SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853-7501

TECHNICAL REPORT NO. 760

November 1987

**FITTING MATHEMATICAL MODELS TO
BIOLOGICAL DATA: A REVIEW OF
RECENT DEVELOPMENTS***

by

David Ruppert

*Research has been supported by NSF grants DMS-8400602 and DMS-8701201.

Abstract

A common problem is to fit a theoretical model, $y = f(x, \beta) + \text{error}$, relating a measured response, y to a measured vector of predictors, x , and an unknown parameter vector, β . Ordinary least-squares is an appropriate fitting method when the variation of y about the model, i.e., the "error", is normally distributed with a conditional variance, given x , that is constant. In many biological problems, y is nonnegative, right-skewed, and has a conditional variance that is an increasing function of $f(x, \beta)$. In this paper, two methods for fitting models to such data are discussed. The methods are (1) transformation of the response and the model and (2) weighted least-squares. The rationale behind these methods, estimation techniques, and statistical inference (testing and confidence intervals) are discussed.

1. Introduction

This paper reviews recent research on fitting theoretical models to data. Though these methods are applicable outside biology, my own experiences applying these techniques have been primarily with biological data, and this paper will concentrate on that field.

By "theoretical model", I mean a model derived from biological assumptions using mathematical techniques. In contrast, an empirical model takes a convenient, flexible class of functions and fits them to the data. Parameters in theoretical models usually have biological interpretation; e.g., they are migration rates, growth rates, probability of infection, etc. The distinction between theoretical and empirical model will be relevant later when data transformations are discussed.

The theoretical model is assumed to be in the form

$$y_i = f(x_i, \beta), \quad (1.1)$$

where (y_i, x_i) , $i = 1, \dots, n$ are independent observations of a scalar response, y , and a vector, x , of independent variables. The model determines f , so f is a known function, but β is a vector of unknown biological or physical parameters. Equation (1.1) holds exactly only in the absence of error, but in reality there are many sources of random variability. For example, y_i and x_i may be measured with error, β may vary between observations, or y_i may be random with $f(x_i, \beta)$ being only its expectation or, perhaps, median. Also, $f(x_i, \beta)$ may be a simple approximation to the complex relationship between x_i and y_i .

Several statistical problems arise immediately. The foremost is to estimate β well. After that, one may wish to test the model, to predict a future y with a known x , or to control a future y by controlling x . This paper is primarily a review of recent research on parameter estimation, but prediction will also be touched upon.

This review will not attempt to be complete. Indeed, the topic of fitting mathematical models is sufficiently broad to cover a sizeable proportion of the entire statistics literature. Rather, I will

concentrate on a related set of statistical methodologies that have been the focus of much of my recent research. Much of this research is discussed much more fully in a book to be published (Carroll and Ruppert 1988), but this paper provides a shorter overview. Also, the later sections of this paper outline recent work not found in the book.

If the conditional distributional of y given x is known, then β can be estimated by maximum likelihood. Since such knowledge is rarely available, biologists traditionally have used nonlinear least-squares where β is estimated by $\hat{\beta}$ that minimizes

$$\sum_{i=1}^n [y_i - f(x_i, \hat{\beta})]^2. \quad (1.2)$$

Nonlinear least-squares is often said to be nonparametric since the method can be applied without knowing the probability distribution of the data. However, least-squares is efficient only if the "errors" defined by

$$\epsilon_i = y_i - f(x_i, \beta) \quad (1.3)$$

are normally distributed and homoscedastic (have a constant variance).

In practice, biological data are often right skewed with the conditional variance of y given x an increasing function of the conditional mean. An alternative to least-squares is robust regression, but robust regression estimators are designed for data where $\epsilon_1, \dots, \epsilon_n$ have a constant variance and are symmetrically distributed, though possibly with tails that are heavier than the normal distribution.

For right-skewed heteroscedastic data, one can postulate that y follows a distribution other than the normal distribution, say the gamma distribution. This is the approach of generalized linear models (McCullagh and Nelder 1983). This technique can be quite useful, especially if the assumed distribution has a "shape parameter" that can be estimated from the data.

In this paper, I will discuss another class of models, transformation and weighting models. These have proved to be effective methods for modeling skewed and heteroscedastic data.

By weighting, I mean estimating the conditional variance of y given x and weighting each observation by the reciprocal of its estimated variance. This technique is sometimes called generalized least-squares or weighted least-squares with estimated weights. Generalized least-squares is suitable for data that are nearly normally distributed but heteroscedastic.

Transformation models assume that a nonlinear function, $h(\cdot)$, of y is normally distributed with a constant variance. The function $h(\cdot)$ is assumed to be in a parametric class, e.g., the power transformation family. Transformation models are suitable for skewed data where the conditional variance of y is a function of its conditional mean.

Transformation/weighting models combine a transformation of y with generalized least-squares, and are flexible enough to model skewed data with the conditional variance of y a function of independent variables.

2. Transformations

The effects of a monotonically increasing transformation $h(\cdot)$ depend upon whether h is concave or convex—we will not be concerned with transformations that are neither.

If h is concave then its first derivative is decreasing. If h is applied to a random variable W , then h contracts the right tail of W while spreading the left tail. As a result $h(W)$ is less right skewed (or more left skewed) than W . Conversely, if h is convex then $h(W)$ is more right skewed than W . These statements are made mathematically precise by van Zwet (1964) in his important study of convex transformations of random variables. Let $\gamma(W)$ be the third-moment skewness coefficient, i.e.,

$$\gamma(W) = E[(W - \mu_W)/\sigma_W]^3, \quad (2.1)$$

where μ_W and σ_W are the expectation and standard deviation of W . Van Zwet proves that if h is convex then $\gamma(W) < \gamma(h(W))$.

Bartlett (1947) shows how transformations affect the relationship between the conditional variance and the conditional mean of y . Suppose that these quantities are related by

$$\text{Var}(y|x) = G[E(y|x)]$$

for some positive function G . Fix x and let $\mu = E(y|x)$. Linearizing h about μ , one has

$$\begin{aligned} \text{Var}(h(y)|x) &= E\{[h(y) - E(h(y)|x)]^2|x\} \\ &\doteq E\{[h(y) - h(\mu)]^2|x\} \\ &\doteq (\dot{h}(\mu))^2 E\{(y-\mu)^2|x\} \\ &= (\dot{h}(\mu))^2 G(\mu). \end{aligned} \tag{2.2}$$

Therefore, $\text{Var}(h(y)|x)$ is approximately constant if

$$\dot{h}(\mu) \propto \frac{1}{\sqrt{G(\mu)}}. \tag{2.3}$$

If G is increasing, then h will be concave.

Much biological data is non-negative, right-skewed, and is more variable when the mean is large. It is fortunate that concave transformation reduce both right skewness and this type of heteroscedasticity.

Power transformations are widely used in statistical analysis. They are simple and flexible. The logarithm can be naturally embedded in the power transformation family (Box and Cox 1964). Define

$$\begin{aligned} y^{(\lambda)} &= (y^\lambda - 1)/\lambda \text{ if } \lambda \neq 0 \\ &= \log(y) \text{ if } \lambda = 0. \end{aligned} \tag{2.4}$$

I will call $y^{(\lambda)}$ the modified power transformation. Notice that $y^{(\lambda)}$ is convex when $\lambda > 1$ and

concave when $\lambda < 1$. Notice that $y^{(\lambda)}$ is a smooth function of λ even at $\lambda = 0$ and

$$(\partial/\partial\lambda)y^{(\lambda)} = y^{\lambda-1} \text{ for all } \lambda.$$

3. Transformations and Theoretical Models

When the theoretical model (1.1) is available, a transformation of y alone is disadvantageous since it can destroy the model. However, the model can be preserved by transforming y and $f(x, \beta)$ in the same way. Carroll and Ruppert (1984) propose the following model for fitting theoretical models to data. Let $\{h_\lambda(\cdot)\}$ be a family of monotonic transformations, e.g., the modified power transformations (2.4). The transformation parameter λ may be scalar or a vector and varies over some parameter space. Carroll and Ruppert's (1984) "transform-both-sides" (TBS) model is

$$h_\lambda(y_i) = h_\lambda(f(x_i, \beta)) + \sigma\epsilon_i, \quad (3.1)$$

where $\epsilon_1, \dots, \epsilon_n$ are independent, standard normal random variables. By estimating λ , we use the data to choose the transformation to normality and homoscedasticity.

It should be emphasized that the purpose of transformation is to induce the proper error structure for least-squares estimation. Simultaneous transformations of y and $f(x, \beta)$ have been used for other purposes, particularly to linearize the model $f(x, \beta)$. However, the transformation that linearizes may exasperate, rather than reduce, skewness and heteroscedasticity. Box and Hill (1974) give an example where the linearizing transformation induces such marked heteroscedasticity that least-squares estimation after linearization gives unacceptable estimates. Box and Hill (1974) correct the heteroscedasticity by weighted least-squares, and Carroll and Ruppert (1984) analyze the same data by estimating the transformation to homoscedastic errors. Since nonlinear least-squares software is now widely available, transforming models solely to linearize them should now be considered poor statistical practice.

Let $\theta = (\beta, \lambda, \sigma)$. Carroll and Ruppert (1984) estimate of θ by maximum likelihood. The log-likelihood is, up to an additive constant,

$$L(\theta) = -n \log(\sigma) - \sum_{i=1}^n \{h_{\lambda}(y_i) - h_{\lambda}(f(x_i, \beta))\}^2 / 2\sigma^2 \quad (3.2)$$

$$+ \sum_{i=1}^n \log(J_i(\lambda))$$

where $J_i(\lambda)$ is the Jacobian of the transformation from y_i to $h_{\lambda}(y_i)$. For general transformation families there are several simple methods for maximizing (3.2). The first is to apply the Newton-Raphson procedure or the Fisher method of scoring directly to (3.2). This, of course, requires special software but such software is readily available, say in IMSL or GAUSS. I have experienced nonconvergence difficulties when attempting to minimize (3.2), and I recommend first eliminating σ^2 by maximizing $L(\theta)$ over σ^2 . For any fixed β and λ , $L(\theta)$ is maximized in σ^2 by

$$\hat{\sigma}^2(\beta, \lambda) = n^{-1} \sum_{j=1}^n \{h_{\lambda}(y_i) - h_{\lambda}(f(x_i, \beta))\}^2,$$

and the maximized value of $L(\theta)$ is

$$L_{\max}(\beta, \lambda) = -\frac{n}{2} \log(\hat{\sigma}^2(\beta, \lambda)) - \frac{n}{2} + \sum_{i=1}^n \log(J_i(\lambda)). \quad (3.3)$$

Note that

$$\sum_{i=1}^n \log(J_i(\lambda)) = -\frac{n}{2} \log(1/J(\lambda)^2) \quad (3.4)$$

where $J(\lambda)$ is the geometric mean of $J_i(\lambda)$, $i = 1, \dots, n$. Substituting (3.4) into (3.3), we have

$$L_{\max}(\beta, \lambda) = -\frac{n}{2} \log \left[\left(\frac{\hat{\sigma}(\beta, \lambda)}{J(\lambda)} \right)^2 \right] - \frac{n}{2}. \quad (3.5)$$

I have only rarely had difficulties maximizing $L_{\max}(\beta, \lambda)$ using MAXLIK on the GAUSS package for PCs, and these rare instances were always with a model that fit poorly or was over-parameterized.

From (3.5) we see that for each fixed λ , $L_{\max}(\beta)$ is maximized in β by minimizing $\hat{\sigma}^2(\beta, \lambda)$, that is, by the least-squares estimate of β , say $\hat{\beta}(\lambda)$. Let

$$L_{\max}(\lambda) = L_{\max}(\hat{\beta}(\lambda), \lambda). \quad (3.6)$$

Computing $L_{\max}(\lambda)$ requires only nonlinear least-squares software. If only such software is available, then $L_{\max}(\lambda)$ can be computed for all λ on some grid and then $\hat{\lambda}$ can be taken as the value where $L_{\max}(\lambda)$ is maximized on this grid. This method is simple but time consuming.

If $h_\lambda(y)$ is $y^{(\lambda)}$, the modified power transformation, then there is a simple method for maximizing $L_{\max}(\beta, \lambda)$ simultaneously over β and λ by nonlinear least-squares. By (3.5), the MLE minimizes the sum of squares

$$\left[\frac{\hat{\sigma}(\beta, \lambda)}{\hat{J}(\lambda)} \right]^2 = \sum_{i=1}^n \left\{ \frac{[h_\lambda(y_i) - h_\lambda(f(x_i, \beta))]}{\hat{y}^{\lambda-1}} \right\}^2$$

where $\hat{y}$ is the geometric mean of $y_1, \dots, y_n$; see Carroll and Ruppert (1988, section 4.3) or Giltinan and Ruppert (1987) for details.

The asymptotic covariance matrix, $V_{\hat{\theta}}$, of $\hat{\theta} = (\hat{\beta}, \hat{\lambda}, \hat{\sigma})$ can be estimated by inverting the Hessian of $-L(\theta)$, also called the observed Fisher matrix. That is,

$$V_{\hat{\theta}} = -\{\nabla^2 L(\theta)\}^{-1}.$$

By a theorem of Patefield (1977), the covariance matrix of $(\hat{\beta}, \hat{\lambda})$ is $V_{(\hat{\beta}, \hat{\lambda})} = -\{\nabla^2 L_{\max}(\beta, \lambda)\}^{-1}$. This is convenient when the Fisher method of scoring is applied to (3.3). Large-sample confidence

intervals and tests can be constructed from $V_{\hat{\theta}}$ or $V_{(\hat{\beta}, \hat{\lambda})}$ by well-known methods that are described in Carroll and Ruppert (1988, section 4.3). For accurate tests and confidence intervals with moderate to small samples, Efron's (1979, 1982) bootstrap is preferred to inference based on $V_{\hat{\theta}}$.

Model (3.1) can be reexpressed as a model for y rather than $h_{\lambda}(y)$. Let $g_{\lambda}(y)$ be the inverse of $h_{\lambda}(y)$ so that $g_{\lambda}[h_{\lambda}(y)] = h_{\lambda}[g_{\lambda}(y)] = y$. Then

$$y_i = g_{\lambda}[h_{\lambda}(f(x_i, \beta)) + \sigma \epsilon_i]. \quad (3.7)$$

Equation (3.7) models the skewness and heteroscedasticity of y_i as arising from a nonlinear transformation of ϵ_i . It follows from (3.7) that the conditional p -th quantile of y_i , given x_i , is

$$q_p(y_i | x_i) = g_{\lambda}[h_{\lambda}(f(x_i, \beta)) + \sigma \Phi^{-1}(1-\alpha)]. \quad (3.8)$$

Since $\Phi^{-1}(1/2) = 0$, (3.8) shows that the conditional median of y_i given x_i is $g_{\lambda}[h_{\lambda}(f(x_i, \beta))] = f(x_i, \beta)$. Because its y_i is skewed, the conditional expectation of y_i will not equal $f(x_i, \beta)$ exactly. Model (3.8) combines the theoretical model $f(x_i, \beta)$ for the median of y_i with an empirical model for the error structure. I say empirical model because in most applications, the family $\{h_{\lambda}\}$ is not derived from biological assumptions and λ and ϵ_i do not have biological interpretations.

$q_p(y_i | x_i)$ can be estimated by substituting estimates of β , λ , and σ into (3.8). By plotting the estimated quartiles (first, third, and median) against x , one can see the heteroscedasticity and skewness of the response. An overlay of the same plot based on the nonlinear regression model will show how far predictions differ between the TBS and the usual regression models. The difference can be substantial; see Carroll and Ruppert (1988, section 4.4) for an example. Estimation of the conditional expectation of y_i is slightly more complicated, but still easy; see Carroll and Ruppert (1988, section 4.4) for a full discussion of inference about y .

Gallagher (1986) has extensively studied the application of TBS to water quality models. He presents the analyses of several real data sets as well as a Monte Carlo study.

Although I have emphasized the transform—both—sides method as a technique for theoretically derived models, Snee (1986) gives a number of interesting examples where transform—both—sides is applied to empirical models.

4. Generalized Least—Squares

If y_i is normally distributed but with a nonconstant variance, then the following model is appropriate:

$$y_i = f(x_i, \beta) + \sigma g(z_i, \beta, \xi) \epsilon_i, \quad (4.1)$$

where $\epsilon_1, \dots, \epsilon_n$ are independent standard normal random variates, g is a known "variance function", z_i is a vector of independent variables possibly equal to x_i or a subvector of x_i , and ξ is an unknown parameter vector. Examples of $g(z_i, \beta, \xi)$ are

$$g(z_i, \beta, \xi) = z_i^\xi, \quad (4.2)$$

where z_i is a scalar and g does not depend on β , and

$$g(z_i, \beta, \xi) = f(x_i, \beta)^\xi \quad (4.3)$$

where $z_i = x_i$ and the conditional variance of y_i is proportional to a power of the mean.

Now let $\theta = (\beta, \xi, \sigma)$. Up to an additive constant, the log—likelihood is

$$L(\theta) = -n \log(\sigma) - \sum_{i=1}^n \log g(z_i, \beta, \xi) - \frac{1}{2} \sum_{i=1}^n \left[\frac{y_i - f(x_i, \beta)}{\sigma g(z_i, \beta, \xi)} \right]^2. \quad (4.4)$$

$L(\theta)$ can be maximized jointly in β , ξ , and σ . Sheiner and Beal (1985) call the MLE "extended least—squares" because, like least—squares, it is consistent even for data that are not normally distributed. The MLE is not equivalent to generalized least—squares. The GLS estimator has certain robustness properties that the MLE lacks (Carroll and Ruppert 1982), but the MLE is more

efficient than GLS when the data are *exactly* normal and the conditional variance of y_i is given *exactly* by $g(z_i, \beta, \xi)$ (Jobson and Fuller 1980).

The GLS can be computed by the following algorithm: (1) Compute the unweighted LS estimate $\hat{\beta}_{(0)}$. (2) Fix β equal to $\hat{\beta}_{(0)}$ in (4.4) and maximize over σ and ξ ; call the maximizers $\hat{\sigma}$ and $\hat{\xi}$; (3) Using $\{g(z_i, \hat{\beta}_{(0)}, \hat{\xi})^{-2}\}$ as weights, compute the weighted least-squares estimator of β and call it $\hat{\beta}$; (4) Let $\hat{\beta}_{(0)} = \hat{\beta}$ and go back to (2). Iterate until convergence. The details of this algorithm are discussed in Giltinan and Ruppert (1987), where sample SAS programs are given.

The theory of estimating variance functions is given by Davidian and Carroll (1987). Full discussion of the available methods and several examples appear in Carroll and Ruppert (1988, chapter 3).

5. Transformation/Weighting Models

The TBS method was applied to stock-recruitment data in Carroll and Ruppert (1984). Later (Ruppert and Carroll 1985) it was realized that a certain transformation/weighting model contains as special cases several standard methods for fitting the Beverton-Holt stock-recruitment model. Interestingly, this model is the same equation as the Michaelis-Menten model of enzyme kinetics. This served as motivation for a general transformation/weighting model

$$h_\lambda(y_i) = h_\lambda(f(x_i, \beta)) + \sigma g(z_i, \beta, \xi) \epsilon_i, \quad (5.1)$$

where $\epsilon_1, \dots, \epsilon_n$ are independent standard normal random variables. According to model (5.1), y_i can be transformed to normally distributed errors by $h_\lambda(\cdot)$, but the errors are heteroscedastic with conditional variance $\sigma^2 g^2(z_i, \beta, \xi)$. Model (5.1) is extremely flexible. It separates the mean, or more precisely the median, structure from distributional shape and the variance structure: The conditional median of y_i is given by $f(x_i, \beta)$, distributional shape is specified by $h_\lambda(\cdot)$, and $g^2(z_i, \beta, \xi)$ specifies the variance after transformation.

In the Michaelis–Menten or Beverton–Holt model x is scalar, $\beta = (\beta_1, \beta_2)$, and

$$f(x, \beta) = \frac{1}{\beta_1 + \beta_2/x} = \frac{x}{\beta_2 + \beta_1 x}. \quad (5.2)$$

Various fitting methods that have appeared in the literature are described in Carroll, Cressie, and Ruppert (1987). Each is a special case of model (5.1) with $h_\lambda(y) = y^{(\lambda)}$, $z_i = x_i$, and

$$h(z_i, \beta, \xi) = x_i^\xi \quad (5.3)$$

for some ξ . The different methods are distinguished simply by different values of λ and ξ .

There has been considerable controversy on how to best fit the Michaelis–Menten model. Some studies (see Currie, 1982) simulate a single, or only a few, error structures. (An error structure is a single value of (λ, ξ) .) Not surprising, the method which is the MLE for the simulated error structure shows up as best in the study. What is surprising is that the author concludes that this method is preferred in practice, although no evidence is given that the error structure used in the simulations actually obtains in practice!

Using (5.1)/(5.3), Carroll, Cressie, and Ruppert (1987) show that error structures vary widely, not only between subject matters but even between similar experiments from the same laboratory. Model (5.1)/(5.3) is effective for identifying error structure and results in an efficient estimate of β . Carroll, Cressie, and Ruppert (1987) give an example showing that the TBS model can estimate β with considerably more accuracy than methods currently in use.

The log–likelihood under (5.1) is

$$L(\beta, \lambda, \xi, \sigma) = -n \log \sigma - \sum_{i=1}^n \log g(z_i, \beta, \xi) \quad (5.4)$$

$$- \frac{1}{2} \sum_{i=1}^n \left[\frac{h_\lambda(y_i) - h_\lambda(f(x_i, \beta))}{\sigma g(z_i, \beta, \xi)} \right]^2 + \sum_{i=1}^n \log(J_i(\lambda)).$$

The log-likelihood can be maximized by methods similar to those in sections 3 and 4. Inferential procedures for model (5.1) are discussed in Carroll and Ruppert (1988, chapter 5).

Kettl (1987) has studied model (5.1) with power transformations and power of the mean (POM) weighting, that is, with $x_i = z_i$ and

$$g(x_i, \beta, \xi) = f(x_i, \beta)^\xi. \quad (5.5)$$

Since a power transformation can remove heteroscedasticity precisely of the form (5.5), the TBS model with POM weighting appears redundant at first. However, combining a power transformation with power of the mean weighting has the advantage of allowing the transformation parameter to be chosen primarily to remove skewness. Often the transformation parameter inducing symmetry is substantially different from the one inducing a constant variance. Then power transformation combined with POM weighting can result in a significantly better fit than either alone, particularly for larger data sets; see Rudemo et al. (1987) for an example with 150 observations. Another advantage of a model combining a power transformation with POM weighting is that a context is provided for testing whether power transformation alone fits better than POM weighting alone, or vice versa.

Rudemo et al. (1987) have introduced two models that derive the function $g(z_i, \beta, \xi)$ from a random-effects model for x . Since their two models are similar to each other, we will only describe the first, which uses Berkson's (1950) "controlled variate" model.

Suppose $x = z$ is univariate and that x is the "target value" of the independent variable, while the actual value of the independent variable is $\tilde{x} = x + \sigma_X \delta$. Here δ is a random variable with mean 0 and variance 1, so that σ_X^2 is the conditional variance of $\tilde{x}$ given x . An example would be bioassay where x is intended dose and $\tilde{x}$ is actual dose that varies from x because of measurement errors. Only x , not $\tilde{x}$, is known. Assume further that

$$\sigma_X = \sigma_1 x^{1-\Delta} \quad (5.6)$$

for some σ_1 and Δ , and that the TBS model

$$h_\lambda(y) = h_\lambda(f(\tilde{x}, \beta)) + \sigma\epsilon \quad (5.7)$$

holds with the independent variable equal to its actual value, $\tilde{x}$. We, of course, do not know $\tilde{x}$ and fit the model with the independent variable equal to x . This introduces another source of error that we will now study. Let $\dot{h}_\lambda(y) = (d/dy)h_\lambda(y)$. Using a Taylor approximation, we have

$$h_\lambda(y) = h_\lambda(f(x, \beta)) + [\dot{h}_\lambda(f(x, \beta))][\frac{\partial}{\partial x} f(x, \beta)]\sigma_X\delta + \sigma\epsilon. \quad (5.8)$$

From (5.6) and (5.8), Rudemo et al. (1987) derive

$$g^2(x, \beta, \Delta) = 1 + Ax^{2-2\Delta}[\dot{h}_\lambda(f(x, \beta)) \frac{\partial}{\partial x} f(x, \beta)]^2$$

where $A = (\sigma_1/\sigma)^2$. The parameters in such theoretical variance models will often have biological interpretations and may allow comparisons between the amounts of variability from different sources (δ versus ϵ). Such comparisons could lead to important biological insight or, at least, may be relevant for future experimental designs.

6. Other Developments

Until now, only maximum likelihood estimation of β and of the transformation and variance parameters λ and ξ has been mentioned. There are several reasons why other estimators might be used. First, MLEs are notoriously sensitive to bad data or slight violations of the model assumptions. For example, a single gross error, say due to a measurement or recording error, can ruin the least-squares estimate of a regression parameter, and this estimator is, of course, the MLE for normal errors. Moreover, there is no guarantee that within a given transformation family there exists a transformation to homoscedastic, normal errors. It might be better to search for a transformation to errors that are either symmetric or homoscedastic, but not necessarily both.

Carroll and Ruppert (1987) introduce a robust estimator for the TBS model. Their estimator is essentially a weighted maximum likelihood estimator. The weights depend on the data in such a way that observations that are outlying or have an unusually large influence on the MLE are automatically downweighted. Giltinan, Carroll, and Ruppert (1986) introduce similar estimators for weighting models of form (4.1).

Ruppert and Aldershof (1987) extend the transform-to-symmetry estimators of Hinkley (1975) and Taylor (1985) to the transform-both-sides model. Such estimators only attempt to remove skewness, not heteroscedasticity. Ruppert and Aldershof also define estimators that attempt to remove only heteroscedasticity, not skewness. These are defined through a measure of correlation between the squared residuals and the estimated mean responses. Using both these "skewness" and "heteroscedasticity" estimators, they develop a test of the null hypothesis that a transformation exists to both symmetric and homoscedastic errors. Finally, in the case that this null hypothesis is true, they show how to optimally combine the skewness and heteroscedasticity estimators.

Acknowledgment

My research has been supported by NSF grants DMS-8400602 and DMS-8701201.

References

- Bartlett, M.S. (1947). The use of transformations. *Biometrics*, 3, 39–52.
- Berkson, J. (1950). Are there two regressions? *Journal of the American Statistical Association*, 45, 164–180.
- Box, G.E.P. and Cox, D.R. (1964). An analysis of transformations. *J. of the Royal Statist. Soc., B*, 26, 211–246.
- Box, G.E.P. and Hill, W.J. (1974). Correcting inhomogeneity of variance with power transformation weighting. *Technometrics*, 16, 359–389.
- Carroll, R.J. and Ruppert, D. (1982). A comparison between maximum likelihood and generalized least-squares in a heteroscedastic linear model. *J. of the Amer. Statist. Assoc.*, 77, 878–882.
- Carroll, R.J. and Ruppert, D. (1984). Power transformations when fitting theoretical models to data. *J. of the Amer. Statist. Assoc.*, 79, 321–328.
- Carroll, R.J. and Ruppert, D. (1987). Diagnostics and robust estimation when transforming the regression model and the response. *Technometrics*, 29, 287–299.
- Carroll, R.J. and Ruppert, D. (1988). *Transformation and Weighting in Regression*. Chapman and Hall (to appear).
- Carroll, R.J., Cressie, N.A.C., and Ruppert, D. (1987). A transformation/weighting model for estimating Michaelis–Menten parameters. Statistical Laboratory Preprint Series, Preprint #87–20. Iowa State University.
- Currie, D.J. (1982). Estimating Michaelis–Menten parameters: bias, variance and experimental design. *Biometrics*, 38, 907–919.
- Davidian, M. and Carroll, R.J. (1987). Variance function estimation. *J. of the Amer. Statist. Assoc.* (to appear).
- Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Stat.* 7, 1–26.
- Efron, B. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans*. CBMS–NSF Monograph 39. SIAM, Philadelphia.
- Gallagher, D. (1986). *The Application of Data Based Transformations to Parameter Estimation in Water Quality Models*. Ph.D. dissertation. Department of Environmental Sciences and Engineering. School of Public Health. University of North Carolina at Chapel Hill.
- Giltinan, D., Carroll, R.J., and Ruppert, D. (1986). Some new methods for weighted regression when there are possible outliers. *Technometrics*, 28, 219–230.
- Giltinan, D. and Ruppert, D. (1987). Fitting Heteroscedastic Regression Models to Individual Pharmacokinetic Data Using Standard Statistical Software. Technical Report No. 759. School of Operations Research and Industrial Engineering, Cornell University.

- Hinkley, D.V. (1975). On power transformations to symmetry. *Biometrika*, 62, 101–111.
- Jobson, F.D. and Fuller, W.A. (1980). Least-squares estimation when the covariance matrix and the parameter vector are functionally related. *J. of the Amer. Statist. Assoc.*, 75, 176–181.
- Kettl, E. (1987). Some Applications of the Transform–Both–Sides Regression Model. Ph.D. Dissertation, Dept. of Statistics, Univ. of North Carolina at Chapel Hill.
- McCullagh, P. and Nelder, J.A. (1983). *Generalized Linear Models*. Chapman and Hall: New York.
- Patefield, W.M. (1977). On the maximized likelihood function. *Sankhya*, B, 39, 92–96.
- Rudemo, M., Ruppert, D., and Streibig (1987). Random effects models in nonlinear regression with applications to bioassay. Mimeo Series #1727. Dept. of Statistics, Univ. of North Carolina at Chapel Hill.
- Ruppert, D. and Aldershof, B. (1987). Transformations to symmetry and homoscedasticity. Manuscript.
- Ruppert, D. and Carroll, R.J. (1985). Data transformations in regression analysis with applications to stock recruitment relationships. In *Resource Management: Lecture Notes in Biomathematics* 61, M. Mangel, editor. Springer Verlag, New York.
- Sheiner, L.B. and Beal, S.L. (1985). Pharmacokinetic parameter estimates from several least-squares procedures: superiority of extended least-squares. *J. Pharmacokin. Biopharm.* 13, 185–201.
- Snee, R.D. (1986). An alternative approach to fitting models when reexpression of the response is useful. *J. of Quality Technology*, 18, 211–225.
- Taylor, J.M.G. (1985). Power transformations to symmetry. *Biometrika*, 72, 145–152.
- van Zwet, W.R. (1964). *Convex Transformations of Random Variables*. Mathematisch Centrum, Amsterdam.