NONLINEAR ESTIMATION

by

Dwight A. Rockwell

Research and Development Branch
Research Division
Statistical Reporting Service

April 1975

NONLINEAR ESTIMATION

Introduction

During recent years, there has been considerable interest in the possibility of using only current year data as a basis for developing yield forecasts during the growing season. To this end, efforts have been directed toward developing "within year" forecast models for which the pertinent parameters can be estimated with acceptable precision from current season data only. Current plans include continuing this research.

Within year models could be a valuable supplement to the "between year" models presently being used in yield forecasting. Between year models assume relationships that are estimated to exist between independent and dependent variables during a base period (usually a three year period preceding the current growing season) are applicable to the current year. The performance of these models has been closely related to the degree that this assumption of applicability from a base period to the current year has been violated. The need for supplemental information from a within year model is most critical in those years when growing conditions differ greatly from those of the base period.

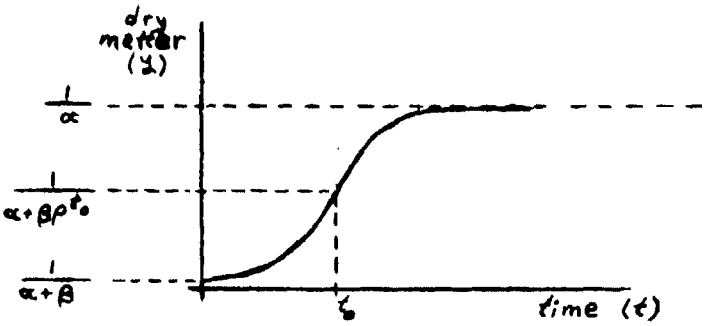
In addition to providing supplemental information to an ongoing yield forecast system, within year models could be very useful when developing a forecast system for a new crop. Usually 3-5 years of information must be gathered before a reliable between year model can be implemented. It is expected that a within year model could be developed in a shorter period of time since a base period is not needed.

Given the need for a within year forecast model, the question becomes what form this model might take and how its parameters can be estimated.

The approach taken thus far with respect to grain crops is that the model should describe the process of kernel dry matter accumulation. With this in mind, an examination of alternative models lead to a special case of what is sometimes called the Logistic Growth Model. Algebraically, this model can be expressed as

$$y = \frac{1}{\alpha + \beta\rho^{t}} \quad , \quad \alpha > 0 \; , \; \beta > 0 \; , \; 0 < \rho < 1 \; ,$$

and it can be represented as shown in the figure below.



This model has been used in several applications, including population growth studies. In population studies ρ is sometimes assumed to be slightly larger than unity.

In applying this model to kernel dry matter accumulation for an individual plant, we are hypothesizing that accumulation begins slowly at first, increases at an increasing rate for a period of time and then increases at a decreasing rate until a maximum (asymptotic) value is reached. This asymptotic value would be the kernel dry matter at harvest.

The point in the phenological development of a plant coinciding with time equal to zero should approximate as closely as possible the inital stages of kernel development (e.g., silk emergence in corn or flowering in wheat).

The value of y at t = o is $\frac{1}{\alpha + \beta}$, since $\rho^{0} = 1$. The asymptotic value of y is $\frac{1}{\alpha}$ and is attained as t becomes large since ρ^{t} and thus $\beta\rho^{t}$

tends to o with $o < \rho < 1$. Assuming the logistic growth model is the appropriate model, the question remaining is whether or not the parameters α, β and ρ can be estimated satisfactorily and early enough in the growing season to provide a useful yield forecasting system.

The objective of this paper is to describe a method for estimating the parameters of a nonlinear model (such as the logistic growth model) based on sample data and to provide examples. Also, two variations of this model which allow for relaxing certain model assumptions concerning residuals are explored.

Estimation

The growth model we are considering is of the form

$$y_i = \frac{1}{\alpha + \beta \rho^{t_i}} + u_i \ , \ i = 1, \ldots, n$$

where y_i is a specific value of the dependent variable, t_i is the corresponding value of a time variable and u_i is the disturbance term for the i^{th} observation. Since this model is intrinsically nonlinear in the unknown parameters α, β and ρ, the method of least squares is not directly applicable for fitting it to sample data.

One method for estimating the parameters of a nonlinear model is the linearization (or Taylor series) method.* In general, we begin by hypothesizing a model of the form

$$y_i = f(\underline{X}_i, \underline{\Theta}) + u_i \ , \ i = 1, \ldots, n$$

where y_i is the value of the dependent variable, $\underline{X}_i' = (X_{i1}, X_{i2}, \ldots, X_{ik})$ is the vector of k independent variables and u_i is the disturbance term for the i^{th} observation and $\underline{\Theta}' = (\Theta_1, \Theta_2, \ldots, \Theta_p)$ is the vector of p unknown parameters to be estimated.

---

*Source: N. R. Draper and H. Smith: Applied Regression Analysis, Wiley, New York, 1966, Chap. 10.

Beginning with an initial estimate of the parameters, $\underline{\theta}_o' = (\theta_{1o}, \theta_{2o}, \ldots, \theta_{po})$, if we carry out a Taylor series expansion of $f(\underline{X}_i, \underline{\theta})$ about the point $\underline{\theta}_o$ and disregard the terms beyond the first derivatives, we can say that, approximately, when $\underline{\theta}$ is close to $\underline{\theta}_o$,

$$y_i \doteq f(\underline{X}_i, \underline{\theta}_o) + \sum_{j=1}^{p} \left[\frac{\partial f(\underline{X}_i, \underline{\theta})}{\partial \theta_j} \right]_{\underline{\theta} = \underline{\theta}_o} (\theta_j - \theta_{jo}) + u_i, \quad i = 1, \ldots, n.$$

All information available from theory and previous survey results concerning the population being sampled would be used in estimating initial values for the parameters. We can re-write the equation above as

$$y_i - f(\underline{X}_i, \underline{\theta}_o) \doteq \left[\frac{\partial f(\underline{X}_i, \underline{\theta})}{\partial \theta_1} \right]_{\underline{\theta} = \underline{\theta}_o} (\theta_1 - \theta_{1o}) + \left[\frac{\partial f(\underline{X}_i, \underline{\theta})}{\partial \theta_2} \right]_{\underline{\theta} = \underline{\theta}_o} (\theta_2 - \theta_{2o}) + \ldots$$

$$\ldots + \left[\frac{\partial f(\underline{X}_i, \underline{\theta})}{\partial \theta_p} \right]_{\underline{\theta} = \underline{\theta}_o} (\theta_p - \theta_{po}) + u_i, \quad i = 1, \ldots, n$$

which can be expressed in matrix notation as

$$(\underline{Y} - \underline{f}_o) \doteq \underline{Z}_o \underline{\gamma}_1 + \underline{U}$$

where

$$(\underline{Y} - \underline{f}_o) = \begin{pmatrix} y_1 - f(\underline{X}_1, \underline{\theta}_o) \\ y_2 - f(\underline{X}_2, \underline{\theta}_o) \\ \vdots \\ y_n - f(\underline{X}_n, \underline{\theta}_o) \end{pmatrix}, \quad \underline{\gamma}_1 = \begin{pmatrix} \theta_1 - \theta_{1o} \\ \theta_2 - \theta_{2o} \\ \vdots \\ \theta_p - \theta_{po} \end{pmatrix}, \quad \underline{U} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix},$$

and

$$\underline{Z}_o = \begin{pmatrix} \left.\frac{\partial f(\underline{X}_1, \underline{\Theta})}{\partial \Theta_1}\right|_{\underline{\Theta}=\underline{\Theta}_o} & \left.\frac{\partial f(\underline{X}_1, \underline{\Theta})}{\partial \Theta_2}\right|_{\underline{\Theta}=\underline{\Theta}_o} & \cdots & \left.\frac{\partial f(\underline{X}_1, \underline{\Theta})}{\partial \Theta_p}\right|_{\underline{\Theta}=\underline{\Theta}_o} \\ \left.\frac{\partial f(\underline{X}_2, \underline{\Theta})}{\partial \Theta_1}\right|_{\underline{\Theta}=\underline{\Theta}_o} & & & \\ \vdots & & \ddots & \vdots \\ \left.\frac{\partial f(\underline{X}_n, \underline{\Theta})}{\partial \Theta_1}\right|_{\underline{\Theta}=\underline{\Theta}_o} & \cdots & & \left.\frac{\partial f(\underline{X}_n, \underline{\Theta})}{\partial \Theta_p}\right|_{\underline{\Theta}=\underline{\Theta}_o} \end{pmatrix}$$

The parameter vector, $\underline{\gamma}_1$, can then be estimated by applying ordinary least squares to obtain

$$\hat{\underline{\gamma}}_1 = (\underline{Z}_o' \underline{Z}_o)^{-1} \quad \underline{Z}_o' (\underline{Y} - \underline{f}_o)$$

where

$$\hat{\underline{\gamma}}_1 = \begin{pmatrix} \Theta_{11} - \Theta_{1o} \\ \Theta_{21} - \Theta_{2o} \\ \vdots \\ \Theta_{p1} - \Theta_{po} \end{pmatrix}$$

The vector $\hat{\underline{\gamma}}_1$ will minimize the error sum of squares

$$SS(\hat{\underline{\gamma}}_1) = \sum_{i=1}^{n} \left\{ \left[y_i - f(\underline{X}_i, \underline{\Theta}_o) \right] - \sum_{j=1}^{p} \left[\left.\frac{\partial f(\underline{X}_i, \underline{\Theta})}{\partial \Theta_j}\right|_{\underline{\Theta}=\underline{\Theta}_o} \right] (\Theta_j - \Theta_{jo}) \right\}^2$$

with respect to the $(\Theta_j - \Theta_{jo})$, $j = 1, \ldots, p$.

Using $\underline{\theta}'_1 = (\theta_{11}, \theta_{21}, \ldots, \theta_{p1})$ as a revised estimate of the unknown parameter vector $\underline{\theta}' = (\theta_1, \theta_2, \ldots, \theta_p)$, we can place the θ_{j1}, $j = 1, \ldots, p$ in the same role as the θ_{jo}, $j = 1, \ldots, p$ in the equations above, repeat the process of deriving the least squares solution and obtain another revised estimate $\underline{\theta}'_2 = (\theta_{12}, \theta_{22}, \ldots, \theta_{p2})$.

This iterative process is continued until the solution converges. The criterion for convergence might be

$$\left| \frac{\theta_{j(k+1)} - \theta_{jk}}{\theta_{jk}} \right| < \delta_1, \quad j = 1, \ldots, p,$$

or alternatively

$$\left| \frac{SS(\hat{\gamma}_{(k+1)}) - SS(\hat{\gamma}_k)}{SS(\hat{\gamma}_k)} \right| < \delta_2$$

in successive iterations k and (k+1), where δ_1 or δ_2 would be predetermined tolerance values.

Note that with the terminating $k^{\underline{th}}$ iteration, the $SS(\hat{\gamma}_k)$ will be the minimum attainable to the accuracy imposed by the termination criterion chosen. One should be aware of the effects of this limitation. For example, even though the error term u of the nonlinear model is assumed to be normally distributed, $\hat{\underline{\theta}}$ is not normally distributed, $\hat{\sigma}^2 = SS(\hat{\gamma}_k)/(n-p)$ is not an unbiased estimate of σ^2 and confidence intervals constructed for population parameters are only approximate. Of course, the more closely the sample data fits the hypothesized model and the smaller the termination criterion, the better the approximation will be.

Example (Grape Data)

The example that follows will illustrate the iterative process. Data used were collected as part of a study of grape growth characteristics in

Michigan during 1962. Although data collection was not designed for fitting the results to a growth model, the calculations will hopefully serve to clarify the linearization procedure.

Weight per Grape (grams)	Time Since Berry Formation Began (days)
1.13	6
1.04	6
1.10	6
1.06	6
1.15	6
1.03	6
1.56	11
1.61	11
1.42	11
1.96	21
1.31	21
1.59	22
2.09	22
1.72	22
1.85	22
1.75	22
2.27	22

The model hypothesized is

$$y_i = \frac{1}{\theta_1 + \theta_2\ \theta_3^{\,t_i}} + u_i\ ,\ i = 1, \ldots, 17$$

where Y_i is the weight per grape, t_i is the time since berry formation began and u_i is the disturbance term for the i^{th} observation and θ_1, θ_2, and θ_3 are the unknown parameters to be estimated.

Initial estimates of the parameters to be used are

$$\underline{\theta}_o = \begin{pmatrix} \theta_{1o} \\ \theta_{2o} \\ \theta_{3o} \end{pmatrix} = \begin{pmatrix} 0.27 \\ 0.73 \\ 0.90 \end{pmatrix} .$$

An initial estimated error mean square will be computed to compare with that which will result after the first iteration.

$$\hat{\sigma}^2_o = \frac{1}{n-p} \sum_{i=1}^{n} \left[y_i - f(t_i, \underline{\theta}_o) \right]^2$$

$$= \frac{1}{17-3}\left\{\left[1.13 - \frac{1}{.27 + .73\,(.90)^{6}}\right]^{2} + \left[1.04 - \frac{1}{.27 + .73\,(.90)^{6}}\right]^{2}\right.$$

$$\left. + \ldots\ldots + \left[2.27 - \frac{1}{.27 + .73\,(.90)^{22}}\right]^{2}\right\}$$

$$= 0.85276$$

Computations for the first iteration are as follows:

$$\underline{Y} - \underline{f}_o = \begin{pmatrix} y_1 - f(t_1, \underline{\theta}_o) \\ y_2 - f(t_2, \underline{\theta}_o) \\ \vdots \\ y_n - f(t_n, \underline{\theta}_o) \end{pmatrix} = \begin{pmatrix} 1.13 - \dfrac{1}{.27 + .73\,(.90)^{6}} \\ 1.04 - \dfrac{1}{.27 + .73\,(.90)^{6}} \\ \vdots \\ 2.27 - \dfrac{1}{.27 + .73\,(.90)^{22}} \end{pmatrix} = \begin{pmatrix} -0.38987 \\ -.47987 \\ \vdots \\ -0.65493 \end{pmatrix}$$

$$\underline{Z}_o = \begin{pmatrix} \left.\dfrac{\partial f(t_1, \underline{\theta})}{\partial \theta_1}\right|_{\underline{\theta} = \underline{\theta}_o} & \left.\dfrac{\partial f(t_1, \underline{\theta})}{\partial \theta_2}\right|_{\underline{\theta} = \underline{\theta}_o} & \left.\dfrac{\partial f(t_1, \underline{\theta})}{\partial \theta_3}\right|_{\underline{\theta} = \underline{\theta}_o} \\ \vdots & \vdots & \vdots \\ \left.\dfrac{\partial f(t_{17}, \underline{\theta})}{\partial \theta_1}\right|_{\underline{\theta} = \underline{\theta}_o} & \left.\dfrac{\partial f(t_{17}, \underline{\theta})}{\partial \theta_2}\right|_{\underline{\theta} = \underline{\theta}_o} & \left.\dfrac{\partial f(t_{17}, \underline{\theta})}{\partial \theta_3}\right|_{\underline{\theta} = \underline{\theta}_o} \end{pmatrix}$$

$$= \begin{pmatrix} \dfrac{-1}{[.27 + .73\ (.90)^{6}]^{2}} & \dfrac{-(.90)^{6}}{[.27 + .73\ (.90)^{6}]^{2}} & \dfrac{-6\ (.73)(.90)^{5}}{[.27 + .73\ (.90)^{6}]^{2}} \\ \vdots & \vdots & \vdots \\ \dfrac{-1}{[.27 + .73\ (.90)^{22}]^{2}} & \dfrac{-(.90)^{22}}{[.27 + .73\ (.90)^{22}]^{2}} & \dfrac{-22\ (.73)(.90)^{21}}{[.27 + .73\ (.90)^{22}]^{2}} \end{pmatrix}$$

$$= \begin{pmatrix} -2.31000 & -1.22763 & -5.97445 \\ \vdots & \vdots & \vdots \\ -8.55527 & -0.84250 & -15.03391 \end{pmatrix}$$

$$\hat{\underline{\gamma}}_1 = (\underline{Z}_o' \underline{Z}_o)^{-1} \quad \underline{Z}_o' (\underline{Y} - \underline{f}_o)$$

$$= \begin{pmatrix} 652.99429 & 90.03984 & 1238.66071 \\ 90.03984 & 19.66104 & 189.70704 \\ 1238.66071 & 189.70704 & 2412.96196 \end{pmatrix}^{-1} \begin{pmatrix} 85.44007 \\ 12.47068 \\ 163.21489 \end{pmatrix}$$

$$= \begin{pmatrix} 0.19016 \\ 0.21683 \\ -0.04751 \end{pmatrix} = \begin{pmatrix} \theta_{11} - \theta_{10} \\ \theta_{21} - \theta_{20} \\ \theta_{31} - \theta_{30} \end{pmatrix}$$

$$\underline{\theta}_1 = \begin{pmatrix} \theta_{11} \\ \theta_{21} \\ \theta_{31} \end{pmatrix} = \begin{pmatrix} \theta_{11} - \theta_{10} \\ \theta_{21} - \theta_{20} \\ \theta_{31} - \theta_{30} \end{pmatrix} + \begin{pmatrix} \theta_{10} \\ \theta_{20} \\ \theta_{30} \end{pmatrix} = \begin{pmatrix} 0.19016 \\ 0.21683 \\ -0.04751 \end{pmatrix} + \begin{pmatrix} 0.27 \\ 0.73 \\ 0.90 \end{pmatrix} = \begin{pmatrix} 0.46016 \\ 0.94683 \\ 0.85249 \end{pmatrix}$$

$$\hat{\sigma}^2_1 = \frac{1}{n-p} \sum_{i=1}^{n} \left[y_i - f(t_i, \underline{\theta}_1) \right]^2$$

$$= \frac{1}{17-3} \left\{ \left[1.13 - \frac{1}{.46016 + .94683\ (.85249)^{6}} \right]^2 + \cdots\cdots \right.$$

$$\left. \cdots\cdots + \left[2.27 - \frac{1}{.46016 + .94683\ (.85249)^{22}} \right]^2 \right\}$$

$$= 0.08297$$

If we set $\delta = 0.0005$, the computation $\left| \frac{0.08297 - 0.85276}{0.85276} \right| = 0.90270 > \delta$ indicates the process should be carried at least one more iteration. The results of five iterations are summarized below.

Iteration	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\sigma}^2$ value	$\hat{\sigma}^2$ relative change
0	0.27000	0.73000	0.90000	0.85276	--------
1	0.46016	0.94683	0.85249	0.08297	0.90270
2	0.53872	1.27950	0.80516	0.04713	0.43196
3	0.53990	1.52010	0.79161	0.04646	0.01422
4	0.53996	1.55280	0.79098	0.04643	0.00065
5	0.53996	1.55283	0.79099	0.04643	0.00000

The fifth iteration results in a relative change in the estimated error mean square far less than δ and thus the process is terminated.

For the fifth iteration,

$$\hat{\underline{Y}}_5 = (\underline{Z}_4' \underline{Z}_4)^{-1} \quad \underline{Z}_4' \quad (\underline{Y} - \underline{f}_4)$$

$$= \begin{pmatrix} 0.02457 & 0.55101 & -0.05254 \\ 0.55101 & 31.57812 & -2.48136 \\ -0.05254 & -2.48136 & 0.20912 \end{pmatrix} \quad \begin{pmatrix} -0.00056 \\ -0.00003 \\ -0.00052 \end{pmatrix}$$

$$\hat{\underline{Y}}_5 = \begin{pmatrix} 0.00000 \\ 0.00003 \\ 0.00001 \end{pmatrix} = \begin{pmatrix} \theta_{15} - \theta_{14} \\ \theta_{25} - \theta_{24} \\ \theta_{35} - \theta_{34} \end{pmatrix}, \text{ and}$$

$$\underline{\theta}_5 = \begin{pmatrix} \theta_{15} \\ \theta_{25} \\ \theta_{35} \end{pmatrix} = \begin{pmatrix} 0.00000 \\ 0.00003 \\ 0.00001 \end{pmatrix} + \begin{pmatrix} 0.53996 \\ 1.55280 \\ 0.79098 \end{pmatrix} = \begin{pmatrix} 0.53996 \\ 1.55283 \\ 0.79099 \end{pmatrix}$$

An estimate of the variance-covariance matrix for the estimated parameters is given by

$$\widehat{\text{cov}}\,(\hat{\underline{\theta}}) = \hat{\sigma}^2 \quad (\underline{Z}_4' \underline{Z}_4)^{-1} = 0.04643 \quad (\underline{Z}_4' \underline{Z}_4)^{-1}$$

$$= \begin{pmatrix} 0.00114 & 0.02558 & -0.00244 \\ 0.02558 & 1.46617 & -0.11521 \\ -0.00244 & -0.11521 & 0.00971 \end{pmatrix} .$$

The accuracy of the parameter and variance estimates with respect to an exact fit to the sample data is limited by lack of precision imposed by the termination criterion used.

The plot on the following page shows the data being fitted (dots) and the function based on the estimated parameters for each iteration. The parameter estimates are so nearly the same for the second through the fifth iteration that only a portion of the second iteration appears and the third and fourth iteration results are completely obscured.

Example (Corn Data)

An analysis of time related growth models in forecasting components of corn yield is presently being conducted by the Yield Forecasting and Estimation Section. Data for this study were collected from three purposely chosen fields in Iowa during the 1973 growing season and from a systematic sample of 10 fields in the Central Crop Reporting district of Iowa during 1974. Current plans are to continue this research in 1975.

As one possible growth model for the 1974 project, it was assumed that the independent time variable was days since silk emergence as of the time a corn plant was sampled. The dependent variable was assumed to be the mean dry grain weight (grams) of all ears per plant for all plants with the same associated value of time and drawn from the same sample field. Days since silk emergence for a plant was taken to be that of the primary ear. It was assumed the residuals in this model are independently distributed with mean zero and a constant variance, σ_u^2. That is,

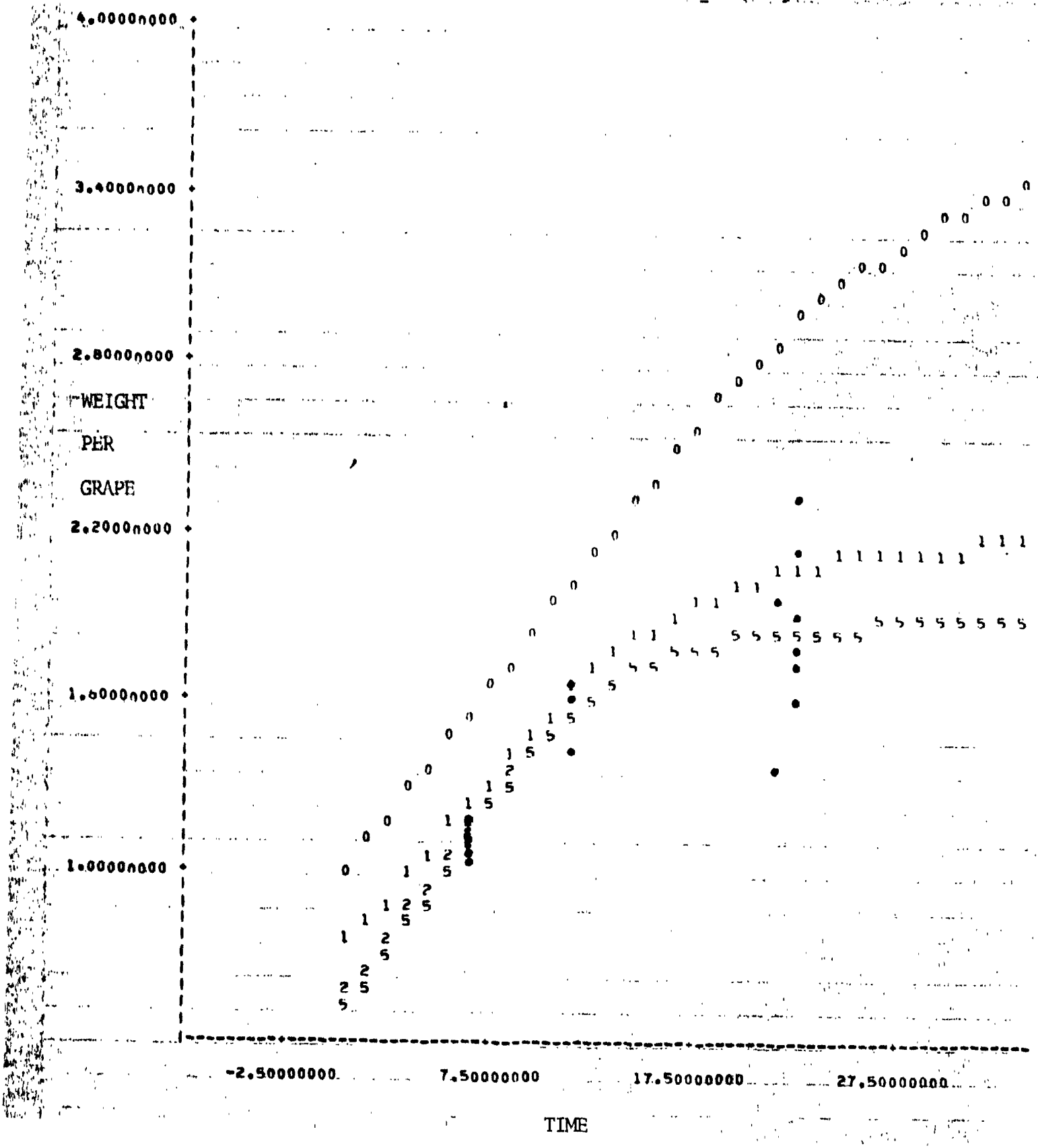
$$E(\underline{U}) = 0 \qquad \text{and}$$

$$E(\underline{U}\,\underline{U}') = \sigma_u^2 I_n$$

The model

$$y_i = \frac{1}{\alpha + \beta \rho^{t_i}} + u_i , \quad i = 1, \ldots, n \tag{1}$$

WEIGHT PER GRAPE
VS
TIME



with y_i and t_i defined as described in the preceding paragraph was fitted to the sample data as it was available up through August 15, September 1, September 15, October 1, October 15, and the end of the growing season. This incrementing of the data was done to provide an indication of how early in the growing season the model could be estimated and how the estimate changed as additional information became available. To evaluate the estimated model, the asymptotic value for each of the six calendar date cut-offs was compared with an estimate of the mean dry grain weight per plant at harvest for the 10 sample fields combined.

The nonlinear least squares option of the Biomedical Computer Programs (BMD) package was used to estimate the model parameters for each of the six cut-offs. This computer program uses a variation of the linearization method. Table 1 shows the estimated value for each parameter, the estimated relative standard error for each estimated parameter, the estimated asymptotic value and the estimated asymptotic value as a percent of the estimated grain weight per plant at harvest for each of the six cut-offs.

Table 1

cut-off	n	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\rho}$	$\hat{\sigma}_{\hat{\alpha}}/\hat{\alpha}$	$\hat{\sigma}_{\hat{\beta}}/\hat{\beta}$	$\hat{\sigma}_{\hat{\rho}}/\hat{\rho}$	$\lim_{t_i \to \infty} y_{t_i}$ est'ed value	% of est'ed hv wt.
all obs.	278	.0061541	.15655	.91866	3.59	34.36	0.98	162.49	106.3
10/15	256	.0058769	.15263	.92037	3.85	31.85	0.90	170.16	111.4
10/1	197	.0053225	.16184	.91977	5.66	29.67	0.90	187.88	123.0
9/15	128	.0063958	.40740	.88626	6.39	38.86	1.34	156.35	102.3
9/1	70	.0063116	.69776	.86809	15.59	46.90	1.91	158.44	103.7
8/15	19	.016119	14.127	.74074	18.90	134.18	6.96	62.04	40.6

These results show considerable variability in the asymptotic value among the six cut-offs. Also, these data indicate little success may be expected, in estimating a reliable model based only on data collected up through mid-August. The plots on the pages that follow show the points being fitted and the estimated model function for each calendar date cut-off.

A Heteroscedastic-error Model

Having completed the fitting of the hypothesized growth model (1) to the six subsets of corn data, an attempt was made to evaluate how well the underlying assumptions concerning the residuals had been met. An examination of the plots showing the fitted model and data points for each cut-off provides an initial indication that there may be a statistically significant relationship between the variation in the residuals and the independent variable, time. The plots show the estimated residuals becoming larger for large values of time. In other words, we may be violating the assumption

$$E(u_i^2) = \sigma_u^2$$

for all i. This condition is commonly referred to as heteroscedasity.

To pursue this possibility, a method suggested by Glejser* was used. Although this procedure was suggested in connection with linear models, its application to a nonlinear model does not seem inappropriate. Accordingly, we begin by assuming that each residual, u_i, can be expressed as

$$u_i = v_i f(t_i), \quad i = 1, \ldots, n$$

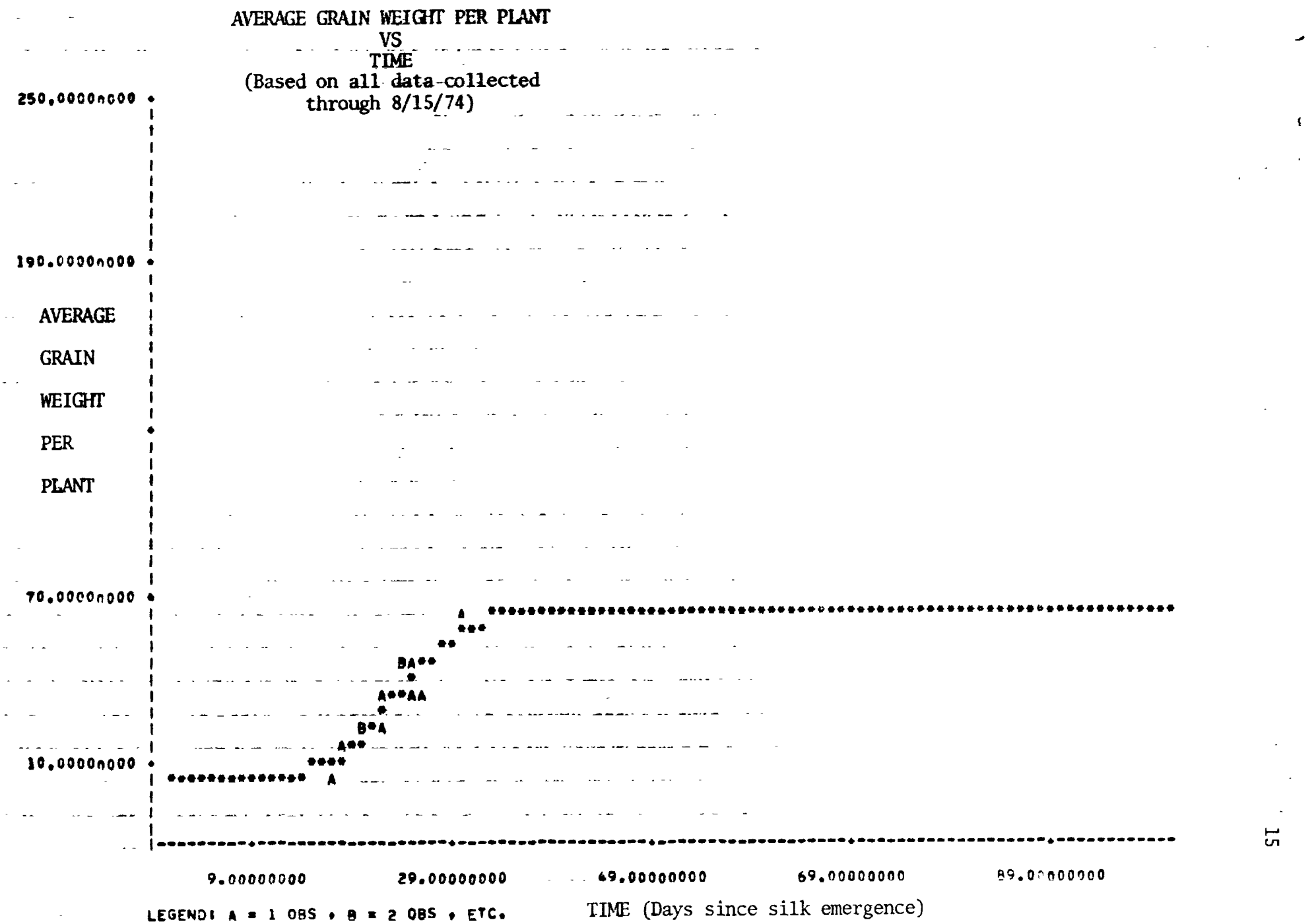
where v_i is a random variable with

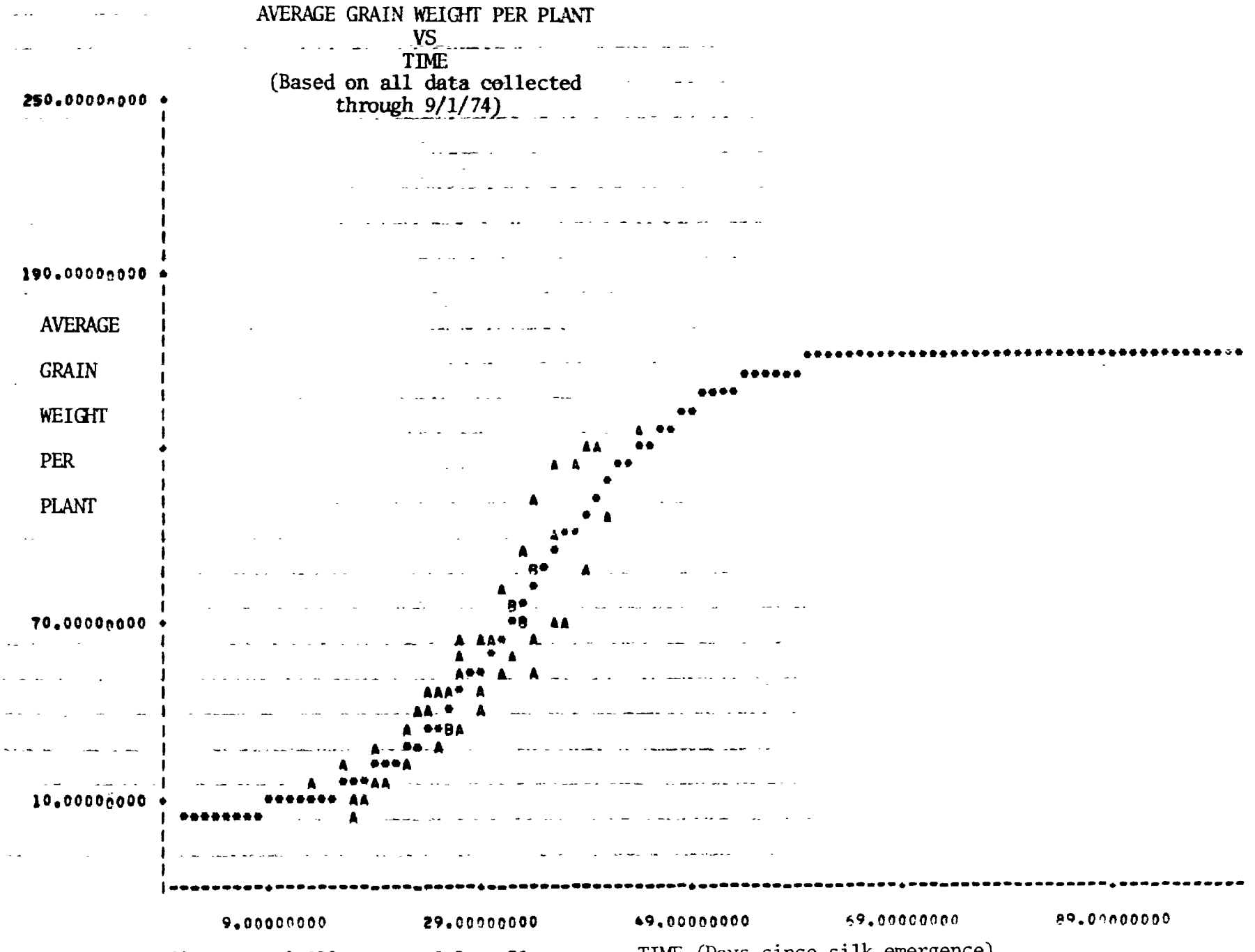
$$E(\underline{V}) = 0 \text{ and}$$

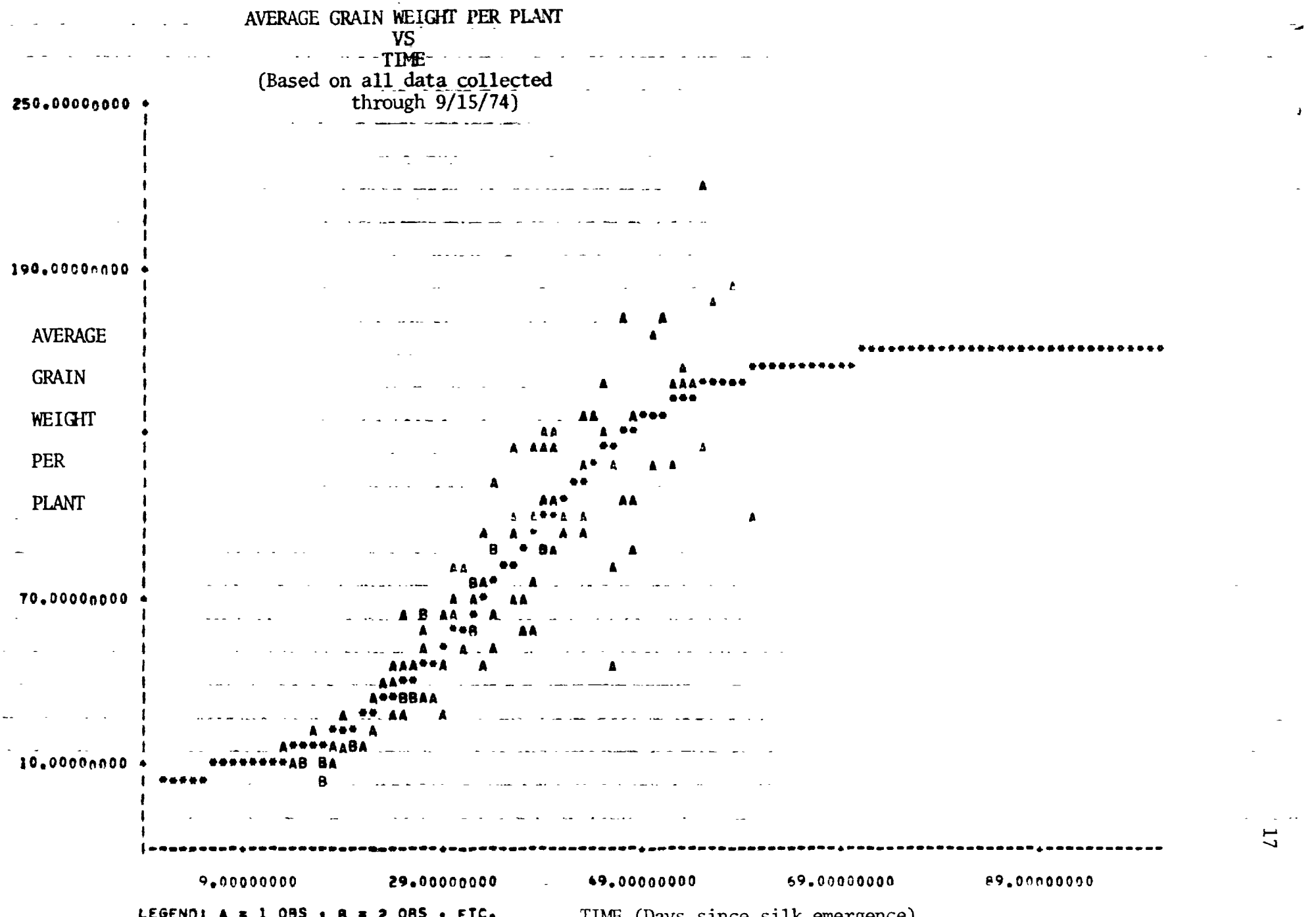
$$E(\underline{V}\,\underline{V}') = \sigma_v^2 I_n.$$

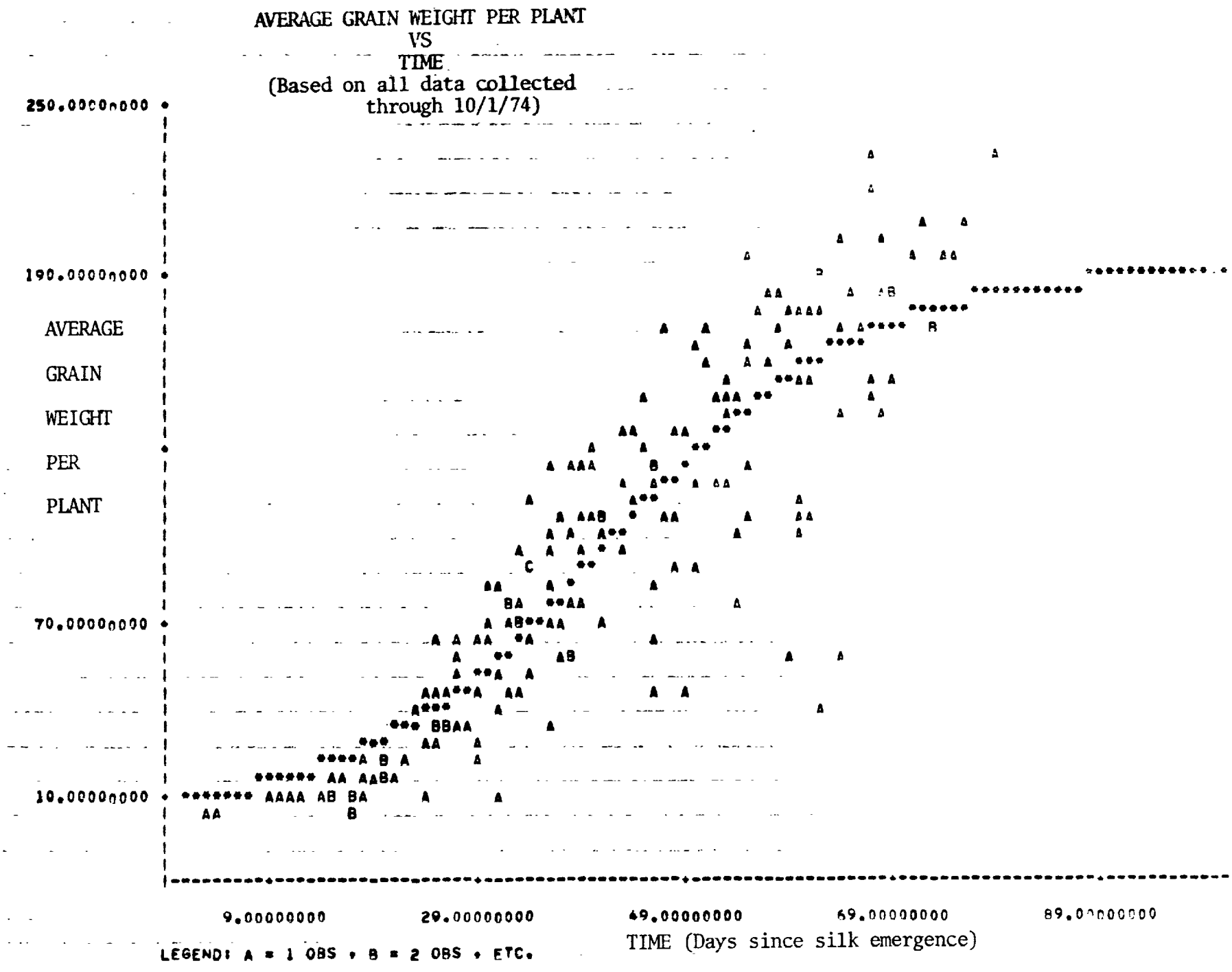
---

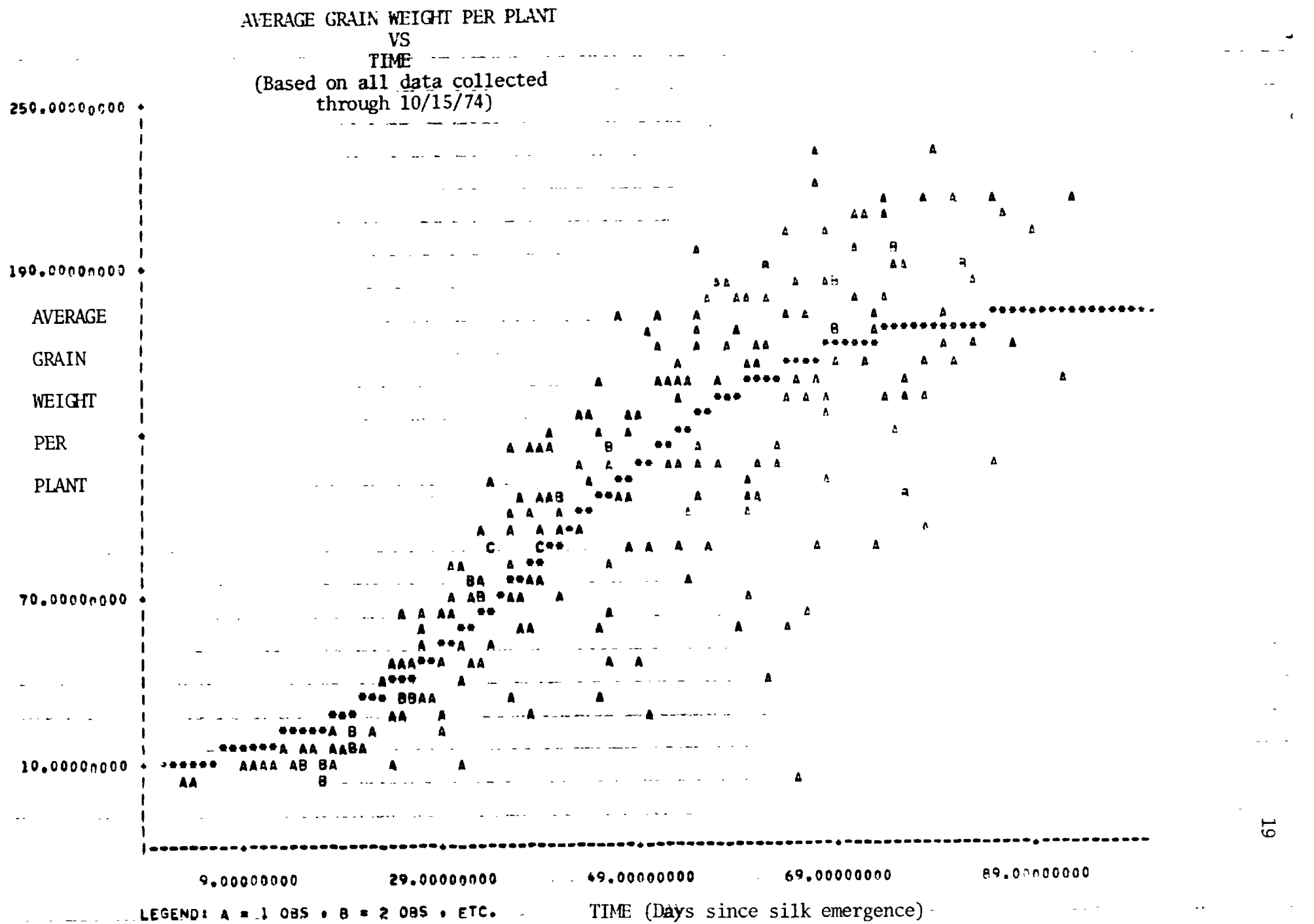
* H. Glejser, "A New Test for Heteroscedasity," Journal of the American Statistical Association, vol. 64, pp. 316-323, 1969.

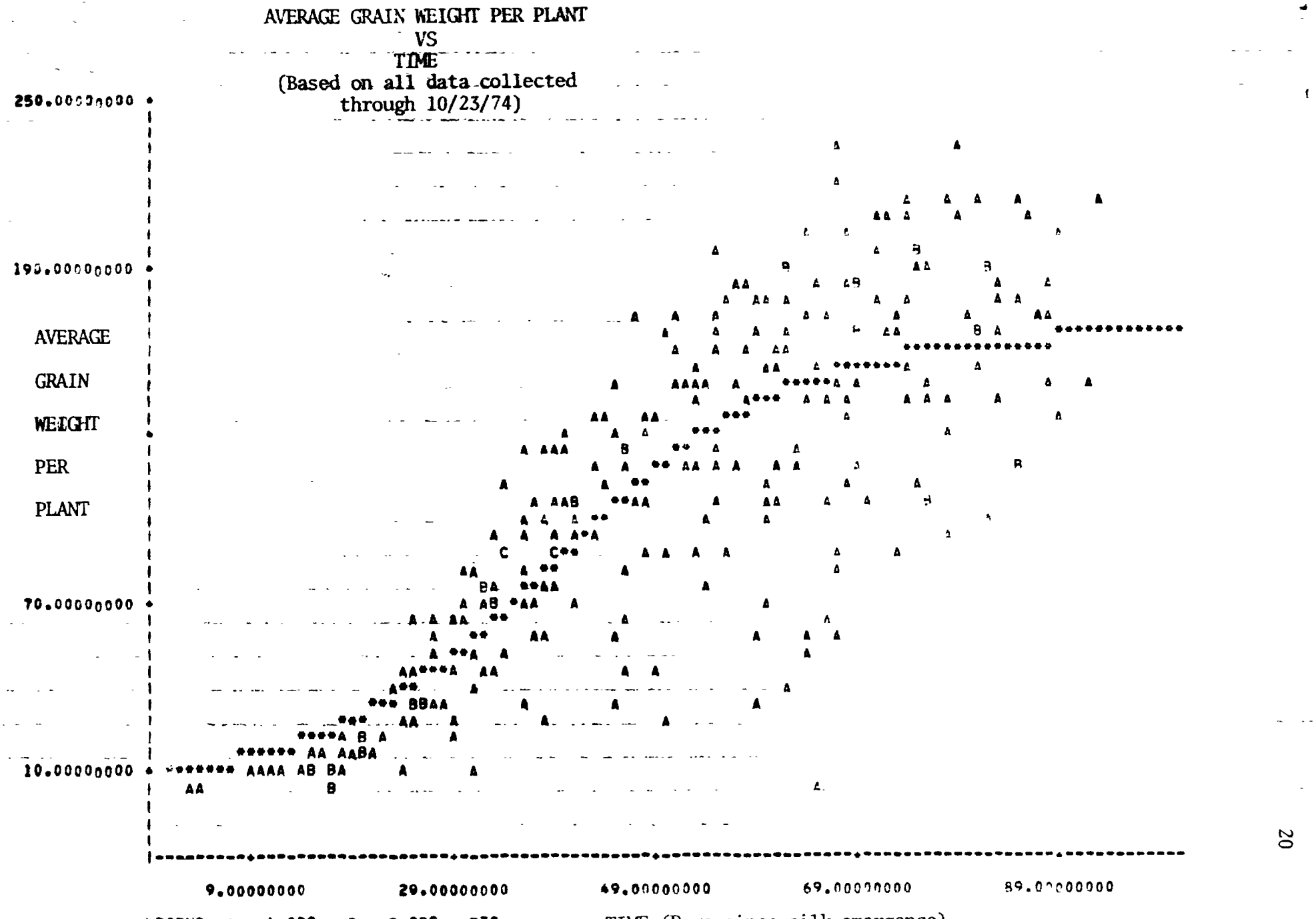

AVERAGE GRAIN WEIGHT PER PLANT
VS
TIME
(Based on all data collected
through 8/15/74)
250.0000n000
190.0000n000
70.0000n000
10.0000n000
AVERAGE
GRAIN
WEIGHT
PER
PLANT
9.00000000
29.00000000
49.00000000
69.00000000
89.00000000
LEGEND: A = 1 OBS , B = 2 OBS , ETC.
TIME (Days since silk emergence)


AVERAGE GRAIN WEIGHT PER PLANT
VS
TIME
(Based on all data collected
through 9/1/74)
250.00000000
190.00000000
70.00000000
10.00000000
AVERAGE
GRAIN
WEIGHT
PER
PLANT
9.00000000
29.00000000
49.00000000
69.00000000
89.00000000
LEGEND: A = 1 OBS , B = 2 OBS , ETC.
TIME (Days since silk emergence)


AVERAGE GRAIN WEIGHT PER PLANT
VS
TIME
(Based on all data collected
through 9/15/74)
250.00000000
190.00000000
70.00000000
10.00000000
AVERAGE
GRAIN
WEIGHT
PER
PLANT
9.00000000
29.00000000
49.00000000
69.00000000
89.00000000
LEGEND: A = 1 OBS , B = 2 OBS , ETC.
TIME (Days since silk emergence)


AVERAGE GRAIN WEIGHT PER PLANT
VS
TIME
(Based on all data collected
through 10/1/74)
250.00000000
190.00000000
70.00000000
10.00000000
AVERAGE
GRAIN
WEIGHT
PER
PLANT
9.00000000
29.00000000
49.00000000
69.00000000
89.00000000
TIME (Days since silk emergence)
LEGEND: A = 1 OBS , B = 2 OBS , ETC.


AVERAGE GRAIN WEIGHT PER PLANT
VS
TIME
(Based on all data collected
through 10/15/74)
250.00000000
190.00000000
70.00000000
10.00000000
AVERAGE
GRAIN
WEIGHT
PER
PLANT
9.00000000
29.00000000
49.00000000
69.00000000
89.00000000
LEGEND: A = 1 OBS, B = 2 OBS, ETC.
TIME (Days since silk emergence)


AVERAGE GRAIN WEIGHT PER PLANT
VS
TIME
(Based on all data collected
through 10/23/74)
250.00000000
190.00000000
70.00000000
10.00000000
AVERAGE
GRAIN
WEIGHT
PER
PLANT
9.00000000
29.00000000
49.00000000
69.00000000
89.00000000
LEGEND: A = 1 OBS, B = 2 OBS, ETC.
TIME (Days since silk emergence)

Further, it is assumed that the form of the function f is known, but at least one of its parameters is unknown. It then follows that

$$E(u_i) = E[v_i f(t_i)]$$
$$= f(t_i) E(v_i)$$
$$= 0 , \qquad i = 1, \ldots, n ,$$
$$E(u_i^2) = [f(t_i)]^2 \sigma_v^2 , \quad i = 1, \ldots, n , \text{ and}$$
$$E(u_i u_j) = E[v_i f(t_i) v_j f(t_j)]$$
$$= f(t_i) f(t_j) E(v_i v_j)$$
$$= 0 , \qquad i \neq j.$$

Instead of the original assumption

$$E(\underline{U}\,\underline{U}') = \sigma_u^2 I_n ,$$

it is now assumed that

$$E(\underline{U}\,\underline{U}') = \sigma_v^2 \Omega$$

where

$$\Omega = \begin{pmatrix} [f(t_1)]^2 & 0 & \cdots & 0 \\ 0 & [f(t_2)]^2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & [f(t_n)]^2 \end{pmatrix}$$

It can be shown that if the assumption of heteroscedasity holds true, using model (1) will give less efficient estimates of the parameters. That is, the estimated relative errors of the parameters will be unnecessarily large. Applying the method of generalized least squares at each iteration of the linearization procedure, an estimate of γ_{k+1} would be given by

$$\hat{\gamma}_{k+1} = (\underline{Z}_k' \underline{\Omega}^{-1} \underline{Z}_k)^{-1} \underline{Z}_k' \underline{\Omega}^{-1} (\underline{Y} - \underline{f}_k) .$$

Alternatively, the same estimate of γ_{k+1} would be obtained if the model

$$\frac{1}{f(t_i)} y_i = \frac{1}{f(t_i)} \frac{1}{\alpha + \beta \rho^{t_i}} + \frac{1}{f(t_i)} u_i, \quad i = 1, \ldots, n \qquad (2)$$

were fitted to the sample data using ordinary least squares at each iteration. Either procedure can be used with BMD. Note that the residuals of model (2),

$$\frac{u_i}{f(t_i)} = v_i, \quad i = 1, \ldots, n$$

do have the desired characteristic of being independently distributed with mean zero and a constant variance, σ_v^2.

Since neither the function f nor its parameters are known, they must be estimated. Following the procedure outlined by Glejser, the absolute value of the estimated residuals obtained from fitting model (1), $|\hat{u}_i|$, $i = 1, \ldots, n$, were regressed on an estimated function of time. An examination of a plot of the absolute value of the residuals against time suggested the function

$$|\hat{u}_i| = f(t_i) = \tau_o + \tau_1 t_i + e_i, \quad i = 1, \ldots, n.$$

Since the estimated value of τ_o was not significantly different from zero for any of the six cut-off dates, the function

$$|\hat{u}_i| = f(t_i) = \tau t_i + e_i, \quad i = 1, \ldots, n$$

was used. Estimates of τ were significant for all cut-offs. The estimates for τ for each cut-off date and the results of fitting model (2) using estimates for f (t) are shown in the Table 2.

Table 2

cut-off	n	$\hat{\tau}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\rho}$	$\hat{\sigma}_{\hat{\alpha}/\hat{\alpha}}$ (%)	$\hat{\sigma}_{\hat{\beta}/\hat{\beta}}$ (%)	$\hat{\sigma}_{\hat{\rho}/\hat{\rho}}$ (%)	$\lim_{t_i \to \infty} y_{t_i}$ est'ed value	% of est'ed hr. wt.
all obs.	278	.491552	.0068499	.50744	.88328	2.99	23.22	0.79	145.99	95.5
10/15	256	.481948	.0066370	.47529	.88618	3.22	22.26	0.76	150.67	98.6
10/1	197	.467133	.0063319	.50529	.88455	3.87	22.03	0.76	157.93	103.4
9/15	128	.447382	.0070626	.76550	.86517	5.32	25.63	1.00	141.59	92.7
9/1	70	.354731	.0069929	.95989	.85613	12.07	29.70	1.33	143.00	93.6
8/15	19	.207561	.017680	28.315	.71327	16.90	128.11	6.89	56.56	37.0

A comparison of Tables 1 and 2 shows the estimated relative error of the estimated parameters is now smaller as expected. The asymptotic values of the estimated heteroscedastic-error model are at a somewhat lower level than those of model (1). Perhaps the most attractive aspect of the model (2) results is that the asymptotic values are less variable among the six cut-offs. Note the October 1 value is substantially more in line with the other cut-offs than before. The August 15 value is still far from being realistic.

An Autocorrelated-error model

Further examination of the plots from model (1) indicated the residuals may not be independently distributed. Specifically, for small values of time most of the data points lie below the function, particularly for the last three cut-offs. This led to hypothesizing a third set of assumptions concerning the residuals.

Let us assume the residuals in model (1), u_i, can be expressed as

$$u_i = v_i \, f(t_i), \quad i = 1, \ldots, n$$

same as for the heteroscedastic-error model, but further assume the v_i

follow a first-order autoregressive scheme*

$$v_i = \lambda\, v_{i-1} + \varepsilon_i$$

where $|\lambda| < 1$ and the ε_i satisfy the assumptions

$$E(\varepsilon_i) = 0$$

$$E(\varepsilon_i\, \varepsilon_{i+s}) = \sigma_\varepsilon^2 \quad , \quad s = 0$$

$$= 0 \quad , \quad s \neq 0$$

for all i. It then follows that

$$v_i = \lambda\, v_{i-1} + \varepsilon_i$$

$$= \lambda\,(\lambda\, v_{i-2} + \varepsilon_{i-1}) + \varepsilon_i$$

$$= \cdots\cdots\cdots$$

$$= \varepsilon_i + \lambda\, \varepsilon_{i-1} + \lambda^2\, \varepsilon_{i-2} + \cdots\cdots\cdots$$

Therefore,

$$E(v_i) = E(\varepsilon_i) + \lambda\, E(\varepsilon_{i-1}) + \lambda^2\, E(\varepsilon_{i-2}) + \cdots\cdots\cdots$$

$$= 0$$

since $E(\varepsilon_i) = 0$ for all i. Accordingly we can write

$$E(v_i^2) = E[(\varepsilon_i + \lambda\varepsilon_{i-1} + \lambda^2\varepsilon_{i-2} + \ldots)(\varepsilon_i + \lambda\varepsilon_{i-1} + \lambda^2\varepsilon_{i-2} + \ldots.)]$$

$$= E(\varepsilon_i^2 + 2\lambda\varepsilon_i\varepsilon_{i-1} + 2\lambda^2\varepsilon_i\varepsilon_{i-2} + \lambda^2\varepsilon_{i-1}^2$$

$$+ 2\lambda^3\varepsilon_{i-1}\,\varepsilon_{i-2} + \lambda^4\varepsilon_{i-2}^2 + \ldots.)$$

---

*J. Johnston, Econometric Methods, McGraw-Hill, New York, 1963, pp. 244-246.

$$= E(\varepsilon_i^2) + 2\lambda E(\varepsilon_i\varepsilon_{i-1}) + 2\lambda^2 E(\varepsilon_i\varepsilon_{i-2}) + \lambda^2 E(\varepsilon_{i-1}^2)$$

$$+ 2\lambda^3 E(\varepsilon_{i-1}\varepsilon_{i-2}) + \lambda^4 E(\varepsilon_{i-2}^2) + \ldots.)$$

$$= \sigma_\varepsilon^2 + \lambda^2\sigma_\varepsilon^2 + \lambda^4\sigma_\varepsilon^2 + \ldots\ldots\ldots$$

$$= (1 + \lambda^2 + \lambda^4 + \ldots..)\ \sigma_\varepsilon^2$$

$$= \frac{\sigma_\varepsilon^2}{1-\lambda^2}$$

for all i. Also,

$$E(v_iv_{i-1}) = E[(\varepsilon_i + \lambda\varepsilon_{i-1} + \lambda^2\varepsilon_{i-2} + \ldots.)(\varepsilon_{i-1} + \lambda\varepsilon_{i-2} + \lambda^2\varepsilon_{i-3} + \ldots.)]$$

$$= E\{[\varepsilon_i + \lambda(\varepsilon_{i-1} + \lambda\varepsilon_{i-2} + \lambda^2\varepsilon_{i-3} + \ldots)][\varepsilon_{i-1} + \lambda\varepsilon_{i-2}$$

$$+ \lambda^2\varepsilon_{i-3} + \ldots.]\}$$

$$= E[\varepsilon_i\varepsilon_{i-1} + \lambda\varepsilon_i\varepsilon_{i-2} + \lambda^2\varepsilon_i\varepsilon_{i-3} + \ldots + \lambda(\varepsilon_{i-1} + \lambda\varepsilon_{i-2}$$

$$+ \lambda^2\varepsilon_{i-3} + \ldots.)^2]$$

$$= E(\varepsilon_i\varepsilon_{i-1}) + \lambda E(\varepsilon_i\varepsilon_{i-2}) + \lambda^2 E(\varepsilon_i\varepsilon_{i-3}) + \ldots\ldots\ldots$$

$$+ \lambda E(\varepsilon_{i-1} + \lambda\varepsilon_{i-2} + \lambda^2\varepsilon_{i-3} + \ldots.)^2$$

$$= \lambda E(\varepsilon_{i-1} + \lambda\varepsilon_{i-2} + \lambda^2\varepsilon_{i-3} + \ldots..)^2$$

$$= \lambda [E(\varepsilon_{i-1}^2) + 2\lambda E(\varepsilon_{i-1}\varepsilon_{i-2}) + 2\lambda^2 E(\varepsilon_i\varepsilon_{i-2})$$

$$+ \lambda^2 E(\varepsilon_{i-2}^2) + 2\lambda^3 E(\varepsilon_{i-2}\varepsilon_{i-3}) + \lambda^4 E(\varepsilon_{i-3}^2) + \ldots]$$

$$= \lambda(\sigma_\varepsilon^2 + \lambda^2\sigma_\varepsilon^2 + \lambda^4\sigma_\varepsilon^2 + \ldots..)$$

$$= \lambda(1 + \lambda^2 + \lambda^4 + \ldots.)\ \sigma_\varepsilon^2$$

$$= \lambda\ \sigma_v^2$$

Similarly,

$$E\ (v_i v_{i-2}) = \lambda^2\ \sigma_v^2$$

and in general,

$$E\ (v_i v_{i-s}) = \lambda^s\ \sigma_v^2 \quad .$$

Returning to the residuals, u_i, from model (1), we have

$$\begin{aligned} E\ (u_i) &= E[v_i\ f(t_i)] \\ &= f(t_i)\ E(v_i) \\ &= 0 \end{aligned}$$

for all i, and

$$\begin{aligned} E\ (u_i^2) &= E\ \{v_i^2\ [f(t_i)]^2\} \\ &= [f(t_i)]^2\ E(v_i^2) \\ &= [f(t_i)]^2\ \sigma_v^2 \end{aligned}$$

$$\begin{aligned} E\ (u_i u_{i-s}) &= E[v_i\ f(t_i)\ v_{i-s}\ f(t_{i-s})] \\ &= f(t_i)\ f(t_{i-s})\ E(v_i v_{i-s}) \\ &= f(t_i)\ f(t_{i-s})\ \lambda^s\ \sigma_v^2 \end{aligned}$$

Summarized in matrix notation, we have

$$E\ (\underline{U}\ \underline{U}') = \sigma_v^2\ \underline{\Omega}$$

where

$$\underline{\Omega} = \begin{pmatrix} [f(t_1)]^2 & f(t_1)\ f(t_2)\ \lambda & f(t_1)\ f(t_3)\ \lambda^2 & \cdots & f(t_1)\ f(t_n)\ \lambda^{n-1} \\ f(t_1)\ f(t_2)\ \lambda & [f(t_2)]^2 & f(t_2)\ f(t_3)\ \lambda & & f(t_2)\ f(t_n)\ \lambda^{n-2} \\ f(t_1)\ f(t_3)\ \lambda^2 & f(t_2)\ f(t_3)\ \lambda & [f(t_3)]^2 & & f(t_3)\ f(t_n)\ \lambda^{n-3} \\ \vdots & & & & \\ f(t_1)\ f(t_n)\ \lambda^{n-1} & f(t_2)\ f(t_n)\ \lambda^{n-2} & f(t_3)\ f(t_n)^{n-3} & \cdots & [f(t_n)]^2 \end{pmatrix}$$

and

$$\underline{\Omega}^{-1} = \frac{1}{1-\lambda^2} \begin{pmatrix} \frac{1}{[f(t_1)]^2} & \frac{-\lambda}{f(t_1)\,f(t_2)} & 0 & \cdots & 0 & 0 & 0 \\ \frac{-\lambda}{f(t_1)\,f(t_2)} & \frac{1+\lambda^2}{[f(t_2)]^2} & \frac{-\lambda}{f(t_2)\,f(t_3)} & 0 & 0 & 0 & \\ 0 & \frac{-\lambda}{f(t_2)\,f(t_3)} & \frac{1+\lambda^2}{[f(t_3)]^2} & 0 & 0 & 0 & \\ \vdots & & & & & & \\ 0 & 0 & 0 \cdots & \frac{-\lambda}{f(t_{n-2})\,f(t_{n-1})} & \frac{1+\lambda^2}{[f(t_{n-1})]^2} & \frac{-\lambda}{f(t_{n-1})f(t_n)} & \\ 0 & 0 & 0 \cdots 0 & \frac{-\lambda}{f(t_{n-1})\,f(t_n)} & \frac{1}{[f(t_n)]^2} & & \end{pmatrix}$$

It can be shown that if the assumption of autocorrelation holds true, using model (2) will underestimate the true sampling variance of the estimated parameters. As in the case of the heteroscedastic-error model, we can apply the method of generalized least squares at each iteration of the linearzation procedure and obtain an estimate of $\underset{\sim}{\gamma}_{k+1}$ by

$$\hat{\underset{\sim}{\gamma}}_{k+1} = (\underset{\sim}{Z}_k' \, \underline{\Omega}^{-1} \, \underset{\sim}{Z}_k)^{-1} \, \underset{\sim}{Z}_k' \, \underline{\Omega}^{-1} \, (\underset{\sim}{Y} - \underset{\sim}{f}_k) \quad .$$

However, this approach cannot be used with BMD when $\underline{\Omega}^{-1}$ is not a diagonal matrix. Therefore, we must use a transformation matrix T such that a new model will be formulated that can be fitted by ordinary least squares and

that will have a scalar dispersion matrix. That is,

$$E\,(\underline{T}\,\underline{U}\,\underline{U}'\,\underline{T}') = \sigma^2\, I_n \quad .$$

It can be verified by multiplying out that if $\underline{T}_1$ is defined as

$$\underline{T}_1 = \begin{pmatrix} \frac{\sqrt{1-\lambda^2}}{f(t_1)} & 0 & 0 \cdots 0 & 0 & 0 \\ \frac{-\lambda}{f(t_1)} & \frac{1}{f(t_2)} & 0 \cdots 0 & 0 & 0 \\ 0 & \frac{-\lambda}{f(t_2)} & \frac{1}{f(t_3)} \cdots 0 & 0 & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 \cdots \frac{-\lambda}{f(t_{n-2})} & \frac{1}{f(t_{n-1})} & 0 \\ 0 & 0 & 0 \cdots 0 & \frac{-\lambda}{f(t_{n-1})} & \frac{1}{f(t_n)} \end{pmatrix}$$

then,

$$E\,(\underline{T}_1\,\underline{U}\,\underline{U}'\,\underline{T}_1') = \sigma_\varepsilon^2\, I_n \quad .$$

The result of applying this transformation to the original model is

$$\frac{\sqrt{1-\lambda^2}}{f(t_1)}\, y_1 = \frac{\sqrt{1-\lambda^2}}{f(t_1)}\, \frac{1}{\alpha + \beta\,\rho^{t_1}} + \frac{\sqrt{1-\lambda^2}}{f(t_1)}\, u_1 \,,$$

and

$$\frac{y_i}{f(t_i)} - \frac{\lambda}{f(t_{i-1})}\, y_{i-1} = \frac{1}{f(t_i)}\, \frac{1}{\alpha + \beta\,\rho^{t_i}} - \frac{\lambda}{f(t_{i-1})}\, \frac{1}{\alpha + \beta\,\rho^{t_i}} + \frac{u_i}{f(t_i)} - \frac{\lambda}{f(t_{i-1})}\, u_{i-1} \,,$$

$$i = 2, \ldots, n. \qquad (3)$$

To verify that the residuals of this autocorrelated-error model are independently distributed with mean zero and have a constant variance, let r_i be the residual for the i^{th} observation. Then,

$$E(r_1) = E\left[\frac{\sqrt{1-\lambda^2}}{f(t_1)} \; u_1\right]$$

$$= \frac{\sqrt{1-\lambda^2}}{f(t_1)} \; E(u_1)$$

$$= 0$$

$$E(r_i) = E\left[\frac{u_i}{f(t_i)} - \frac{\lambda}{f(t_{i-1})} \; u_{i-1}\right]$$

$$= \frac{1}{f(t_i)} \; E(u_i) - \frac{\lambda}{f(t_{i-1})} \; E(u_{i-1})$$

$$= 0, \qquad i = 2, \ldots, n.$$

$$E(r_1^2) = E\left\{\left[\frac{\sqrt{1-\lambda^2}}{f(t_1)} \; u_1\right]^2\right\}$$

$$= (1-\lambda^2)\,\sigma_v^2$$

$$= \sigma_\varepsilon^2 \; .$$

$$E(r_i^2) = E\left\{\left[\frac{u_i}{f(t_i)} - \frac{\lambda}{f(t_{i-1})} \; u_{i-1}\right]^2\right\}$$

$$= E\left\{\frac{u_i^2}{[f(t_i)]^2} - 2\,\frac{u_i}{f(t_i)}\;\frac{\lambda}{f(t_{i-1})}\;u_{i-1} + \frac{\lambda^2}{[f(t_{i-1})]^2}\;u_{i-1}^2\right\}$$

$$= \sigma_v^2 - 2\lambda^2\sigma_v^2 + \lambda^2\sigma_v^2$$

$$= (1-\lambda^2)\;\sigma_v^2$$

$$= \sigma_\varepsilon^2 , \qquad i = 2, \ldots, n.$$

$$E(r_i r_1) = E\left\{\left[\frac{u_i}{f(t_i)} - \frac{\lambda}{f(t_{i-1})} u_{i-1}\right]\left[\frac{\sqrt{1-\lambda^2}}{f(t_1)} u_1\right]\right\}$$

$$= E\left\{\frac{\sqrt{1-\lambda^2}}{f(t_i)\, f(t_1)} u_i u_1 - \frac{\lambda\sqrt{1-\lambda^2}}{f(t_{i-1})\, f(t_1)} u_{i-1} u_1\right\}$$

$$= \sqrt{1-\lambda^2}\ \lambda^{i-1} \sigma_v^2 - \lambda\sqrt{1-\lambda^2}\ \lambda^{i-2} \sigma_v^2$$

$$= 0, \qquad i = 2, \ldots, n.$$

$$E(r_i r_{i-s}) = E\left\{\left[\frac{u_i}{f(t_i)} - \frac{\lambda}{f(t_{i-1})} u_{i-1}\right]\left[\frac{u_{i-s}}{f(t_{i-s})} - \frac{\lambda}{f(t_{i-s-1})} u_{i-s-1}\right]\right\}$$

$$= E\left\{\frac{u_i u_{i-s}}{f(t_i)\, f(t_{i-s})} - \frac{\lambda u_i u_{i-s-1}}{f(t_i)\, f(t_{i-s-1})} - \frac{\lambda u_{i-1} u_{i-s}}{f(t_{i-1})\, f(t_{i-s})} + \frac{\lambda^2 u_{i-1} u_{i-s-1}}{f(t_{i-1})\, f(t_{i-s-1})}\right\}$$

$$= \lambda^s \sigma_v^2 - \lambda\lambda^{s+1} \sigma_v^2 - \lambda\lambda^{s-1} \sigma_v^2 + \lambda^2\lambda^s \sigma_v^2$$

$$= 0, \qquad i = 2, \ldots, n.$$

Thus, the residuals, r_i, have the desired characteristics.

Having formulated this autocorrelated-error model which retains the assumption of heteroscedasity, the next step was to test for first-order autocorrelation for each of the six cut-offs. The test used was the von Neumann ratio.* This is a large sample test and is made by comparing the computed value obtained with a pre-selected critical region from the normal distribution with the appropriate mean and variance.

For each cut-off the von Neumann ratio, δ^2/s^2, was computed by

*J. Johnston, op. cit., 1963, p. 250.

$$\frac{\delta^2}{s^2} = \frac{\sum_{i=2}^{n} (\hat{v}_i - \hat{v}_{i-1})^2 / (n-1)}{\sum_{i=1}^{n} \hat{v}_i^2 / n}$$

where

$$\hat{v}_i = \hat{u}_i / f(t_i) \quad , \quad i = 1, \ldots, n.$$

The expected value and variance of δ^2/s^2 are given by

$$E\left(\frac{\delta^2}{s^2}\right) = \frac{2n}{n-1}$$

$$V\left(\frac{\delta^2}{s^2}\right) = \frac{4n^2 (n-2)}{(n+1) (n-1)^3}$$

An estimate of λ for each cut-off was obtained by

$$\hat{\lambda} = \frac{\sum_{i=2}^{n} \hat{v}_i \hat{v}_{i-1}}{\sum_{i=2}^{n} \hat{v}_i^2}$$

which is a regression of the $\hat{v}_i$ on $\hat{v}_{i-1}$ with no intercept. The results of the von Neumann tests and the estimates for λ are summarized in Table 3.

Table 3

cut-off	computed von Neumann ratio	$E(\delta^2/s^2)$	$V(\delta^2/s^2)$	Estimated 95% confidence interval limits for δ^2/s^2	$\hat{\lambda}$
all obs.	1.70	2.00722	0.01439	(1.77, 2.24)	0.169891
10/15	1.74	2.00784	0.01562	(1.76, 2.25)	0.152777
10/1	1.53	2.01020	0.02030	(1.73, 2.29)	0.253208
9/15	1.76	2.01575	0.03125	(1.67, 2.36)	0.126061
9/1	1.99	2.02899	0.05714	(1.56, 2.50)	0.021710
8/15	1.95	2.11111	0.21046	(1.21, 3.01)	0.053826

For those cut-offs where the computed von Neumann ratio lies outside the estimated confidence interval, we can, based on this sample data, reject the null hypothesis of non-autocorrelated residuals in favor of the hypothesis of positive first-order autocorrelation. Table 3 shows there is evidence of autocorrelation for the last three cut-offs, although it is only for the October 1 cut-off that the evidence is particularly convincing. For the first three cut-offs, we cannot reject the null hypothesis of non-autocorrelated residuals. Accordingly, model (3) was fitted to the last three cut-offs.

As is commonly done, the first observation was deleted in fitting the model. The results are presented in Table 4.

Table 4

cut-off all	n	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\rho}$	$\hat{\sigma}_{\hat{\alpha}}/\hat{\alpha}$ (%)	$\hat{\sigma}_{\hat{\beta}}/\hat{\beta}$ (%)	$\hat{\sigma}_{\hat{\rho}}/\hat{\rho}$ (%)	$\lim_{t_i \to \infty} y_{t_i}$ est'ed value	% of est'ed hv. wt.
obs.	277	.0068072	.46516	.88591	3.71	29.25	1.00	146.90	96.1
10/15	255	.0065846	.43236	.88903	3.94	27.50	0.93	151.87	99.4
10/1	196	.0062570	.45860	.88753	5.41	31.02	1.07	159.82	104.6

A comparison of Tables 2 and 4 shows the estimated relative error of the estimated parameters for model (3) are larger than for model (2) as expected. However, the asymptotic values changed only slightly. This is consistent with the relatively small degree of estimated autocorrelation as reflected in the small values of the estimated λ's.

Summary

We have shown the fit of an intrinsically nonlinear model to sample data by the linearization method is only an approximation of an exact fit in the sense that an absolute minimum of the error sum of squares is not attained. In spite of this limitation, the results of its use in fitting the growth model appear satisfactory.

The results of the heteroscedastic-error model are encouraging. However, the assumption of a first-order autoregressive scheme is not particularly convincing. Further investigation of alternative autocorrelation schemes might be warranted.