CALIBRATION AND REGRESSION WITH NON-CONSTANT ERROR VARIANCE

NO. 1978 July 1990

Marie Davidian and Perry D. Haaland

Calibration and Regression with Nonconstant Error Variance

MARIE DAVIDIAN
Department of Statistics
Box 8203, North Carolina State University
Raleigh, NC 27695-8203

PERRY D. HAALAND
Becton Dickinson Research Center
P.O. Box 12016
Research Triangle Park, NC 27709-2016

CONTENTS

ABSTRACT

Ordinary least squares regression analysis is generally inappropriate for calibration and regression problems when the usual assumption of constant variance across all observations doesn't hold. Estimators of regression parameters are of relatively poor quality and the resulting inference can be misleading. The use of standard data transformations is a common alternative but may not provide enough flexibility for some cases. The use of weighted regression with weights estimated from replicates is generally unreliable for reasonable sample sizes. However, when the error variance changes systematically with the mean response or other variables, generalized least squares (GLS) and variance function estimation (VFE) methods can be used. The GLS-VFE approach allows the experimenter to specify a model for the systematic change in variance, estimate unknown parameters, and to use this information to provide more efficient estimates of the regression parameters. In this tutorial, GLS-VFE methods are introduced and described in the context of calibration and regression. An example of calibration for a chemical assay is used to motivate discussion and illustrate the implementation of these methods using standard software packages.

1 INTRODUCTION

In many areas of application in physical and biological science, such as chemical kinetics, clinical and analytical chemistry, biochemical and chemical assay development, enzyme kinetics, and pharmacokinetics, regression analysis is a common tool for characterizing the relationship between a response Y and a set of predictor variables **x**. A model for the mean relationship is postulated, parameters of the model are estimated, and the fitted model is often used for prediction of future observations and calibration.

The "textbook" (e.g., [1]) fitting method is that of unweighted least squares (LS), which is based on assumptions that observations on Y are independent, follow a normal distribution, and have the same variance regardless of the level of the response or settings of the predictor variables. A formal model for this situation can be written as

$$Y_{ij} = f(\mathbf{x}_i,\beta) + \sigma\,\varepsilon_{ij},\ i = 1, ..., N,\ j = 1, ..., m_i \geq 1,\ n = \sum_{i=1}^{N} m_i \qquad (1)$$

where Y_{ij} is the response for the jth replicate at the ith setting of the (k×1) vector of predictor variables $\{\mathbf{x}_i\}$ and $f(\mathbf{x}_i,\beta)$ is the regression or mean function, which we suppose to be possibly nonlinear in the elements of the (p×1) vector β of regression parameters. Throughout this article, we will often use the notation $\mu_i = f(\mathbf{x}_i,\beta)$.

In the context of the model, the classical assumptions are that $\{Y_{ij}\}$ are independent, and the independent random errors $\{\varepsilon_{ij}\}$ have a standard normal distribution with mean $E(\varepsilon_{ij}) = 0$ and variance $Var(\varepsilon_{ij}) = 1$. In (1), the term $\sigma\,\varepsilon_{ij}$ describes the nature of the experimental error associated with the response, and implies the classical assumption of constant variance, or homoscedasticity, for all $\{Y_{ij}\}$. Under these assumptions, $E(Y_{ij}) = f(\mathbf{x}_i,\beta)$, $Var(Y_{ij}) = \sigma^2$, the response is normally-distributed, and the LS estimate of the regression parameter β is the value $\hat{\beta}_{LS}$ minimizing

$$\sum_{i=1}^{N}\sum_{j=1}^{m_i} \{Y_{ij} - f(\mathbf{x}_i, \beta)\}^2. \qquad (2)$$

The results cited in the remainder of this article do not depend on the assumption of normality of the response unless explicitly mentioned. However, the assumption of independence is assumed to hold throughout.

Often, $\{\mathbf{x}_i\}$ are scalars, and the single predictor may be substrate concentration, mass, temperature, etc. The $\{Y_{ij}\}$ may be peak area, velocity, radioactive count, or some physical property. Common regression functions are

- the simple linear regression model:

$$f(x,\beta) = \beta_0 + \beta_1 x,$$

- compartment models such as the biexponential model:

$$f(x,\beta) = \beta_1 e^{-\beta_2 x} + \beta_3 e^{-\beta_4 x},$$

- logistic models such as the sigmoid:

$$f(x,\beta) = \beta_1 - (\beta_2 - \beta_1)/\{1 + \exp\{\beta_3(\log x - \beta_4)\}\}, \text{ and}$$

- the Michaelis-Menten relation:

$$f(x,\beta) = \beta_0 x/(\beta_1 + x).$$

Examples in physical science indicate that the constant variance assumption of (1) may often be inappropriate [2], [3], [4]. For example, Figure 1 is a scatterplot of data from a calibration curve which relates a reading from an HPLC assay to the blood concentration (ng/ml) of a drug. The data are given in Table 1. As a first approximation, a reasonable regression model for the mean is $f(x,\beta) = \beta_0 + \beta_1 x$. However, variability increases systematically with response level so the usual assumption of constant variance across the response range is not satisfied. Also note that there is a very large value of the response at the highest concentration level. No physical justification could be given for excluding this value from the analysis. In the subsequent analysis of the data set, we will use a method that accommodates the influence of this point on both the choice of model and the quality of the estimated parameters.

In order to illustrate the effects of incorrectly assuming constant variance, LS was used to obtain estimates $\hat{\beta}_0$=0.0033, $\hat{\beta}_1$=0.0014, and $\hat{\sigma}$=0.0265. This fit is superimposed on Figure 1. To gain insight into the nature of the variability in the data, we can examine the relationship between the predicted values $\hat{\mu}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ and the residuals $r_{ij} = Y_{ij} - \hat{\mu}_i$. Model (1) implies that $(Y_{ij} - \mu_i)/\sigma = \varepsilon_{ij}$, thus a plot of standardized residuals $r_{ij}/\hat{\sigma}$ versus predicted values should show variability within a band of constant width if the model is appropriate. An advantage of using standardized residuals is that the vertical scale is in units of the assumed standard deviation of the data. Note that in Figure 2 the "fan-shaped" pattern shows that experimental error increases with mean response so the constant variance assumption in (1) is inappropriate. Another problem with the LS fit is that there is not very good agreement between $\hat{\beta}_0$=0.0033 which should be estimating the mean response at zero dose and the observed mean of the corresponding 13 replicates, 0.0021. This shows that a consequence of ignoring the nonconstant variance in this example is a poorly fitting mean model.

The general phenomenon of nonconstant variance is known as heteroscedasticity. One approach to regression with such data is to ignore the heteroscedasticity and undertake estimation and inference as if (1) were correct. A second common approach is to transform the response so that the variance of transformed response appears constant and then proceed as usual; for example, one might model and fit the relationship between log Y_{ij} and log x_i for a single predictor. As will be shown, the first approach can lead to misleading conclusions based on the fitted curve. The second approach is reasonable, but the analyst may overlook the useful information about the process being studied that is implied by the nonconstant variability of the response on the original scale. In addition, choice of a standard transformation may not provide the desired correction.

A scatter plot of the data from the example after using the log transformation of the response is shown in Figure 3. (Due to zero doses, we cannot use the transformation log x_i.) This transformation obviously affects the pattern of variability. However, it seems to overcorrect for nonconstant variance at small concentrations, and it does not preserve the apparent linear relationship. As an alternative, Figure 4 shows a scatterplot of the data after a square root transformation of the response. The square root transformation does not seem to go far enough in correcting for nonconstant variance at low concentrations although it is better at preserving the apparent linear relationship. It is clear from these plots that a simplistic approach to the data based on standard transformations is not satisfactory. However, there are more sophisticated methods using transformations that might be appropriate for this data set [5].

In this article, we describe an approach to dealing with nonconstant variance which does not involve transformations. This approach incorporates knowledge about the form of the response variance into modeling and fitting and has the additional advantage of permitting reasoning on the original scale. To begin our formal discussion, suppose that experimental error is proportional to known constants w_i^{-1}. (For example is some chemistry applications, there may be some theoretical justification for the assumption that $w_i^{-1} = x_i$ or x_i^2 where x_i is a scalar.) Then a modification of (1) more accurately reflecting the character of the response is

$$Y_{ij} = f(\mathbf{x}_i,\beta) + \sigma w_i^{-1/2}\varepsilon_{ij},\ i = 1, ..., N,\ j = 1,..., m_i \qquad (3)$$

so that the assumption is that $\mathrm{Var}(Y_{ij}) = \sigma^2 w_i^{-1}$.

The LS estimate given by (2) assigns equal weight to all observations in determining the fit. However, if variance is nonconstant, the quality of information about the response in regions where variance is large is inferior to that where variance is small. In this case, it makes more sense to give those responses with small associated variability relatively more influence in determining the fit.

In terms of (3), a fitting procedure reflecting this is weighted least squares (WLS), in which β is estimated by minimizing

$$\sum_{i=1}^{N} \sum_{j=1}^{m_i} w_i \{ Y_{ij} - f(x_i, \beta) \}^2. \tag{4}$$

The $\{w_i\}$ are called weights.

In the remainder of the paper we describe methods that are applicable when the form of the variability is more complex and the w_i's are unknown. In Section 2 we introduce the idea of modeling the variance as a function of the mean response and other parameters. In Sections 3 and 4 we introduce the use of generalized least squares and variance function estimation for fitting these more complex models for variability. In Section 5 we consider the problem of prediction and the consequences of incorrect modeling of the variance. We have already introduced an example of data from a chemical assay and we will be referring to this data throughout to illustrate the methods being presented.

2 VARIANCE FUNCTION MODELS

When the weights $\{w_i\}$ are unknown, one simple approach is to estimate them using the data at hand; namely, if $m_i>1$ at all x_i, calculate the sample variance s_i^2 of the responses at x_i, and emulate (4) by replacing w_i by s_i^{-2}. This approach can be unreliable if the number of observations, $\{m_i\}$, at each x_i are small because these estimated weights can be poor characterizations of the truth [2]. This danger can be avoided if a function can be found which models the variance.

2.1 Models for the variance

For many physical processes, variance seems to be systematically related to the mean or other variables. The explicit description of this relationship by a function leads to the variance function (VF) model. For example, a common model for variance is the constant coefficient of variation (CV) model. The function $\mathrm{Var}(Y_{ij}) = \sigma^2 \mu_i^2$ yields a constant CV equal to σ and leads to weights $w_i^{-1} = \mu_i^2$. Other models may assume $\mathrm{Var}(Y_{ij}) = \sigma^2 \mu_i$ or $\mathrm{Var}(Y_{ij}) = \sigma^2 (\alpha + \mu_i^2)$ for α known [2], [3]. In each case, the values of the weights $\{w_i\}$ are completely determined by the values $\{\mu_i\}$ of the mean response at x_i, so the weights are often estimated by inserting an estimate for μ_i. Typical estimates for μ_i are the sample mean $\bar{Y}_i$ based on the replicate responses at x_i or the predicted value $\hat{\mu}_i$ based on an initial fit to the data by LS.

In some applications, however, the variance function may not be completely known. For example, in radioimmunoassay the response is a count for which a Poisson assumption implies that $Var(Y_{ij}) = \mu_i$. Often, however, variance at $\mathbf{x}_i$ is larger than the Poisson variance would indicate, a phenomenon known as "overdispersion" [6]. A general model for the variance that accommodates overdispersion is the power variance function $Var(Y_{ij}) = \sigma^2\mu_i^{2\theta}$ which is the Poisson model if $\sigma = 1$, $\theta = 0.5$. This model allows the relative size of variance to vary through σ and its increase with mean to be governed by θ. When $\sigma = 1$, $\theta > 0.5$ indicates overdispersion. In our experience, θ often falls in the range $0.6 \le \theta \le 0.9$. An important feature of this general model is that it can accommodate cases in which the size of the variability relative to the range of the mean response is rather small [3], [7].

For assays under development, θ may be unknown. Since the $\{w_i\}$ depend on both μ_i and θ, the true value of θ must be estimated. When $m_i > 1$, a simple approach to estimation is to note that $\log\{Var(Y_{ij})\}^{1/2} = \log\sigma + \theta \log\mu_i$. Then the slope of LS regression of $\log s_i$ on $\log \bar{Y}_i$ provides an estimate $\hat{\theta}$. The estimated weights would then be $\{\bar{Y}_i^{\hat{\theta}}\}$ [3]. However, other methods will be discussed in this article that do not require replication and have better statistical properties.

All of the methods discussed so far model the variability as a function of the mean response. However, this is not a restriction of the variance function approach. In some situations it is realistic to think about variability as a function of a predictor or some other variable not necessarily in the mean model. For example, in the analysis of fractional-factorial experiments variance may be modeled as a log-linear function of the factor levels [8].

This discussion and comments of other authors [2] suggest that the most flexible approach to modeling variability should allow the variance function to depend on the particular experimental situation. In this spirit, consider a general model for the response

$$Y_{ij} = f(\mathbf{x}_i,\beta) + \sigma\, g(\mu_i,\mathbf{z}_i,\theta)\, \varepsilon_{ij},\ i = 1, \ldots, N,\ j = 1,\ldots, m_i \qquad (5)$$

where g is a function which may depend on the mean response, on a vector $\mathbf{z}$ of known variables containing possibly some or all of the values in $\mathbf{x}$, and on a $(q\times1)$ vector of parameters θ that may be known or unknown. The implied assumption is that $Var(Y_{ij}) = \sigma^2\, g^2(\mu_i,\mathbf{z}_i,\theta)$ so that appropriate weights would be $w_i = 1/g^2(\mu_i,\mathbf{z}_i,\theta)$. Hence, the weights $\{w_i\}$ in (4) are functionally related to the mean and other variables. The function g is called the variance function. This framework includes all of the individual situations we have discussed. For example, the constant CV model has $g(\mu_i,\mathbf{z}_i,\theta) = \mu_i^{\theta}$ with $\theta = 1$.

The advantage of a general model like (5) is that it encourages analysts to think about the relationship between the variance and other factors for a given experimental situation instead of relying on a few fixed relationships for describing variance with little basis beyond standard usage; e.g., pharmacokineticists have often chosen the constant CV model when another model may be appropriate [9]. The general model also allows additional flexibility because the values of the components of θ can be estimated from the data when necessary.

2.2 *Choosing a variance function*

To choose an appropriate variance function, the analyst must consider features of the process under study. For example, consider an assay in which the variance of the response increases with the mean. Now suppose that due to imprecision of measurement there is a minimum level of variability no matter how small the response becomes. In the flexible framework of (5), an appropriate variance function would be $g(\mu_i,z_i,\theta) = \left(\theta_1 + \mu_i^{\theta_2}\right)^{1/2}$, so that $\mathrm{Var}(Y_{ij}) = \sigma^2\left(\theta_1 + \mu_i^{\theta_2}\right)$.

This model can be thought of as a "components of variance" model, where the component θ_1 describes the imprecision of measurement that dominates at small response values and $\mu_i^{\theta_2}$ describes the relationship between mean and variance that dominates at larger response values. Although the analyst may have evidence and knowledge to support this characterization, he/she may not be sufficiently confident to choose values for θ_1 and θ_2 *a priori*. We discuss the problem of parameter estimation in Section 4.

Along these same lines, additional models can be postulated to describe different functional relationships between the mean response and the variability that would be difficult to capture using standard transformations. For example, a model for the variance that allows the variability to increase very quickly with the mean is $\mathrm{Var}(Y_{ij}) = \sigma^2 \exp(2\theta\mu_i)$. A model in which the standard deviation is thought to increase as a quadratic function of a predictor is $g(\mu_i,z_i,\theta) = 1 + \theta_1 x_i + \theta_2 x_i^2$ (in this case $z_i = x_i$). Thus, variance models can be proposed to accommodate the scientist's insight into a wide range of physical processes resulting in improved inference and estimation.

2.3 *Graphical analysis*

We consider now the example introduced in Section 1 and discuss some graphical displays that are useful in selecting an appropriate variance model. If there is replication ($m_i > 1$), a plot of log s_i against log $\bar{Y}_i$ should have the form of a straight line as long as the power variance function

$\text{Var}(Y_{ij}) = \sigma^2 \mu_i^{2\theta}$ is suitable. For the assay example, Figure 5 shows that the power variance model is reasonable.

A graphical display that can be used for cases with no replication ($m_i = 1$), is a plot of the logarithms of the absolute residuals $|r_{ij}|$ versus the logarithms of the predicted values $\hat{\mu}_i$ from the LS fit. This plot is similar in spirit to the plot of log s_i against log $\overline{Y}_i$ because the absolute residuals contain information about the standard deviations. This plot also should show a straight line relationship as in Figure 6. We prefer this plot to that of log s_i against log $\overline{Y}_i$ even for cases with replication since $\overline{Y}_i$ and s_i and are not very good estimates of the mean and standard deviation for small values of m_i.

Graphical displays can be constructed using the same approach to investigate other variance function models. For example, a plot of absolute residuals $|r_{ij}|$ versus scalar x_i that showed a quadratic relationship would suggest the use of the quadratic variance function $\text{Var}(Y_{ij}) = 1 + \theta_1 x_i + \theta_2 x_i^2$. Similarly a plot of log $|r_{ij}|$ versus predicted values $\hat{\mu}_i$ that showed a linear relationship would suggest the exponential model $\text{Var}(Y_{ij}) = \sigma^2 \exp(2\theta\mu_i)$.

These plots along with the researcher's understanding of the underlying physical process should be useful in identifying an appropriate variance function model. However, we recommend against using the relationships demonstrated by these plots to estimate the parameters θ and σ of the variance function; for example, using the estimated slope from the regression of log s_i against log $\overline{Y}_i$ as an estimate of θ. The methods in Section 4 have better statistical properties.

3 GENERALIZED LEAST SQUARES

Generalized least squares (GLS) is a method for fitting model (5). This method incorporates variance function estimation (VFE) for unknown variance parameters. In this section we describe the GLS method, its mathematical basis, and its implementation using standard nonlinear regression programs assuming θ is known. Estimation of θ is incorporated into the method in Section 4. A more comprehensive treatment of these issues is given by Carroll and Ruppert [5].

3.1 Motivation and description

The terminology generalized least squares refers to use of WLS (4) with estimated weights $\{\hat{w}_i\}$. When the VF model (5) holds and θ is known, then the weights $w_i = 1/g^2(\mu_i, z_i, \theta)$ are known except for the values of the $\{\mu_i\}$. If the mean response model is well-established, it makes sense to

construct estimates of the weights using estimates $\{\hat{\mu}_i\}$. In this case, an initial LS fit ignoring the variance function gives a reasonable starting point. Thus, an initial set of estimated weights is

$$\hat{w}_i = 1/g^2(\hat{\mu}_i^{(0)}, z_i, \theta) \text{ where } \hat{\mu}_i^{(0)} = f(x_i, \hat{\beta}_{LS}) \tag{6}$$

where $\hat{\beta}_{LS}$ is obtained from the initial LS fit. Minimization of (4) using these weights will produce a new estimate $\hat{\beta}_{GLS}$ that makes more efficient use of the data, since it incorporates the nature of variability into the fit. Such an estimate is a GLS estimate.

As an alternative to making use of the mean response model, some analysts have used $\bar{Y}_i$ in place of $\hat{\mu}_i^{(0)}$ when all $m_i > 1$. However, we generally prefer $\hat{\mu}_i^{(0)}$ because it incorporates the information in the mean response model. Use of Y_{ij} in place of $\hat{\mu}_i^{(0)}$ has also been advocated. This method can be quite unreliable because an individual value Y_{ij} can often be a very poor characterization of the mean response at x_i. This practice is highly discouraged.

3.2 Mathematical justification

The use of estimated weights makes intuitive sense and may in fact be formally justified by an appeal to statistical theory. Consider the usual linear model with constant variance and normal data, i.e., (1) with $f(x_i,\beta) = x_i^T\beta$ for $k = p$. It is known [1], that as long as the classical assumptions are true, $\hat{\beta}_{LS}$ has a multivariate normal distribution with mean β and covariance matrix

$$\Sigma_{LS} = \sigma^2 (X^T X)^{-1} \tag{7}$$

where the matrix X is defined as the $(n \times p)$ matrix with its first m_1 rows equal to x_1^T, its second m_2 rows equal to x_2^T, and so on.

When the assumption of normality is not made, large sample theory can be used to show that $\hat{\beta}_{LS}$ has approximately this multivariate normal distribution. Such theory uses mathematical arguments based on letting the sample size $n \to \infty$ to gain insight into properties of estimators that are too complicated to directly evaluate. The implications are often relevant to practical situations with smaller sample sizes.

For general nonlinear f, the same large sample theory results hold as long as we define $f_\beta(x_i,\beta)$ to be the $(p \times 1)$ vector $[\partial/\partial\beta_1\{f(x_i,\beta)\}, \ldots, \partial/\partial\beta_p\{f(x_i,\beta)\}]^T$ and redefine X to be the $(n \times p)$ matrix

with its first m_1 rows equal to $f_\beta^T(x_1,\beta)$ and so on. (For the linear model, $f_\beta(x_i,\beta) = x_i$ and X is the usual design matrix with rows x_i^T.)

The relationship (7) is used to construct confidence intervals for the elements of β since the diagonal elements of Σ_{LS} are the variances of $\hat{\beta}_{LS}$. Since σ^2 is unknown it is replaced by the estimate

$$\hat{\sigma}^2_{LS} = (n-p)^{-1} \sum_{i=1}^{N} \sum_{j=1}^{m_i} \{Y_{ij} - f(x_i, \hat{\beta}_{LS})\}^2 \tag{8}$$

when estimated variances are desired. Linear regression programs construct standard errors and confidence intervals for β using this method. If f is nonlinear, standard nonlinear regression programs use the diagonal elements of $\hat{\Sigma}_{LS}$ in the same way except that X is replaced by $X(\hat{\beta}_{LS})$, where $X(\beta)$ is defined to be X evaluated at its argument.

Large sample approximation theory states that when (5) is true, $\hat{\beta}_{GLS}$ has approximately a multivariate normal distribution with mean and covariance matrix

$$\Sigma_{GLS} = \sigma^2 (X^T G^{-1} X)^{-1}. \tag{9}$$

where G is a (n×n) diagonal matrix with its first m_1 diagonal elements equal to $g^2\{f(x_1,\beta),z_1,\theta\}$, second m_2 diagonal elements equal to $g^2\{f(x_2,\beta),z_2,\theta\}$, and so on.

An estimate $\hat{\Sigma}_{GLS}$ of Σ_{GLS} is constructed using (9) where σ^2 is replaced by

$$\hat{\sigma}^2 = (n-p)^{-1} \sum_{i=1}^{N} \sum_{j=1}^{m_i} \{Y_{ij} - f(x_i, \hat{\beta}_{GLS})\}^2 / g^2\{f(x_i, \hat{\beta}_{GLS}),z_i,\theta)\}. \tag{10}$$

In constructing this estimate, G is replaced by $G(\hat{\beta}_{GLS},\theta)$, where $G(\beta,\theta)$ is defined to be G evaluated at its arguments. If f is nonlinear, X is replaced by $X(\hat{\beta}_{GLS})$.

It is important to note that if (5) is true, $\hat{\beta}_{LS}$ has approximately a multivariate normal distribution with mean β but covariance matrix

$$\Sigma^*_{LS} = \sigma^2 \Sigma^{-1}_{LS} Q \Sigma^{-1}_{LS}, Q = X^T G X. \tag{11}$$

Thus, inference based on estimated standard errors for $\hat{\beta}_{LS}$ rather than for $\hat{\beta}_{GLS}$ will be incorrect. This shows that a penalty for not accommodating nonconstant variance is unrealistic inference about β.

There is a further implication of (9) and (11). Standard results for matrix theory imply that Σ^*_{LS} - Σ_{GLS} is a nonnegative definite matrix. The practical implication is that when (5) is true, the variability associated with $\hat{\beta}_{LS}$ as an estimator of β will be greater than that associated with $\hat{\beta}_{GLS}$. Thus the GLS estimator will make more efficient use of data; i.e., it can do as good a job as LS with less data. An example of numerical evidence dramatically illustrating this contention is given in [10].

Going back to the WLS estimates given by (4), it can be shown that the "ideal" WLS estimator that could be calculated if $w_i = g^{-2}\{f(\mathbf{x}_i,\beta),\mathbf{z}_i,\theta\}$ were known has approximately a multivariate normal distribution with mean β and covariance matrix Σ_{GLS}. This shows that the GLS method using estimated weights based on a previous estimate of β approaches what we would do if the weights were known.

Thus, statistical theory suggests that GLS is preferred to LS for heteroscedastic data as long as a suitable g can be determined, that failure to account for heteroscedasticity will lead to misleading assessment of the quality of estimation, and that GLS is a reasonable approximation to WLS when the true weights are unknown. The benefits of using GLS over LS for assessing prediction of a future observation are discussed in Section 5. The above results hold even if the assumption of normally distributed observations does not, implying that the properties of the GLS estimator should not be affected adversely if the normality assumption is not correct. Practically speaking, this implies that GLS estimation is relatively insensitive to anomalies such as outliers.

3.3 Implementation

Since GLS is more efficient than LS, it makes sense to consider further refinement. In particular, the weights (6) are based on $\hat{\beta}_{LS}$, an inefficient estimate of β, so intuition would suggest that new weights based on $\hat{\beta}_{GLS}$ may be a more efficient representation of $\{w_i\}$. This suggests forming new estimated weights

$$\hat{w}_i = 1/g^2\,(\hat{\mu}_i^{(1)},\mathbf{z}_i,\theta) \text{ where } \hat{\mu}_i^{(1)} = f(\mathbf{x}_i,\hat{\beta}_{GLS}) \qquad (12)$$

and then again using (4) to update estimation of β. This yields the following algorithm for GLS when θ is known [5]:

(1) Obtain the LS estimator $\hat{\beta}_{LS}$. Let $\hat{\beta}^{(0)} = \hat{\beta}_{LS}$ and set k = 0.

(2) Form estimated weights $\hat{w}_i = 1/g^2\,(\hat{\mu}_i^{(k)},\mathbf{z}_i,\theta)$ where $\hat{\mu}_i^{(k)} = f(\mathbf{x}_i,\hat{\beta}^{(k)})$.

(3) Use the estimated weights from (2) to obtain $\hat{\beta}_{GLS}$ by minimizing (4).

(4) Set k = k + 1, let $\hat{\beta}^{(k)} = \hat{\beta}_{GLS}$, and return to (2).

The algorithm as stated is incomplete because it does not specify the number of times, C, to repeat the process. For large samples any of the GLS estimates obtained in step (3) will emulate the ideal WLS estimator. Thus, the theory is not useful for deciding upon a "best" value C. For many problems, only 2 or 3 iterations are necessary for the algorithm to converge within a reasonable tolerance. However, at least 2 iterations must be taken to wash out the effect of the inefficient initial estimate $\hat{\beta}_{LS}$. Full iteration is a version of iteratively reweighted least squares. Final estimation of σ is accomplished after the last iteration using $\hat{\sigma}^2$ in (10). If g does not depend on μ_i, note that only one pass through the algorithm is required.

In the general case, implementation of this algorithm can be accomplished by repeated use of a nonlinear regression program that allows known weights to be specified is required; e.g., PROC NLIN in the Statistical Analysis System (SAS) [11]. Details of this application are given in [12]. If f is linear, a linear regression program may be used (e.g., PROC REG [11]). For certain standard values of θ and power of the mean variance function, Genstat [13] or the Generalized Linear Interactive Modeling System (GLIM) [14] provide estimates based on iteratively reweighted least squares.

3.4 Example

For purposes of illustration, we carry out the GLS procedure for the assay data assuming the power variance function with $\theta = 1$, the constant CV model. The results in Table 2 were obtained using a standard linear regression package and C=3. This fit could also be accomplished in GLIM [14] using gamma error structure and the identity link function.

As in the LS case, residuals can be used to assess how well the model characterizes the observed variability. Define the final residuals as $r_{ij}^{(C)} = Y_{ij} - \hat{\mu}_i^{(C)}$ and let $\hat{\sigma}^2 = (n-p)^{-1} \sum_{i=1}^{N} \sum_{j=1}^{m_i} \{Y_{ij} - \hat{\mu}_i^{(C)}\}^2 / g^2\{\hat{\mu}_i^{(C)}, z_i, \theta)\}$ where $g(\hat{\mu}_i^{(C)}, z_i, \theta) = \hat{\mu}_i^{(C)}$. Since (5) implies $(Y_{ij} - \mu_i)/\{\sigma\, g(\mu_i, z_i, \theta)\} = \varepsilon_{ij}$, if the variance model is adequately characterizing the variability, a plot of the standardized residuals $r_{ij}^{(C)} / \hat{\sigma}\, g(\hat{\mu}_i^{(C)}, z_i, \theta)$ versus predicted values as in Figure 7 should show a band of constant width. Note that the vertical scale is in units of assumed standard deviation so a direct comparison

with Figure 2 can be made. Figure 7 obviously shows a great improvement over Figure 2. We consider using the data to estimate θ in the next Section.

4 VARIANCE FUNCTION ESTIMATION

We now relax our assumption that θ is known and consider estimation of θ and how to incorporate this into the GLS algorithm.

4.1 Motivation and description

In order to motivate the estimation of σ and θ, we cast the problem in terms of a regression of the squared residuals against the variance function. For any estimate $\hat{\beta}$, form the residuals $r_{ij} = Y_{ij} - \hat{\mu}_i$ where $\hat{\mu}_i = f(x_i,\hat{\beta})$. Note that the general VF model for the response (5) implies that $E\{(Y_{ij} - \mu_i)^2\} = \sigma^2 g^2(\mu_i,z_i,\theta)$ since $E(\varepsilon_{ij}^2) = 1$. Thus we might expect that

$$E(r_{ij}^2) \approx \sigma^2 g^2(\hat{\mu}_i,z_i,\theta) \tag{13}$$

where $\hat{\beta}$ is viewed as if it were known. Now (13) is reminiscent of a "regression" relationship with "response" r_{ij}^2 and "regression function" $\sigma^2 g^2(\hat{\mu}_i,z_i,\theta)$. Thus, given $\hat{\beta}$, (13) suggests estimation of σ and θ could be accomplished by some form of "regression" of r_{ij}^2 on $\sigma^2 g^2(\hat{\mu}_i,z_i,\theta)$.

For normal data, $Var(\varepsilon_{ij}^2) = 2$ so $Var\{(Y_{ij} - \mu_i)^2\} = 2\sigma^4 g^4(\mu_i,z_i,\theta)$. This suggests that

$$Var(r_{ij}^2) \approx (2\sigma^4)\, g^4(\hat{\mu}_i,z_i,\theta). \tag{14}$$

Thus, variance of the "response" for the "regression" (13) is proportional to g^4. Appealing to the spirit of WLS, this suggests estimation of σ and θ by minimizing

$$\sum_{i=1}^{N}\sum_{j=1}^{m_i} v_i \{r_{ij}^2 - \sigma^2 g^2(\hat{\mu}_i,z_i,\theta)\}^2, \tag{15}$$

where $v_i = g^{-4}(\hat{\mu}_i,z_i,\theta)$. Emulating the GLS approach, one could obtain an initial estimate for θ by the equivalent of LS with $v_i \equiv 1$ and use these weights to estimate σ and θ by minimizing (15). Using this estimate of θ, new estimated weights $\hat{v}_i$ could be constructed, a new estimate of θ obtained, and the process iterated. In Section 4.4 we discuss an equivalent method that is simpler to implement.

Note that viewing estimation of σ and θ as a regression problem is in the spirit of the method based on $\log s_i$ and $\log \bar{Y}_i$ described in the Section 2. The method here relies on residuals and pre-

dicted values, the other on standard deviations and means based on replication. Note that the former uses the information in the postulated mean function while the $\{s_i\}$ can be poor assessments of variance for small $\{m_i\}$. Hence, the method based on (15) is preferred [15].

Generally, one would construct a final estimator of σ using (10) with θ replaced by its final estimate, so interest in VFE focuses on estimation of θ. Estimation of θ based on squared residuals from a previous fit as described above can be shown to be theoretically sound [15], so intuition leads to a mathematically justifiable method. It may also be established that estimation based on $\{s_i\}$ is inefficient relative to that based on residuals, as intuition suggests [15].

4.2 *Accounting for the effects of outliers*

Often it is not reasonable to assume that the data are normally distributed because of outlying observations. Unlike those of the GLS estimator for β, the properties of estimators of σ and θ will be sensitive to this form of nonnormality. If outliers are present in the data, the theory can be used to show that the quality of estimation based on $\{r_{ij}^2\}$ can deteriorate [15]. Various alternative methods have been proposed for VFE in the presence of outliers, many based on the idea that squaring the residuals magnifies the effect of outliers since outliers will often generate residuals that are large in magnitude. One method uses an argument similar to the above, but based on the observation that $E|Y_{ij} - \mu_i| = \eta\, g(\mu_i,z_i,\theta)$, where $\eta = \sigma\, E(|\varepsilon_{ij}|)$. Then $E(|r_{ij}|) \approx \eta\, g(\hat{\mu}_i,z_i,\theta)$ and $Var(|r_{ij}|) \propto g^2(\hat{\mu}_i,z_i,\theta)$, suggesting minimization in η and θ of

$$\sum_{i=1}^{N}\sum_{j=1}^{m_i} u_i \{|r_{ij}| - \eta\, g(\hat{\mu}_i,z_i,\theta)\}^2 \text{ where } u_i = g^{-2}(\hat{\mu}_i,z_i,\theta). \qquad (16)$$

Estimation η of is of little interest but is an adjunct to estimating θ. We describe a simple method of implementation shortly. If data are relatively outlier-free, use of squared residuals will be preferred, but when outliers are known to be present, this method based on absolute values of residuals is better, because the absolute value function will not magnify the effect of large residuals as severely.

4.3 *Consequences of estimating* θ

In Section 3.2 we discussed the multivariate normality of $\hat{\beta}_{GLS}$ when θ is known. These same results can be shown to hold if the estimated weights in Step 2 of the GLS algorithm in Section 3.3 are constructed using either of the estimates $\hat{\theta}$ described in Sections 4.1 and 4.2 [15]. Estimates of the variance of $\hat{\beta}_{GLS}$ using these weights will be found from (9) using $\hat{\sigma}^2$ in (10) with $\hat{\theta}$ in place

of θ and $\mathbf{G}$ replaced by $\mathbf{G}(\hat{\beta}_{GLS}, \hat{\theta})$, replacing $\mathbf{X}$ by $\mathbf{X}(\hat{\beta}_{GLS})$ if f is nonlinear. Modification of the GLS algorithm to included estimation of θ is described in the next section.

4.4 Implementation

Estimation of θ can be accomplished jointly with that for σ or η as above, using a nonlinear regression program. However, by slightly different arguments it can be shown that a more easily implemented scheme described in this section is equivalent to those described above for large samples [15]. For ease of computation the following methods are useful; justification can be found in [12].

Estimation of θ based on squared residuals, by iteration of (15) is equivalent in large samples to minimizing the sum of squares [12]

$$\sum_{i=1}^{N}\sum_{j=1}^{m_i}\left\{\frac{r_{ij}^2\dot{g}}{g(\hat{\mu}_i, z_i, \theta)}\right\}^2 \quad \text{where } \dot{g} = \prod_{i=1}^{N} g(\hat{\mu}_i, z_i, \theta)^{m_i/n} \tag{17}$$

which may be accomplished via a nonlinear regression program by regressing a dummy variable identically equal to 0 on the quantity in braces in (17). Details of practical implementation using PROC NLIN in SAS are given in [12]. Note that (17) depends only on θ -- σ may be estimated after final estimation of β and θ by substituting the final estimates $\hat{\beta}$, $\hat{\theta}$ in (10).

As explained in [12], the nonlinear regression program must allow multiple passes through the data for general g in order to compute $\dot{g}$. However, for common g such as μ_i^{θ}, $\dot{g} = \left(\prod \hat{\mu}_i^{m_i/n}\right)^{\theta}$ so the expression in parentheses is known and only one pass through the data is needed. For estimation based on absolute residuals, replace (17) by

$$\sum_{i=1}^{N}\sum_{j=1}^{m_i}\left\{\left(\frac{|r_{ij}|\dot{g}}{g(\hat{\mu}_i, z_i, \theta)}\right)^{\frac{1}{2}}\right\}^2 \quad \text{where } \dot{g} = \prod_{i=1}^{N} g(\hat{\mu}_i, z_i, \theta)^{m_i/n}. \tag{18}$$

which may be accomplished as above via a nonlinear regression program by regressing a dummy variable identically equal to 0 on the quantity in braces in (18).

These developments suggest the following modification of the GLS algorithm to include estimation of unknown variance parameters: Replace (2) by

(2)* Given $\hat{\beta}^{(k)}$, use (17) or (18) to obtain the estimate $\hat{\theta}^{(k)}$ and form estimated weights

$\hat{w}_i = 1/g^2(\hat{\mu}_i^{(k)}, z_i, \hat{\theta}^{(k)})$ where $\hat{\mu}_i^{(k)} = f(x_i, \hat{\beta}^{(k)})$.

Giltinan and Ruppert [12] outline this in detail and illustrate implementation in SAS.

4.5 Example

For purposes of illustration, we now carry out the GLS procedure for the assay data assuming the power variance function estimating θ from the data. The results in Tables 3 and 4 were obtained using a standard nonlinear regression package to implement the methods described by (17) and (18), respectively. Methods (17) and (18) give similar results because the outlier remarked on earlier occurs at the highest concentration level where variability is expected to be greatest. Thus the outlier is downweighted in such a way that it plays little role in the analysis. In Section 3.4, we carried out the GLS procedure using the value θ=1.00 for the constant CV model. However, the estimated value of shown in Table 3 is 0.845. It makes sense that the estimated value of θ is less than 1.00 because the observed CVs as shown in Table 1 are clearly not constant.

Figure 8 shows the standardized residuals from method (17) versus predicted values where θ is replaced by $\hat{\theta}$ in g when constructing the standardized residuals. While Figures 7 and 8 are very similar, it could be argued that the constant CV model used for Figure 7 slightly overcorrects for nonconstant variance at low concentrations. The impression given by Figure 8 is that the VF power model with estimated θ provides a more satisfactory representation of the observed variability. In any event, doing some form of weighting is better than ignoring the nonconstant variance (see Figure 2). In particular, the intercept for either of the weighted fits agrees very closely with the observed mean at zero dose in contrast to results from the LS fit.

5 PREDICTION

While GLS is generally preferred to LS on grounds of efficiency, another reason for careful modeling and estimation of nonconstant variance is to facilitate appropriate inference based on the fitted curve. The VF approach gives a characterization of variance even for settings of the predictors not in the experiment, so that correct inference on future observations at new predictor settings can be accomplished. Detailed discussion is given in [5] and [16]; we illustrate the major point here for prediction of a new observation.

5.1 Motivation and description

Given a new predictor setting x_0, we would like to predict the associated response Y_0. The usual estimator for this response is $f(x_0, \hat{\beta})$. To correctly assess the quality of this estimate, the error in prediction should be reported as a prediction or uncertainty interval based on some level of confidence. If model (1) is assumed, one would obtain $\hat{\beta}_{LS}$, estimate σ^2 by (8) and construct a predic-

tion interval for $f(\mathbf{x}_0, \hat{\beta})$. Defining $f_0(\beta)$ to be $f_\beta(\mathbf{x}_0, \hat{\beta})$ as in Section 3.3 and assuming that the normal assumption is correct, the form of such an interval with confidence coefficient $100(1 - \alpha)\%$ is [5]

$$f(\mathbf{x}_0, \hat{\beta}_{LS}) \pm t \{\hat{\sigma}^2_{LS} + n^{-1} f_0^T(\hat{\beta}_{LS}) \hat{\Sigma}_{LS} f_0(\hat{\beta}_{LS})\}^{1/2} \tag{19}$$

with t equal to the $1 - \alpha/2$ percentage point of the t distribution with $n - p$ degrees of freedom. Now if the true model is (5), the estimator $\hat{\sigma}^2_{LS}$ is an attempt to assess variability by averaging over the entire data set. Thus $\hat{\sigma}^2_{LS}$ represents an underestimate of variance in regions where variance is large and an overestimate where it is small. Under (5), using the final estimates from the GLS algorithm $\hat{\beta}_{GLS}$, $\hat{\theta}$, and $\hat{\sigma}^2$, the form of a prediction interval will be [5]

$$f(\mathbf{x}_0, \hat{\beta}_{GLS}) \pm t \{\hat{\sigma}^2 g^2(f(\mathbf{x}_0, \hat{\beta}_{GLS}), z_i, \hat{\theta}) + n^{-1} f_0^T(\hat{\beta}_{GLS}) \hat{\Sigma}_{GLS} f_0(\hat{\beta}_{GLS})\}^{1/2}. \tag{20}$$

This interval correctly assesses variance in a future observation at all regions.

To gain insight, assume that the sample size is large. Since in both (19) and (20) the terms containing $\hat{\Sigma}_{LS}$ and $\hat{\Sigma}_{GLS}$ depend on n^{-1}, we expect that for large samples these would be negligible. Hence prediction intervals would behave approximately like

$$f(\mathbf{x}_0, \hat{\beta}_{LS}) \pm t \, \hat{\sigma}_{LS} \tag{21}$$

and

$$f(\mathbf{x}_0, \hat{\beta}_{GLS}) \pm t \, \hat{\sigma} \, g(f(\mathbf{x}_0, \hat{\beta}_{GLS}), z_i, \hat{\theta}) \tag{22}$$

respectively. Interval (21) is of constant width regardless of x_0, while the width of (22) changes with x_0 to reflect the varying degree of quality of estimation across its range. Thus, ignoring nonconstant variance and basing inference on a LS fit and its assumptions, yields an interval of uncertainty that is too narrow, or optimistic, where the true variance is large and too wide, or conservative, where true variance is small. If one wishes to attach an uncertainty interval for predictions based on the fit, one can be grossly misled about the quality of prediction for certain settings of the predictors.

5.2 Example

Figures 9 and 10 show prediction intervals using LS and GLS with estimated θ, respectively, constructed using (19) and (20). In calculating the prediction limits, we used $\hat{\Sigma}_{LS}=$

$\begin{bmatrix} 2.343 \times 10^{-5} & -3.524 \times 10^{-7} \\ -3.524 \times 10^{-7} & 1.279 \times 10^{-8} \end{bmatrix}$ and $\hat{\Sigma}_{GLS} = \begin{bmatrix} 2.073 \times 10^{-7} & -1.025 \times 10^{-8} \\ -1.025 \times 10^{-8} & 1.793 \times 10^{-8} \end{bmatrix}$. It is clear that ignoring heteroscedasticity by using LS assuming constant variance leads to the situation described above.

6 DISCUSSION

Modeling the variance can lead to a more meaningful analysis when the variance changes systematically with the mean response or other variables. The idea of postulating a variance model depending on unknown parameters, and then estimating these parameters from the data via the GLS-VFE algorithm, provides a flexible framework for analysis and insight into the form of experimental error. These methods can be implemented using standard linear and nonlinear regression packages.

The methods we describe are appropriate for data that have nonconstant variance and are approximately normally distributed or for the most part normally distributed with some outliers. We have not discussed the issue of nonnormal data that is skewed. In this case, data transformation can often be used to make the data appear more normal and homoscedastic on the transformed scale. More complex data-analytic tools are available which include the idea of using data transformation and weighted regression together. An illustration of these methods and considerations for the Michaelis-Menten model are given in [17].

Acknowledgments

The data for the example used in this article were provided by R. Dabrowsky of the College of Veterinary Medicine at North Carolina State University, Raleigh, NC.

REFERENCES

[1] N.R. Draper and H. Smith, *Applied Regression Analysis*, Second Edition, Wiley, New York, 1981.

[2] A.C. Storer, M.G. Darlinson, A. Cornish-Bowden, The nature of experimental error in enzyme kinetic measurements, *Biochemistry Journal*, **151** (1975) 361-367.

[3] D. Rodbard, R.H. Lenox, H.L. Wray, D. Ramseth, Statistical characterization of the random errors in the radioimmunoassay dose-response variable, *Clinical Chemistry*, 22/3 (1976) 350-358.

[4] L. Oppenheimer, T.P. Capizzi, R.M. Weppelman, H. Mehta, Determining the lowest limit of reliable assay measurement, *Analytical Chemistry*, **55** (1983) 638-643.

[5] R.J. Carroll and D. Ruppert, *Transformation and Weighting in Regression*, Chapman and Hall, New York, 1988.

[6] N. E. Breslow, Extra-Poisson variation in log-linear models, *Applied Statistics*, **33** (1984) 38-44.

[7] M. Davidian, R.J. Carroll, and W. Smith, Variance functions and the minimum detectable concentration in assays, *Biometrika*, **75** (1988) 549-556.

[8] G.E.P. Box and R.D. Meyer, Dispersion effects from fractional designs, *Technometrics*, **16** (1986) 19-28.

[9] S.L. Beal and L.B. Sheiner, Heteroscedastic nonlinear regression, *Technometrics*, **30** (1988) 327-338.

[10] A.H. Kalantar, Large inefficiencies of unweighted least squares treatment of logarithmically transformed Aexp(-kt) data, *International Journal of Chemical Kinetics*, **19** (1987) 923-927.

[11] SAS User's Guide Version 5, Statistics.

[12] D.M. Giltinan and D. Ruppert, Fitting heteroscedastic regression models to individual pharmacokinetic data using standard statistical software, *Journal of Pharmacokinetics and Biopharmaceutics*, **17** (1989) 601-614.

[13] Genstat User's Manual.

[14] GLIM User's Manual Release 3.7.

[15] M. Davidian and R.J. Carroll, Variance function estimation, *Journal of the American Statistical Association*, **82** (1987) 1079-1091.

[16] R.L. Watters, R.J. Carroll, and Speigelman, C.H. Error modeling and confidence interval estimation for inductively coupled plasma calibration curves, *Analytical Chemistry*, **59** (1987) 1639-1643.

[17] D. Ruppert, N. Cressie, and R.J. Carroll, A transformation/weighting model for estimating Michaelis-Menten parameters, *Biometrics*, **45** (1989) 637-656.

Table 1. Experimental data and summary statistics

dose	0	5	15	45	90
response	0.0016	0.0118	0.0107	0.106	0.106
	0.0019	0.0139	0.0670	0.026	0.158
	0.0002	0.0092	0.0410	0.088	0.272
	0.0030	0.0033	0.0087	0.078	0.121
	0.0042	0.0120	0.0410	0.029	0.099
	0.0006	0.0070	0.0104	0.063	0.116
	0.0006	0.0025	0.0170	0.097	0.117
	0.0011	0.0075	0.0320	0.066	0.105
	0.0006	0.0130	0.0310	0.052	0.098
	0.0013	0.0050			
	0.0020	0.0180			
	0.0050				
	0.0050				
m_i	13	11	9	9	9
mean	0.0021	0.0094	0.0288	0.0672	0.1324
sd	0.0017	0.0048	0.0193	0.0282	0.0554
CV	0.8115	0.5148	0.6173	0.4191	0.4181

Table 2. Results of GLS algorithm with θ=1

Parameter	k=0	k=1	k=2	k=3
β_0	0.00325	0.00209	0.00209	0.00209
Std. Error	0.00484	0.000457	0.000346	0.000346
β_1	0.00144	0.00153	0.00153	0.00153
Std. Error	0.000113	0.000142	0.000166	0.000166
σ	0.0265	0.520	0.604	0.604

Table 3. Results of GLS algorithm with θ estimated based on squared residuals

Parameter	k=0	k=1	k=2	k=3
β_0	0.00325	0.00211	0.00210	0.00210
Std. Error	0.00484	0.000413	0.000455	0.000455
β_1	0.00144	0.00152	0.00151	0.00151
Std. Error	0.000113	0.000143	0.000134	0.000134
σ	0.0265	0.381	0.309	0.308
θ		0.897	0.846	0.845

Table 4. Results of GLS algorithm with θ estimated based on absolute residuals

Parameter	k=0	k=1	k=2	k=3
β_0	0.00325	0.00212	0.00210	0.00210
Std. Error	0.00484	0.000468	0.000475	0.000475
β_1	0.00144	0.00151	0.00151	0.00151
Std. Error	0.000113	0.000132	0.000131	0.000131
σ	0.0265	0.296	0.285	0.285
θ		0.836	0.826	0.826

Figure Legends

Figure 1. Scatter plot of with regression line.

Figure 2. Plot of LS residuals showing fan shape.

Figure 3. Scatter plot for log(response).

Figure 4. Scatter plot for sqrt(response).

Figure 5. Log(sd) versus log($\overline{Y}_i$)

Figure 6. Log(absolute LS residuals) versus log(LS predicted values).

Figure 7. GLS residuals with $\theta = 1$.

Figure 8. GLS residuals with θ estimated based on squared residuals.

Figure 9. LS regression line with prediction limits.

Figure 10. GLS regression line with prediction limits.

Figure 1. Scatter plot with regression line

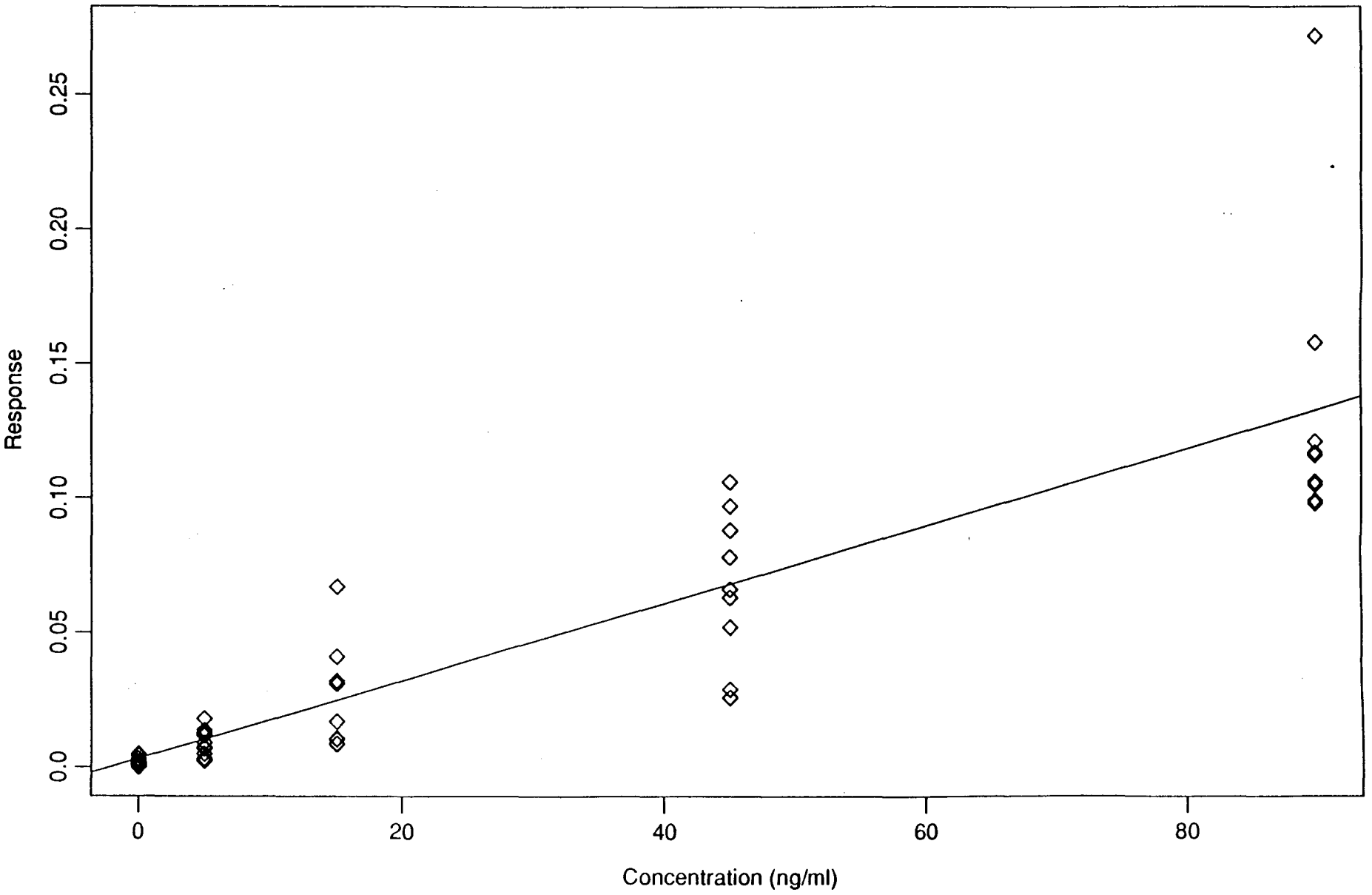

Figure 2. Plot of LS residuals showing fan shape

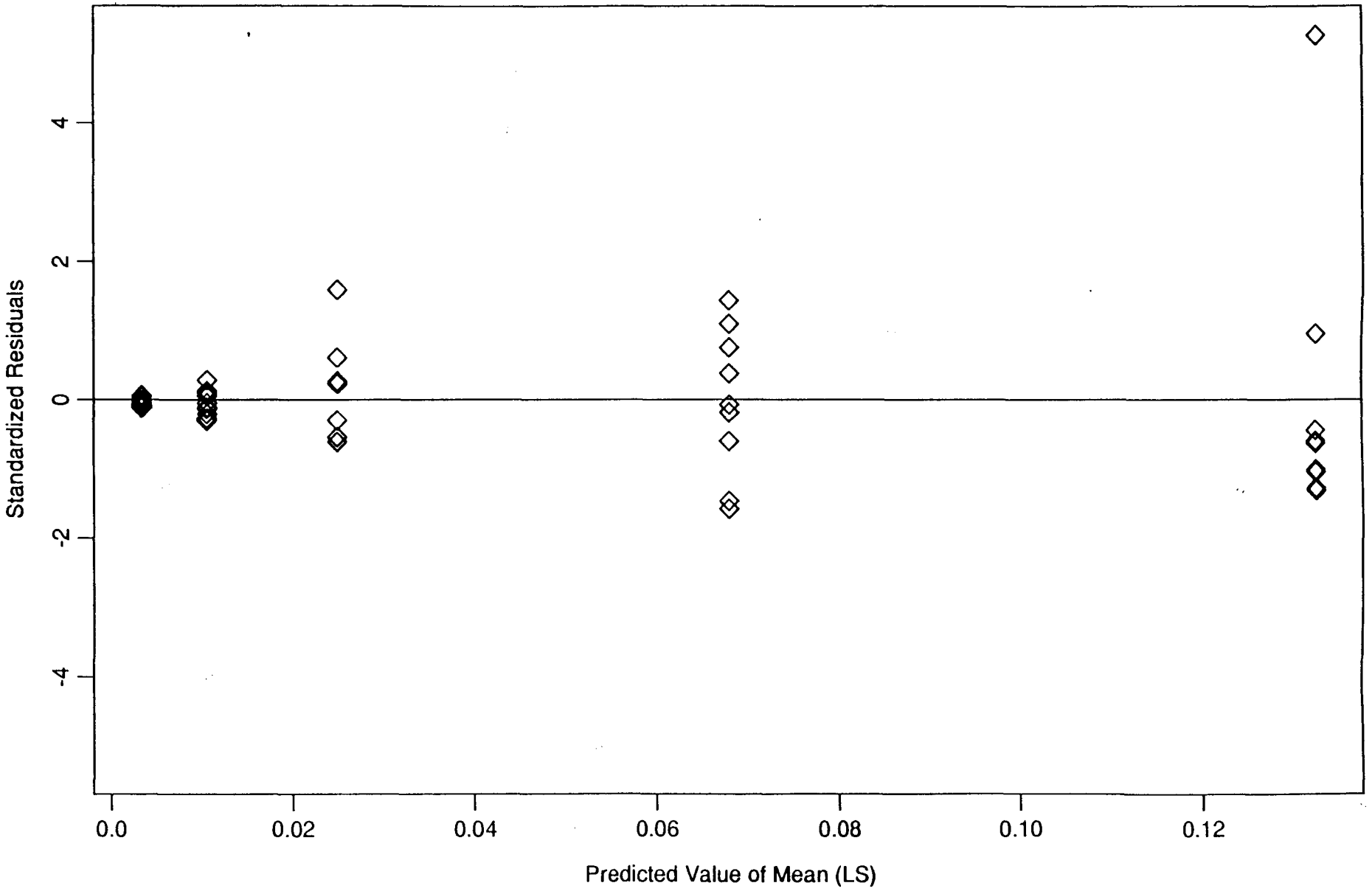

Figure 3. Scatter plot for log(response)

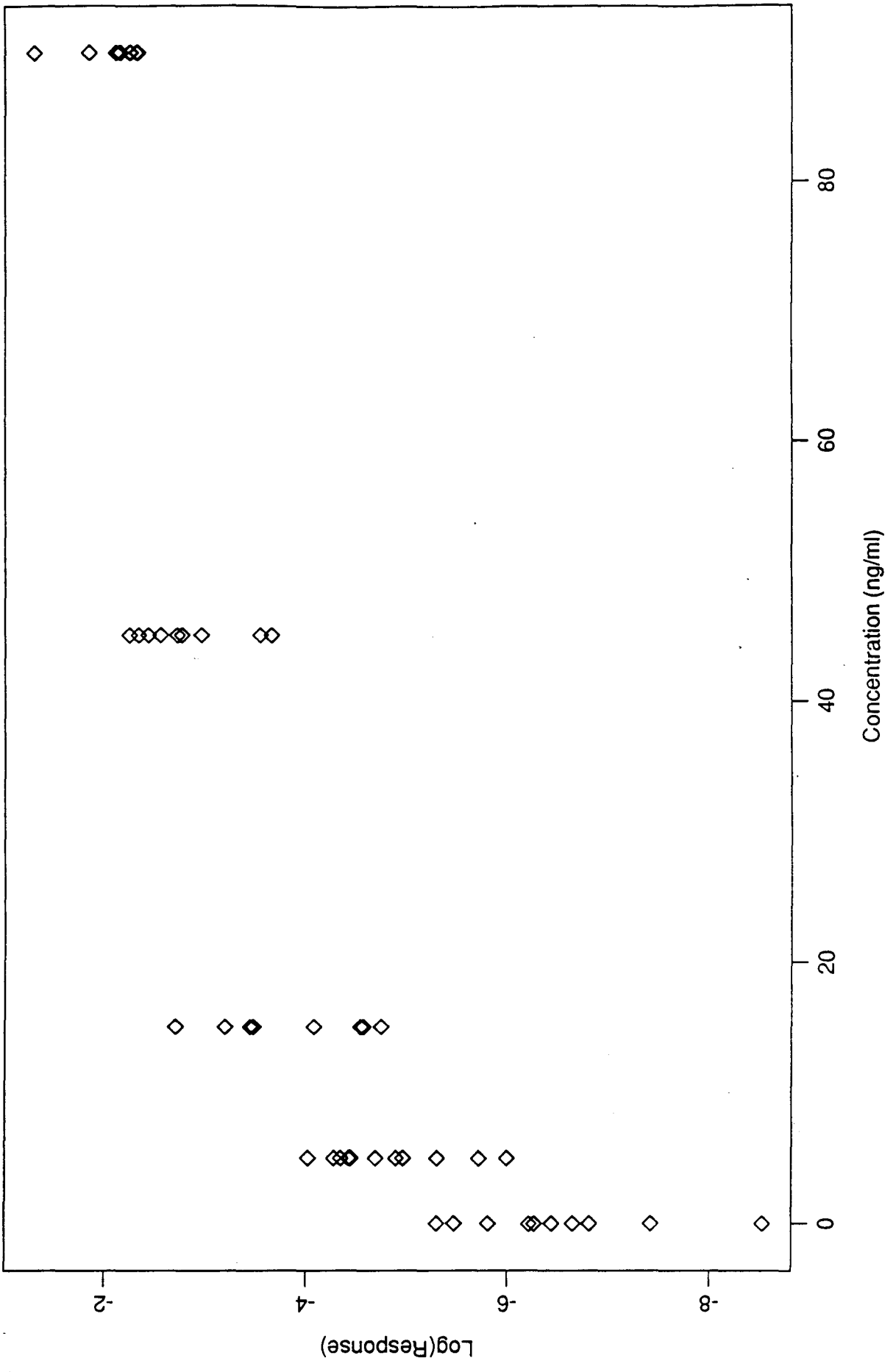

Figure 4. Scatter plot for sqrt(response)

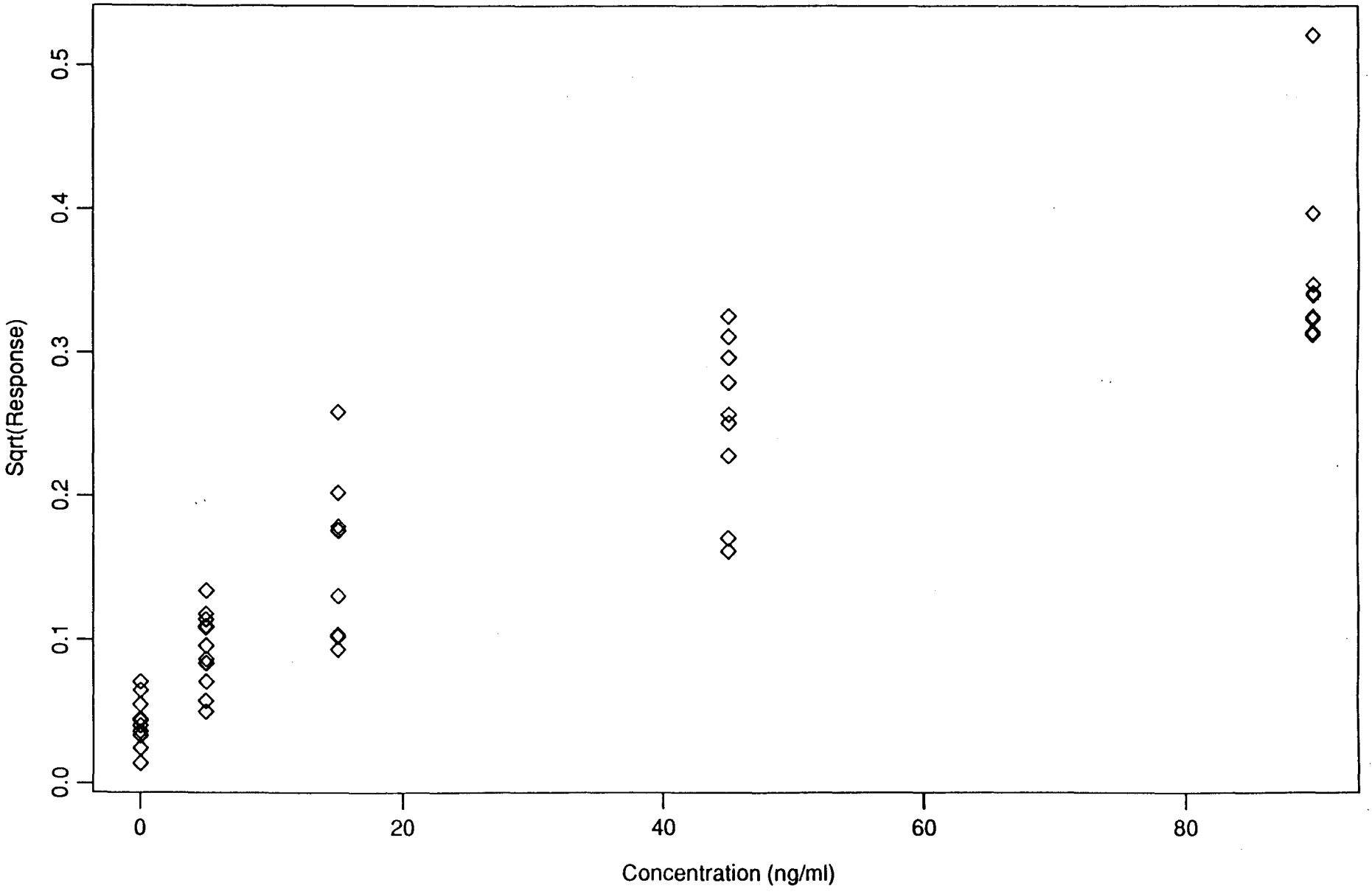

Figure 5. Log(sd) versus log(ybari)

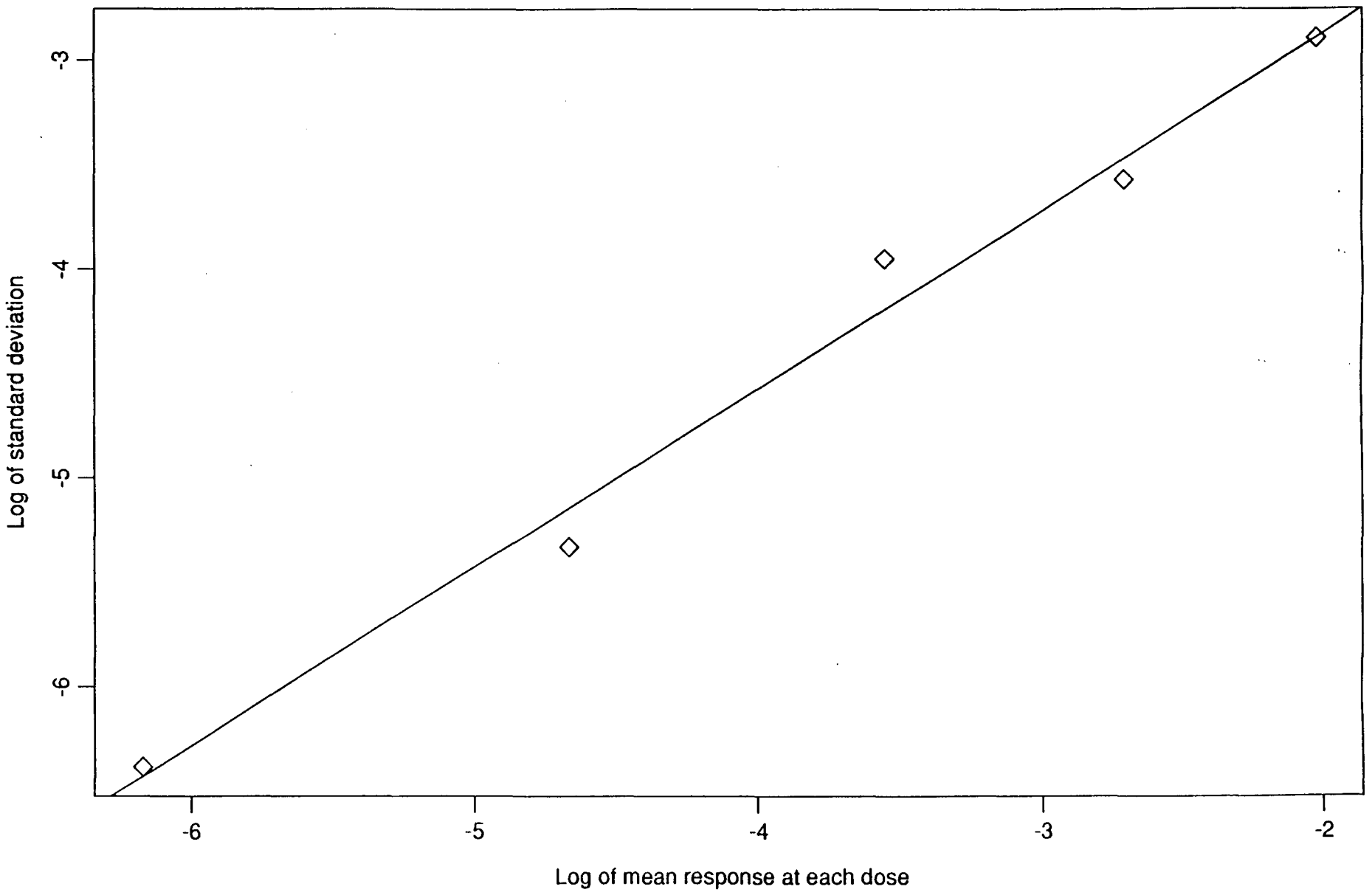

Figure 6. Log(abs LS resids) versus log(LS pred)

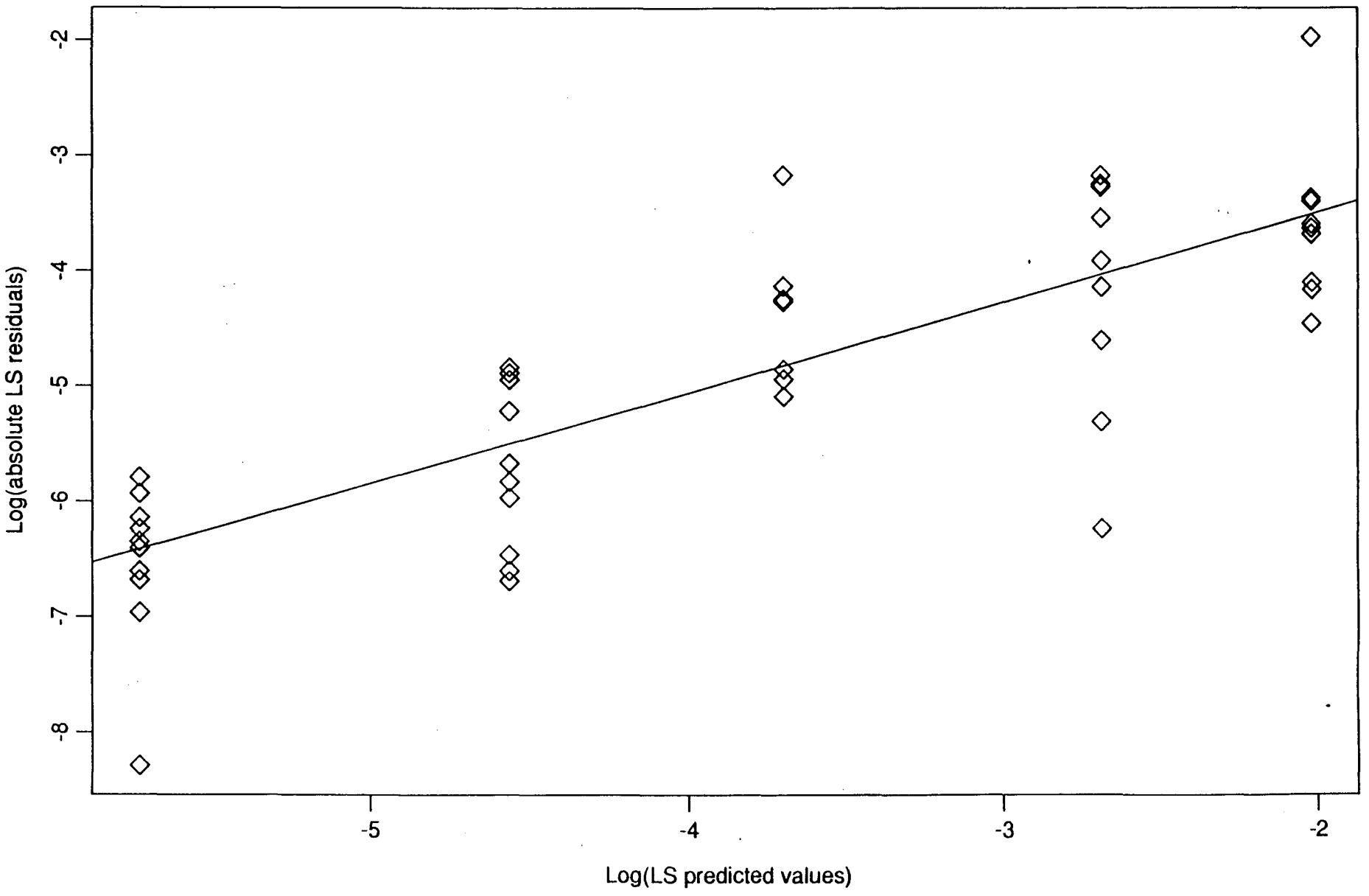

Figure 7. GLS residuals with theta=1

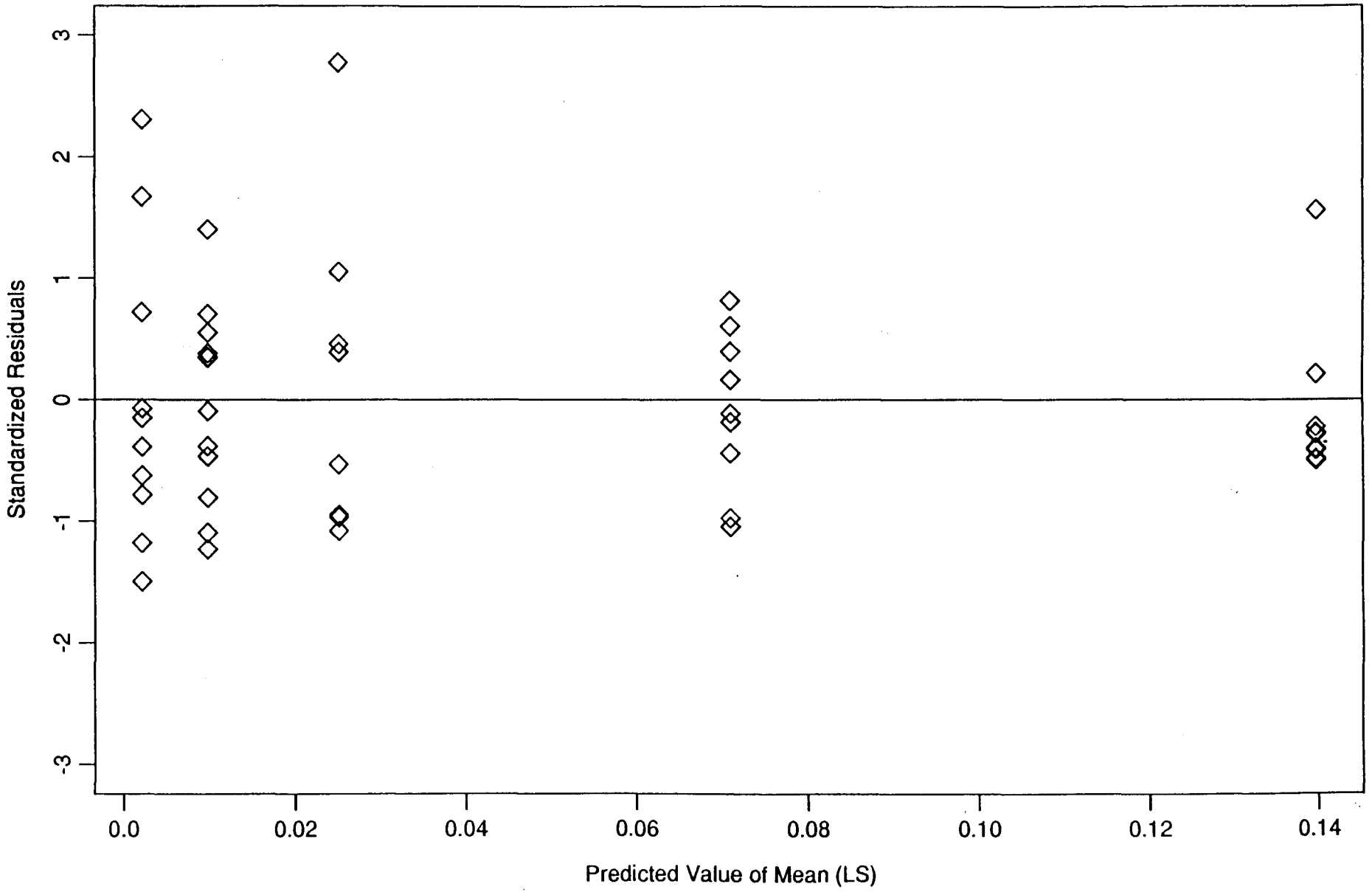

Figure 8. GLS residuals with theta estimated based on squared residuals

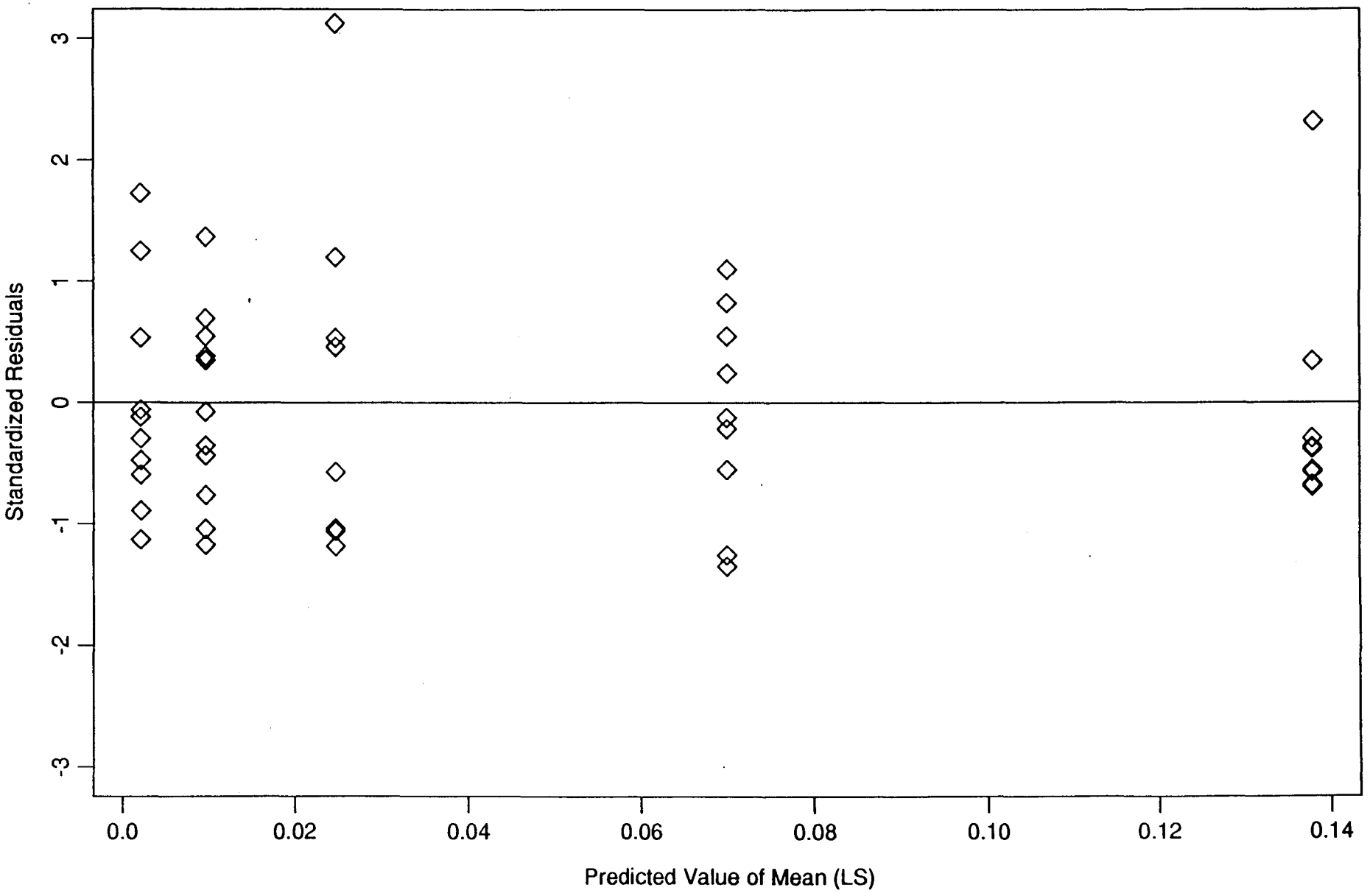

Figure 9. LS regression line and prediction limits

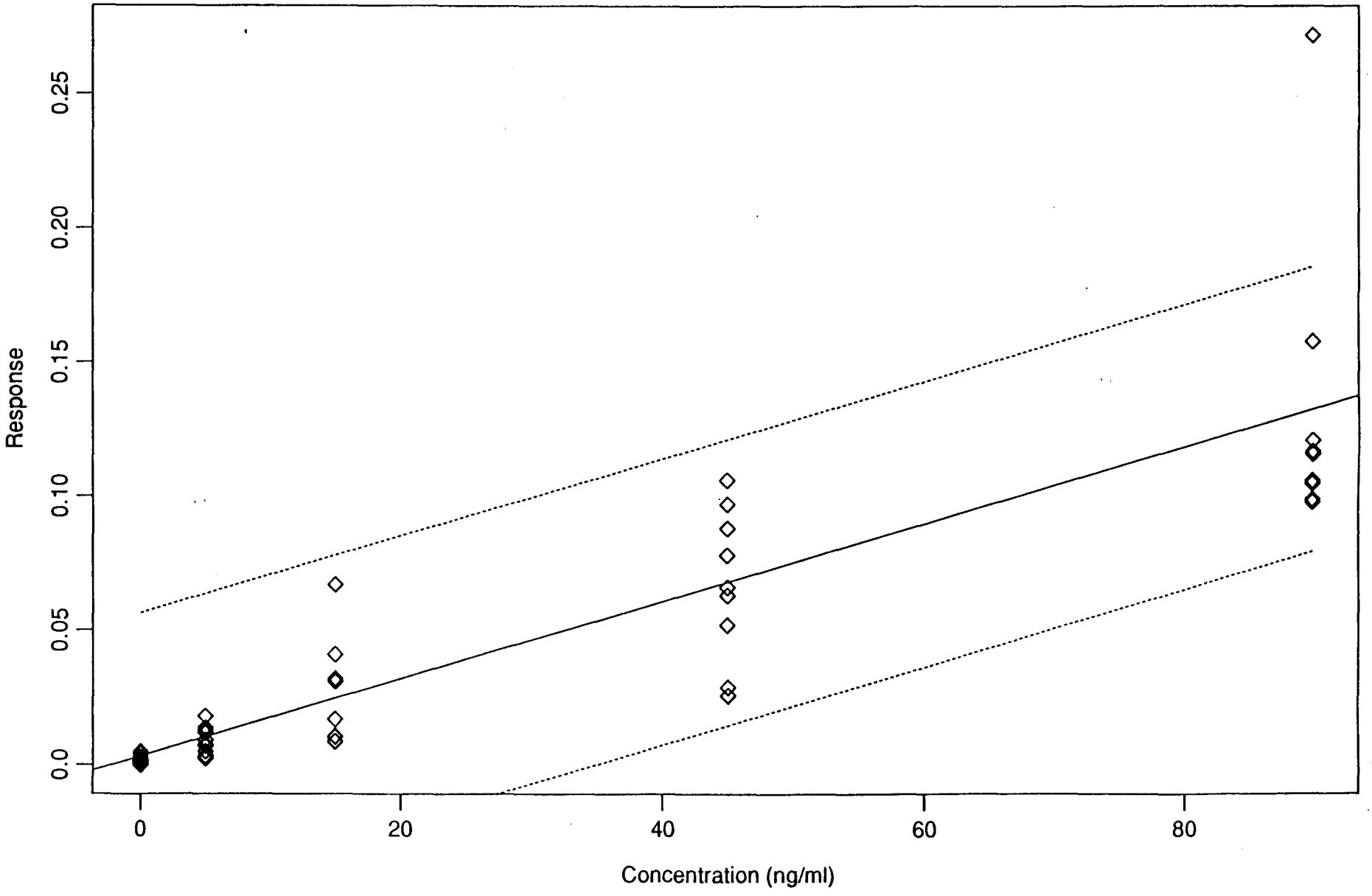

Figure 10. GLS regression line and prediction limits