University of Tennessee, Knoxville
Trace: Tennessee Research and Creative Exchange

Research Reports   AgResearch

4-1986

A Method, SAS Program, and Example for Fitting the Broken-Line to Growth Data

University of Tennessee Agricultural Experiment Station

Kelly R. Robbins

Follow this and additional works at: http://trace.tennessee.edu/utk_agresreport

Part of the Agriculture Commons

Recommended Citation

University of Tennessee Agricultural Experiment Station and Robbins, Kelly R., "A Method, SAS Program, and Example for Fitting the Broken-Line to Growth Data" (1986). *Research Reports.*
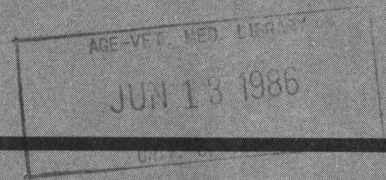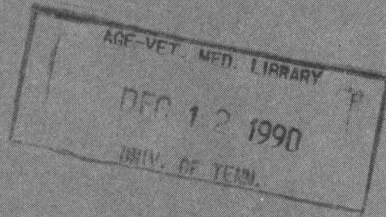http://trace.tennessee.edu/utk_agresreport/70

A *Method,* *SAS Program,* and *Example* for Fitting the Broken-Line to Growth Data

Kelly R. Robbins

Department of
Animal Science

A METHOD, SAS PROGRAM,

AND EXAMPLE FOR FITTING THE BROKEN-LINE

TO GROWTH DATA

Kelly R. Robbins

Associate Professor, Department of Animal Science

The University of Tennessee

Knoxville, Tennessee

A Method, SAS Program, and Example

for Fitting the Broken-Line to Growth Data

The broken-line model, fitted by the method of least squares, yields an objective estimate of the level of a dietary nutrient that is fully adequate, also called the nutrient "requirement." Although for many data a sigmoidal function can more accurately describe the nutrient response, any response that approaches an asymptote can be satisfactorily fitted by a broken line for a narrow range of doses and a large enough error variance (Robbins et al., 1979). Given the simplicity of the broken-line model, its use for such data is sometimes preferable since it is often easier to interpret.

A broken line model consists of two parts: a straight line with an increasing or decreasing slope and a horizontal line. Their point of intersection is the breakpoint. This one-slope broken-line model is often adequate for fitting growth data, especially if the levels tested cover a rather narrow range. For other types of biological data, a broken-line equation describing two intersecting straight lines, both with non-zero slopes (i.e., a two-slope model) may be needed.

The general model of the one-slope broken line is: $Y = L + U(R-X_{LR})$. For the two-slope broken line, the model is: $Y = L + U(R-X_{LR}) + V[X_{GR}-R]$. In these equations, L is the ordinate and R the abscissa of the breakpoint in the curve. U is the slope of the line for $X<R$, and in the two-slope equation, V is the slope of the line at $X>R$. Thus by definition $(R-X_{LR})$ is zero when $X>R$, and $[X_{GR}-R]$ is zero when $X<R$.

Both of these models may be fitted using ordinary least squares via the following matrices:

1. One-slope model

$$
\underline{X'X} \qquad\qquad \underline{X'Y}
$$

$$
\begin{bmatrix} n & \Sigma(R-X_{LR}) \\ \Sigma(R-X_{LR}) & \Sigma(R-X_{LR})^2 \end{bmatrix} \qquad \begin{bmatrix} \Sigma Y \\ \Sigma(R-X_{LR})Y \end{bmatrix}
$$

$$
\text{where } (X'X)^{-1}X'Y = B = \begin{bmatrix} L \\ U \end{bmatrix}
$$

2. Two-slope model

$$
\underline{X'X} \qquad\qquad\qquad \underline{X'Y}
$$

$$
\begin{bmatrix} n & \Sigma(R-X_{LR}) & \Sigma[X_{GR}-R] \\ \Sigma(R-X_{LR}) & \Sigma(R-X_{LR})^2 & 0 \\ \Sigma[X_{GR}-R] & 0 & \Sigma[X_{GR}-R]^2 \end{bmatrix} \qquad \begin{bmatrix} \Sigma Y \\ \Sigma(R-X_{LR})Y \\ \Sigma[X_{GR}-R]Y \end{bmatrix}
$$

$$
\text{where } (X'X)^{-1}X'Y = B = \begin{bmatrix} L \\ U \\ V \end{bmatrix}
$$

The model sum of squares is equal to $B'X'Y$. Using an iterative process we can arrive at the maximum likelihood estimate of R, i.e., the value of R that maximizes the model sum of squares.

The approximate sampling variances of L, U, V, _and_ R may be estimated from the matrix of sums of squares and products of first derivatives. These

matrices are:

1. One-slope model

$$\underline{D'D}$$

$$\begin{bmatrix} n & \Sigma(R-X_{LR}) & n_{LR}U \\ \Sigma(R-X_{LR}) & \Sigma(R-X_{LR})^2 & U\Sigma(R-X_{LR}) \\ n_{LR}U & U\Sigma(R-X_{LR}) & n_{LR}U^2 \end{bmatrix}$$

2. Two-slope model

$$\underline{D'D}$$

$$\begin{bmatrix} n & \Sigma(R-X_{LR}) & \Sigma[X_{GR}-R] & n_{LR}U - n_{GR}V \\ \Sigma(R-X_{LR}) & \Sigma(R-X_{LR})^2 & 0 & U\Sigma(R-X_{LR}) \\ \Sigma[X_{GR}-R] & 0 & \Sigma[X_{GR}-R]^2 & -V\Sigma[X_{GR}-R] \\ n_{LR}U - Vn_{GR} & U\Sigma(R-X_{LR}) & -V\Sigma[X_{GR}-R] & n_{LR}U^2 + n_{GR}V^2 \end{bmatrix}$$

In these matrices n_{LR} is the number of doses with values less than R, and n_{GR} is the number of doses with values greater than R. The variance covariance matrix is obtained by multiplying the inverse of the D'D matrix by an appropriate error term, e.g., the residual mean square.

An alternative and more efficient method of fitting the broken line is to iteratively adjust a set of initial model parameter estimates until the vector of adjustments approaches zero. For the one-slope model, given observed

values of X $(X_1 \ldots X_m \ldots X_n)$ and Y $(Y_1 \ldots Y_m \ldots Y_n)$ and a vector of

initial parameter estimates, $Q_1 = \begin{bmatrix} L_1 \\ U_1 \\ R_1 \end{bmatrix}$,

the jth iteration is:

$$
\overset{D_j}{\begin{bmatrix} 1 & (R_j - X_1) & U_j \\ \vdots & \vdots & \vdots \\ 1 & (R_j - X_m) & U_j \\ \vdots & \vdots & \vdots \\ 1 & (R_j - X_n) & U_j \end{bmatrix}}
\qquad
\overset{Y - \hat{Y}_j}{\begin{bmatrix} Y_1 - \hat{Y}_{X_1} \\ \vdots \\ Y_m - \hat{Y}_{X_m} \\ \vdots \\ Y_n - \hat{Y}_{X_n} \end{bmatrix}}
$$

$$
(D_j{}'D_j)^{-1} D_j{}'(Y - \hat{Y}_j) = B_{j+1} = \begin{bmatrix} L_{j+1} - L_j \\ U_{j+1} - U_j \\ R_{j+1} - R_j \end{bmatrix} \quad \text{(i.e., the vector of adjustments)}
$$

$$
Q_j + B_{j+1} = Q_{j+1} = \begin{bmatrix} L_{j+1} \\ U_{j+1} \\ R_{j+1} \end{bmatrix}
$$

Similarly, for the two-slope model:

$$
\overset{D_j}{
\begin{bmatrix}
1 & (R_j-X_1) & [X_1-R_j] & U_j-V_j \\
\vdots & \vdots & \vdots & \vdots \\
1 & (R_j-X_m) & [X_m-R_j] & U_j-V_j \\
\vdots & \vdots & \vdots & \vdots \\
1 & (R_j-X_n) & [X_n-R_j] & U_j-V_j
\end{bmatrix}
}
\qquad
\overset{Y-\hat{Y}_j}{
\begin{bmatrix}
Y_1-\hat{Y}_{x_1} \\
\vdots \\
Y_m-\hat{Y}_{x_m} \\
\vdots \\
Y_n-\hat{Y}_{x_n}
\end{bmatrix}
}
$$

$$
(D_j{}'D_j)^{-1}D_j{}'(Y-\hat{Y}_j) = B_{j+1} =
\begin{bmatrix}
L_{j+1}-L_j \\
U_{j+1}-U_j \\
V_{j+1}-V_j \\
R_{j+1}-R_j
\end{bmatrix}
\quad \text{(i.e., the vector of adjustments)}
$$

$$
Q_j + B_{j+1} = Q_{j+1} =
\begin{bmatrix}
L_{j+1} \\
U_{j+1} \\
V_{j+1} \\
R_{j+1}
\end{bmatrix}
$$

In the above, $\hat{Y}_j$ is the value of Y estimated at an observed value of X using the parameter estimates calculated in the J-1 iteration.

The method of iteratively adjusting a set of initial parameter estimates is the basic procedure used in the SAS (1982) procedure NLIN. An example program is shown in Table 1.

The pertinent portions of the printed output from the program is shown in Table 2. For this example, the fitted two-slope broken line model was:

$$Y = 117.22 - 395.35 (.52 - X_{LR}) + 11.43 [X_{GR} - .52]$$

However, V was estimated to be 11.43 ± 51.18, which is not significantly different from zero, thus indicating that the one-slope model is sufficient for these data. The fitted one-slope model was:

$$Y = 118.33 - 395.35 (.52 - X_{LR})$$

This one-slope model accounted for 99.8% of the uncorrected total sum of squares (i.e., 6795 x 100/68073), and the requirement was estimated to be $.52 \pm .019$.

The use of SAS PROC NLIN is a simple, quick, and efficient means of fitting the broken line to growth data. Although this requires the user to supply a set of initial model parameter estimates, they generally need not be very accurate.

REFERENCES

Robbins, K. R., H. W. Norton and D. H. Baker (1979) Estimation of nutrient and requirements from growth data. J. Nutr. 109:1710-1714.

SAS (1982) SAS User's Guide: Statistics. SAS Institute Inc., Cary, NC 584 pp.

TABLE 1. SAS program for calculating the least squares parameter estimates for both the one-slope and two-slope broken-line models.[a/]

Program	Notes
DATA ONE; INPUT X Y; CARDS; .25 15 .30 23 .40 71 .45 95 .50 106 .55 119 .60 116 .70 120	These data (Robbins et al., 1979) are average body weight gains (15 to 120 g) of chicks fed diets containing varied levels of sulfur amino acids (.25 to .70%).
PROC NLIN DATA=ONE; PARAMETERS L=100 U=-500 V=10 R=.5;	Initial parameter estimates for the two-slope model.
Z1=(X < R)*(R - X);	This statement evaluates the expression $(R-X_{LR})$.
Z2=(X > R)*(X - R);	This statement evaluates the expression $[X_{GR}-R]$.
MODEL Y=L + U*(Z1) + V*(Z2);	The two-slope model
PROC NLIN DATA=ONE; PARAMETERS L=100 U=-500 R=.5;	Initial parameter estimates for the one-slope model.
Z1=(X < R)*(R - X);	This statement evaluates the expression $(R-X_{LR})$.
MODEL Y=L + U*(Z1);	The one-slope model

[a/] SAS PROC NLIN (SAS, 1982).

TABLE 2.   Pertinent printed output from program listed in Table 1.

NON-LINEAR LEAST SQUARES SUMMARY STATISTICS   DEPENDENT VARIABLE Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE
REGRESSION	4	67950.78737542	16987.69684385
RESIDUAL	4	122.21262458	30.55315615
UNCORRECTED TOTAL	8	68073.00000000	
(CORRECTED TOTAL)	7	12794.87500000	

PARAMETER	ESTIMATE	ASYMPTOTIC STD. ERROR	ASYMPTOTIC 95% CONFIDENCE INTERVAL LOWER	UPPER
L	117.22499135	6.06060476	100.39828260	134.05170010
U	-395.34883721	26.66449368	-469.38033846	-321.31733596
V	11.42857143	51.18376877	-130.67843004	153.53557290
R	0.51968674	0.01922224	0.46631797	0.57305551

NON-LINEAR LEAST SQUARES SUMMARY STATISTICS   DEPENDENT VARIABLE Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE
REGRESSION	3	67949.26356589	22649.75452196
RESIDUAL	5	123.73643411	24.74728682
UNCORRECTED TOTAL	8	68073.00000000	
(CORRECTED TOTAL)	7	12794.87500000	

PARAMETER	ESTIMATE	ASYMPTOTIC STD. ERROR	ASYMPTOTIC 95% CONFIDENCE INTERVAL LOWER	UPPER
L	118.33333333	2.85832972	110.98586936	125.68079731
U	-395.34883721	24.32855443	-457.88647088	-332.81120354
R	0.52249020	0.01276597	0.48967470	0.55530570