. (1987), "Variance Function Estimation," *Journal of the American Statistical Association,* 82, 1079-1091.

Davidian, M., Carroll, R.J., and Smith, W. (1988), "Variance Functions and the Minimum Detectable Concentration in Assays," *Biometrika* 75, 549-556.

Giltinan, D.M., Carroll, R.J., and Ruppert, D. (1986), "Some New Methods for Weighted Regression When There Are Possible Outliers," *Technometrics,* 28, 219-230.

Giltinan, D.M. and Ruppert, D. (1989), "Fitting Heteroscedastic Regression Models to Individual Pharmacokinetic Data Using Standard Statistical Software," *Journal of Pharmacokinetics and Biopharmaceutics,* to appear.

Hinkley, D.V. and Runger, G. (1984), "Analysis of Transformed Data," *Journal of the American Statistical Association,* 79, 302-308.

McCullagh, P. and Nelder, J.A. (1983), *Generalized Linear Models,* New York: Chapman and Hall.

Morton, R. (1987), "A Generalized Linear Model with Nested Strata of Extra-Poisson Variation," *Biometrika,* 74, 247-257.

Oppenheimer, L., Capizzi, T.P., Weppelman, R.M., and Mehto, H. (1983), "Determining the Lowest Limit of Reliable Assay Measurement," *Analytical Chemistry,* 55, 638-643.

Raab, G.M. (1981), "Estimation of a Variance Function, With Application to Radiomimmunoassay," *Applied Statistics,* 30, 32-40.

Rodbard, D. and Frazier, G.R. (1975), "Statisticial Analysis of Radioligand Assay Data," *Methods of Enzymology*, 37, 3-22.

Rudemo, M., Ruppert, D., and Streibig, J.C. (1989), "Random Effects Models in Nonlinear Regression With Applications to Bioassay," *Biometrics,* to appear.

Ruppert, D. and Aldershof, B. (1988), "Transformations to Symmetry and Heteroscedasticity," Preprint.

Table 1. *Assay data of Section 4.*

x	0.476	0.924	1.905	3.696	7.619	14.874	30.474[a]	59.134
y	0.05706	0.11781	0.25071	0.49596	1.03928	2.14635	4.24397	8.53848
	0.05700	0.11615	0.25398	0.48070	1.03659	2.09495	·	8.41333
	0.06363	0.12587	0.24552	0.49442	1.12641	2.24941	4.70110	9.01437
	0.05566	0.12308	0.24889	0.52321	1.10456	2.19937	4.44709	8.73544
	0.05449	0.11629	0.24858	0.49931	1.03184	2.16042	4.42707	8.33862
	0.06153	0.11878	0.24657	0.50210	1.02598	2.09198	4.39725	8.33347
	0.05837	0.11869	0.24212	0.48860	0.98267	2.07686	4.35511	8.35725
	0.05388	0.11886	0.25975	0.48158	1.04321	2.06961	4.37357	8.39123
	0.05618	0.12492	0.25311	0.48270	1.03838	2.12548	4.32040	8.39010

[a] The response for x = 30.474 for the second group of samples assayed is missing.

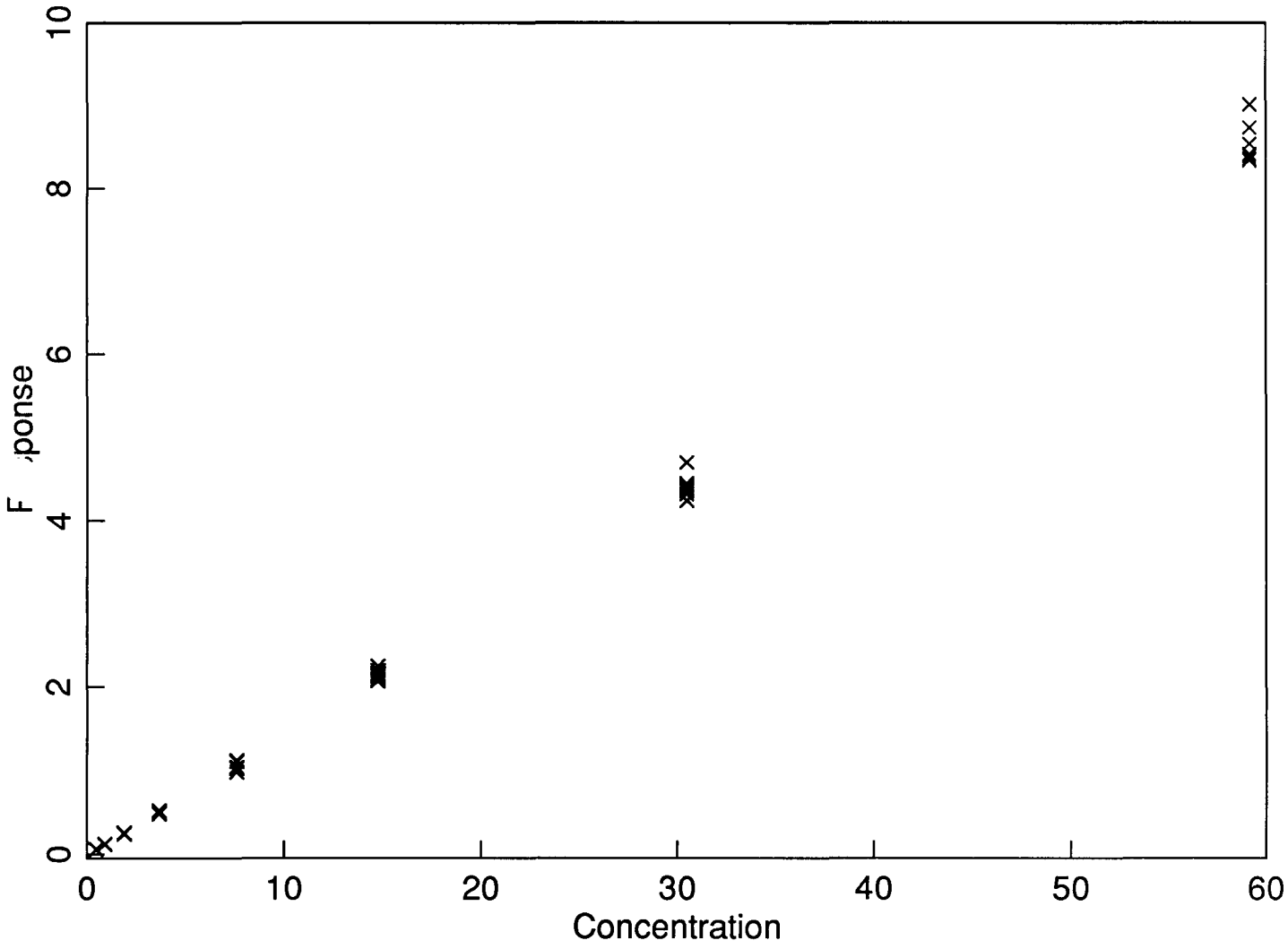

Figure 1. Plot of the Data of Section 4.

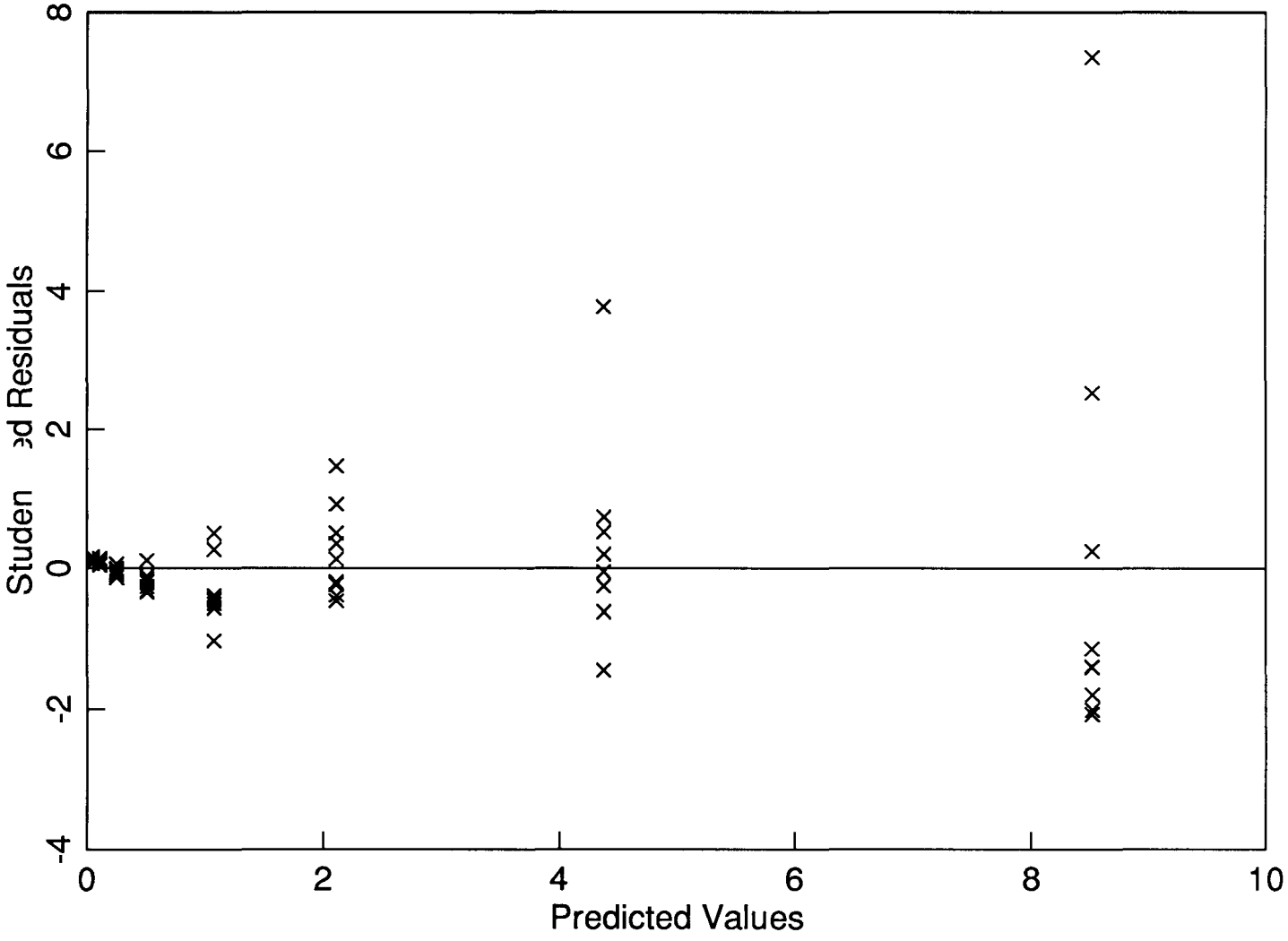

Figure 2. Studentized Residuals vs. Predicted Values for the Ordinary Least Squares Fit.

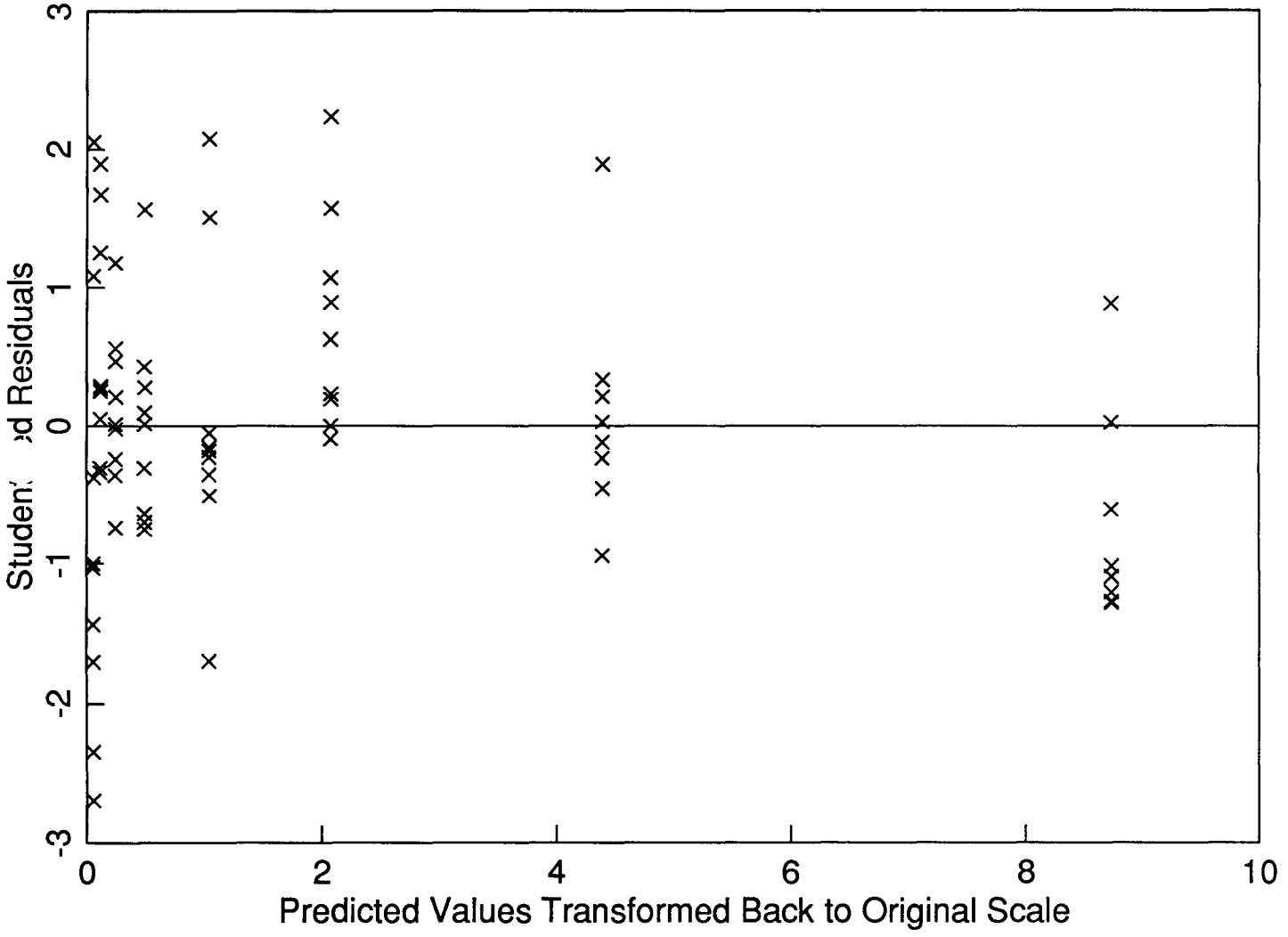

Figure 3. Studentized Residuals (on Transformed Scale) vs. Predicted Values ("Back-Transformed" to the Original Scale) for the Ordinary Least Squares Fit Based on log(y) and log(x).

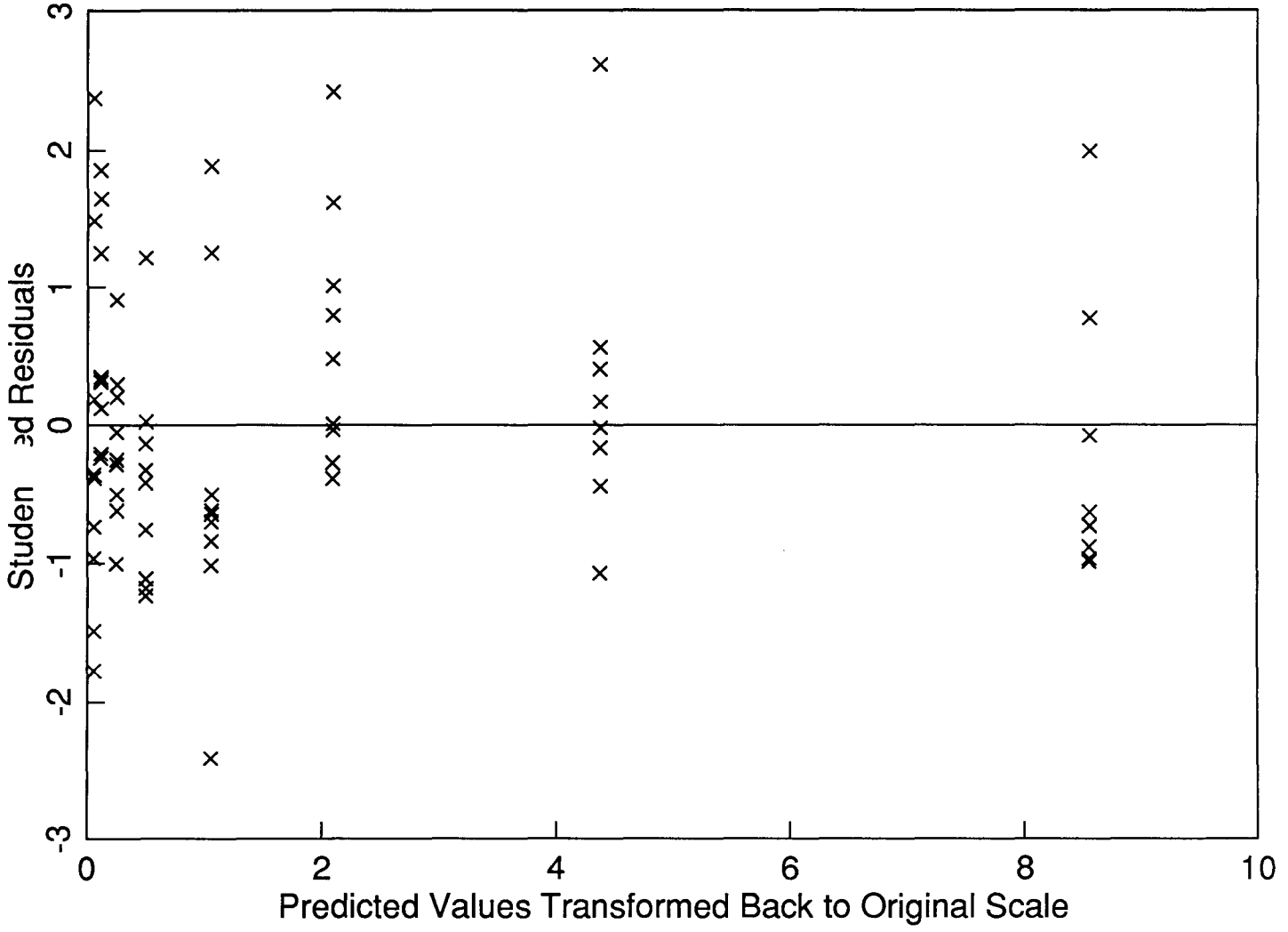

Figure 4. Studentized Residuals (on Transformed Scale) vs. Predicted Values ("Back-Transformed" to the Original Scale) for the Estimated Transformation Model (2.5).