COMMUN. STATIST. -SIMULA., 20(4), 843-886 (1991)

Strongly Heteroscedastic Nonlinear Regression

J. Ross Macdonald and William J. Thompson

Department of Physics and Astronomy
University of North Carolina
Chapel Hill, NC 27599-3255

Key Words and Phrases: nonlinear regression; heteroscedasticity; extended least squares; generalized least squares; error-variance models; weighted least squares; Monte-Carlo simulation; transformations; exponential models

ABSTRACT

We describe novel, analytical, data-analysis, and Monte-Carlo-simulation studies of strongly heteroscedastic data of both small and wide range. Many different types of heteroscedasticity and fixed or variable weighting are incorporated through error-variance models. Attention is given to parameter bias determinations, evaluations of their significances, and to new ways to correct for bias. The error-variance models allow for both additive and independent power-law errors, and the power exponent is shown to be able to be well determined for typical physical-sciences data by the rapidly-converging, general-purpose, extended-least-squares program we use. The fitting and error-variance models are applied to both low- and high-heteroscedasticity situations, including single-response data from radioactive decay. Monte-Carlo simulations of data with similar parameters are used to evaluate the analytical models developed and the various minimization methods employed, such as extended and generalized least squares. Logarithmic and inversion transformations are investigated in detail, and it is shown analytically and by simulations that exponential data with constant percentage errors can be logarithmically transformed to allow a simple parameter-bias-removal procedure. A more-general bias-reduction approach combining direct and inversion fitting is also developed. Distributions of fitting-model and error-variance-model parameters are shown to be typically non-normal, thus invalidating the usual estimates of parameter bias and precision. Errors in conventional confidence-interval estimates are quantified by comparison with accurate simulation results.

1. INTRODUCTION

"In a world in which the price of calculation continues to decrease rapidly, but the price of theorem proving continues to hold steady or increase, elementary economics indicates that we ought to spend a larger and larger fraction of our time on calculation." (Tukey 1986).

Nonlinear regression with heteroscedasticity (nonuniform error variance) and the use of weighting in nonlinear-least-squares (NLLS) fitting are of increasing interest (Ratkowsky 1983, Gallant 1987, Bates and Watts 1988, Davidian and Carroll 1987, Carroll and Ruppert 1988, Beal and Sheiner 1988, Seber and Wild 1989), especially in the analysis of data from the life sciences. The range of such data is seldom greater than two orders of magnitude, often 10 or fewer data are available, and large errors are frequent, so that analyses usually show little dependence on the type of heteroscedasticity, or even on whether homoscedasticity (uniform error variance) is assumed (Giltinan and Ruppert 1989). By contrast, in the physical sciences data typically range over three of more orders of magnitude and may encompass a 10^{12} range (Norman *et al.* 1988); usually 25 or more data are available, and errors are relatively small. For example, when the errors are proportional to the magnitude of the dependent variable, they rarely exceed 15%. Further, the appropriate fitting model is often known.

In NLLS analyses the fitting method may significantly affect the parameter estimates obtained. Here we show that the choice of error-variance model (EVM) substantially influences the accuracy and precision of parameter estimates for typical physical-sciences data. We use a powerful general-purpose fitting program that accommodates arbitrary nonlinear fitting models; the data may range from homoscedastic to highly heteroscedastic and may be of very large range; and the fitting models may require many parameters. By using this fitting program, we investigate various fitting strategies, such as extended least squares, ELS, and generalized least squares, GLS (Jobson and Fuller 1980, Davidian and Carroll 1987, Beal and Sheiner 1988, Giltinan and Ruppert 1989). The program allows simultaneous or sequential maximum-likelihood estimation of parameters in both the fitting model and in an EVM of quite general form. When feasible, following Tukey's advice, we compare these estimates with analytical interval estimates and with our MC simulations.

High heteroscedasticity and ways to obtain optimum parameter estimates are important in statistical analyses of wide-ranging data, but they have been seldom investigated. Therefore, major emphases in the present work are: exploration of several models of heteroscedasticity; analysis of their asymptotic properties; effects of transformations of data and fitting model; new bias-reduction possibilities; and extensive Monte-Carlo (MC) simulations against which the theoretical analyses are rigorously tested. High-precision Monte-Carlo simulations (typically 200,000 samples) also allow us the examine the robustness of parameter estimates for various transformation and fitting approaches.

The outline of the paper is as follows. After presenting definitions and models in Section 2, we discuss in Section 3 the details of the fitting methods used, and our MC simulation procedures. In Section 4 we present two data and model transformations relevant to data of very large range: logarithmic and inversion (reciprocation). Section 5 describes our analysis of several heteroscedastic data sets, either drawn from experiment or simulated, and described by exponential models.

2. DEFINITIONS AND MODELS

We first define general notation and models. Principal acronyms and symbol definitions are given at the end of the paper. Let x_i be an exact element of the independent-variable vector data, $\boldsymbol{x}$, with $i = 1,2,\ldots,N$, and let the corresponding dependent-variable vector be $\boldsymbol{y}$, having general element y_i. The fitting model is denoted $Y(x, \boldsymbol{\theta}) \equiv Y$, with representative element Y_i. Here $\boldsymbol{\theta}$ is the converged set of fitting parameter estimates whose mth element is θ_m. The set of exact-model parameter values is $\boldsymbol{\theta}_o$, with components θ_{om}, $m = 1,2,\ldots,P$. Since we are not concerned with errors arising from incorrect choice of fitting model, we have

$Y(x, \theta_o) \equiv Y_o$. We designate exact values, as used in MC simulations, with a subscript "o", and denote single-fit or MC estimates by the parameter itself. Wherever such notation is ambiguous, a caret is used to distinguish an estimate from its exact value.

2.1 Error Models

We define the error model as $\varepsilon(x, Y_o) \equiv \varepsilon$ with representative element $\varepsilon(x_i, Y_{oi}) \equiv \varepsilon_i$, where $Y_{oi} \equiv Y(x_i, \theta_o)$. The ith data element is then

$$y_i = Y(x_i, \theta_0) + \varepsilon(x_i, Y_{0i}) \equiv Y_{0i} + \varepsilon_i \approx Y(x_i, \theta) \equiv Y_i \tag{2.1}$$

Fitting the y data with the Y model yields the estimated parameter set θ. The error model is intrinsically unknown except in simulation studies, whereas an error-variance model (Section 2.2) is our best guess to account for the unknown errors, ε_i. Although these are taken as just $\varepsilon(x_i)$ in homoscedastic linear-least-squares fitting, the present more general dependence on Y_{oi} is necessary to allow adequate treatment of heteroscedasticity. We always ensure that $E[\varepsilon(x_i, Y_{oi})] = 0$, and thus

$$E[y_i] = Y(x_i, \theta_0) \tag{2.2}$$

There are two types of unknowns in Eq. (2.1): the θ_o vector and the error model, $\varepsilon(x_i, Y_{oi})$. Our fitting method, described in Section 3.2, automatically takes account of a common type of heteroscedasticity and is therefore robust with respect to heteroscedasticity in the sense of Beal and Sheiner (1988). The particular ELS fitting method we use allows one to obtain NLLS estimates of the θ_m parameters that are as close as possible to the unknown exact θ_{om}, at least for normally-distributed errors, where the solution is a maximum-likelihood one.

We now specify an error model appropriate for both low and high heteroscedasticity. Data of extensive range often have proportional errors (constant percentage errors), so that the errors are associated with a probability distribution whose standard deviation is proportional to true model values, Y_{oi}. More generally, the proportionality may be a positive power of the $|Y_{oi}|$ (Finney and Phillips 1977, Beal and Sheiner 1988, Carroll and Ruppert 1988). There will probably also be an independent, minimum set of additive errors from limited measurement resolution and other random effects. Such errors will dominate proportional errors for sufficiently small $|Y_{oi}|$.

In order to account for both possibilities and thus to allow more-realistic situations to be considered, we take an element of the general error model for generating errors in our MC simulations as

$$\varepsilon_i = \alpha_r P_1(0,I_i) + \sigma_r |Y_{0i}|^{\xi_0} P_2(0,I_i) \tag{2.3}$$

where $P_1(0,I_i)$ and $P_2(0,I_i)$ are random variables with values drawn from independent, uncorrelated probability distributions $P_1(0,I)$ and $P_2(0,I)$ with zero means and unity standard deviations. We use uniform, normal, or Poisson distributions for P_1 or P_2, but we select normal distributions unless stated otherwise. In Eq. (2.3) I_i is an element of the unit vector $\mathbf{I}$, so that $I_i = 1$ for all i, while α_r, σ_r, and ξ_0 (with ξ_0 usually in the range 0.5 to 1.5) are known, positive constants. Estimates of α_r and σ_r will be denoted by $\widehat{\alpha}_I$ and $\widehat{\sigma}_I$. To ensure accurate bias estimates in simulations, we enforce standardization on each sample of N random numbers. In Beal and Sheiner and in earlier treatments, no α_r term was included. When $\sigma_r = 0$ and $\alpha_r \neq 0$, the error distribution is homoscedastic and additive, as it is for $\alpha_r = 0$, $\sigma_r \neq 0$, and $\xi_0 = 0$.

2.2 Least Squares and Maximum Likelihood

We now discuss the relation between least-squares minimization and maximum-likelihood criteria for defining a best fit. Although this relation is well-known for weights that are the inverses of (presumed-determined) variances at each point, the connection is less clear when one allows the weights to contain fitting parameters, as we do.

Jobson and Fuller (1980), Beal and Sheiner (1988) and others, have shown that for normally-distributed errors a maximum-likelihood estimate of all parameters may be obtained by minimizing the objective function

$$\mathbf{O} \equiv \sum_{i=1}^{N} [\ln(V_i) + (y_i - Y_i)^2 / V_i] \tag{2.4}$$

where the error variance $V_i = (\sigma T_i)^2$. This weighting factor for the ith datum is written with the (usually unknown) variance σ^2 as a scaling factor and with all other dependencies assumed to be included in the function T_i. This function will often depend upon the fitting parameters. For example, if σ represents a scale factor for proportional errors in the model function, then $T_i = Y_i$, and this function therefore varies during iterative convergence of the parameter values.

The procedure of minimizing the objective function, Eq. (2.4), has been termed *extended least squares* (ELS) by Beal and Sheiner. As Giltinan and Ruppert (1989) point out, it is expected to have good properties when the data are normally distributed and the form of T_i is correctly specified. ELS has been criticized by van Houwelingen (1988) when these conditions do not apply. As shown in Section 5, ELS is a powerful and appropriate method for fitting typical physical-sciences data, even when the data errors are not normal. Thus, it is the method we use in most of the present work.

We note that Eq. (2.4) differs from the usual least-squares criterion because of the logarithm terms. However, suppose that we set

$$T_i = \tau_i / \sqrt[N]{\Pi \ \tau_i} \tag{2.5}$$

in which each τ_i is non-zero and the product is over i from 1 to N. Then, the T_i do not contribute to the sum over logarithms, and only if σ is a fitting parameter is there any distinction between log-likelihood for normal errors and least-squares. The normalization procedure of Eq. (2.5) has been used in statistics in another context (Hinkley and Runger 1984), and has been discussed for least-squares fitting by Carroll and Ruppert (1988), and by Giltinan and Ruppert (1989). These authors did not, however, cite any fitting results using this approach, and they considered life-sciences data of quite limited range.

The estimator of the variance is obtained directly by minimizing the objective function with respect to σ^2:

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{N} \left[\frac{y_i - Y_i}{T_i} \right]^2 \tag{2.6}$$

The least-squares equations to be solved for each parameter, ψ_j, are:

$$\sum_{i=1}^{N} \frac{1}{T_i}\left[\frac{y_i - Y_i}{T_i}\right]\left\{\frac{\partial Y_i}{\partial \psi_j} + \left[\frac{y_i - Y_i}{T_i}\right]\frac{\partial T_i}{\partial \psi_j}\right\} = 0 \tag{2.7}$$

It remains to specify a model for the T_i in Eqs. (2.6) and (2.7). Therefore, a model is needed for the τ_i in Eq. (2.5).

2.3 Error-Variance Models

By specifying the τ_i in Eq. (2.5) we are choosing an error-variance model (EVM), which is equivalent to making a model for the relative weights of the data points in the objective function Eq. (2.4). Our model function is

$$\tau_i = \sqrt{U^2 + |Z_i|^{2\xi}} \tag{2.8}$$

in which U and ξ are parameters of the EVM, and Z_i is either the data value y_i or the model value Y_i.

Although we use Y_{oi} in Eq. (2.8) in generating errors for simulation studies, these values are unknown when analyzing experimental data. Then the EVM can only involve the estimates, Y_i, or y_i. Often, however, as iterative fitting progresses, the Y_i will approach their true values, so that the EVM should progressively improve the parameter estimates, a beneficial feedback process.

The justification for the dependence of τ_i in Eq. (2.8) on U, Z_i and ξ is as follows. The quantity U is an estimator of errors that are independent of the data or fitting model. If only U is present, then it is has no effect on the T_i and therefore no effect on the fitting, which becomes a unity-weighting (UWT) situation. The power ξ determines the relative influence of the magnitude of the data (for $Z_i = y_i$) or of the model function (for $Z_i = Y_i$). For example, for $U = 0$, $\xi = 1$ gives proportional errors and $\xi = 1/2$ gives Poisson errors. The present ξ parameter corresponds to the θ parameter of Davidian and Carroll (1987), and to $\zeta/2$ in Beal and Sheiner (1988). The parameters of the EVM, here U and ξ, should be fitting parameters in order that their optimum values may be estimated from the data.

For convenience in reference, we name the weighting choices associated with Eq. (2.8). Eight such possibilities are defined in Table I. Prefixes are D for *data* if the weighting involves the y_i values and F for *function* if the weighting involves

TABLE I

Definitions of some specific weighting models discussed in Section 2.3. A * indicates that this variable is not applicable.

Line	Name	U	ξ	Z_i
1	UWT	fixed	*	*
2	DPWT	*	1, fixed	y_i
3	DFWT	fixed	fixed	y_i
4	DGWT	arb.	arb.	y_i
5	FPWT	*	1, fixed	Y_i
6	FFWT	fixed	fixed	Y_i
7	FGWT	arb.	arb.	Y_i
8	FPLWT	*	arb.	Y_i

the Y_i. The weighting schemes are FWT for *fixed* weighting, GWT for *general* weighting, and PWT for *proportional* weighting, in which the T_i are directly proportional to y_i or to Y_i. More generally, PLWT stands for *power-law* weighting, termed the power-function model by Beal and Sheiner (1988). In Table I "arb." indicates that the quantity may be arbitrary and either fixed or free during fitting.

Since the ε_i generated in our simulations, as well as errors in actual data, may involve Y_i, as in Eq. (2.3), but do not involve the data y_i (which already contain errors), it is clear that the choice $Z_i = y_i$ in Eq. (2.5) leads to an incorrect EVM and is thus inappropriate. This choice was, however, used previously (Macdonald, Hooper, and Lehnen 1982, Macdonald and Potter 1987), since the DPWT and DFWT variance models in Table I, like the UWT model, involve weighting that remains unchanged during iteration, thus simplifying the fitting program. In spite of the theoretical inadequacy of the $Z_i = y_i$ choice, we shall compare some DPWT and FPWT fitting results to quantify their differences and to discover to what degree they are significant.

The ELS fitting method has seldom been used previously, probably in part because it cannot be directly implemented by standard statistical software (Giltinan and Ruppert 1989). Until the present work, ELS fitting results have appeared only for data of limited range (Beal and Sheiner 1988) without using the geometric-mean Eq. (2.5), but omitting U in Eq. (2.8). We find that the combination of Eqs. (2.5) and (2.8) greatly improves convergence of the ELS procedure, so we use it in the following. Our ELS realization was developed independently of that formulated but not implemented by Ruppert and co-workers.

Another approach, *generalized least squares* (GLS), a staged, sequential fitting procedure, has been more popular than ELS because it can be implemented with standard commercial software and because it has theoretical advantages over ELS when errors are not normal and models are mis-specified (Jobson and Fuller 1980, Davidian and Carroll 1987, Carroll and Ruppert 1988, Giltinan and Ruppert 1989, Davidian 1990). Some GLS and ELS simulation results are compared in the following. Our realization of GLS involves first fitting with fixed weighting, then fitting by ELS FGWT with only the EVM parameters free, then fitting with fixed weighting (including fixed variance parameters). This sequence is repeated until fractional changes in the variance parameters are less than 10^{-5}, although most other implementations of GLS do not continue to such convergence. When the EVM parameters are not well-determined, as often happens for small-range, life-sciences data (Giltinan and Ruppert 1989), such convergence is unwarranted, but for physical-sciences data it is warranted because the EVM parameters can usually be well determined. Davidian and Carroll (1987) discussed a variety of methods for estimating EVM parameters but did not provide numerical comparisons of them. The method we use for both ELS and GLS has proved very satisfactory, as judged by our present MC-simulation parameter bias estimates (Section 4.1.2).

When convergence has been attained in a NLLS fit, one may calculate S_F, the standard deviation of the overall fit to the data. We use the converged values of Y_i and T_i in

$$S_F^2 = \frac{\Lambda}{D} \sum_{i=1}^{N} \left[\frac{y_i - Y_i}{T_i} \right]^2 \tag{2.9}$$

Here

$$\Lambda = \sqrt[-2/N]{\prod T_i} \tag{2.10}$$

when the geometric normalization in Eq. (2.5) is used, and $\Lambda = 1$ otherwise. In

Eq. (2.9) D is the number of degrees of freedom, just N minus the number of free parameters for single-response situations. In Eq. (2.10) the final converged values of all free parameters are used when $\Lambda \neq 1$. Since the value of Λ is then unknown until convergence, it cannot be used to normalize T_i^2 during iteration. But, unlike the Beal and Sheiner algorithm, S_F need only be calculated at final convergence. Since the choice of Λ then cancels out the effects of geometric normalization in Eq. (2.5), S_F is independent of Λ. Further, when $\alpha_r = 0$ in Eq. (2.3), S_F and the σ estimator in Eq. (2.6) differ only by the known factor $(N - p)/N$, so S_F is usually an excellent estimator of σ_r whatever the value of ξ.

Although an estimator for σ_r does not appear explicitly in our form for τ_i, because its value is unknown until final convergence, the quantity U is actually an estimate of α_r/σ_r when $\sigma_r \neq 0$. Thus, when an estimate of σ_r (such as S_F) is available, then that of α_r may be obtained from the U and S_F estimates.

3. PARAMETER ESTIMATION METHODS

Since the fitting algorithm and its computer implementation are important for efficient parameter estimation, we describe in this section the fitting program and procedures used. Then we summarize our general approach and notations used for the Monte Carlo simulations that we used to validate the regression models introduced in Section 2 and developed in practical transformations in Section 4.

3.1 Specifics of the Fitting Procedure

The nonlinear-least-squares minimization procedure we use is based on the robust Levenberg-Marquardt NLLS program described by Moré (1978), but generalized for variable weighting and to allow complex data (two separate dependent variables). We have used it since 1982 with $U = 0$ and with ξ taken as a fixed input parameter in a complex-nonlinear-least-squares (CNLS) data-fitting program named LOMFP that handles complex, real, or imaginary data (Macdonald and Potter 1987, Macdonald 1987). In the current version, LEVM, both U and ξ may be fixed or free to vary during fitting. Like most NLLS programs, the modified Moré procedure uses as input only the components of the weighted residual vector and the Jacobian matrix, and it ignores second-derivative terms in the Hessian matrix.

For our program to be readily usable with any response function involving a single $\boldsymbol{x}$ vector and one or two associated $\boldsymbol{y}$ vectors (multiple-response), and because analytical differentiation is usually intractable, we use numerical differentiation to calculate the derivatives in Eq. (2.7). All calculations are carried out in double-precision arithmetic, and the relative numerical derivative step size is set at 10^{-8} times the value of the component whose derivative is to be calculated (or 10^{-8} if the value is zero). Conventional iteration stopping criteria are used, with final convergence assumed when the relative change of all parameters is less than 10^{-8} or when the relative change of the SD of the fit is less than 10^{-8}.

3.2 Monte Carlo Simulation Procedures

Our simulations use NLLS fitting of K replicate sets of data with errors, and thus require N independent random errors, ε_i, for each value of $k = 1,2,\ldots,K$. The simulations are restricted in several ways. First, we use only fitting results that converge in 91 or fewer iterations of the NLLS fitting in the statistical calculations. The actual number of iterations required for an individual fit to converge is I, and its maximum allowed value is I_{max}, here 91. Second, we consider only those data sets and model values for which all $y_i > 0$. Finally, if negative values of $Y_i = Y_{oi} + \varepsilon_i$ are generated, they are replaced by $Y_i = Y_{oi} + |\varepsilon_i|$. Unless α_r in Eq. (2.3) is non-zero and U in Eq. (2.8) is free to vary, or if σ_r or ξ_0 is very large, no fits are eliminated by these restrictions. Thus, there is usually no censoring present except for a few results reported in Section 5.3.2.

We characterize the MC simulation results as follows. Define the error of the jth parameter in the kth fit (ψ_{jk}), as $E_{jk} \equiv \psi_{jk} - \psi_{0j}$, where ψ_{0j} is the exact value of the jth parameter. Then the corresponding relative error is

$$e_{jk} = E_{jk}/\psi_{0j} \tag{3.1}$$

The estimated relative bias of the jth parameter is then the mean of the relative errors

$$b_j \equiv B_j/\psi_{0j} = \frac{1}{K}\sum_{k=1}^{K} e_{jk} \tag{3.2}$$

where ψ_j is the estimated value of ψ_{0j}, and we usually take K sufficiently large that the standard deviation (SD) of b_j is less than $0.01b_j$. Quoted values of b_j are at most slightly uncertain in their last place. To allow easy comparison between parameters and between fits of different data, we use b_j rather than ψ_j and ψ_{oj} . If the exact value of a fitting parameter is zero, we quote B_j. The relative bias without regard to sign is

$$b_{abj} = \frac{1}{K}\sum_{k=1}^{K} |e_{jk}| \tag{3.3}$$

In work that we are currently doing, we find that the quantities b_j and b_{abj} can be used to estimate the SD of the e_j distribution, σ_{ej}, for a specified form of the distribution. Instead of using 95% confidence intervals, we use standard deviation estimates, since this is conventional in the physical sciences. It is also useful to know b_j relative to σ_{ej}, since when b_j/σ_{ej} is sufficiently small, bias effects can be ignored.

We use several different SD estimates for the e_j distribution, so it is convenient to omit the "e" subscript and to denote the estimate s_{ej} of σ_{ej} by just s_j. We then distinguish the various estimates by additional subscripts. The most direct estimate of σ_{ej} is just the *central* SD for the usual unbiased estimator,

$$s_{jC} = \sqrt{\frac{1}{K-1}\sum_{k=1}^{K} (e_{jk} - b_j)^2} \tag{3.4}$$

For a single NLLS fit of data with normally-distributed errors, the 68.3% confidence interval involving s_{jC} extends around $\widehat{\psi_{0j}}$ from $[1 - s_{jC}]\,\widehat{\psi_{0j}}$ to $[1 + s_{jC}]\,\widehat{\psi_{0j}}$, where $\widehat{\psi_{0j}}$ is an estimate of ψ_{0j}, corrected for bias, if known. This result assumes that the $\widehat{\psi_{0j}}$ are sampled from a normal distribution, but results of the present work show that this is not generally true and that an asymmetric confidence interval is needed. When b_j is very small, a large number of samples, up to $K \geq 10^6$, may be required to estimate it accurately to two significant figures. Much smaller K suffice for the same accuracy of s_j. For $K \geq 10^6$, computer internal memory size limitations may become significant, so to avoid such limitations in calculating sums in Eqs. (3.2) – (3.4), we use blocking. Then, K is replaced by K/J (with $J > 1$), and the J results for each calculation are averaged. We used J in the range of 2 – 10 if K was very large.

Each NLLS fit of a simulation calculation yields a *linearized* estimate of σ_{ej} calculated at convergence from the Jacobian of the fitting function. We denote by s_{jL} the average of K such estimates. Donaldson and Schnabel (1987) have shown that it often appreciably underestimates σ_{ej} for NLLS with additive errors. We also find problems with this estimator for proportional errors. The probability interpretation of σ_{jC} and σ_{jL} assumes that the parameter-error distribution is normal. But for NLLS, even with normal errors in data, the parameter errors are generally not normal, and one is also dealing with a discrete distribution rather than a continuous one. With simulation, we can examine both effects directly. In our MC runs with $K \geq 2\text{x}10^5$, for each value of j we save $K_l = 2\text{x}10^5$ values of e_{jk}, allowing accurate plots to be made of the error distribution for each fitted parameter.

The distribution of e_{jk} values allows one to test directly the adequacy of the various estimates of σ_{ej}. To do so, we calculate accurate confidence-interval values of the distribution (including separate left- and right-hand estimates) to indicate possible asymmetry of the distribution, as follows. After obtaining the mean, SD, skewness, and kurtosis of the e_{jk} values, they are sorted by increasing algebraic size, and are then sequentially allocated to 800 bins, each of whose width is 1/800th of the total finite-distribution width. The bin values to the left and right of the mean value are then treated separately. For each such set, bin values away from the mean are summed until they exceed 0.68269 of the total count for that set. Finally, by rational function approximation, we estimate the value of e_{jk} corresponding to 68.269% of the probability. The resulting values are then referenced to the mean, b_j, so that they measure the distance from the mean to the 68.3% probability point on either side. The results, defined as s_{jLH} and s_{jRH}, for the *left-hand* and *right-hand* parts of the distribution, respectively, thus estimate the 68.3% confidence interval around the mean.

The *average* of s_{jLH} and s_{jRH}, denoted by s_{jAV}, may be directly compared to the other dispersion measures, s_{jL} and s_{jC}. Finally, the *normalized block count* in each block (NBC), the actual count normalized by the maximum block count present, is plotted at the center of each block, normalized by s_{jC}. The resulting distribution plot has a maximum height of unity and an abscissa measured in units of s_{jC}, termed the *normalized block value* (NBV). In each distribution plot we include a central vertical line at the mean position and shorter adjoining ones that define the true 68.3% confidence interval, all normalized by s_{jC}.

4. PRACTICAL TRANSFORMATIONS

In this section we derive and test properties of two "transform-both-sides" approaches, the logarithmic and inversion transformations. The discussion of the logarithmic transformation is divided into an analytical part (Section 4.1.1) and a simulation part (Section 4.1.2) where we validate the new analytical results. In Section 4.2 we consider, again from analytical and simulation viewpoints, use of the inversion (reciprocation) transformation.

4.1 Logarithmic Transformation

This transformation is particularly appropriate for monoexponential response, where the fitting model is

$$Y(x_i, \boldsymbol{\theta}_0) = Y_{0i} = \theta_{01} \exp(\theta_{02} x_i) \tag{4.1}$$

and the transformed model is just

$$Y^*(x_i, \boldsymbol{\theta}_0) = Y^*_{0i} = \theta^*_{01} + \theta^*_{02} x_i \tag{4.2}$$

a linear model in terms of the parameters $\theta^*_{01} = \ln(\theta_{01})$ and $\theta^*_{02} = \theta_{02}$. The logarithmic transform assumes that both data and model values are positive. If the original error distribution were normal, then the transformed one will not be so. Logarithmic transformation is most appropriate when the fitting model is a single exponential and the errors are of the form of Eq. (2.3) with $\alpha_r = 0$ and $\xi_0 = 1$ (constant percentage errors). It has a long history and has been used for other models besides single exponentials (Carroll and Ruppert 1988, Chap. 4.) Although many of our new results below also apply to such cases, we apply them here only to the single-exponential case.

4.1.1 *Bias from logarithmic transformation*

We now derive expressions for the intrinsic asymptotic bias induced by a logarithmic transformation having data errors given by Eq. (2.3) with $\alpha_r = 0$ and $\xi_0 = 1$. Such errors are common in experimental situations, at least over a limited range. We begin with a general model, then specialize to the monoexponential case. Before transformation, the appropriate NLLS weighting model is that of line 5 of Table I, FPWT. Logarithmic transformation of Eq. (2.1) yields

$$y_i^* = Y_{0i}^* + \ln [1 + (\varepsilon_i / Y_{0i})] \tag{4.3}$$

which is clearly applicable only if

$$1 + (\varepsilon_i / Y_{0i}) > 0 \tag{4.4}$$

precluding such a transformation for relatively large negative errors. For arbitrary ξ_0 in Eq. (2.3), the errors may be written as

$$\varepsilon_i = \sigma_r Y_{0i}^{\xi_0} P_2 (0, I_i) \tag{4.5}$$

and thus

$$\mathrm{var} (y_i) = \sigma_r^2 Y_{0i}^{2\xi_0} \tag{4.6}$$

On substituting Eq. (4.5) in Eq. (4.3), we obtain

$$y_i^* = Y_{0i}^* + \ln [1 + \sigma_r Y_{0i}^{\xi_0 - 1} P_2 (0, I_i)] \tag{4.7}$$

Only for the common proportional-errors case, $\xi_0 = 1$, is the logarithmic term in Eq. (4.7) independent of Y_{0i}. When $\xi_0 < 1$, as in Poisson statistics ($\xi_0 = 1/2$), the inequality in Eq. (4.4) must fail for sufficiently small Y_{0i} even if P_2 is truncated so that arbitrarily large negative values of ε_i are deleted. For $\xi_0 = 1$ we rewrite Eq. (4.7) as

$$y_i^* = Y_{0i}^* + \ln [1 + \sigma_r P_2 (0, I_i)] \tag{4.8}$$

which becomes, on specializing to the monoexponential situation,

$$\begin{aligned} y_i^* &= \theta_{01}^* + \ln [1 + \sigma_r P_2 (0, I_i)] + \theta_{02} x_i \\ &= \theta_1^* + \theta_2 x_i + \varepsilon_i^* \end{aligned} \tag{4.9}$$

A closely related expression was given by Cook and Weisberg (1982).

An ordinary, unweighted least-squares (OLS) fit of the transformed data to the RHS of Eq. (4.9), directly involving the parameters θ_1^* and θ_2, leads to an entirely unbiased estimate of the slope θ_2 for the case of proportional errors. The intercept is biased by

$$L = E\{\ln[1 + \sigma_r P_2(0, \boldsymbol{I})]\} \tag{4.10}$$

Although L depends on the error distribution, it is independent of the x_i and parameter values. It can be estimated by series expansion of the logarithm and subsequent term-by-term evaluation of each expectation value. Because P_2 has zero mean and unity variance, the lowest-order approximation to L, L_l, is independent of the type of standardized distribution assumed and is given by

$$L \approx L_l = -\sigma_r^2/2 \tag{4.11}$$

In this approximation the pre-exponential parameter in Eq. (4.1) is underestimated by a factor of about $(1 - \sigma_r^2/2)$ for errors proportional to model values. Term-by-term evaluation of the series expansion of Eq. (4.10) leads to the asymptotic bias estimate for the continuous, normal, distribution

$$L \approx L_n^c = -0.5\,\sigma_r^2[\,1 + (3/2)\,\sigma_r^2\,\{\,1 + (10/3)\,\sigma_r^2 + (35/4)\,\sigma_r^4 + \ldots\}\,] \tag{4.12}$$

Formally, this is a divergent series in which the ratio of successive terms exceeds unity after about $[\,2 + 1/(2\sigma_r^2)]$ terms. In practice, a MC simulation so rarely samples the extreme wings of the distribution that moments higher than the 8th moment included in Eq. (4.12) have negligible effect for $\sigma_r > 0.5$. For the uniform distribution (u), a similar analysis yields the bias estimate

$$L \approx L_u^c = -0.5\,\sigma_r^2[\,1 + (9/10)\,\sigma_r^2\,\{\,1 + (10/9)\,\sigma_r^2 + (10/4)\,\sigma_r^4 + \ldots\}\,] \tag{4.13}$$

which diverges for $\sigma_r > 1/\sqrt{3} \approx 0.58$, but is still sufficiently accurate, for the number of terms given, if $\sigma_r < 0.5$. The negative signs of the bias in these equations indicate that the true values of the pre-exponential parameters, θ_{0l} and θ_{0l}^*, are always larger than their estimated values. This is intuitively clear because

TABLE II

Comparisons of theoretical logarithmic bias estimates for continuous distributions, "c", with the results of direct calculation of discrete-distribution expectation values, "d", for normal (n) and uniform (u) distributions with errors proportional to model values ($\xi_0 = 1$, FPWT in Table I). A * indicates that the value is unreliable because of censoring.

		σ_r				
Line	Bias estimate	0.1	0.2	0.3	0.4	0.5
1	$-L_l$	0.00500	0.0200	0.045	0.080	0.125
2	$-L_n^d$	0.00508	0.0214	0.053	0.11	0.24
3	$-L_n^a$	0.00508	0.0214	0.054	*	*
4	$-L_u^c$	0.00505	0.0208	0.0492	0.0949	0.168
5	$-L_n^d$	0.00505	0.0208	0.0492	0.0952	0.173

the logarithmic transformation maps the interval 0 to 1 into $-\infty$ to 0, but the interval 1 to ∞ has the same absolute range after mapping. Our definition of relative bias now leads, for proportional errors, to

$$b_l^* = L/\theta_{0l}^* \quad (\theta_{0l}^* \neq 0), \qquad b_l^* = L \quad (\theta_{0l}^* = 0) \tag{4.14}$$

and thus a relative bias in the pre-exponential parameter of

$$b_l = \exp(L) - 1 \tag{4.15}$$

4.1.2 *Monte Carlo comparison of bias*

For comparison with results in Section 5.2.1 in a MC analysis of Eq. (4.9), it is useful to determine some expectation values for the logarithmic transformation directly. Table II compares the above continuous-distribution bias estimates with MC results. The simulation results in lines 3 and 5 were determined by direct averaging, as in Eq. (4.10), for discrete normal and uniform distributions, respectively. Up to 10^8 separate values were averaged for errors taken from discrete distributions with zero mean. The values of the bias function for the discrete distribution, L^d, are significant in the last decimal place.

As expected, for $\sigma_r \leq 0.2$ the L^c values in Table II do not differ much from the first-order approximations, L_l, and there are only small differences between values for the two different distributions, so L_l is distributionally robust. In obtaining the $\sigma_r = 0.3$ result for L_n^d in line 3 we eliminated fewer than 100 of 10^6 error values that led to divergent logarithms. Because such truncation renders the original error distribution less normal, however, no L_n^d results are included for $\sigma_r > 0.3$ (30% error), which is an uncommon percentage error in the physical sciences.

Figure 1 shows plots of the distribution of $\ln(1 + \varepsilon_i)$ and a normal distribution for comparison, all with $2\text{x}10^5$ samples. Values for plotting were calculated using the binning procedure discussed in Section 3.2. For $\sigma_r = 0.2$, very long, thin, and asymmetrical tails appear in the logarithmic distributions. The means of these distributions show excellent agreement between predictions and the MC estimates. For example, with $\sigma_r = 0.2$ the MC result for the mean of the logarithmic distribution was again 0.0214, as in Table II. Similar expansions for the skewness, γ_l, produce only order-of-magnitude agreement because of strong sensitivity to outliers. For example, for $\sigma_r = 0.2$, we predict $\gamma_l \approx -3\sqrt{\sigma_r}/2 \approx -0.67$, whereas the MC value was -0.74.

Thus, fitting exponential-response data after logarithmic transformation of both data and model allows one to obtain nearly zero bias for both fitting parameters if $\alpha_r = 0$ and $\xi_0 = 1$, that is, if errors are proportional to model values. The exponent parameter is unbiased for this error model, and the bias of the pre-exponent can be readily estimated, leaving a residual bias perhaps even smaller than in NLLS fitting of the untransformed system. This is further illustrated in the fitting results presented in Sections 5.1 and 5.3. As a usually adequate approximation, the value of S_F, Eq. (2.9), obtained from the fit may be used to estimate σ_r. As shown in Sections 5.1 and 5.3, the S_F value obtained from a UWT fit of Eq. (4.9) has a larger bias than that of a FPWT NLLS fit of the untransformed data. Therefore, a more accurate result will generally be produced by taking S_F from such a FPWT fit. A similar fit, but with ξ a free parameter (FPLWT), can also yield valuable information on the appropriateness of assuming proportional errors and making a logarithmic transformation.

4.2 Inversion Transformations

Inversion is important in NLLS fitting when both the distribution P_2 and the error-weight exponent ξ_o are arbitrary, but $\alpha_r = 0$ in Eq. (2.3). There are two interesting possibilities. In the first (Type I), appropriate only for MC simulations, we invert the exact values, Y_i, before adding errors. In the second possibility (Type II), the errors are already present, as for data in real situations, so the inversion approach that is appropriate is to form the y_i before inverting data and model. In the following, we consider these possibilities in turn. As common notation, we write for inverted variables $x^\# \equiv 1/x$, and for fitting parameters obtained by inversion a similar notation is used.

4.2.1 *Inverting before including errors (Type I)*

For the untransformed situation we have, from Eqs. (2.1) and (2.3),

$$y_i = Y_{0i}\left[1 + \sigma_r P_2(0, I_i)\, Y_{0i}^{\xi_0 - 1}\right] \tag{4.16}$$

The corresponding transform following from $y_i^\# = Y_{0i}^\# + \varepsilon_i^\#$ is thus

$$y_i^\# = Y_{0i}^\#\left[1 + \sigma_r P_2'(0, I_i)\,(Y_{0i}^\#)^{\xi_0^\# - 1}\right] \tag{4.17}$$

where the prime on P_2 indicates that the distribution in Eq. (4.17) is not necessarily the same as that in Eq. (4.16). Comparison of Eqs. (4.16) and (4.17), with the two distributions assumed equal, suggests the relation $\xi_0 + \xi_0^\# \approx 2$, indicating reflection symmetry around $\xi_0 = 1$. Thus, for MC results an estimate of ξ obtained from untransformed data whose errors involve $\alpha_r = 0$ should be simply related to the $\xi^\#$ estimate found from Type-I inversion of the same data and model. The relative biases should therefore be related by

$$\xi_0\, b_\xi \approx -\,\xi_0^\#\, b_\xi^\# \tag{4.18}$$

For practical situations, where ξ_o is unknown but an estimate of b_ξ may be available, Eq. (4.18) may be used to relate ξ, b_ξ, and $b_\xi^\#$.

4.2.2 *Inverting after errors are included (Type II)*

Next, consider the inversion appropriate to data, where model and errors are combined before inversion. We consider the case of relatively small errors. On using Eq. (2.1), we may write $y_i^{\#} = 1/y_i = 1/(Y_{0i} + \varepsilon_i)$, which is, to first order in ε_i / Y_{0i}, just $y_i^{\#} \approx Y_{0i}^{\#}(1 - \varepsilon_i Y_{0i}^{\#})$. By using Eq. (2.3) for the error model with $\alpha_r = 0$ we obtain

$$y_i^{\#} \approx Y_{0i}^{\#}\left[1 - \sigma_r P_2(0, I_i)\left(Y_{0i}^{\#}\right)^{1-\xi_0}\right] \tag{4.19}$$

Thus, if $\xi_0 + \xi_0^{\#} \approx 2$ and $P_2 = P_2'$, this result is the same as that for Type-I inversion, Eq. (4.17), except for the sign, which is irrelevant for a symmetrical error distribution. Thus, for such errors, Type-II inversion with small σ_r should lead to essentially the same results as Type I inversion.

For proportional errors, a connection can be made between Eqs. (4.17) for Type-I inversion and the approximate (4.19) for Type-II inversion. Inverting Eq (4.16) with $\xi_0 = 1$, gives

$$y_i^{\#} = Y_{0i}^{\#}\left[1 - \frac{\sigma_r P_2(0, I_i)}{1 + \sigma_r P_2(0, I_i)}\right] \tag{4.20}$$

which identifies the P_2' distribution in Eq. (4.17) with the second term in the brackets of this equation. When $\sigma_r P_2$ is small compared to unity and is symmetric about the origin, there is thus little difference between the fitting expressions Eqs (4.17) and (4.19). When the error term is not negligible compared to unity, the error distribution of the inverted data will be appreciably skewed even if P_2 is symmetric.

To illustrate the skewing effect in an inversion transformation, we show in Figure 1 plots of the distribution of $\varepsilon_i/(1 + \varepsilon_i)$, where $\varepsilon_i = \sigma_r P_2(0, I_i)$ is drawn from a normal distribution and $\sigma_r = 0.2$. A total of 2×10^5 samples was used. By making a Taylor expansion of the skewed distribution about $\sigma_r = 0$, we predict a mean of $\sigma_r^2\left[1 + 3\sigma_r^2 + 15\sigma_r^4 + \ldots\right]$ and a skewness $\gamma_l \approx 3\sqrt{\sigma_r}$.

If the logarithmic transformation is applied to Type-II inversion with proportional errors, one will obtain the same results as in Section 4.1, except that the sign of the bias in the transformed pre-exponent will be reversed.

4.2.3 *Monte Carlo tests of inversion*

We carried out several MC simulations for monoexponential and other data fitting models with $0.5 \leq \xi_o \leq 1.5$ in order to investigate the effects of inversion. As expected from the above analysis, Type-I and small-error Type-II fits yielded very nearly the same estimates. We found that Eq. (4.18) is satisfied very well, implying that when ξ or $\xi^{\#}$ is a fitting parameter, the relation $\xi + \xi^{\#} = 2$ is an excellent approximation. Further, within statistical variability the various standard-deviation estimates of ξ are the same as the corresponding estimates for the inversion transformation.

In monoexponential MC fits we found that the relative biases in the pre-exponentials are nearly independent of whether inversion is performed, while the biases in the exponents are just reversed in sign by inversion, as expected. When MC fitting was carried out with ξ or $\xi^{\#}$ fixed at their exact values, results were not quite so clearcut. Our MC results suggest that the average of the fitted parameters, $\theta_m = (\theta_m + 1/\theta_m^{\#})/2$, usually gives a closer approximation to the true parameter value than does either separately.

For the $\varepsilon_i/(1 + \varepsilon_i)$ distribution shown in Figure 1, the predicted mean value from the formula in Section 4.2.2 is 0.0458, compared with 0.0463 for the MC result. The predicted skewness is about 1.3, but our MC result was 3.1.

5. EXPONENTIAL MODELS

The ubiquitous presence of exponential response in science makes it important and justifies studying its fitting properties. It has long been known that NLLS fitting generally leads to biased parameter estimates, but there has been little quantitative study of this problem, although there are complicated theoretical expressions for such bias, assuming normally-distributed parameter errors (Seber and Wild 1989). Further, although an asymptotic theory of NLLS estimation demonstrating the inconsistency of exponential-model parameter estimates has been developed (Wu 1981), it provides no quantification of the inconsistencies. The present results, however, yield information about typical parameter bias levels for several fitting and variance models. They thus guide selecting an approach to yield accurate parameter estimates and show that parameter-error distributions are not usually closely normal, but have long tails similar to those shown in Fig. 1 for logarithmic and inversion transformations of a normal distribution.

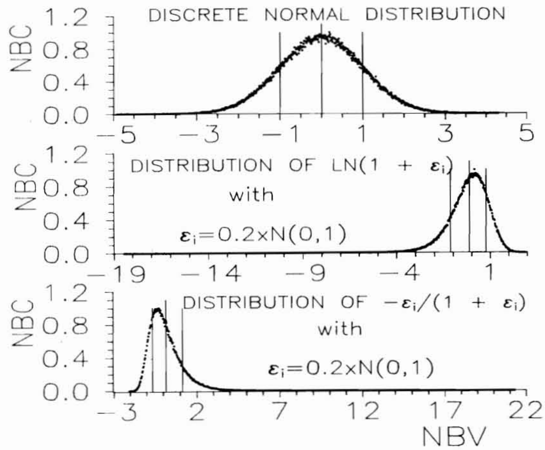

FIG. 1. Normalized, 800-point distributions, each based on $2\text{x}10^5$ samples in 800 bins. *Top*: independent, random samples from a normal distribution, N(0,1); *middle*: log transformation, $\ln(1 + \varepsilon_i)$; *bottom*: $\varepsilon_i/(1 + \varepsilon_i)$, both for $\varepsilon_i = 0.2\,N(0,I_i)$. The block value is normalized with S_C, the calculated SD, to yield the NBV scale, and the normalized block count, NBC, is the ratio of the count in a bin to the maximum such count. The mean value is denoted by the longer vertical line, and the two shorter vertical lines show the positions of the 68.3% probability points.

5.1 Analysis of the Beal-Sheiner Monoexponential Model

Beal and Sheiner (1988) discussed analysis of a small-range, nearly homoscedastic, data set by a monoexponential model. We use it to illustrate several results from Section 4. The parameter values that they used are $\theta_{O1} = 2$ and $\theta_{O2} = -0.693 \approx -\ln(2)$. The 10 x_i values are 0.1, 0.2, 0.3, 0.5, 0.75, 1.0, 1.5, 2.0, 2.5, and 3.0. In the present MC study of this exponential-decay fitting model, we follow them and assume proportional errors in Eq. (2.1).

Table III summarizes MC simulation results obtained with from $5\text{x}10^5$ to $2\text{x}10^6$ replications. The P_2 error distribution was taken normal (as in Beal and Sheiner) for lines 1 through 4 of the table and was taken uniform for line 5. For the FPLWT fits of line 1 in Table III, ξ was free to vary, and we found $b_\xi = 0.225$, $s_{\xi L} = 0.68$, and $s_{\xi C} = 0.61$. These values show that ξ is strongly biased for the present analysis of their data and can be only very poorly determined by such fitting.

TABLE III

MC simulation results for the Beal-Sheiner monoexponential model, $\sigma_r = 0.15$. The notation for weights is in Table I, and for other column headings is in Section 3.2. The -n and -u indicate normally- and uniformly-distributed errors. The /LT in lines 4 and 5 indicates that a logarithmic transformation of the data was made. A * indicates that the values are not significantly different from zero.

Line	Weight	S_F	$b_1 x 10^3$	$s_{1L} x 10^2$	$s_{1C} x 10^2$	$b_2 x 10^3$	$s_{2L} x 10^2$	$s_{2C} x 10^2$
1	FPLWT-n	0.1530	2.80	8.09	7.44	1.92	7.25	8.10
2	FPWT-n	0.1496	2.91	7.47	5.77	2.87	7.00	7.01
3	UWT-n	0.1724	4.00	6.31	7.56	8.86	14.0	12.5
4	UWT/LT-n	0.1525	-15.0	11.0	8.54	*	11.0	7.20
5	UWT/LT-u	0.1518	-14.9	10.9	8.48	*	7.14	7.15

Plots of the relative error distributions of the parameters are shown in Fig. 2. Because distribution plots of e_{jk} and of the actual, unnormalized errors differ only in the position of their zero values on the NBV abscissa scale, we generally do not distinguish between them. In Fig. 2 the distribution of the errors of θ_l has a very long, thin, right tail and a large kurtosis (excess) of 3.3; that of ξ also has an appreciable right tail but its kurtosis is 1.6. The distribution of θ_2 errors is clearly closest to normal (kurtosis $\approx$ 0.1). The biased value of ξ is 1.225, and the 68.3% confidence interval around the true mean, $\xi_o = 1$, extends from 0.47 to 1.6.

For the other runs summarized in Table III, ξ was either fixed or not present (as in UWT). Note that the constant-variance FPWT weighting model of line 2 is fully consistent with the error model selected. Runs like that in line 2, but with ξ fixed at a value different from unity, gave results comparable to those of line 2. For example, that with ξ = 1.225 led to somewhat worse $b_{\xi j}$ estimates and that with $\xi = 0.5$ to slightly better ones. However, the b_j values in line 2 are sufficiently smaller than the corresponding s_{jC} ones that for most purposes bias can be neg-

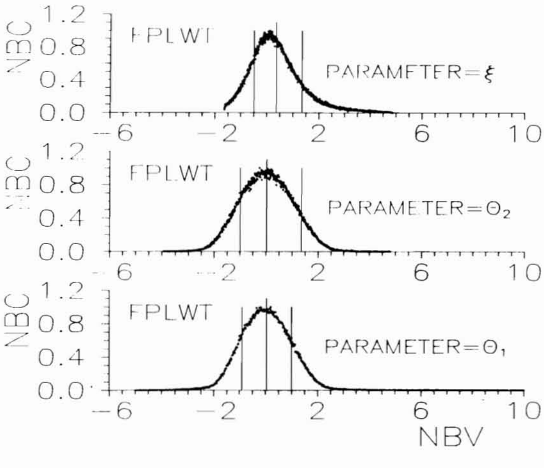

FIG. 2. Normalized distributions of the relative errors of the ξ, θ_1, and θ_2 parameters for the FPLWT MC results in line 1 of the Table III Beal-Sheiner small-range monoexponential model and data.

lected. It is consistent that, although ξ cannot be well determined here, its value makes little difference to other parameter estimates. This conclusion is somewhat counter to that of Beal and Sheiner who state that "there is considerable benefit in letting ζ (twice our ξ) be estimated rather than fixed." Even though no significant parameter estimation benefit appears in the present example, we agree with their conclusion for highly heteroscedastic data, such as those discussed in Section 5.2.

For FPWT our result for the b_{ab1} mean of Eq. (3.3) agrees with the comparable ELS, 500-sample result in Beal and Sheiner, but our value for b_{ab2} agrees with their "iteratively reweighted least squares" result, itself smaller than their ELS value of b_{ab2}. In Table III the line-3 UWT results are significantly worse than those in line 2, as one might expect, since the weighting model is here inconsistent with the error model. Nevertheless, for the present small data range and mild heteroscedasticity, results are clearly not strongly sensitive to a particular choice of weighting model. The results of lines 4 and 5 of Table III apply for fitting after the logarithmic transformation discussed in Section 4.1. Only small differences are evident between the results in lines 4 and 5. Also, although their original data-error distributions differed, those of the parameters are both nearly normal, as expected from the central-limit theorem.

We do not list estimates of b_2 because, although the MC values were of order 10^{-5}, their uncertainties were sufficiently large that they could not be well distinguished from zero even with K = $2x10^6$. The calculation, using double-precision arithmetic, required 9.5 hours on a Compaq 386-20 computer with Weitek co-processor board. The S_F value in line 2 is substantially closer to the σ_r = 0.15 value used to generate input data errors than are any other S_F estimates of σ_r in Table III. Although for lines 4 and 5 of Table III the bias of θ_2 is consistent with zero, as we expect from Section 4.1, the relative bias of θ_l^*, which is listed in the b_l column, but it is actually b_l^*, is quite large. The values of L predicted in Section 4.1.1 for σ_r = 0.15 are used for normal or uniform distributions to correct the b_2 values in lines 4 and 5, respectively. This yields approximate residual relative bias values of $1.8x10^{-3}$ or $1.7x10^{-3}$ in the log-transformed pre-exponent and about $1.3x10^{-3}$ or $1.2x10^{-3}$ for the relative bias of the pre-exponent itself, appreciably smaller in magnitude than any other such estimates in Table III, but still not zero.

Thus, we have produced a very nearly zero-bias fitting approach for mono-exponentials that is robust with respect to the data-error distribution. It is tested for a strongly heteroscedastic situation in Section 5.3. Although all the fitting results of Table III use Eqs. (2.5) - (2.8), a few simulations were carried out using the Beal-Sheiner ELS approach. Both FPLWT and FPWT runs showed that fitting by this ELS approach slowed convergence. In fact, many of the replication fits failed to converge even after many iterations. For example, for the FPWT simulation in line 2 of Table III all fits converged in three or fewer iterations using θ_{oj} values as initial guesses for the θ_j. The Beal-Sheiner approach led to 11% non-convergent replications after a maximum of 91 iterations, 9.5% non-convergence for 270, and did not further decrease when the maximum was allowed to increase. Analysis by Beal-Sheiner ELS with up to 91 iterations required about four times more computer time than did our method and about eight times more was needed for 273 iterations.

Although very few fits failed to converge in ELS for σ_r = 0.01, analyses with such small errors nevertheless still took about 35% more computer time than did comparable ones using our method. As expected, both approaches give essentially the same results when there is no censoring. For wide-range data with appreciable errors there will be even more difference between the convergence properties of the two approaches than illustrated here; thus we used only our FPLWT or FGWT with ELS fitting method in the rest of our work.

5.2 Analysis of a Strongly Heteroscedastic Monoexponential Model

Again using the monoexponential model, Eq. (4.1), to describe the data, we now consider two extensive MC simulations involving the general error model, Eq. (2.3), and the EVM, Eqs. (2.5) and (2.8). We use 31 data points, selected with a ratio of adjacent x values of $10^{1/10}$. Thus, the x values are distributed uniformly on a logarithmic scale. The Y_i fitting-model values are then calculated from Eq. (4.1) with $\theta_{o1} = \theta_{o2} = 1$ and they range from 1.01 to 2.2×10^4, an exponential growth model with a data range exceeding 10^4.

5.2.1 $\alpha_r = 0$ *situations*

Here in generating the y_i we choose $\alpha_r = 0$, $\xi_o = 1$, and $\sigma_r = 0.2$ in Eq. (2.3) for the error model. This compromise choice for σ_r produces errors larger than usual in physical-sciences data but smaller than in much life-sciences data.

Table IV presents simulation results for a variety of weighting and fitting models. For these runs the number of replications ranged from 2×10^5 to 2×10^6. Lines 1 and 4–13 do not include transformation before fitting. The first three lines are for ξ free and involved 5×10^5 replications each.

The exponent in the error model Eq. (2.3), ξ_o, was relatively well determined in the line-1 model, with bias of <4%. Lines 2 and 3 present results for FPLWT with Type-I and Type-II inversion, in which we found comparable results to those in line 1. In particular, the predictions of Section 4.2 are very well borne out for Type-I inversion. Because of the large value of σ_r used here, Type-II inversion results are not very similar to those for Type-I, as line 3 shows, and the bias in θ_1 is much greater.

Line 4 in Table IV shows results for $U = 0$ and ξ fixed at 0.9637, as estimated from the b_ξ value of the line-1 fit. For comparison, the results in line 6 are for ξ fixed at unity, the value of ξ_o. It is evident that although the line-6 fit yields a better estimate of σ_r than does S_F in line 4, the line-4 bias estimates are appreciably smaller. Surprisingly, a fixed value of ξ unequal to the correct value, ξ_o leads to smaller bias than found with ξ fixed at ξ_o. In consonance with the normal-uniform

TABLE IV

MC simulation results for large-range monoexponential model with $\sigma_r = 0.2$. The notations -n or -u denote normally- or uniformly-distributed errors. The notations iI and iII indicate Type-I and Type-II inversion transformations (Section 4.2). A * indicates that the value is not statistically different from zero.

Line	Weight	s_F	$b_1 \times 10^3$	$s_{1L} \times 10^2$	$s_{1C} \times 10^2$	$b_2 \times 10^3$	$s_{2L} \times 10^2$	$s_{2C} \times 10^2$
1	FPLWT-n	0.2115	0.37	4.35	2.50	-0.61	1.31	1.53
2	FPLWT/iI-n	0.2117	0.25	4.35	2.51	0.61	1.31	1.53
3	FPLWT/iII-n	0.2385	-0.389	4.65	2.85	-0.75	1.43	1.77
4	FFWT-n	0.2120	-1.2	4.35	2.25	0.26	1.19	1.43
5	GLS-n	0.2420	-77	4.39	2.94	0.2	1.41	1.99
6	FPWT-n	0.2000	1.85	4.23	2.20	-1.42	1.41	1.41
7	FPWT-u	0.2001	1.91	4.23	2.20	-1.46	1.41	1.41
8	FFWT-u	0.2104	-0.77	4.34	2.25	0.05	1.21	1.43
9	GLS-u	0.2258	-75	4.20	2.54	0.4	1.41	1.73
10	DPWT-n	0.2188	-87.2	4.33	3.30	2.02	1.61	1.94
11	DPWT-u	0.2061	-80.0	4.11	2.40	1.76	1.51	1.60
12	UWT-n	39.1	746	14.6	230	2.47	0.656	13.0
13	UWT-u	40.0	701	13.3	188	2.06	0.665	12.8
14	UWT/LT-n	0.2101	-20.6	4.45	2.36	*	1.51	1.51
15	UWT/LT-u	0.2060	-20.1	4.36	2.32	*	1.48	1.48

comparison in Table III, we find that in the four such comparisons included in Table IV the differences between corresponding results are generally quite small.

Table IV also allows comparisons between ELS and GLS results for the present fitting model and data errors. First, comparison of the results in lines 4 and 5 is appropriate since they both use fixed ξ values. The GLS MC estimate was $\xi = 0.9578$, a somewhat worse estimate than the line-4 ELS value of 0.9637. As expected for normal-distributed errors, nearly all the other GLS results are also worse than those using ELS. Comparison of the results of lines 8 and 9 with data errors drawn from a uniform distribution shows, surprisingly, that again the ELS results are superior to the GLS ones. Here, the ξ in line 8 was 0.9679, nearly identical to that in line 9, 0.9692. The relative bias estimates from ELS are 50 to 100 times smaller than those from GLS for both normal and uniform errors.

These new results and those of Section 5.3.2, for Poisson-distributed errors, justify our recommendation to use the faster ELS fitting rather than slower GLS fitting for most work involving wide-range, physical-sciences data and errors. Thus, the criticisms of ELS by van Houweligen (1988) are unwarranted for such data.

Lines 10–13 of Table IV show results for various inappropriate weightings. Although both DPWT and UWT lead to more bias, the increase is particularly strong for the bias in θ_l. Figure 3 shows some of the relevant normalized distributions with very thin and long tails for the DPWT θ_2 and θ_l error distributions. For plotting resolution, the right-hand θ_2 tail, which extends to 6.9, was cut off at NBV = 3, as was the UWT θ_l error distribution, which is clearly very far from normal, with a skewness parameter of about 4, a kurtosis of 33, and extending up to 26.6. In spite of this pathological behavior, the UWT θ_2 distribution (not shown) is quite close to normal. The results shown in lines 10–13 in Table IV demonstrate the severe problems that arise from using incorrect weighting of two common types.

Lines 14 and 15 in Table IV show logarithmic transformation results. Again, b_2 is not statistically different from zero, and the relative bias in the transformed pre-exponential (in the b_l column) is dominated by the transformation bias. Upon subtracting the bias estimates in lines 3 and 5 of Table II, the residual bias estimates are 8×10^{-4} and 7×10^{-4}. These results and comparable ones in Table III, show that when only OLS fitting (UWT) is available, logarithmic transformation of monoexponential data and subsequent transformation-bias correction of the resulting θ_l estimate from Eq. (4.14) will yield essentially unbiased parameter estimates. This procedure is proper, however, only when ξ_o is unity, which may not be appropriate for the data considered. On the other hand, weighted NLLS fitting, as in lines 1 through 4 of Table IV, is more general and flexible since it is not limited to monoexponential response with $\xi_o = 1$, and NLLS should therefore be used for general-purpose fitting when available.

The results in Tables III and IV indicate that the linearized estimate of the SD of the θ_2 error distribution, s_{2L}, is an adequate to excellent approximation for s_{2C}

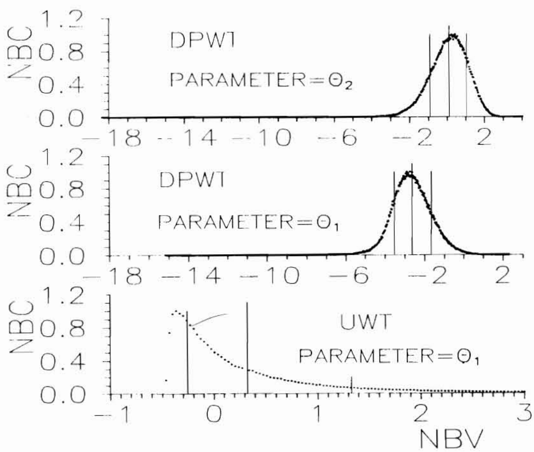

FIG. 3. Normalized distributions of the relative errors of θ_2 and θ_1 for the DPWT MC results in Table IV line 10, and the UWT MC fits in Table IV, line 12. Wide-range monoexponential model and strongly heteroscedastic data.

when the appropriate weighting is used, but that this is certainly not so for the corresponding SD estimates associated with the pre-exponential parameter θ_1, s_{1L} and s_{1C}. In particular, for the UWT response in Table IV, s_{1L} and s_{2L} values are exceedingly misleading and should be given no credence. Finally, use of FPLWT and FFWT, with the value of ξ found from the former weighting used in the latter, appears appropriate for the present high-heteroscedasticity data. The bias is closely proportional to σ_r^2 (or even closer to S_F^2 for large σ_r, where S_F becomes larger than σ_r), while quantities such as s_{jC} are nearly proportional to σ_r or to S_F. Therefore, whenever b_j is nonzero it may grow to at least as large as s_{jC} as σ_r increases. Thus, when errors in the data are appreciable, it is dangerous to neglect bias correction in exponential-fitting problems.

Thus far we have dealt only with the $\xi_o = 1$ error situation in Eq. (2.3). But, how well can ξ be estimated when ξ_o is not unity? Some answers to this question are provided by the results in Fig. 4. For FPLWT applied to MC simulation of the monoexponential model, we found an interesting decoupling between the results for ξ and σ_r when $\alpha_r = 0$. Although the parameter biases depend strongly on σ_r, the values of ξ do not. In fact, quantities related to ξ were found to change by only 1

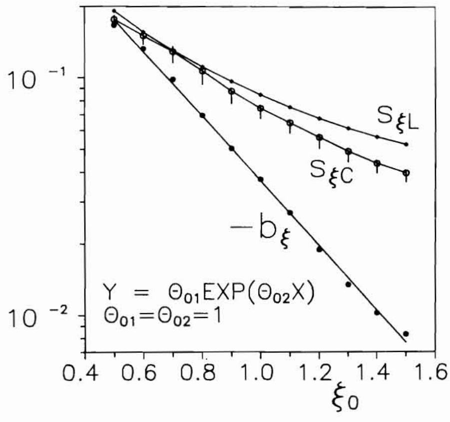

FIG. 4. Dependences of the dispersion measures, $s_{\xi L}$ and $s_{\xi C}$, and of the bias b_ξ on ξ_o for FPLWT and the wide-range monoexponential model. The short vertical lines on the $s_{\xi C}$ curve extend between the values of the actual 68.3% probability values $s_{\xi RH}$ (bottom) and $s_{\xi LH}$ (top).

to 3% as σ_r changed from 10^{-4} to 0.2. For this reason, the FPLWT fittings that led to the Fig. 4 results were carried out for both α_r and U set to zero and $\sigma_r = 10^{-4}$. In Fig. 4 the short vertical lines on the $s_{\xi C}$ curve are drawn between an upper value of $s_{\xi LH}$ and a lower value of $s_{\xi RH}$, showing how the conventional $s_{\xi C}$ dispersion measures differs from the true confidence interval values for. Thus $s_{\xi L}$ is a poor approximation to $s_{\xi C}$ or to $s_{\xi AV}$ for $\xi_o > 0.8$. For $\xi_o = 1.5$ $s_{\xi L}$ is over 30% too large.

The line associated with the b_ξ points in Fig. 4 is from exponential fitting. The various dispersion measures also decrease as ξ_o increases, but more slowly than does the bias. Therefore b_ξ may be neglected for large ξ_o, but should not be ignored for ξ_o small. On fitting the b_ξ results to $-A_o \exp(-A_1 \xi_o)$ using FPLWT, we obtained $A_o = 0.840 \mid 0.04$, $A_1 = 0.320 \mid 0.014$, and $\xi \approx 0.86 \mid 0.3$. Here the combination $A \mid \Delta A$ indicates a parameter estimate, A, and its estimated relative standard deviation, ΔA. When the inversion transformation is applied, so that we have exponential decay instead of growth, the results agree with the predictions of Section 4.2. In particular, the relation $\xi + \xi^\# \approx 2$ holds well, as does Eq. (4.18).

Some Type-II inversion results were obtained for the extreme choice $\xi_o = 2$, using $\sigma_r = 10^{-5}$ in order to avoid generating any negative y_i. Using $2\text{x}10^5$ samples we obtained $b_\xi = 4.6\text{x}10^{-3}$ and $s_{\xi C} = 0.029$, while $b_\xi = 6.8\text{x}10^{-3}$ and $s_{\xi C} = 0.062$ were found for the corresponding inversion-transformed fits. Since $\xi_0^{\#} = 0$, these latter results are direct, not relative quantities, and they show that the bias in the inverse-transformed ξ cannot be distinguished from zero on a statistical basis and thus $\xi^{\#} \approx 0$ as expected. Nevertheless, the biases themselves are determined to better than 5%, and lead to $\xi + \xi^{\#} = 1.998$, satisfactorily close to 2.

5.2.2 *Situations with $\alpha_r \neq 0$*

Thus far we have set α_r to zero in the error model, Eq. (2.1). Here we present a MC simulation study where this is not so. We begin with the specific wide-range monoexponential growth model and data in Section 5.2.1 and convert them to a decay model by direct Type-I inversion. For the error model we take $\xi_o = 1$ and α_r of the order of magnitude of the smallest data value, here $\exp(-10) \approx 5\text{x}10^{-5}$. We therefore select $\alpha_r = 10^{-5}$ and find MC estimates of the relative biases b_U and b_ξ for various σ_r choices. Such results allow us to evaluate how well ξ, U, and α_r can be estimated for the present situation. For sufficiently large σ_r, the proportional errors should dominate, while for small enough σ_r the additive ones should do so.

Table V summarizes results obtained with $2\text{x}10^5$ replications. No bias estimates for θ_1 and θ_2 are included because even with $\sigma_r = 0.01$ the fitting model parameter biases were found to be entirely negligible. Since fitted values of U can be of either sign without affecting weighted fit results, for simplicity we take U_o positive and calculate b_U using $e_{Uk} = (|U_k| - U_o)/U_o$. When U and ξ are both free to vary, we find more frequent fit convergence failure in these MC simulations. To eliminate such non-converging fits early and thus save computer time, we immediately terminated all searches for which $|U/U_o| > 4$. For the line-4 run, about 3% of the fit trials were so eliminated, but the percentage was somewhat greater for other σ_r values.

Consider first the FGWT results in Table V. We see that S_F is only slightly biased until σ_r approaches α_r. As usual, S_F is a good estimate of σ_r in its region of main interest. We are particularly concerned with the biased estimates ξ and $\widehat{\alpha}_r$, and with the ξ_o and α_r distributions SD's, which allow us to evaluate how well our fitting method estimates the error-model parameters if there are both additive and power-law errors.

TABLE V

MC simulation results for a large-range monoexponential decay model with $\alpha_r = 10^{-5}$ and $\xi_0 = 1$. The notation $A{:}\Delta A$ denotes an estimate A and the standard deviation of its associated distribution, ΔA. All the simulations are for normal error distributions. Entries indicated by --- are for quantities not relevant to the analysis.

Line	Weight	σ_r	S_F/σ_r	$b_U{:}s_{UC}$	$b_\xi : s_{\xi C}$	$\xi : S_{\xi C}$	$\hat{\alpha}_r : S_{\alpha C}/\alpha_r$
1	FPLWT	$1\text{x}10^{-2}$	0.920	---	−0.20:0.09	0.80:0.09	---
2	FGWT	$1\text{x}10^{-2}$	1.080	−0.09:0.53	0.07:0.14	1.07:0.14	0.90:0.53
3	FGWT	$3\text{x}10^{-3}$	1.077	−0.11:0.45	0.08:0.17	1.08:17	0.89:0.45
4	FPLWT	$1\text{x}10^{-3}$	0.848	---	0.45:0.08	0.55:0.08	---
5	FGWT	$1\text{x}10^{-3}$	1.076	−0.10:0.40	0.097:0.23	1.10:0.23	0.90:0.40
6	FGWT	$3\text{x}10^{-4}$	1.075	−0.10:0.35	0.14:0.33	1.14:0.33	0.90:0.35
7	FGWT	$1\text{x}10^{-4}$	1.073	−0.11:0.33	0.19:0.50	1.19:0.50	0.89:0.33
8	FGWT	$3\text{x}10^{-5}$	1.000	−0.43:0.47	−0.20:0.74	0.80:0.74	0.57:0.47
9	FGWT	$1\text{x}10^{-5}$	1.315	−0.93:0.19	−0.91:0.24	0.09:0.24	0.073:0.19
10	FPLWT	$1\text{x}10^{-5}$	1.300	---	−0.95:0.06	0.05:0.06	---

The SD values in Table V are of s_{jC} or S_{jC} type, which we present as $A_j{:}S_{jC}$, where lower-case letters are for relative quantities and upper-case letters are for the quantities themselves, and the colon divider identifies a distribution SD. If A_j is the bias in a single measurement of value a_j from a particular distribution, the nominal 68.3% confidence-interval estimate around the bias-corrected value of a_j would extend from $(a_j - A_j) - S_{jC}$ to $(a_j - A_j) + S_{jC}$. The number of MC replications was always taken large enough that values of such quantities as b_U are estimated to $\leq 1\%$; therefore, their estimated SDs are not presented; those of the underlying distribution are generally much larger and are of primary interest here.

In the rightmost column of Table V are values of $\widehat{\alpha_r} / \alpha_r = 1 + b_U$. Note that the bias and uncertainty of an ξ_o estimate increase as σ_r decreases, while those of α_r estimates are substantially constant or decrease. The results in line 9 show, however, that when σ_r and α_r are comparable, a meaningful estimate of α_r can no longer be obtained. Instead, ξ becomes very small, a condition that yields nearly UWT, which is then the preferred choice. The line-10 FPLWT results again yield a very small ξ, further indicating the appropriateness of UWT, and the S_F estimate is no longer close to the proper, but very small, σ_r value. Not surprisingly, the $|U|$ and ξ distributions are far from normal. For example, for $\sigma_r = 10^{-3}$, the skewness and kurtosis are, respectively, about 0.56 and 0.91 for the $|U|$ distribution and about 1.7 and 5.6 for the ξ distribution. For $\sigma_r = 10^{-5}$, the corresponding values are about 7 and 64, and 8 and 81, respectively.

In Table V the results in line 1 should be compared to those in line 2. First, we see that the S_F biases are comparable, but of opposite sign, for the two MC results. Second, the use of FPLWT, as in line 1, leads to a much greater bias of the ξ_o estimate and to a smaller estimate of the SD of its distribution, making the poorly-estimated ξ value appear much more accurate than it is. But, by using FGWT (and thus allowing U to be free), one takes proper account of the additive errors and obtains reasonable estimates of σ_r, ξ_o, and α_r. Similar conclusions follow when we compare the FPLWT results in line 4 with the FGWT ones in line 5. Although smaller SD values of the ξ_o and α_r distributions than those found would be desirable, there are appreciable regions of σ_r in the present case for which the SD values are small enough to make it worthwhile to fit with both U and ξ varying.

5.3 Analysis of Radioactive Decay by a Sum-of-Exponentials Model

In order to demonstrate how useful the ELS method is for analyzing physical-sciences data, we now consider the radioactive decay data in Table VI. To obtain these results we irradiated a sample of ^{103}Rh with neutrons, and monitored the gamma decay of ^{104}Rh with a scintillation detector. Irradiation produced two different radioactive states of ^{104}Rh and, to a very good approximation, they decay independently. The appropriate NLLS fitting model is thus

$$Y_i = \theta_1 + \theta_2 \exp(-t_i/\theta_3) + \theta_4 \exp(-t_i/\theta_5) \tag{5.1}$$

TABLE VI

Radioactive decay data for ^{104}Rh. Here t is time in seconds and y is the number of counts in the interval centered on t.

t	y	t	y	t	y	t	y	t	y	t	y
2	864	58	462	114	245	170	154	270	117	802	70
6	800	62	418	118	225	174	164	290	120	902	74
10	715	66	371	122	215	178	166	310	92	1002	76
14	705	70	346	126	189	182	155	330	91	1102	70
18	697	74	362	130	192	186	129	362	112	1202	68
22	685	78	307	134	207	190	141	402	102	1302	86
26	665	82	330	138	195	194	153	442	79	1402	85
30	621	86	311	142	183	198	151	482	87	1502	59
34	606	90	285	146	178	202	134	502	101	1602	68
38	541	94	292	150	178	206	137	542	98	1702	79
42	522	98	289	154	193	214	142	582	76	1802	82
46	510	102	273	158	167	230	117	622	77	1902	80
50	469	106	271	162	175	250	142	702	68	2002	66
54	423	110	241	166	184						

with all θ_i positive, θ_l a background count, and t_i the time from the start of counting. We now use this expression for several different EVM analyses of the data. We begin by analyzing the actual data, then continue in Section 5.3.2 with a MC study of comparable synthetic data.

5.3.1 *Analysis of decay of* ^{104}*Rh*

It has been shown that there is negligible $1/f$ noise in alpha decay (Kennett and Prestwich 1989), so it is plausible to expect none for the present situation. Since there does not appear to be any reason for additive errors in the data, any weighting involving U in Eq. (2.8) is inappropriate. Nevertheless, for comparison purposes we start with UWT and compare fitting results with DPLWT, FPLWT, and FPWT. In addition, two variance models especially appropriate for Poisson statistics will be defined. For arbitrary ξ, define a Poisson form of FPLWT, called PFPLWT, by replacing the τ in Eq. (2.8) by

$$\tau_i^2 = \left[\theta_1\right]^{2\xi} + \left[\theta_2 \exp(-t_i/\theta_3)\right]^{2\xi} + \left[\theta_4 \exp(-t_i/\theta_5)\right]^{2\xi} \tag{5.2}$$

Each term in Eq. (5.1) is associated with a different and independent Poisson process for which the variance of a data value is equal to that value, therefore $\xi = 0.5$. When this particular ξ value is used in Eq. (5.2), it leads to just FPWT. Although there is thus no need to distinguish between the two variance models if $\xi = 0.5$, it is worthwhile to compare their predictions, which should differ when $\xi \neq 0.5$. Finally, let PFFWT denote PFPLWT with ξ fixed.

An independent measurement of the background term θ_l in Eq. (5.1) was also made during the experiment and yielded $\theta_l = 70 \mid 0.0053$. Although it is appropriate to use this *a priori* information in NLLS fits of the decay data, we also investigate how well the fits estimate it. Previous measurements (Blachot *et al.* 1984) led to quite precise estimates, $\theta_3 = 61.03 \mid 0.0095$ sec and $\theta_5 = 373.95 \mid 0.012$ sec.

Table VII summarizes the results of fitting the data with many different EVM's and of several long MC simulation runs. It is a notoriously difficult and ill-conditioned problem to estimate adequately the parameters from real data involving two or more exponentials whose time constants are not very much different or when there is parameter redundancy (Lanczos 1956, Seber and Wild 1989, pp. 118-119). Thus, even with the most appropriate EVM, we expect appreciable uncertainties in parameter estimates. The values shown in line 10 of the table were produced by fixing all parameters except θ_2 and θ_4 at appropriate values, as discussed above. Their values in line 10 are probably best estimates.

The use of UWT fitting produces parameter estimates that are largely determined by the region where the y_i are largest, especially when their range is large. The UWT results in line 1 confirm this expectation, and show that θ_2 and θ_3, associated with the short-time region of the data, are far better determined than are θ_4 and θ_5, associated with the long-time region. Although a good estimate of θ_l is obtained, its RSD estimate is large. As Table VII shows, there are no significant differences between the results in line 2 (where $Z_i = y_i$) and line 3 (for which $Z_i = Y_i$). The FPPLWT results of line 4 seem to be slightly inferior to those of line 3. All the estimates of lines 1-4, however, are within one SD of the values in line 10.

TABLE VII

Individual fit results (lines 1-10), and simulation results with θ_l fixed at 70 (lines 11-13) for radioactive-decay data. A quantity written $A \mid \Delta A$ represents a parameter estimate and its estimated relative standard deviation (RSD). Entries with --- indicate quantities that are not relevant to analysis. The suffix -p in lines 11-13 denotes a Poisson error distribution.

	Weight	S_F	θ_1		θ_2		θ_3		θ_4		θ_5		ξ	
1	UWT	16.2	71.5	0.11	758	0.035	65.1	0.050	49.9	0.49	445	0.91	---	
2	DPLWT	0.847	71.8	0.042	736	0.043	60.3	0.065	83.7	0.41	284	0.38	0.54	0.27
3	FPLWT	0.859	72.8	0.041	737	0.043	60.0	0.065	84.6	0.41	284	0.38	0.53	0.28
4	PFPLWT	0.597	72.9	0.040	735	0.043	59.2	0.068	89.4	0.40	275	0.36	0.60	0.29
5	FPLWT	0.967	70		747	0.028	61.1	0.051	75.3	0.30	348	0.27	0.51	0.28
6	PFPLWT	0.842	70		745	0.029	60.3	0.054	79.7	0.29	335	0.26	0.58	0.29
7	FPWT	0.084	70		741	0.040	57.7	0.071	93.6	0.27	305	0.20	1.0	
8	PFFWT	0.110	70		735	0.043	55.0	0.068	114	0.25	264	0.19	1.0	
9	PFFWT	1.008	70		747	0.027	61.1	0.050	74.9	0.29	350	0.27	0.5	
10	PFFWT	0.995	70		752	0.016	61.03		72.0	0.06	373.95		0.5	
11	FPLWT	1.272	b_j:		- 0.0043		-0.0045		0.055		0.046		-0.024	
	-p/MC		s_jC:		0.0282		0.0487		0.315		0.293		0.230	
			s_jLH:		0.0275		0.0487		0.269		0.253		0.230	
			s_jRH:		0.0249		0.0469		0.318		0.298		0.226	
12	FFWT	0.997	b_m:		-0.0041		-0.0043		0.053		0.047		---	
	-p/MC		s_mC:		0.0279		0.0480		0.309		0.292		---	
			s_mLH:		0.0273		0.0482		0.266		0.251		---	
			s_mRH:		0.0248		0.0463		0.315		0.297		---	
13	GLS	1.331	b_j:		-0.0081		-0.0094		0.089		-0.023		-0.041	
	-p/MC		s_jC:		0.0298		0.0492		0.334		0.266		0.22	
			s_jLH:		0.0290		0.0493		0.282		0.230		0.22	
			s_jRH:		0.0260		0.0472		0.336		0.269		0.22	

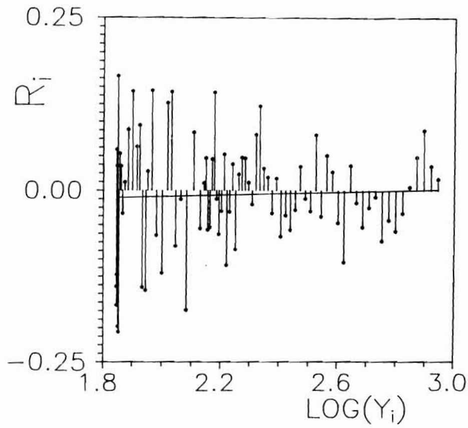

FIG. 5. Weighted residuals for the FPWT, $\xi = 1$, fit in line 7, Table VII. Radioactive decay data and sum-of-exponentials model. The straight-line fit is discussed in Section 5.3.1.

Although the background value θ_l is well estimated by any of the weighting schemes, the results in lines 5 and 6 of Table VII show that better estimates of the other parameters and their RSD values are obtained when θ_l is fixed at its measured value. Of particular note is the best-fit ξ estimate in line 5, very close to the Poisson-statistics value of 0.5.

A FPLWT Type-II inversion fit yielded $\xi^{\#} = 1.46 \mid 0.10$; thus, here $\xi + \xi^{\#} = 1.97$, rather close to 2. Averaging of untransformed and transformed parameter estimates, as discussed in Section 4.2.3, led to $< \theta_3 > = 61.036$, very close to the accepted value of 61.03 for this parameter in line 10. Such averaging for the other free parameters did not yield results closer to those in line 10. It is not surprising that averaging helps reduce bias for the θ_3 estimate, since this is the time constant of the dominant exponential-decay term, and we thus expect from the analysis in Section 4.2 that b_3 and $b_3^{\#}$ should be nearly equal.

We examine in lines 7 and 8 what happens with a fixed but wrong value of ξ. With $\xi = 1$, a plausible choice if one did not know that the data arose from Poisson processes, we see that all parameter estimates are worse than those in line 5 or in line 9. Figure 5 shows the weighted residuals for the line-7 fit *vs* log (Y_i). The vertical lines extend from 0 to R_i. This figure shows very appreciable heterosce-

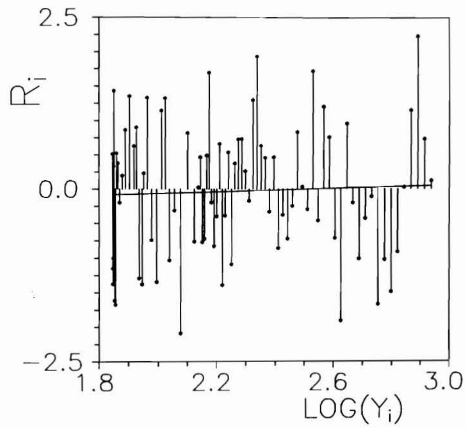

FIG. 6. Weighted residuals for the PFFWT, $\xi = 0.5$, fit of line 9, Table VII. Radioactive decay data and sum-of-exponentials model. The straight-line fit is discussed in Section 5.3.1.

dasticity even of weighted residuals. But Fig. 6, for line 9 ($\xi = 0.5$), appears to be very nearly homoscedastic, as it should be for the appropriate ξ choice. A quantitative measure is afforded by the results in Fig. 7 for $|R_i|$. Completely homoscedastic data should yield a linear-fit straight line of slope zero within the limits of statistical variability. Here the line is nearly horizontal and its equation is $|R_i| = 1.007 \mid 0.42 - (0.079 \mid 2.4) \log(Y_i)$. The equation for the line-7 $|R_i|$ data is $|R_i| = 0.230 \mid 0.15 - (0.072 \mid 0.21) \log(Y_i)$. Since the mean of $|R_i|$ is about 0.066 for the $\xi = 1$ fit and about 0.83 for the $\xi = 0.5$ fit, the effective slope is reduced by about a factor of 10 on reducing ξ. Further, the RSD values of the slopes indicate that although the $\xi = 1$ slope estimate is significantly different from zero, that for $\xi = 0.5$ is not. Finally, in Table VII the line-9 R_i values mostly lie very close to a straight line on a cumulative normal probability plot.

The above results indicate that ξ can be well estimated from the present data, that proper weighting can reduce a heteroscedastic situation to a homoscedastic one, and that use of the appropriate EVM leads to optimum results for parameter estimates. Nevertheless, the ill-conditioned nature of the problem shows up when one compares the θ_2 and θ_3 estimates and their estimated RSD values with those for θ_4 and θ_5. In order to obtain more-accurate estimates of the latter parameters, one

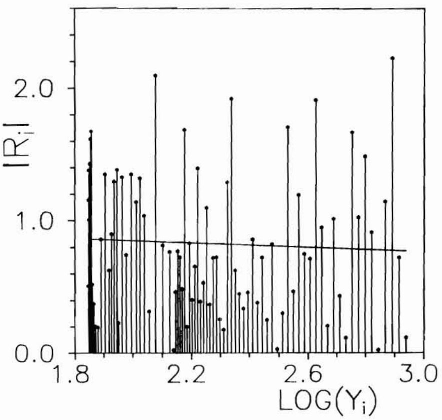

FIG. 7. Absolute values of the weighted residuals in Fig. 6. The straight-line fit is discussed in Section 5.3.1.

would either need to be able to obtain data for the decay of this species separately or to analyze data for which θ_4/θ_l was much larger.

5.3.2 *MC study of two-component exponential decay*

We further investigated two-component exponential decay by MC simulation. In order to obtain exact and consistent values of parameters and data for simulations as close as possible to the experimental results in Section 5.3.1, the following procedure was used. A data set was generated using as θ_m values: 70, 752, 61, 72, and 374, essentially the Table-VII, line-10 results. The $y_i = Y_i$ values were then rounded to integers. Fitting these data with fixed $\theta_l = 70$ and $\xi = 0.5$ values yielded new parameter estimates very close to those cited. These estimates and the integer data were then used as input for subsequent MC simulation involving independent pseudo-random Poisson-distributed errors (identified by p in Table VII). Then, statistically independent pseudo-random errors given by Eq. (2.3) having $\alpha_r = 0$, $\sigma_r = 1$, and $\xi_0 = 0.5$ were added to the exact data values calculated from the Eq. (5.1) model for each replication. Fitting was done with $U = 0$ and with ξ either free to vary or fixed at 0.5 (line-12 results).

In Table VII the fits in lines 11 and 12 were obtained with 500,000 replications and yielded relative bias estimates all of whose RSDs were less than 1%. The

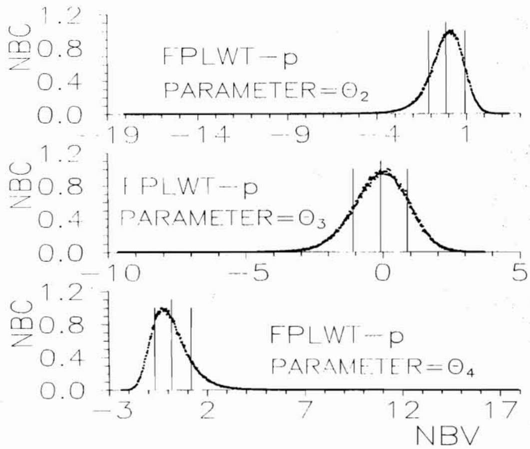

FIG. 8. Normalized distributions of the relative errors of the θ_2, θ_3 and θ_4 parameters for the FPLWT-p MC results in line 11 of Table VII. Radioactive decay data and sum-of-exponentials model.

results in line 13 for GLS fitting are based on only 200,000 samples. These simulation runs were very lengthy; for example, that in line 12 required about 10 hours on a Cray Y-MP supercomputer and that in line 13 used about 16 hours. Thus, the GLS runs required about four times more computation time than did comparable ELS ones. As expected, the relative bias estimates of θ_2 and θ_3 are much smaller than those of θ_4 and θ_5, and, most important, they are small compared to the relative uncertainty of the parameters, as indicated by the s_{jC}, s_{jLH}, and s_{jRH} values shown. Comparison of lines 11 and 13 shows that for all but parameter θ_5 the ELS results are substantially superior to the GLS ones, even though the data errors are Poisson-distributed.

The results in lines 11 and 12 show, as expected from the single-data-set fit results in Table VII, that both the bias and the RSD values of the parameters of the short-time exponential decay are much smaller than those for the long-time decay. These results suggest that the biases of θ_4 and θ_5 are not negligible. For example, when the line-9 value of θ_4 is corrected using the line-12 bias value, one obtains an estimate of 71.1, appreciably closer to the line-10 value. The line-11 and line-12 results show that, depending on the particular parameter, one or the other of s_{jLH} or

s_{jRH} is usually close to s_{jC}. Further, all s_{jL} estimates are smaller than corresponding s_{jC} ones and lie between the corresponding s_{jLH} and s_{jRH} values. The correspondence of the MC dispersion s_{jC} values in lines 11 and 12 with the individual RSD values in line 9 is particularly striking. The individual RSD values are evidently excellent estimators here of relative dispersion in parameter error distributions.

A MC run like that of line 12 was also carried out with normally- rather than Poisson-distributed errors. A value of $\sigma_r = 1.0$ was used and the results were very close to those of line 12, as expected from the central-limit theorem. In Fig. 8 we show the error distributions found for three of the five free parameters of the line-11 MC simulation. Although that for the θ_2 errors appears nearly normal except for its long, thin, left-side tail, its skewness parameter was about 1.1 and its kurtosis was 4.9. In contrast, the θ_3 error distribution, that of the dominant time constant, is much closer to normal and has skewness and kurtosis values of 0.3 and 0.54, respectively. The third distribution, with skewness and kurtosis of 1.4 and 5.8, respectively, is similar to that of the bottom distribution of Fig. 1, that associated with large-error inversion. The distribution of the θ_5 errors was quite similar to that for θ_4. Surprisingly, the error distribution for the ξ estimate was very close to normal, with skewness and kurtosis of about 0.06 and 0.1, respectively.

ACKNOWLEDGMENTS

We are grateful for the Convex C240 computer time provided through Academic Computing Services at UNC and for the Cray Y-MP computer time provided by the North Carolina Supercomputing Center.

PRINCIPAL ACRONYMS AND SYMBOLS

Acronyms

ELS	Extended least squares
EVM	Error-variance model
GLS	Generalized least squares
LEVM	Fitting program used herein
MC	Monte Carlo
NBC	Normalized block count
NBV	Normalized block value
NLLS	Nonlinear least squares
OLS	Ordinary least squares
RSD	Relative standard deviation
SD	Standard deviation

Weighting: See Table I for definitions

Major Subscripts		*Superscripts*	
i	data label	c	continuous distribution
j, m	parameter label	d	discrete distribution
k	replication label	*	logarithmic transformation
n	normal distribution	#	inversion transformation
o	exact value of parameter		
u	uniform distribution		

BIBLIOGRAPHY

Bates, D. M., and Watts, D. G. (1988). *Nonlinear Regression Analysis and its Applications*, New York: John Wiley.

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

Blachot, J., Husson, J. P., Oms, J., and Berrier, G. (1984). *Nuclear Data Sheets*, 41, 325-412, and references therein.

Brownlee, K. A. (1965). *Statistical Theory and Methodology*, 2nd. Ed., New York: John Wiley.

Carroll, R. J., and Ruppert, D. (1988). *Transformation and Weighting in Regression*, New York: Chapman and Hall.

Cook, R. D., and Weisberg, S. (1982). *Residuals and Influence in Regression*, New York: Chapman and Hall, p. 59.

Davidian, M. (1990). Private communication.

Davidian, M., and Carroll, R. J. (1987). 'Variance Function Estimation,' *J. Am. Stat. Assoc.*, 82, 1079-1091.

Donaldson, J. R., and Schnabel, R. B. (1987). 'Computational Experience with Confidence Intervals for Nonlinear Least Squares,' *Technometrics*, 29, 67-82. See also Gross, A. M. (1976). Confidence Interval Robustness with Long-Tailed Symmetric Distributions,' *J. Am. Stat. Assoc.*, 71, 409-416.

Finney, D. J., and Phillips, P. (1977). 'The Form and Estimation of a Variance Function, with Particular Reference to Radioimmunoassay,' *Appl. Statis.*, 26, 312-320.

Gallant, A. R. (1987). *Nonlinear Statistical Methods*, New York: John Wiley.

Giltinan, D. M., and Ruppert, D. (1989). 'Fitting Heteroscedastic Regression Models to Individual Pharmacokinetic Data Using Standard Statistical Software,' *J. Pharmacokin. Biopharm.*, 17, 601-614.

Hinkley, D. V., and Runger, G. (1984). 'The Analysis of Transformed Data,' *J. Am. Stat. Assoc.*, 79, 302-309.

Jobson, J. D. and Fuller, W. A. (1980). 'Least Squares Estimation when the Covariance Matrix and Parameter Vector are Functionally Related,' *J. Am. Stat. Assoc.* 75, 176-181.

Kennett, T. J., and Prestwich, W. V. (1989). 'Limit on the Existence of 1/f Noise in Decay,' *Phys. Rev.*, A40, 4630-4640.

Lanczos, C. (1956). *Applied Analysis*, Englewood Cliffs, N.J.: Prentice Hall; pp. 272-280.

Macdonald, J. R., editor (1987). *Impedance Spectroscopy Emphasizing Solid Materials and Systems*, New York: John Wiley.

Macdonald, J. R., Hooper, A., and Lehnen, A. P. (1982). 'Analysis of Hydrogen-Doped Lithium Nitride Admittance Data,' *Solid State Ionics*, 6, 65-77.

Macdonald, J. R., and Potter, L. D., Jr. (1987). 'A Flexible Procedure for Analyzing Impedance Spectroscopy Results: Description and Illustrations,' *Solid State Ionics*, 23, 61-79.

Moré, J. J. (1978). 'The Levenberg-Marquardt Algorithm: Implementation and Theory,' in G. A. Watson, editor, *Numerical Analysis, Lecture Notes in Mathematics*, Vol. 630, pp. 105-116, Berlin: Springer Verlag.

Nielsen, K. (1977). 'A Method of Optimizing Relative Weights in Least-Squares Analysis,' *Acta Crystallogr.*, A33, 1009-1010. See also: Celmins, A. (1984), 'Analysis of Residuals from Multidimensional Model Fitting,' *Comp. & Chem.* 8, 81-89.

Norman, E. B., Gazes, S. B., Crane, S. G., and Bennett, D. A. (1988). 'Tests of the Exponential Decay Law at Short and Long Times,' *Phys. Rev. Lett.*, 60, 2246-2249.

Ratkowsky, D. A. (1983). *Nonlinear Regression Modeling*, New York: Marcel Dekker.

Seber, G. A. F., and Wild, C. J. (1989). *Nonlinear Regression*, New York: John Wiley.

Tukey, J. W. (1986). *The Collected Works of John W. Tukey*, ed. Jones, L. V, Wadsworth and Brooks, Monterey, California, p. 1010.

van Houwelingen, J. C. (1988). 'Use and Abuse of Variance Models in Regression,' *Biometrics* 44, 1073-1081.

Wu, C.-F (1981), 'Asymptotic Theory of Nonlinear Least Squares Estimation,' *Annals of Statistics,* 9, 501-513.

Received February 1991; Revised April 1991.

Recommended by R. F. Gunst, Southern Methodist University, Dallas, TX.

Refereed by Daniel J. Sehnell; Centers for Disease Control; Atlanta, GA.