Bias and Inconsistency in Linear Regression

Joel Tellinghuisen†

Department of Chemistry, Vanderbilt University, Nashville, Tennessee 37235

Received: June 30, 2000; In Final Form: October 12, 2000

Linear regression yields minimum-variance, unbiased estimates of the adjustable parameters, provided only that the analyzed data be unbiased and of finite variance. If further the data are normally distributed, then so will be the estimated parameters. But frequently data are transformed before fitting, and if the original data are normal, the transformed data may not be. In particular, inversion and logarithmic conversion yield biased, non-Gaussian distributions, so least-squares analysis of such data yields biased, nonnormally distributed parameters, even when the transformed data are properly weighted in accord with the transformation. Monte Carlo calculations are used to study the effects of such nonnormal data in cases of relevance to the analysis of equilibrium and kinetics data (exponential decay, binding constants, enzyme kinetics, fluorescence quenching, adsorption). Typically 10^5 equivalent data sets are processed to obtain precise information about the parameter biases and distributions. The biases generally persist in the limit of an infinite number of data values, which means that the estimators are not only biased but also inconsistent.

Introduction

The statistical properties of linear least-squares (LLS) estimators are well established:[1,2] If the data are distributed independently and normally (i.e., with the Gaussian distribution) about their true values, the estimators of the parameters will be *unbiased* and *minimum-variance* and will themselves be normally distributed about their true values, with variances that can be calculated from the *matrix of the normal equations*. These results hold also if the data are heteroscedastic, i.e., of unequal uncertainty (σ_{yi}), provided they are weighted properly,

$$w_i \propto \sigma_{yi}^{-2} \tag{1}$$

If the weights are neglected, the parameter estimates remain unbiased, but are no longer minimum-variance.[2] In other words, neglect of weights biases the error estimates but not the parameter estimates.

Actually, the data need not be normally distributed in order for least-squares to yield unbiased, minimum-variance estimates of the parameters. But it *is* necessary for them to be finite-variance and unbiased.[2] There are many situations in experimental physical science where data are transformed to facilitate analysis by linear regression. It is by now well recognized that such transformations alter the relative weights of the data, so that usually the resulting LS analysis becomes a weighted fit.[2−6] However, it has not been appreciated that they can *bias* the data and hence also the parameter estimates. Moreover, this "data bias" does not diminish with increasing numbers of data values n in the data set; in fact, it can actually *increase* with n. The resulting estimators are *inconsistent*, and at some n the biases in the parameters must exceed their standard errors. To illustrate this point, Figure 1 shows two nominally equivalent least-squares analyses of a single set of data having constant, normally distributed error in y. The unweighted nonlinear fit and weighted linear fit yield parameter values that disagree by $>3\sigma$ and hence fail many consistency tests.

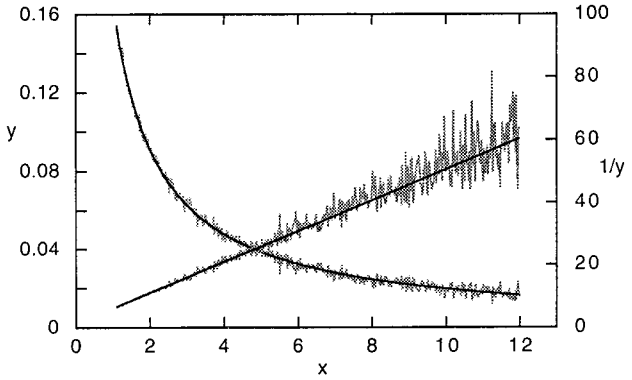

Figure 1. Comparison nonlinear and weighted linear fits to a data set of 501 points spread evenly between $x = 1.1$ and $x = 12$, with constant error ($\sigma_y = 0.002$) in y. The nonlinear fit is an unweighted fit to $y = (A + Bx)^{-1}$ and yields $A = 0.977(41)$, $B = 5.011(21)$, and $\chi^2 = 514$. The weighted linear fit yields $A = 1.091(40)$, $B = 4.932(21)$, and $\chi^2 = 500$. The "true" parameter values are $A = 1$ and $B = 5$, and the values in parentheses represent 1 standard error in terms of the final digits.

Figure 1 illustrates effects of one of two frequently used data transformations, reciprocation. The other, logarithmic conversion, is often used to render first-order kinetics data and thermodynamic temperature dependence into linear functions of the independent variables.[4,5] Reciprocation is employed in a variety of situations, including analysis of equilibrium and binding constant data,[6−11] enzyme kinetics,[12−14] adsorption isotherms,[15,16] and fluorescence quenching.[17−19] If we represent the transformed dependent variable by u, then simple error propagation yields, for logarithmic conversion, $\sigma_u = \sigma_y/y$, and for reciprocation, $\sigma_u = \sigma_y/y^2$. The weights thus become functions of the measured y values, and this can contribute significantly to the bias in the LS parameters.[20]

There is another potentially serious consequence of the inversion transformation. If the original variate y is normal, then its inverse $u = 1/y$ not only is nonnormal, but also has *infinite variance*. Thus, the transformed data violate one of the prime preconditions for LS, which in turn means that the resulting

† E-mail: tellinjb@ctrvax.vanderbilt.edu, FAX: 615-343-1234.

10.1021/jp002332w CCC: $19.00
Published on Web 11/22/2000

LS parameters are also characterized by infinite variance. In practice, of course, such divergence will be impossible to recognize in a single data set. And as long as the data are not excessively uncertain, the variances of their inverses will be well-defined in an asymptotic sense, leading to similarly well-defined LS parameters.

To demonstrate the significance of these problems, I have carried out Monte Carlo (MC) calculations on the linear regression of logarithmically and reciprocally transformed data. The computations employed methods like those used in my recent study of the nonlinear analysis of normal data.[20] As will be shown below, there is nothing pathological about the particular data set that produced the results in Figure 1; rather, this behavior is a predictable consequence of the "data bias" resulting from the inversion of the y_i values to yield the straight-line relationship. In simplest terms the bias in this case just reflects the inequality, $\langle y^{-1}\rangle \neq \langle y\rangle^{-1}$.

Background

The theory of least-squares has been summarized recently[20] and so will be covered only briefly here. The LS equations are obtained by minimizing the sum of weighted squared residuals S,

$$S = \Sigma w_i \delta_i^2 \tag{2}$$

with respect to a set of adjustable parameters $\boldsymbol{\beta}$, where δ_i is the residual (observed-calculated mismatch) for the ith point, w_i is its weight, and the column vector $\boldsymbol{\beta}$ contains p elements, one for each adjustable parameter. The problem is a linear one if the measured values of the dependent variable (y) can be related to those of the independent variable(s) (x, u, ...) and the adjustable parameters through the matrix equation,[2,20,21]

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\delta} \tag{3}$$

where $\mathbf{y}$ and $\boldsymbol{\delta}$ are column vectors containing n elements (for the n measured values) and the *design matrix* $\mathbf{X}$ has n rows and p columns and depends only on the values of the independent variable(s) (assumed to be error-free) and not on the parameters $\boldsymbol{\beta}$ or dependent variables $\mathbf{y}$. The solution to the minimization problem is the set of equations

$$\mathbf{X^T W X}\boldsymbol{\beta} \equiv \mathbf{A}\boldsymbol{\beta} = \mathbf{X^T W y} \tag{4}$$

where the square weight matrix $\mathbf{W}$ here is diagonal, with n elements $W_{ii} = w_i$. Equations 4 are solved for the parameters $\boldsymbol{\beta}$, e.g.,

$$\boldsymbol{\beta} = \mathbf{A}^{-1}\mathbf{X^T W y} \tag{5}$$

where $\mathbf{A}^{-1}$ is the inverse of $\mathbf{A}$ (the matrix of the normal equations). Knowledge of the parameters permits calculation of the residuals $\boldsymbol{\delta}$ from eq 3 and thence of S, which in matrix form is

$$S = \boldsymbol{\delta}^{\mathbf{T}}\mathbf{W}\boldsymbol{\delta} \tag{6}$$

If the parent distributions for the data are normal and the proportionality constant in eq 1 is taken as 1.00, then the quantity S is distributed as a χ^2 variate for $\nu = n - p$ degrees of freedom.[1,2] Correspondingly, the quantity S/ν follows the reduced chi-squared (χ_ν^2) distribution, given by

$$P(z)\,dz = Cz^{(\nu-2)/2}\exp(-\nu z/2)\,dz \tag{7}$$

where $z = \chi_\nu^2$ and C is a normalization constant. A χ^2 variate has a mean of ν and a variance of 2ν,[22] which means that χ_ν^2 has a mean of unity and a variance of $2/\nu$.

Note that $\boldsymbol{\beta}$ (eq 5) is independent of an arbitrary scale factor in the weights, hence the simple proportionality in eq 1. Also, $\mathbf{X}$ is completely determined by the selection of values for the independent variables. Thus, if the fitted quantities are the measured data $\mathbf{y}$ themselves (rather than any transforms thereof), and if their error structure is known at least relatively, $\mathbf{A}$ ($=\mathbf{X^T W X}$) is completely known before any measurements of $\mathbf{y}$ have been made. Since $\mathbf{V} \propto \mathbf{A}^{-1}$, it too is known to within a scale factor, permitting its use in *experimental design*.

If the data error structure is known absolutely, then the proportionality constant in eq 1 can be taken as unity, whence the proportionality constant connecting $\mathbf{V}$ and $\mathbf{A}^{-1}$ is likewise unity, giving

$$\mathbf{V} = \mathbf{A}^{-1} \tag{8}$$

If additionally, the parent data distributions are normal, the parameter distributions are also normal, as already noted. Thus, $\mathbf{V}$ is known exactly at the outset, and the confidence intervals for the parameters can be evaluated from standard error function tables. Any failure to observe this result in MC calculations on LLS represents a flaw in the computational procedures.

Since the use of eq 8 implies prior knowledge of the statistics of the y_i, the corresponding weights may be designated as *a priori* weights.[23,24] At the other extreme is the situation where nothing is known about the statistics of the y_i, except that the parent distributions are assumed to be normal, with constant variance, independent of y_i. In this case the weights w_i can be set to 1.00, giving unweighted regression. The variance in y is then estimated from the fit itself, as

$$\sigma_y^2 \approx s_y^2 = \frac{\Sigma\delta_i^2}{n-p} = \frac{S}{\nu} \tag{9}$$

This represents an *a posteriori* assessment of the variance in y_i. Correspondingly,

$$\mathbf{V} = \frac{S}{\nu}\mathbf{A}^{-1} \tag{10}$$

Under the same conditions as stated just before eq 7, s_y^2 is distributed as a scaled χ^2 variate.

Given the probability distribution function $P(y)$, one can obtain the distribution function $Q(u)$ for a second random variable u related to y by the function $u(y)$ using[1]

$$Q(u) = \sum P[y_i(u)]\left|\frac{dy_i(u)}{du}\right| \tag{11}$$

where the sum is over all points y_i that solve $u = u(y_i)$. If data y_i are normally distributed about their true values and are inverted to $u_i = 1/y_i$, eq 11 yields

$$Q(u) = \frac{C}{u^2}\exp\left[-\frac{1}{2\sigma_y^2}\left(\frac{1}{u} - y_0\right)^2\right] \tag{12}$$

where y_0 is the true value of y (for the ith point) and C is a normalizing constant. This function has Lorentzian wings; therefore, as already noted, the variance of u is infinite. However, as was also noted previously, this divergence will be of no practical concern provided $|\sigma_y/y|$ is sufficiently small, whereupon the variance of u will be well-defined in an asymptotic sense.[20] On the other hand, when $|\sigma_y/y_0| \gtrsim 1/3$,

TABLE 1: Monte Carlo Statistics of $A = 1$ and Its Reciprocal, from 10^5 Values with Normally Distributed Random Error of Specified σ_A[a]

σ_A	$\langle A\rangle$	$\langle A^2\rangle - \langle A\rangle^2$	$\langle A^{-1}\rangle$	$\langle A^{-2}\rangle - \langle A^{-1}\rangle^2$	$\langle A^{-1}\rangle_{true}$[b]
0.05	0.99996	2.491×10^{-3}	1.00255	2.542×10^{-3}	1.00252
0.10	0.99992	9.965×10^{-3}	1.01036	1.083×10^{-2}	1.01032
0.15	0.99988	2.242×10^{-2}	1.02425	2.745×10^{-2}	1.02422
0.20	0.99984	3.986×10^{-2}	1.04629	6.447×10^{-2}	1.04623
0.20	1.00116	3.997×10^{-2}	1.04473	5.997×10^{-2}	1.04623
0.35	0.99961	1.220×10^{-1}	1.14868	1.367×10^{2}	1.20093
0.35	1.00202	1.224×10^{-1}	1.15218	2.569×10^{2}	1.20093
0.35	0.99972	1.221×10^{-1}	0.89179	1.158×10^{4}	1.20093
0.35	1.00107	1.222×10^{-1}	1.05755	3.191×10^{3}	1.20093

[a] Same seed used for first four σ_A values, to illustrate the effects of scaling for a given set of unit-variance normal deviates. [b] Obtained by numerical integration.

sampling statistics for u show erratic behavior for both u_0 and σ_u^2. Although $\langle u\rangle$ *is* mathematically defined in this case (in the sense of the Cauchy principal value), the sampling estimates of $\langle u\rangle$ do not converge, because the central limit theorem does not apply.

Even if the data are narrowly enough defined to avoid catastrophic divergence problems in inversion, the transformed variates u_i remain nonnormal and are biased estimators of the true $u_{0,i}$. Thus, even a linear LS fit can be expected to yield biased estimates of the parameters $\boldsymbol{\beta}$. This result assumes proper weighting of the data following the transformation, and as already noted, the weighting actually contributes to the bias. Likewise, data transformed using $u_i = \ln y_i$ will yield biased parameter estimates. In this case we have the additional problem that the transformation fails for negative y_i, though with the restriction to positive y_i the variance of u_i does remain finite.

From a purely phenomenological standpoint, slightly nonnormal distributions can be represented as skewed Gaussians,

$$P(x) = C\,[1 + q(x - x_0)]\,\exp\left[-\frac{(x - x_0)^2}{2\sigma_x^{\,2}}\right] \qquad (13)$$

where $C^{-1} = \sqrt{2\pi}\sigma_x$ and the skew parameter q is determined empirically. This distribution yields for the bias in x,

$$\langle x - x_0\rangle = q\sigma_x^{\,2} \qquad (14)$$

The bias thus scales as the variance, a dependence which holds quite generally for modest bias.[20]

Results and Discussion

Statistics of Reciprocals. The anomalous statistics of reciprocals can be demonstrated with the simplest kind of Monte Carlo calculation: Generate normal error of specified magnitude on a quantity and then examine the MC statistics of it and its reciprocal. Some results of such calculations are given in Table 1 and displayed in Figure 2.

For small relative σ_A the reciprocal seems well-behaved; but its average is biased by an amount that scales roughly with σ_A^2, as predicted by eq 14, and its estimated variance is systematically larger than the "true" value (the asymptotic value from error propagation), the disparity increasing with σ_A. The last four averages clearly reflect the divergence. At the same time the estimates of A and its variance behave as expected: The scatter in the four $\langle A\rangle$ values for $\sigma_A = 0.35$ is consistent with the standard error, $\sigma_A/N^{1/2} = 0.0011$, while that in the estimated variance is well within the predicted relative standard error of $(2/N)^{1/2} = 0.0045$.

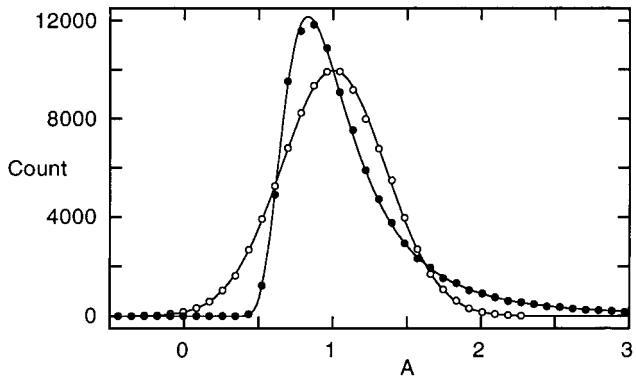

Figure 2. Histogrammed results of 10^5 Monte Carlo estimates of a constant $A = 1$ (open points) and its reciprocal, with random normal error of $\sigma_A = 0.35$ superimposed upon A. The curves are the LS fits of the properly weighted histogram counts to a Gaussian for A and eq 12 for A^{-1}. (The statistical errors in the counts are smaller than the plotted points in this and most subsequent figures; this may be assumed in cases where they are not shown.)

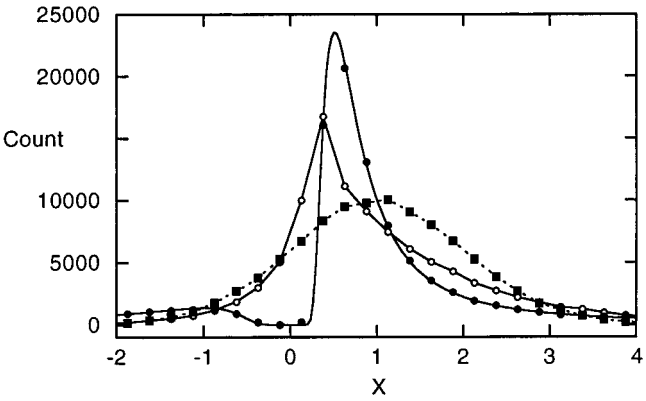

Figure 3. Histogrammed results of 10^5 Monte Carlo values of products and ratios of normal variates. In all cases the mean values are $X = 1.0$. Open points: $X = A \times B$, $\sigma_A = \sigma_B = 1$. Filled circles: A/B, $\sigma_A = 0.1$, $\sigma_B = 1$. Squares: A/B, $\sigma_A = 1$, $\sigma_B = 0.1$. For the filled circles, the smooth curve represents a fit to eq 12; in the other cases the lines are just linear connections between adjacent points.

A useful variation on this theme is the question of distributions of products and ratios of independent normal variates. Neither of these is normally distributed, and the deviations from normal are immediately obvious if both variates have standard deviations comparable to their means in magnitude (see Figure 3). However, as both variates become relatively well defined ($\sigma \approx |\mu|/10$), the distributions of products and ratios become at least roughly normal, with standard deviations that follow the normal rules of error propagation (e.g., if q represents either A/B or $A \times B$, then $\sigma_q/q = [(\sigma_A/A)^2 + (\sigma_B/B)^2]^{1/2}$). In fact one can verify that this holds quite well if even one of the two variates has a narrow distribution, but with one exception: The ratio A/B will demonstrate "reciprocal statistics" if the distribution of B is broad, no matter how narrowly defined A is. Recognition of this behavior is important for understanding the anomalous statistics reported in some MC studies.[20]

Despite their poor behavior when subjected to sampling statistics, reciprocal normal variates can still be characterized with respect to confidence limits from MC calculations, through simple sorting procedures. Also, because of the one-to-one mapping between quantities and their inverses, the range bracketing a given fraction of a normal variate yields directly, through inversion, the equivalent range for the reciprocal variate, and vice versa.

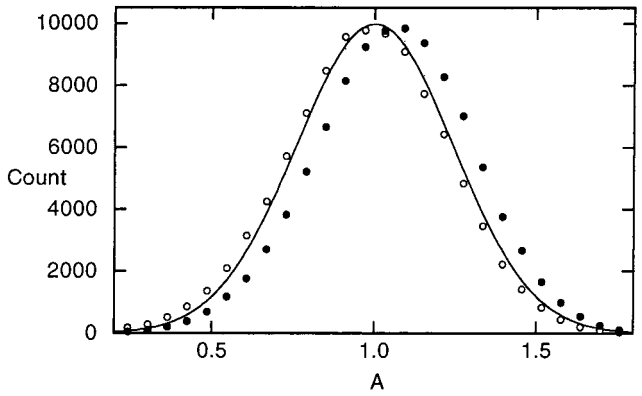

Figure 4. Histogrammed results of 10^5 LS estimates of the intercept A from fits of 5-point data sets to the linear model $1/y = A + Bx$. The smooth curve is a Gaussian that represents the results predicted for normally distributed error in $(1/y)$. The points were obtained for normally distributed error in y, with weights calculated using the theoretical y values (open points) or the "observed" y values (filled).

Fitting Reciprocal Data to a Linear Model. To illustrate the effects of bias from inversion, I have used the straight-line model, $1/y = A + Bx$, with $A = 1$, $B = 5$, and the five x values 1.1, 3.3, 5.5, 8.3, and 12.0. As a benchmark for the results obtained when the nonnormal inverted data are fitted, I have also run MC calculations for normally distributed error in the fitted quantities $(1/y)$, but having magnitude proportional to y^{-2}, as evaluated on the theoretical line. For this example, the error in y was taken as 0.002, which is about one-eighth the magnitude of the smallest value (at $x = 12$).

Results obtained for normally distributed error in the fitted quantity $(1/y)$ bore out expectations: normally distributed, unbiased parameter estimates having standard errors as predicted by eq 8, and S/ν values that follow the reduced chi-square distribution of eq 7. When the weights were neglected, the parameters remained normal and unbiased, but their standard errors increased by factors of 3 (for B) and 9 (A). For comparison, eq 10 predicts an even larger σ_A but a smaller σ_B, which illustrates the unreliability of the variance-covariance matrix when heteroscedastic data are not properly weighted. Neglect of weights also yielded S/ν values whose distribution was grossly in disagreement with the χ_ν^2 distribution for $\nu = 3$ (after a necessary rescaling to an average value of unity in this case).

When the errors are normal in y instead of $1/y$, the parameter distributions are no longer Gaussian, as is illustrated in Figure 4. The proper weights in this case are y_i^4/σ_y^2. As the "rollers of the dice" in this Monte Carlo game, we have the choice of evaluating the weights using either the "true" or the "observed" y_i values. The former choice results in a negative bias in A of $-2.55(8)\%$, and a positive bias in B of about the same absolute magnitude. The latter weighting choice doubles the biases in magnitude but reverses their signs. The dispersion parameters are also biased by statistically significant amounts, as compared with the "true" values (calculated for normal error in $1/y$): by $+1.1(2)\%$ (σ_A, theoretical weighting), $+0.3(2)\%$ (σ_A, observed), $+1.5(2)\%$ (σ_B, theoretical) and $+0.4(2)\%$ (σ_B, observed). Thus, despite the larger biases in the parameters, the "observed" weighting (which is the only one available for the linear analysis of actual data) yields smaller parameter variances. Both distributions of S/ν values resemble the theoretical χ_ν^2 distribution for $\nu = 3$ but still deviate from it by statistically significant amounts (Figure 5).[25]

Figure 6 illustrates that the bias scales roughly with the data variance, as predicted by eq 14. To obtain the n dependence it

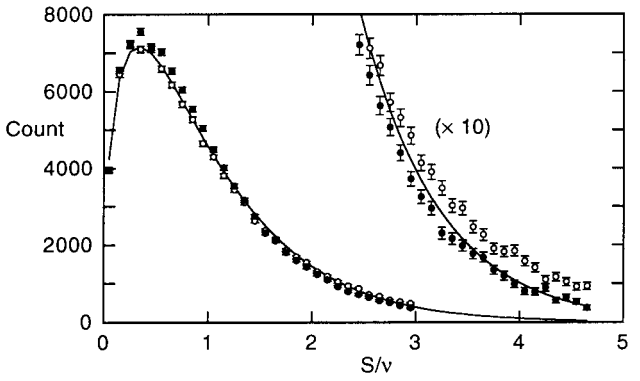

Figure 5. Histogrammed S/ν values from same fits that produced Figure 4. Error bars represent 1 σ. The smooth curve is the theoretical χ_ν^2 distribution for $\nu = 3$. In both cases, weighted fits of the binned values to eq 7 yielded values of $\chi^2 > 300$, for $\nu = 46$.

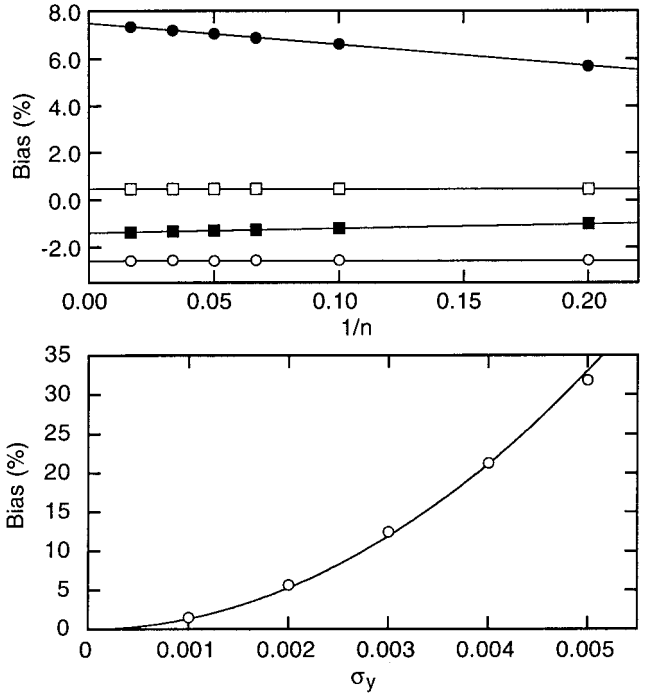

Figure 6. Dependence of parameter bias (%) on σ_y (lower) and on the number of data points. The upper plot displays the bias in the intercept A (circles) and slope B (squares) for theoretical (open points) and "observed" weighting, with all results obtained using $\sigma_y = 0.002$. The lower plot shows the bias in A ("observed" weighting) for $n = 5$, with the smooth curve being a fitted quadratic through the origin.

is important to preserve the x-structure of the data; this was done by simply doubling, tripling, etc. the number of points at each x_i. The key result here is that none of the biases vanishes in the limit $n \rightarrow \infty$, which demonstrates that both estimators are inconsistent for both weighting choices. Moreover, since the parameter standard errors scale as $n^{-1/2}$, it is clear that the bias must exceed the standard error for suitably large n. It was this realization that led to the preparation of Figure 1, which illustrates statistically significant bias for the weighted linear fit of the transformed data. By contrast, the nonlinear estimators from an unweighted fit of the y values are consistent (though biased for finite n).[20]

The parameter standard errors and biases both depend on the x-structure of the data. For example, if the first point is moved from $x = 1.1$ to $x = 0$, the statistics for the intercept become essentially those of $1/y$ at that point: The bias practically vanishes, and the standard error drops to 0.002. At the same

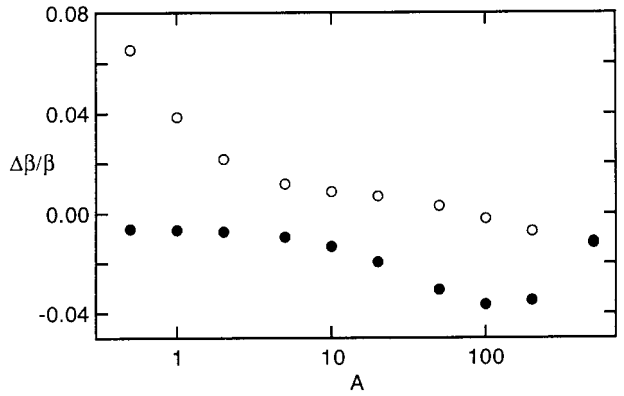

Figure 7. Relative parameter bias $\Delta\beta/\beta = (\langle\beta\rangle - \beta_{true})/\beta_{true}$, as a function of *A*, for LLS model $1/y = A + Bx$, with $B = 5$: (open points) *A*; (filled points) *B*. Standard errors (not shown) are mostly smaller than the displayed points. For this and the following figure σ_y is fixed at 10% of the smallest *y* value, and the weights are assessed using "observed" y_i.

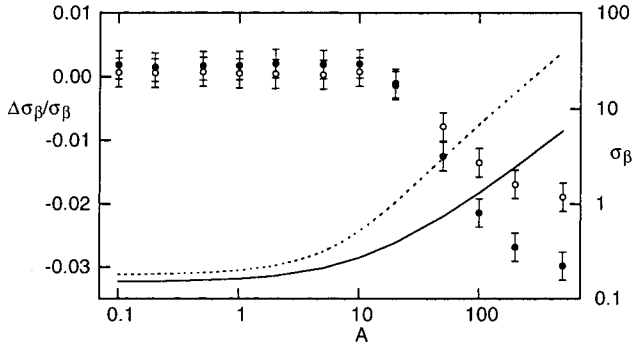

Figure 8. Relative bias in parameter errors, $\Delta\sigma_\beta/\sigma_\beta = (\sigma_\beta - \sigma_{\beta,true})/\sigma_{\beta,true}$, as a function of *A*, for same linear model. Here $\sigma_{\beta,true}$ is the value calculated from eq 8. Error bars represent one standard error. Also shown are the parameter standard errors (curves, ordinate scale to right). Open points and dashed curve: *A*. Filled points and solid curve: *B*.

time, the bias and precision of the slope are hardly affected. Thus, in certain situations (Stern−Volmer quenching data, for example), where data are easily recorded at $x = 0$, intercept bias may not be a problem.

Similarly, the bias and precision are also functions of the *x* range of the data, or equivalently, of the relative magnitudes of *A* and *B* for a given *x* range. For example, in the limit where the *y* range of the data is small compared with *A*, the fit becomes de facto unweighted, and the bias is attributable to just the non-Gaussian nature of the data. Quantitative results for the bias in both the parameters and their errors are shown in Figures 7 and 8. For the purpose of these illustrations, σ_y was fixed at 10% of the smallest *y* value, which means that it varies with *A*. The relative bias in *A* rises to 0.34 at $A = 0.1$ (not shown). Expressed as a fraction of σ_β, the largest error bias under these conditions is $+0.19\ \sigma_A$ for *A* and $-0.25\ \sigma_B$ for *B*.

The results in Figures 7 and 8 were all obtained for five-point data sets. The previously noted $1/n$ dependence of the bias in both *A* and *B* holds for all *A*; however, the slopes of this dependence and the infinite-*n* limiting values both vary with *A*, and in no simple way.

Although the bias is statistically significant in both the parameters and their standard errors, the latter bias should rarely be a source of concern. Stated differently, any hypothesis that is dependent on changes in the confidence limits as small as those illustrated in Figure 8 is in need of additional data rather than additional statistical interpretation. In short, nothing in these

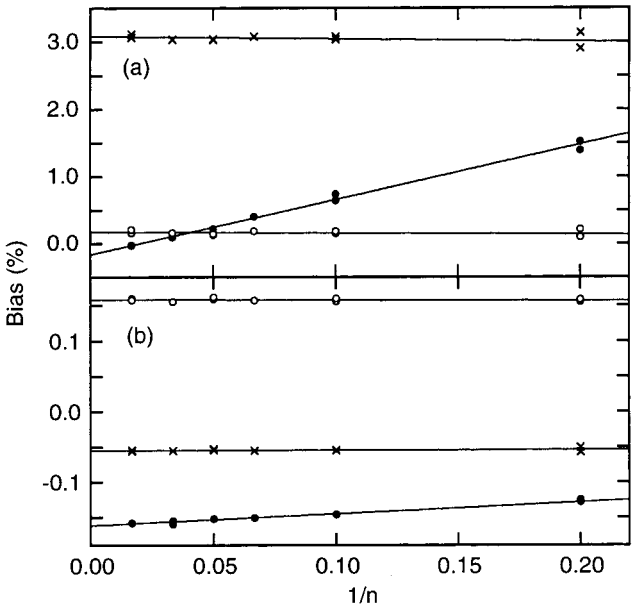

Figure 9. Percent bias in intercept *A* (a) and slope *B* (b) as functions of $1/n$, for 4% proportional error in *y* (circles), and for $\sigma_{yi} = 0.5\ y_i^2$ (×). For proportional error, open points represent results obtained evaluating the weights using "true" y_i, while filled were obtained by weighting on "observed" y_i. For most values of $1/n$, two points are plotted; MC statistical errors (not shown) are comparable to the spread in these values. For reference, the "exact" standard errors for five-point data sets are $\sigma_A = 0.336$ and $\sigma_B = 0.1355$ (4% error) and $\sigma_A = 0.419$ and $\sigma_B = 0.0586$ ($\sigma_{1/y} = 0.5$).

tests indicates a need to go beyond eqs 8 and 10 to assess parameter errors for the linear model.

The observation of opposite signs for the biases from the two weighting options in Figure 6 suggests that the total bias includes contributions that work in opposite directions from the data themselves and from their variance transformation. For a closer look at these two contributions, I examined two other error structures: 4% relative error in *y*, and $\sigma_y \propto y^2$. The latter represents an unlikely structure for experimental data, but it is useful here since it yields an unweighted fit in the transformed data ($1/y$); the former yields 4% relative error in the transformed data also, and requires weights $\propto y_i^2$ (cf. y_i^4 for constant σ_y). The results (Figure 9) show that the biases are mostly smaller in magnitude than for constant σ_y; however, the estimators for *A* and *B* remain inconsistent in every case. It is noteworthy also that the largest bias and inconsistency occur for *A* in the case of $\sigma_y \propto y^2$, where there can be no "weighting" contribution.

Linear Fitting of Logarithmically Transformed Data. Logarithmic conversion is often used to achieve a linear dependence for background-free exponential data. As was noted earlier, if $z = \ln y$, $\sigma_z = \sigma_y/y$. Thus, if σ_y is constant, the linear log fit will require weights proportional to y_i^2, while if $\sigma_y \propto y$, the log fit is properly an unweighted one. I have examined both of these cases here, using the model, $y = ae^{-bx}$, with $a = 1.5$ and $b = 1$.

Consider first the case of constant σ_y. The magnitude of σ_y is limited by practical considerations, because if it is too large compared with the smallest *y* value, negative values of y_i will occur. Keeping $\sigma_y \lesssim {}^1/_5$ of the smallest *y* makes such occurrences adequately rare. In most of the calculations this ratio was $\sim{}^1/_{10}$, but in some it was increased to ${}^1/_5$. The reference data set had 5 (sometimes 4) points covering from $x = 0$ to 1.5−2.4.

Qualitatively, the results of these calculations resembled those reported for inverse data, including nonnormal parameter

distributions and S/ν distributions that did not quite fit the χ_ν^2 distribution. For five x values spread evenly over the range $x = 0-2$, with $\sigma_y = 0.02$, the bias in b was -0.24%, or -10% of σ_b. With the upper limit extended to $x = 2.4$ (where this σ_y represents 15% of y_{min}), the bias was -0.33%, while a decrease in x_{max} resulted in a smaller absolute bias. Thus, at least roughly the bias tracks with σ_y/y_{min}. The bias again scales as σ_y^2. It is also linear in $1/n$ and *increases* in magnitude with increasing n. Thus, as before, for sufficiently large n the bias in b must exceed σ_b. In the case of data spanning $x = 0-2$ and having $\sigma_y = 0.04$ (20% of y_{min}), this occurs for <100 points, where the bias is -1.25%.

To check the effects of proportional uncertainty, I again took σ_y to be 4% of y. The now unweighted linear fits yielded parameter distributions that were borderline normal, and the S/ν histogram was similarly close to the χ_ν^2 distribution. The bias in $\langle b \rangle$ was not significant at the 2σ level, and that in $\langle \ln a \rangle$ was small, with no statistically significant dependence on n. Increasing the data error to 10% confirmed the σ_y^2 scaling of the bias in the intercept but still gave no statistically significant bias in the slope. With 4% error, the MC statistical estimates of σ_b^2 were consistent with the "exact" values.

These results may be compared with those obtained from the direct nonlinear analysis of normal data.[20] For the distributions, the only noteworthy difference was closer agreement with the χ_ν^2 distribution for the S/ν values in the nonlinear fits, especially for constant σ_y. The parameter biases were very small at constant σ_y for finite n but extrapolated to zero as $n^{-1} \rightarrow 0$. For proportional error there was no statistically significant bias in b for any n. However, in this case there *was* significant bias in a, and it remained finite at $n = \infty$. The MC and "exact" parameter variances were usually statistically consistent and showed a maximum spread of 1%. With addition of a background, the exponential parameter remained consistent for all weighting choices, but the preexponential and background parameters were inconsistent for some error structures.

Conclusion

The biases identified in this work are significant enough to become obvious in certain circumstances, as illustrated in Figure 1. Since the biases persist in the limit of an infinite number of data points, the pertinent estimators are also *inconsistent*. By comparison, nonlinear LS fits of normal data usually display parameter bias for finite n; however, in the cases examined to date, the key nonlinear estimators are *consistent*.[20] Inconsistency is a more serious problem than bias. Put simply, inconsistency means that for sufficiently large n, the estimator is *guaranteed* to yield the "wrong" answer by any statistical test.

Inconsistency seems to be a given in the fitting of transformed data, so a general recommendation of the present study is to apply nonlinear models to normal data rather than linear models to nonnormal (transformed) data. The one exception to this rule is the case of background-free exponential data with proportional y error, for which the transformed fit is unweighted and yields unbiased estimates of the exponential parameter for all n (provided the data error is small enough to ensure the success of the transformation). However, one cannot generalize from this result, because the corresponding case of an unweighted fit in the reciprocally transformed data ($\sigma_y \propto y^2$; $\sigma_{1/y} =$ constant) yielded strong bias and inconsistency in the intercept A (Figure 9).

Parameter bias has been noted previously in two different cases of straight-line fitting: (a) when there is error in both variables;[26] and (b) in "inverse calibration," where the error-free variable is treated as uncertain while the response variable is taken as error-free.[27-29] The former case is more properly considered as nonlinear fitting, because the equations cannot be expressed in the algebraically linear form of eqns 3−5. A closer examination of the dependence of the biases on n for the cases considered in ref 26 indicates that the slope and intercept estimators are both consistent.[30] In inverse calibration, on the other hand, both estimators are inconsistent;[27] and when used to estimate an unknown concentration, they give a bias that is actually worse for large n than for small n.[28,29]

Regarding the estimation of parameter errors and confidence limits, both the linear analysis of nonnormal data and the nonlinear analysis of normal data yield nonnormal parameter distributions. Accordingly, the standard errors from the variance−covariance matrix cannot fully characterize these distributions. However, the deviations from normality are seldom significant enough to yield errors greater than ~5% in these parametric error estimates. Thus, with exclusion of the special problems of "reciprocal parameters," confidence intervals should be assignable without need to resort to Monte Carlo calculations.[31]

A fundamental assumption of this work has been that the "raw data" possess normally distributed error. The fact that this is also an article of faith in most experimental work does not really excuse us from a long-standing neglect, namely the quest for real information about the error structure of the data. Fortunately there are situations where there is good reason to trust the normal assumption. For example, when each recorded datum is itself the result of averaging—in the computer logging of data, for instance—the central limit theorem guarantees the normal distribution in the large-n limit, no matter what the parent distribution for individual measurements (as long as the latter has finite variance). It does not take very much averaging to produce fairly normal means. For example, a sum of 12 uniform variates is often used as a simple yet reliable way to generate Gaussian variates having $\sigma = 1$ in Monte Carlo calculations!

References and Notes

(1) Mood, A. M.; Graybill, F. A. *Introduction to the Theory of Statistics*, 2nd ed.; McGraw-Hill: New York, 1963.
(2) Hamilton, W. C. *Statistics in Physical Science: Estimation, Hypothesis Testing, and Least Squares*; The Ronald Press Co.: New York, 1964.
(3) Deming, W. E. *Statistical Adjustment of Data*; Dover: New York, 1964.
(4) Cvetanovic, R. J.; Singleton, D. L.; Paraskevopoulos, G. *J. Phys. Chem.* **1979**, *83*, 50−60.
(5) Kalantar, A. H. *J. Phys. Chem.* **1986**, *90*, 6301−6303.
(6) Ramachandran, B. R.; Allen, J. M.; Halpern, A. M. *Anal. Chem.* **1996**, *68*, 281−286.
(7) Benesi, H. A.; Hildebrand, J. H. *J. Am. Chem. Soc.* **1949**, *71*, 2703−2707.
(8) Hoenigman, S. M.; Evans, C. E. *Anal. Chem.* **1996**, *68*, 3274−3276.
(9) Loukas, Y. L. *Analyst* **1997**, *122*, 377−382.
(10) Bowser, M. T.; Chen, D. D. Y. *J. Phys. Chem. A* **1998**, *102*, 8063−8071.
(11) Bowser, M. T.; Chen, D. D. Y. *J. Phys. Chem. A* **1999**, *103*, 197−202.
(12) Ritchie, R. J.; Prvan, T. *J. Theor. Biol.* **1996**, *178*, 239−254.
(13) Deng, Q.; Li, B.; Dong, S. *Analyst* **1998**, *123*, 1995−2000.
(14) Nakaminami, T.; Ito, S-i.; Kuwabata, S.; Yoneyama, H. *Anal. Chem.* **1999**, *71*, 1068−1076.
(15) Dickert, F. L.; Haunschild, A.; Kuschow, V.; Reif, M.; Stathopulos, H. *Anal. Chem.* **1996**, *68*, 1058−1061.
(16) O'Kelly, J. P.; Forster, R. J. *Analyst* **1998**, *123*, 1987−1994.
(17) Di Marco, G.; Lanza, M.; Mamo, A.; Stefio, I.; Di Pietro, C.; Romeo, G.; Campagna, S. *Anal. Chem.* **1998**, *70*, 5019−5023.
(18) Clements, J. H.; Webber, S. E. *J. Phys. Chem. A* **1999**, *103*, 2513−2523.
(19) Analysis of fluorescence quenching via the Stern−Volmer method usually involves plotting I_0/I as a function of quencher pressure or

concentration, with intercept fixed at unity. However, in most such cases the unquenched intensity I_0 is also subject to experimental error and is measured separately from I, so should properly be included among the data (as I at $x = 0$, multiple times if so measured) and as a fit parameter, giving intercept I_0^{-1}.

(20) Tellinghuisen, J. *J. Phys. Chem. A* **2000**, *104*, 2834−2844.

(21) Albritton, D. L.; Schmeltekopf, A. L.; Zare, R. N. An introduction to the least-squares fitting of spectroscopic data. In *Molecular Spectroscopy: Modern Research II*; Narahari Rao, K., Ed.; Academic Press: New York, 1976; pp 1−67.

(22) Abramowitz, M.; Stegun, I. A. *Handbook of Mathematical Functions*; Dover: New York, 1965.

(23) Tellinghuisen, J. *J. Mol. Spectrosc.* **1996**, *179*, 299−309.

(24) Hayashi, Y.; Matsuda, R.; Poe, R. B. *Analyst* **1996**, *121*, 591−599.

(25) There is another weighting option for the analysis of experimental data: Assess the weights using the *calculated* y_i values within each fit. However, this requires that the weights be adjusted iteratively, which makes the analysis nonlinear, in which case a nonlinear fit of the original data is normally preferable. This weighting was considered in ref 20.

(26) Kalantar, A. H.; Gelb, R. I.; Alper, J. S. *Talanta* **1995**, *42*, 597−603.

(27) Berkson, J. *Technometrics* **1969**, *11*, 649−660.

(28) Shukla, G. K. *Technometrics* **1972**, *14*, 547−553.

(29) Tellinghuisen, J. *Fresenius J. Anal. Chem.* **2000**, *368*, 585−588.

(30) Tellinghuisen, J. Unpublished.

(31) In most cases the parameter distributions are close enough to normal about the biased values to permit reliable estimation of confidence limits using the normal approximation. However, in cases involving relatively large σ_β/β, it may be necessary to quantify the asymmetry in the distribution.