AD NUMBER

AD464183

NEW LIMITATION CHANGE

TO

Approved for public release, distribution unlimited

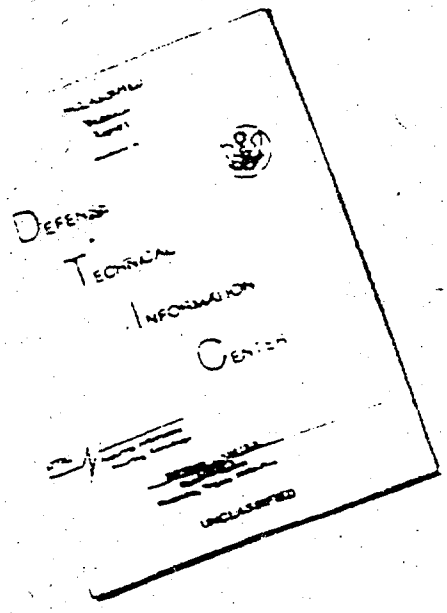
FROM

Distribution authorized to U.S. Gov't. agencies and their contractors; Administrative and Operational Use; March 1965. Other requests shall be referred to the Air Force School of Aerospace Medicine, Brooks AFB, TX 78235.

AUTHORITY

AFSAM, per ltr dtd 20 Apr 1967

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

SAM-TR-65-4

SOME PROBLEMS ON THE USE OF NEGATIVE EXPONENTIAL CURVES IN BIOLOGY

M. BRYAN DANFORD, Ph.D.

March 1965

USAF School of Aerospace Medicine
Aerospace Medical Division (AFSC)
Brooks Air Force Base, Texas

Qualified requesters may obtain copies of this report from DDC. Orders will be expedited if placed through the librarian or other person designated to request documents from DDC.

When U. S. Government drawings, specifications, or other data are used for any purpose other than a definitely related government procurement operation, the government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

SOME PROBLEMS ON THE USE OF NEGATIVE EXPONENTIAL CURVES IN BIOLOGY

M. BRYAN DANFORD, Ph.D.

FOREWORD

This report was prepared in the Biometrics Branch, under task No. 631901. It was submitted for publication on 29 December 1964.

This report has been reviewed and is approved.

Harold V. Ellingson

HAROLD V. ELLINGSON
Colonel, USAF, MC
Commander

ABSTRACT

Many biologic processes give responses that decrease rapidly over time to some asymptote $C \geq 0$. The mathematical expressions that describe the various phenomena vary in complexity and form. To fit a curve to data in any such situation, one must consider both the formula for the trend and the nature of the deviations or error terms. A discussion of these problems is given to indicate to the data analyst the possible choices that are his, relative to assumptions about error terms and relative to technics or methods of estimation. No new analytic procedures are given.

It is noted that several methods of estimation are being studied for certain assumed model equations. The results of these random sampling experiments will be reported in subsequent papers.

SOME PROBLEMS ON THE USE OF NEGATIVE EXPONENTIAL CURVES IN BIOLOGY

1. INTRODUCTION

Many biologic processes give responses that decrease rapidly over time to some asymptote $C \geqq 0$. The mathematical expressions that describe the various phenomena vary in complexity and form. To fit a curve to data in any such situation, one must consider both the formula for the trend and the nature of the deviations or error terms. The intent of the discussion that follows will be to help indicate to the data analyst the possible choices that are his, relative to assumptions about error terms and relative to technics or methods of estimation. No new analytic procedures are proposed. First, there will be a consideration of equation motivation. This will be followed by a discussion of the error terms in the model equations. Possible extensions of certain models will be noted, followed by a discussion of various estimation methods. Hypothesis testing is briefly discussed, with a final section devoted to concluding remarks.

2. MOTIVATION OF THE MODEL EQUATIONS

For a continuous process such as an organism's blood-sugar level, a scientist might simply observe that following the ingestion of a meal, there is an increase and a subsequent decrease of the level of the sugar in the blood. It might be further observed that the shape of the declining curve from the highest point attained appears to be of a negative exponential form, $\eta_t = \alpha \exp(-\beta t)$. By trying various transformations, a straight line might be obtained when the logarithm of the response is plotted against time, or the ratio of the observation at time t to that at time $t-1$ might give a constant for all values of t. This is empirical curve fitting or modeling. No rationale is presented for the choice of the equation. Such empirical curve fitting may in turn suggest mechanisms and thus lead to better understanding, of course.

In opposition to empirical curve fitting is the mechanistic approach. From first principles, the scientist tries to predict the equation or the form of the response as a function of time. For a simplified example, consider the rate of the nitrogen washout of the lung for an animal breathing room air and placed on pure oxygen. The room air concentration of nitrogen is about 79%. Respiration depth is kept at a constant level; rate is allowed to vary. The nitrogen concentration of each expired volume of air is determined. If the total resting lung volume is denoted by v and the volume of inspired air at each breath by Δv, then initially, or at time zero,

$F_{N_2}(0)$ = concentration of nitrogen = .79
$F_{O_2}(0)$ = concentration of oxygen = .21
$V_{N_2}(0)$ = volume of nitrogen = .79v
$V_{O_2}(0)$ = volume of oxygen = .21v

and after the first breath of pure oxygen,

$$F_{N_2}(1) = \frac{V_{N_2}(0)}{V_{N_2}(0) + V_{O_2}(0) + \Delta v} = \frac{.79v}{.79v + (.21v + \Delta v)} = .79\frac{v}{v + \Delta v} = .79\omega,$$

$$F_{O_2}(1) = \frac{.21v + \Delta v}{.79v + .21v + \Delta v} = \frac{.21v + \Delta v}{v + \Delta v}$$

$$V_{N_2}(1) = F_{N_2}(1)v = .79\omega v$$

$$V_{O_2}(1) = F_{O_2}(1)v = \frac{.21v + \Delta v}{v + \Delta v}(v) = v - .79\omega v,$$

where

$$\omega = \frac{v}{v + \Delta v}.$$

The concentration after the second breath will be

$$F_{N_2}(2) = \frac{V_{N_2}(1)}{V_{N_2}(1) + V_{O_2}(1) + \Delta v} = \frac{.79\omega v}{v + \Delta v}$$
$$= .79\omega\left(\frac{v}{v + \Delta v}\right) = .79\omega^2 .$$

Continuing in the same manner, it can be shown that

$$F_{N_2}(t) = .79\omega^t ;$$

ω is called the dilution ratio. Thus, the concentration at breath t is found by multiplying the concentration at breath t — 1 by the dilution ratio ω, or $F_{N_2}(t) = \omega F_{N_2}(t-1)$. The equation is generated from mechanical principles that are peculiar to this biologic process.

Parenthetically, it should be observed that for many applications, the implication is that observations are made at discrete and usually equidistant time points. For some situations, the response cannot be measured and may possibly not even be defined at nonintegral values of t. For others, the process may be truly continuous but with sampling performed at fixed time intervals. In still other situations, the independent variable may not be chronologic time, but only time related. Thus, in respiration studies, the concentration of a gas in each volume of expired air may be determined. If respiration rate is not fixed, the actual time between breaths is not fixed. In this case, it may be desirable to view breath number as the time metameter.

More complicated equations, such as the sum of two or more negative exponential terms, may be found in similar ways: by trial and error and past experience, a curve may be fitted empirically; or from a knowledge of the mechanisms involved, an equation may be derived to describe the response. In both approaches, there is probably a tendency toward parsimony—i.e., one attempts to explain or fit the data with the simplest expression. And in most curve-fitting situations the distinction between arbitrary curve fitting and mechanistic modeling is not as clear cut as indicated above. There is very likely a bit of empiricism as well as theory generation in nearly every curve-fitting situation.

Not only might the equation to be fitted be arrived at in many different ways, but different-appearing equations are used to describe the same process. That is, the same process may be represented by different equations. For example, in the nitrogen washout example above, the equation $F_{N_2}(t) = \omega F_{N_2}(t-1)$ led to $F_{N_2}(t) = .79\omega^t$. In general, an m^{th} order difference equation may be thought of as the model equation. Yet, its solution will be the sum of m exponential terms. Which equation one uses to fit to the data will depend largely on the assumptions about the error terms, a consideration of which will be the subject of the next section.

That a given process can be described in different ways is rather obvious; the point is that since biologists present equations in many forms, the data analyst should be able to recognize the "best" equation to use in curve fitting, regardless of how the process is described.

3. THE NATURE OF THE ERROR TERM

When a random error term is added to the mathematical equations, they are then referred to as stochastic or model equations. It is at this point that the curve fitter needs to take note, for there are many assumptions which can be made about the error terms. Consider the model equation

$$\begin{aligned} y_t &= \eta_t + \epsilon_t \\ &= \alpha e^{\beta t} + \epsilon_t \\ &= \alpha\omega^t + \epsilon_t , \end{aligned} \qquad (1)$$

where y_t is the observation at time t; α is the true time-zero or initial value; β, the rate constant; $\omega = \exp(-\beta)$; and ϵ_t, the error term.

Consideration will first be given to the possible components of ϵ_t. It may consist of a combination of several sources or kinds of "error"; e.g., model error, measurement error, and a kind of replication-by-time interaction error. For brevity, and for reasons which will be given later, the last kind of error will be referred to as process-control error.

A model error may derive from an improper choice of a mathematical expression, even though the fit to the data using the incorrect equation may be considered adequate. The model error will most likely constitute a bias, an overfitting or underfitting in various regions of the curve. Its major effect will be to increase the mean square deviation about the fitted line. It is assumed in what follows that there is no model error in ϵ_t.

The measurement error, referred to herein as δ_t, may be viewed in the usual way: it is the technic error introduced into the observation when the measurement is being obtained. It may involve only sensing error, the observation being automatically recorded, or it may consist of a combination of errors, a human reading error, the error in sensing, plus an error in one or more analytic procedures. It is usually assumed to have zero mean for each time t.

For certain problems, the most important source of error is that due to nonregular or random behavior of the response from time to time. This deviation or error is considered to be random over experiments. It arises when the organism invokes some complicated control to correct for overproduction or underproduction of the measured quantity. For example, the metabolic rate of the quantity studied may be considered a constant, on the average. The true amount present at time t, however, may be $a\omega^t + \nu_t$, where ν_t is the process control error. For the same animal, in a replicated experiment, ν_t may vary so that the average of ν_t, t fixed, over all these re-runs may reasonably be assumed to be zero. The sign and size of ν_t is determined by a host of factors in the organism, so that it is truly a random element at any point t. Thus, the term ϵ_t will be written as the sum of two elements, ν_t and δ_t; i.e.,

$$\epsilon_t = \nu_t + \delta_t. \tag{2}$$

The next question to be considered is whether the error terms are time dependent. The nature of the biologic process being studied and frequency of sampling are relevant factors. If the process is continuous with measurements being continuously made, then both ν_t and δ_t will be continuous curves. In this case, one can imagine that the error deviations form an undulating curve, weaving around the average response curve. Thus, the deviation at time $t + \Delta t$ will be functionally dependent on the deviation at time t. The commonest departure from a continuous process, continuously observed, is obtained in situations where sampling is at equally spaced time intervals. In what follows, Δt is taken to be greater than zero and constant. A measure of the time dependency between two observations in time is the covariance of the two observations. First, the covariance of the ν_t will be discussed. If the sampling interval, Δt, is small, then as observed above, ν_t and $\nu_{t+\Delta t}$ are functionally dependent and would behave as positively correlated quantities. If Δt is sufficiently large, one might observe a zero correlation between ν_t and $\nu_{t+\Delta t}$; an overyield at time t in no way affects the response at time $t + \Delta t$. If the overyield at time t gave rise to an undercorrection at time $t + \Delta t$, a negative correlation would obtain.

Correlations among the δ_t may be quite different from those among the ν_t. For example, δ_t and $\delta_{t+\Delta t}$ may be positively correlated for all Δt. This could obtain if an observer is recording the response of a system which has been showing a steady but definite decrease in time and if the response tends to level out or asymptote, the observer is likely to remember the response at a previous reading and unconsciously round to effect a nonincreasing response. This kind of behavior essentially creates a moving average of the δ_t. Thus, for a particular system the covariance of ν_t and $\nu_{t+\Delta t}$ may be negative and the covariance of δ_t and $\delta_{t+\Delta t}$, positive. For most systems, it seems reasonable to assume that all ν_t are independent of all δ_t.

As mentioned earlier, the assumptions about the error terms are made with some particular equation in mind. If the comments regarding the ϵ_t are made for equation 1, the next problem is to study the resulting effect on the error terms when one uses another representation of the process. For example, $\eta_t = \alpha \exp(-\beta t)$

satisfies the equation $\eta_t = \omega\eta_{t-1}$, where ω is equal to exp $(-\beta)$ and where $\eta_t = \alpha$ exp $(-\beta t) = E(y_t)$. Then the equation

$$y_t = \omega\eta_{t-1} + \xi_t \tag{3}$$

is another way of writing equation 1. For this situation, it is apparent that the assumptions made about ϵ_t are the same as those for ξ_t since ϵ_t and ξ_t are identical.

Next, consider a variation of equation 3, the first order stochastic difference equation

$$y_t = \omega y_{t-1} + \xi_t. \tag{4}$$

The solution of 4 is also equation 1, but now the error term ϵ_t of equation 1 is a function of $\xi_0, \xi_1, \ldots, \xi_t$. It is instructive to study the process, starting at time zero. The time-zero reading, y_0, consists of a constant, α, say, plus a random error ξ_0, or $y_0 = \alpha + \xi_0$, where α is the true time-zero value. For the nitrogen-washout study mentioned earlier, $\alpha = .79$, the nitrogen concentration of air near sea level. Building up equation 1 from 4,

$$\begin{aligned} y_1 &= \omega y_0 + \xi_1 \\ &= \omega(\alpha + \xi_0) + \xi_1 \\ &= \alpha\omega + \omega\xi_0 + \xi_1 \\ y_2 &= \omega y_1 + \xi_2 \\ &= \alpha\omega^2 + \omega^2\xi_0 + \omega\xi_1 + \xi_2 \\ &\cdots \\ y_t &= \alpha\omega^t + \sum_{i=0}^{t} \omega^{t-i}\xi_i. \end{aligned} \tag{5}$$

Equating 5 with 1, indicates that

$$\epsilon_t = \sum_{i=0}^{t} \omega^{t-i}\xi_i = \omega\epsilon_{t-1} + \xi_t. \tag{6}$$

With an ID$(0,\sigma^2)$ assumption (independently distributed with zero mean, variance σ^2) for the ξ_t, then

$$\text{Cov}\,(\epsilon_t, \epsilon_{t-k}) = \omega^k\sigma^2 \left(\frac{1-\omega^{2(t-k+1)}}{1-\omega^2}\right), \quad k = 0, 1, 2, \ldots, t. \tag{7}$$

Model equation 4 is commonly referred to as a simple autocorrelation model (see Anderson (1)). Note that as t becomes large, the variance approaches $\sigma^2(1-\omega^2)^{-1}$ and for k small relative to t, the k^{th} lag correlation coefficient approaches ω^k.

Instead of writing the autocorrelation model as 4 or 5, it is frequently written as a pair of equations,

$$\begin{aligned} y_t &= \alpha\omega^t + \epsilon_t \\ \epsilon_t &= \rho\epsilon_{t-1} + \xi_t. \end{aligned} \tag{8}$$

If $\rho = \omega$, equation 8 is equivalent to equation 5; if $\rho = 0$, equation 8 is equivalent to equation 3. In this more general formulation, one can see that if ρ is negative, adjacent observations are negatively correlated. This becomes somewhat clearer if one writes $-\rho$ for ρ in 8. Then the single equation analogous to equation 5 is

$$y_t = \alpha\omega^t + \sum_{i=0}^{t} (-\rho)^{t-i}\xi_i. \tag{9}$$

If the ξ_t are ID$(0,\sigma^2)$, then

$$\text{Cov}\,(y_t, y_{t-k}) = (-\rho)^k\sigma^2 \left(\frac{1-\rho^{2(t-k+1)}}{1-\rho^2}\right), \quad k = 0, 1, \ldots, t. \tag{10}$$

As t increases and k is relatively small, the k^{th} lag correlation coefficient very nearly becomes $(-\rho)^k$. Thus, all odd lag correlations are negative and all even are positive. This model may be appropriate when the system corrects itself for overproductions and underproductions as the time course proceeds. Again, it should be noted that the observed lag correlations will be as indicated only if the sampling period coincides with the "correcting" period. The importance of the frequency of this time sampling in relation to the system's assumed behavior cannot be over emphasized.

A model equation that is similar to equations 4 and 8 is the moving average model,

$$\begin{aligned} y_t &= \alpha\omega^t + \epsilon_t \\ \epsilon_t &= \sum_{i=0}^{k} m_i\,\xi_{t-i}. \end{aligned} \tag{11}$$

For example, if

$$m_i = \begin{cases} 1 & \text{for } i = 0 \\ -\rho & \text{for } i = 1 \\ 0 & \text{otherwise,} \end{cases} \tag{12}$$

then

$$\epsilon_t = -\rho\,\xi_{t-1} + \xi_t.$$

Assuming the ξ_t are ID$(0, \sigma^2)$, then using the values of m_i from equation 12,

$$\text{Cov}\,(y_t, y_{t-k}) = \begin{cases} (1+\rho^2)\sigma^2, & \text{for } k = 0 \\ -\rho\sigma^2, & \text{for } k = 1 \\ 0, & \text{for } k = 2, \ldots, t. \end{cases} \tag{13}$$

From equation 13, note that the first lag correlation coefficient is $-\rho/(1+\rho^2)$ and all

higher lag correlation coefficients are zero. This will provide for a simpler kind of corrective action by the system than does equation 9.

Model equation 3 may be described as a mechanical system. It is appropriate when the response at time t is made up of two parts: a constant times the true response at time t − 1 plus a random error term. The error term at time t is independent of all other error terms. Thus, this model may be used when there is only an independent measurement error. Equations 4, 8, and 11, on the other hand, are more properly called feedback or historical models for the errors on previous occasions affect the observation at time t. Most biologic systems are probably of this latter type. A historical model that allows for situations where the process control and measurement errors are correlated in different patterns, and, thus, is perhaps more realistic than equation 8, is

$$y_t = a\omega^t + \epsilon_t = a\omega^t + \nu_t + \delta_t ,$$

where

$$\nu_t = -\rho\,\nu_{t-1} + \gamma_t = \sum_{i=0}^{t} (-\rho)^{t-i}\gamma_i \qquad (14)$$

and

$$\delta_t = \sum_{i=0}^{k} m_i x_{t-i} .$$

If the measurement errors are independent, then $m_0 = 1$ and all other m_i are zero. As stated earlier, if there is no replication, then ν_t and δ_t are inseparable. Even though there is no hope of separation of the two, it may be helpful to imagine the errors as behaving in this fashion. Depending on the relative size of the variances of ν_t and δ_t, one may be able to predict what the net effect will be.

Another whole class of models are those where the error is proportional to the level of the measured response. These models have utility, especially when the range of y_t is several fold. Furthermore, they are more manageable under logarithmic transformation, allowing for simple estimators for the parameters of the equation. The model equation is

$$y_t = a e^{-\beta t} + \epsilon_t . \qquad (15)$$

One can take the logarithm of both sides of the equation and if the ϵ_t are $ID(0,\sigma^2)$, proceed to estimate the parameters in the usual manner since the variance of $\ln(y_t)$ is σ^2 and the covariances are zero. To see how different autocorrelation patterns among the ϵ_t may affect the model, however, equation 15 is written in another form. For ϵ_t small, $e^{\epsilon_t} \approx 1 + \epsilon_t$; and writing, as before, $\eta_t = \alpha e^{-\beta t}$, then y_t can be expressed approximately as

$$y_t = \alpha e^{-\beta t} + \eta_t \epsilon_t . \qquad (16)$$

This equation may be written as

$$y_t = \omega\eta_{t-1} + \eta_t \epsilon_t . \qquad (17)$$

If the ϵ_t in 15 are $ID(0,\sigma^2)$, then the error terms in 16 are $ID(0,\eta_t^2\sigma^2)$. Equation 17 corresponds to the mechanical system as given in equation 3, with error proportional to the true response at time t. The difference equation analogous to equation 17, corresponding to the autocorrelation model equation 4, is

$$y_t = \omega y_{t-1} + y_t \xi_t . \qquad (18)$$

By assuming that the process starts at time zero so that $y_0 = \alpha + y_0\xi_0$, then the solution of equation 18 is

$$y_t = \frac{a\omega^t}{\prod_{i=0}^{t}(1-\xi_i)} . \qquad (19)$$

To a first degree of approximation, $\frac{1}{1-\xi_i} = 1 + \xi_i$, so that

$$\frac{1}{\prod(1-\xi_i)} \cdots \prod(1+\xi_i) = 1 + \Sigma\xi_i + \sum_{i<j}\sum \xi_i\xi_j + \cdots \cdots 1 + \Sigma\xi_i .$$

Thus, equation 19 can be written as

$$y_t = \frac{a\omega^t}{\prod_{i=0}^{t}(1-\xi_i)} \cdots a\omega^t + \eta_t\left(\sum_{i=0}^{t}\xi_i\right) .$$

Another way of arriving at this same form is to assume that the error term in 18 is proportional to the true response at time t rather than to the observed response, i.e.,

$$y_t = \omega y_{t-1} + \eta_t \xi_t . \qquad (20)$$

Then the solution is

$$y_t = a\omega^t + \sum_{i=0}^{t} \omega^{t-i}\eta_i\xi_i$$
$$\cdots a\omega^t + \eta_t\left(\sum_{i=0}^{t}\xi_i\right) . \qquad (21)$$

If the ξ_t in equation 21 are assumed to be $ID(0,\sigma^2)$, then $E(y_t) = a\omega^t$ and

$$\text{Cov}(y_t, y_{t-k}) = (t - k + 1)\, a^2 \omega^{2t-k} \sigma^2. \quad (22)$$

The k^{th} lag correlation coefficient is $(1 - \frac{k}{t+1})^{1/2}$; for k small relative to t it approaches unity as t becomes large. This result states that once the response is far enough along in time and veers to a particular side of the average response curve, then the observational curve would tend to stay on the same side for the remainder of the experiment. There appears to be no way one can distinguish this kind of anomaly from organism (animal-to-animal) variability, especially if the experiment cannot be repeated on the same organism. To allow for continued recrossing of the average response curve, negative correlation between adjacent error deviations may be postulated. Thus, model equations may be written for the proportional error models that allow for the control of the system similar to that allowed in equations 4 and 11.

For model equation 21, the variance of $\ln(y_t)$ is $(t + 1)\sigma^2$ for an $ID(0,\sigma^2)$ assumption on the ξ_t. If the variance of $\ln(y_t)$ is not an increasing function of t, one may either postulate

$$\epsilon_t = \rho\epsilon_{t-1} + \xi_t, \quad (23)$$

where ρ may be negative, or specify a moving average relationship for the ξ_t, such as

$$\epsilon_t = \sum_{i=0}^{k} m_i \xi_{t-i}. \quad (24)$$

For these two cases, the model equations corresponding to 21 may be written as

$$y_t = \eta_t (1 + \epsilon_t) = \eta_t (1 + \sum_{i=0}^{t} \rho^{t-i} \xi_i) \quad (25)$$

and

$$y_t = \eta_t (1 + \epsilon_t) = \eta_t (1 + \sum_{i=0}^{k} m_i \xi_{t-i}), \quad (26)$$

respectively. Note that with restrictions 23 and 24 on the error terms, and with the $ID(0,\sigma^2)$ assumption of the ξ_t, the variances and covariances of $\ln(y_t)$ will have the same pattern as for the nonproportional error models. Thus, for equation 25, variance $\ln(y_t)$ is approximately $\frac{\sigma^2}{1 - \rho^2}$ for t large, and for equation 26, the variance is $\sigma^2 \Sigma m_i^2$.

If the range of y_t is small, the fitted equations for the proportional error models differ little from their nonproportional counterparts; for these situations one might prefer to use the proportional error models for estimation purposes, since the logarithmic transformation linearizes the mathematical expressions. A discussion of estimation problems is given in a later section.

4. INCREASED COMPLEXITY OF THE MODEL

Greater flexibility in the model is obtained by extending equation 1 to the sum of two or more exponentials,

$$y_t = \sum_{i=1}^{m} a_i \omega_i^t + \epsilon_t. \quad (27)$$

For the mechanical model, the ϵ_t are independent; for the autocorrelation models, they are functions of previous random error terms. For both types of models, m^{th} order difference equations may be written similar to those equations involving one exponential term. Furthermore, the ϵ_t may be assumed to be proportional to the level of the response and with different autocorrelation patterns as discussed in the previous section.

When the number of exponential terms is three or more, there arises the question of the determination of the number of true terms in the model equation. For three exponential terms, there are 6 constants to be fitted to the data, usually assuring one a reasonable fit. This will be discussed more fully in the next section, but it does lead one to consider whether it is reasonable to postulate that ω is continuously distributed. For some biologic applications ω can assume only a finite number of values, for others (see reference 2, for example), a continuous distribution is quite attractive. An example of this is the nitrogen washout problem discussed above; one may consider the smallest volumes v_i, as those for the individual alveoli. Recall that the resting and expanded volume of the alveoli are used to determine the ω_i. From the thousands of ω_i, one can visualize the resulting distribution formed by grouping all those that fall in the interval $\omega_i + \Delta\omega$. If one denotes this frequency distribution by

$f(\omega)d\omega$, then the expression for y_t corresponding to equation 27 is

$$y_t = \int \omega^t f(\omega)d\omega + \epsilon_t , \qquad (28)$$

where the ϵ_t in this equation may be assumed to be independent or autocorrelated. Furthermore, either of these two assumptions on the error terms may be used when the ϵ_t are proportional to the response at time t. The model equation is not specified, of course, until the expression for $f(\omega)d\omega$ is named. In this laboratory, the normal probability density has been used for $f(\omega)d\omega$ in studying equation 28.

To either equation 27 or 28, a term for the asymptote may be added. For some applications considered in this laboratory, it has seemed desirable to modify equation 28 by multiplying the integral by a positive constant < 1.

5. ESTIMATION PROBLEMS

If the variance-covariance matrix for the y_t is known aside from a constant multiplier σ^2, then the Gauss-Newton iterative scheme may be used to estimate the parameters in the model equations discussed. For the model equations with proportional errors, which are clearly more manageable after logarithmic transformation, a knowledge of the weights or covariances for the $\ln(y_t)$ is required. Most techniques that are available for estimating the parameters either assume the weights are known or implicitly assume that the y_t, or the $\ln(y_t)$, have equal weights. But the covariances are generally not known a priori. One can estimate the autocorrelation structure, but it is doubtful if this approach is very practical; it may not give good estimates for the parameters. R. L. Anderson (1) indicates some of the problems of estimation when the errors are correlated. Also Zellner and Tiao (3) show that the variances of the estimates may be very large unless knowledge of the correlations are employed in the estimation procedure.

Consider the autocorrelation model where there is $m = 1$ exponential term. One possible approach is to estimate ω, the lag 1 autocorrelation coefficient by one of the standard estimators. For certain applications, ω is the only parameter of interest. For situations where knowledge of α is also desired, one could use this estimate of ω in the equation $y_t = \alpha\omega^t + \epsilon_t$ to estimate α. Thus if ϵ_t is as given in equation 6, and the ξ_t are $ID(0,\sigma^2)$, then from equation 7 one can get the weighting functions for the y_t. In general, this approach would have limited usefulness unless one were quite sure that $m = 1$ and he also had a very good estimate of ω.

For $m \geq 2$ exponential terms, Cornell (4) has proposed a scheme whereby one groups the data into 2m equal sample-size categories, which are nonoverlapping in time, and for each category calculates the sum of the responses. By using the fact that each partial sum is a geometric series with an equal number of terms, he is able, through algebraic manipulation, to estimate the 2m parameters. The scheme appears to work for well-conditioned data. If further refinement is desired in the estimates of the parameters obtained, they may be used as initial or starting values for the more tedious Gauss-Newton iterative scheme.

For the proportional error models, various biologic research workers (see, for example, reference 5) have proposed a "peel-off" method, using a plot of the logarithm of the response versus the metameter as a working graph. This procedure implicitly assumes equation 15 or an extension thereof as the basic model. For $m = 1$, the resulting scatter of points lies approximately on a straight line. For $m \geq 2$, the procedure is started by observing that the points at the right-hand end of the scatter diagram lie approximately on a straight line. To this portion of the data, a straight line is fitted and extrapolated back to the response axis. Deviations are then calculated between the observed and predicted y_t. If the logarithm of the deviations appear to be linear, a line is fitted to these data and the procedure is completed, with $m = 2$. If the logarithm of the second set of deviations has a definite curvature, the second line is fitted only to the straight-line portion of the deviations, and this line is extrapolated back to the response axis.

The process is continued until all data points are fitted.

The rationale for the procedure is clear enough: as t becomes large, the effect of the smallest ω_i, $i = 1, 2, \ldots, m-1$, are washed out, leaving only the effect of ω_m, the largest ω in the far right-hand end of the curve. From this portion of the data ω_m and the associated α_m are estimated. One then calculates $y_t(d_1) = y_t - \hat{\alpha}_m\hat{\omega}_m^t$ and plots $\ln[y_t(d_1)]$ versus t. The next step allows one to estimate ω_{m-1} and α_{m-1}. This scheme is followed until the parameters of all exponential terms are estimated. If the ω_i are few (m no greater than 8, say) and well separated, and further, if each α_i is a reasonable proportion of $\Sigma\alpha_i$, this procedure gives quite satisfactory estimates. This technic has been programmed for a digital computer; some random sampling experiments performed in this laboratory will be reported in a later paper. As for other similarly obtained estimates, the values found in this manner may be used as initial or starting values for the Gauss-Newton iterative procedure. As for all methods, the assumptions about the error terms must be considered. If the proportional errors are independent, this method gives good estimates of the parameters.

Another estimation scheme for the parameters in equation 27 was given a number of years ago by Prony and discussed by Whitaker et al. (6). Prony formally treated the difference equation

$$y_t + \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + \cdots + \gamma_m y_{t-m} = \epsilon_t \quad (29)$$

as a multiple regression problem, considering y_t as the dependent variable and $y_{t-1}, y_{t-2}, \ldots y_{t-m}$ as m independent variables. The γ_s may be estimated by ordinary least squares and are, apart from sign, the elementary symmetric functions of the ω_i, — i.e., $\gamma_1 = -\Sigma\omega_i$, $\gamma_2 = \sum_{i<j}\omega_i\omega_j$, etc. Then by using the estimates of the γ_s in a polynomial of degree m, the ω_i can be obtained as the m roots. They can then be substituted in equation 27 and the α_i obtained by ordinary least squares. Householder (7) has observed that the procedure has two serious drawbacks: "... it provides no means for weighting the observations in accordance with their supposed precision. Second, it provides no criterion for determining the number of exponentials required for the fitting ...." Assuming known weights for the y_t, he gives an iterative technic for getting valid least squares estimates and provides, also, a criterion for deciding on the number of exponential terms needed for an adequate fit. For $m = 1$, Prony's method provides an estimator of $\gamma_1 = -\omega_1$, which is equivalent to one of the unsophisticated estimates of the lag 1 autocorrelation coefficient—namely, $\Sigma y_t y_{t-1} / \Sigma y_{t-1}^2$.

The estimation of the parameters of equation 28 is somewhat more difficult than for those of 27. By using the normal probability density function as the law to describe the distribution of the random variable ω, the problem is to estimate the mean and variance of this distribution from the data. Since the right-hand side of equation 28 gives the t^{th} moment of ω about the origin, one can get preliminary estimates of μ and σ^2 by using the method of moments thus,

$$y_1 \doteq \mu$$
$$y_2 \doteq \sigma^2 + \mu^2,$$

where $\doteq$ stands for "is an estimate of."

These initial estimates can be used in the Gauss-Newton procedure to obtain more refined estimates. A requirement of knowledge of the weights for the y_t exists here as in all estimation problems, of course. The results of empirical sampling studies using a modified version of model equation 28, with errors proportional to the true response, will be reported in a subsequent paper.

For most of the Gauss-Newton estimation approaches, it has been the experience of this laboratory that considerable improvement in convergence as well as speed of convergence is obtained if one adds to the Gauss-Newton scheme the method of the path of steepest ascent. The reader is referred to D. W. Marquardt's discussion (8) of this modification. Yet, even this approach sometimes fails owing to the high correlation of the estimates of the parameters. In some data the correlation is in

excess of $\pm$ 0.98. Thus, convergence is not always possible even for the most sophisticated iterative procedures.

6. THE NUMBER OF EXPONENTIAL TERMS

Determining the number of exponential terms is a serious problem when $m \geq 3$—i.e., distinguishing between $m = 3$ and $m = 4$ is not easy when the covariances of the y_i must be estimated from the data. Carried one step further, deciding whether one has a finite but large number of terms as opposed to an infinite number is indeed difficult. The problem is especially complicated by the fact that the estimates of the parameters are highly correlated. The situation is roughly analogous to two other statistical procedures: determining the degree of a polynomial when there is no replication error and determining the number of common factors in factor analysis studies. In the latter situation, the "tests" are usually administered only once and the factoring is performed on a correlation matrix; thus, there is no external estimate of error available for assessing statistical significance.

As indicated earlier, Householder's scheme (7) for determining the number of exponential terms requires knowledge of the proper weights for the y_i; i.e., it requires the known variance-covariance matrix of the ϵ_i, aside from the constant multiplier σ^2. Thus, this technic is of little use in most practical situations. Watson (9) has discussed the problems of estimating regression coefficients when an incorrect transformation is used on the ϵ_i; that is, when assumed weights for the y_i are in error. He is not too hopeful of the approach of transforming the ϵ_i to remove effect of the autocorrelation in least squares analysis when the covariance of the y_i must be estimated from the data. Thus, until better approaches are found, the data processor must proceed with the curve fitting even though the proper weights are not known. About all one can do is either assume that the y_i are independent and have equal variances, or use some transformation that will at least give equal variances for the y_i, and look at the mean square deviation about the fitted line. Corresponding to polynomial curve fitting where the statistical significance of the coefficient of the last fitted term is assessed after each step, one can add exponential terms until the mean square of the residuals is satisfactorily small. However, there is no test of statistical significance available for determining the number of exponential terms when the weights of the y_i are unknown. The reader is referred to a discussion by Siddiqui (10) of the problems of significance testing of regression coefficients in linear models, when the errors are correlated. Even if the weights of the y_i are known, however, this step-by-step fitting scheme may not distinguish between the m term model 27 and the continuous model 28. By using the normal law for the density of the $f(\omega)d\omega$ in model 28, only two constants are fitted to the data; for model 27, 2m constants are fitted. If m is large, then in general one would expect a better fit for model 27 than for the two-parameter continuous models. Therefore, other information must be brought to bear on model choice; the choice at any point in time will be based on intuition coupled with one's understanding of the biologic process under study.

Results of random sampling studies are planned in this laboratory for models with relatively small m. Attempts will be made to fit $m + n$, $m + n - 1, \ldots, m$, $m - 1, \ldots, m - n$ exponential terms to the artificially generated data for various error patterns. It is hoped that from these studies, some notion can be obtained about the number of terms necessary for an "adequate" fit as well as some feeling about the effect of unequal weights for the y_i on the estimates of the parameters when the weights are not assumed known.

7. CONCLUDING REMARKS

For exponential-decay data, the major problem facing a curve fitter is in proper model choice. This includes consideration of not only the proper mathematical expression to be fitted but also rather intimate knowledge of the nature of the error terms. These problems have been discussed and comments have been

made about a variety of models and error assumptions that one may consider.

It has also been emphasized that no really good procedures are available for curve fitting if the variance-covariance matrix of the error terms is not known or assumed to be known.

For certain assumed model equations, various methods of estimation of the parameters are being studied in this laboratory. Under investigation is the model equation 27, $m = 2$ and 3, with the ϵ_i independent and proportional to the true response at time t. The performance of the peel-off procedure is under study. The estimates from this procedure are used as starting values for a modified Gauss-Newton iterative scheme. Also planned are studies where m is fixed and an attempt is made to fit $m + n$, $m + n - 1, \ldots, m$, $m - 1$, $m - 2, \ldots, m - n$ terms. The effect of sample size on estimating procedures is alsd to be looked into. Finally, a modified version of model 28 is under study and will be fitted to some empirical data once good estimating procedures are developed for the parameters of that model equation. The results of these investigations will be presented in subsequent reports.

The computer for which programs have been and are being written is a Philco 2000, 8K memory.

REFERENCES

1. Anderson, R. L. The problem of autocorrelation in regression analysis. J. Amer. Stat. Ass. 49:113-129 (1954).

2. Creese, R., M. W. Neil, and G. Stephenson. Effect of cell variation on potassium exchange of muscle. Trans. Faraday Soc. 52:1022-1032 (1956).

3. Zellner, A., and G. C. Tiao. Bayesian analysis of the regression model with autocorrelated errors. J. Amer. Stat. Ass. 59:763-778 (1964).

4. Cornell, R. G. A method for fitting linear combinations of exponentials. Biometrics. 18:104-113 (1962).

5. Fowler, W. S., E. R. Cornish, Jr., and S. S. Kety. Lung function studies. VIII. Analysis of alveolar ventilation by pulmonary N_2 clearance curves. J. Clin. Invest. 31:40-50 (1952).

6. Whittaker, Sir E., and G. Robinson. The calculus of observations, 4th ed. New York, N.Y.: D. Van Nostrand Co., Inc., 1944.

7. Householder, A. S. On Prony's method of fitting exponential decay curves and multiple-hit survival curves. ORNL-455, Feb. 1950.

8. Marquardt, D. W. An alogorithm for least-squares estimation of nonlinear parameters. J. of Soc. Indust. and App. Math. 11:431-441 (1963).

9. Watson, G. S. Serial correlation in regression analysis. Institute of Statistics Mimeo. Series No. 49, North Carolina State of the Univ. of North Carolina, Raleigh, N.C., 1951.

10. Siddiqui, M. M. Test for regression coefficients when errors are correlated. Ann. Math. Stat. 31:929-938 (1960).

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1 ORIGINATING ACTIVITY (Corporate author)
USAF School of Aerospace Medicine
Aerospace Medical Division (AFSC)
Brooks Air Force Base, Texas

2a REPORT SECURITY CLASSIFICATION
Unclassified

2b GROUP

3 REPORT TITLE
Some Problems on the Use of Negative Exponential Curves in Biology

4 DESCRIPTIVE NOTES (Type of report and inclusive dates)
Technical Report July 1964 - December 1964

5 AUTHOR(S) (Last name, first name, initial)
Danford, M. Bryan

6 REPORT DATE	7a TOTAL NO OF PAGES	7b NO OF REFS
March 1965	10	10

8a CONTRACT OR GRANT NO.	9a ORIGINATOR'S REPORT NUMBER(S)
b PROJECT NO. Task No. 631901	SAM-TR-65-4
c	9b OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d	

10 AVAILABILITY/LIMITATION NOTICES
Qualified requesters may obtain copies of this report from DDC.

11 SUPPLEMENTARY NOTES	12 SPONSORING MILITARY ACTIVITY
	USAF School of Aerospace Medicine Aerospace Medical Division (AFSC) Brooks Air Force Base, Texas

13 ABSTRACT

Many biologic processes give responses that decrease rapidly over time to some asymptote $C \geq 0$. The mathematical expressions that describe the various phenomena vary in complexity and form. In order to fit a curve to data in any such situation, one must consider both the formula for the trend and the nature of the deviations or error terms. A discussion of these problems is given so as to indicate to the data analyst the possible choices that are his, relative to assumptions about error terms and relative to technics or methods of estimation. No new analytical procedures are given.

It is noted that several methods of estimation are being studied for certain assumed model equations. The results of these random sampling experiments will be reported in subsequent papers.

DD FORM 1473

Unclassified
Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Statistics Exponential Curves Non-Linear Equations Autocorrelation						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year, or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, &, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through _____."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through _____."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through _____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph represented as *(TS)*, *(S)*, *(C)* or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.