A Within-Year Growth Model Approach to Forecasting Corn Yields

Crop Reporting Board

Economics, Statistics, and Cooperatives Service
U.S. Department of Agriculture

Washington, D.C. 20250

A WITHIN YEAR GROWTH MODEL APPROACH TO FORECASTING CORN YIELDS

By

CAROL C. HOUSE

MATHEMATICAL STATISTICIAN
RESEARCH AND DEVELOPMENT BRANCH
STATISTICAL RESEARCH DIVISION

ECONOMICS, STATISTICS AND COOPERATIVES SERVICE

Washington, D. C. September, 1977

TABLE OF CONTENTS

PAGE

TABLES OF GRAPHS AND FIGURES

PAGE

ACKNOWLEDGEMENTS

Thanks are expressed to the Iowa and Texas State Statistical Office staffs for an excellent job in supervising the data collection phase of this research effort.

Special appreciation is extended to Gary Elder who managed the project throughout the planning and data collection phases. Gary worked on both the 1975 and 1976 Corn Research Projects, supplying much of the expertise and hard work involved in collection and analysis of the data discussed in this report.

INTRODUCTION

The Statistical Reporting Service (SRS) presently forecasts grain crop yields with models whose parameters have been estimated by use of historic data. These "between-year models" depend on a base period of time, usually three years, to supply data on the relationships of plant measurements to final yield. The estimated yield at harvest for each plot in a probabilistic sample is regressed against preharvest plant counts and fruit measurements, and the model parameters are estimated. An assumption is made that the present year is a part of the composite population of these base period years, and the model parameters that have been estimated are then used in conjunction with current year counts and measurements to forecast current year yield.

Since 1973, The Yield Assessment Section of SRS has been involved in the development of within-year forecasting models for various grain crops. These models are being developed to provide forecasts of pertinent components of crop yield by relying entirely on growth and survival data collected from plant observations during the current growing season.

Although the value of historic information in crop forecasts is apparent, "within-year" models could be a valuable supplement to the between-year models presently used by SRS. A model which uses data from the current growing season only may be beneficial in improving forecasts during a year with atypical growing conditions. Recent speculation of worldwide climatic changes indicates the possibility of an increasing number of atypical years. A within-year model could also be used effectively in developing an objective yield program for a crop or state previously excluded from objective forecasting. In such a situation, there would be no base period of growth data from which to estimate parameters.

The 1976 Corn Growth Research Project was carried out as part of the continuing effort to develop feasible within-year forecasting techniques. The objectives of this paper are to:

1) Discuss the form of the logistic growth model and its applicability to forecasting corn yields.

2) Describe the sampling, data collection and data handling techniques involved in the 1976 Corn Research Projects.

3) Discuss the analysis that was performed on this data. This discussion will include:

    a) The performance of the logistic model and two homoscedastic modifications when fitted using 1976 corn growth data from Iowa and Texas.

    b) The use of simulated growth data to observe various model characteristics.

    c) The use of the logistic model to forecast plant survival.

4) Make recommendations for future research.

LOGISTIC GROWTH AND SURVIVAL MODELS

A logistic model is a non-linear model having a single dependent variable and an independent "time" variable. The model uses repeated observations from the current year to estimate the parameters needed to predict the dependent variable at maturity. The logistic model has been shown by previous studies to accurately describe a growth process as well as a survival process in corn kernel formation. (XII, X)[1/]

The form of the logistic model is:

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

α, β, ρ = non-negative parameters, $o < \rho < 1$

ε_i = disturbance term

t_i = independent time variable

y_i = dependent growth or survival variable

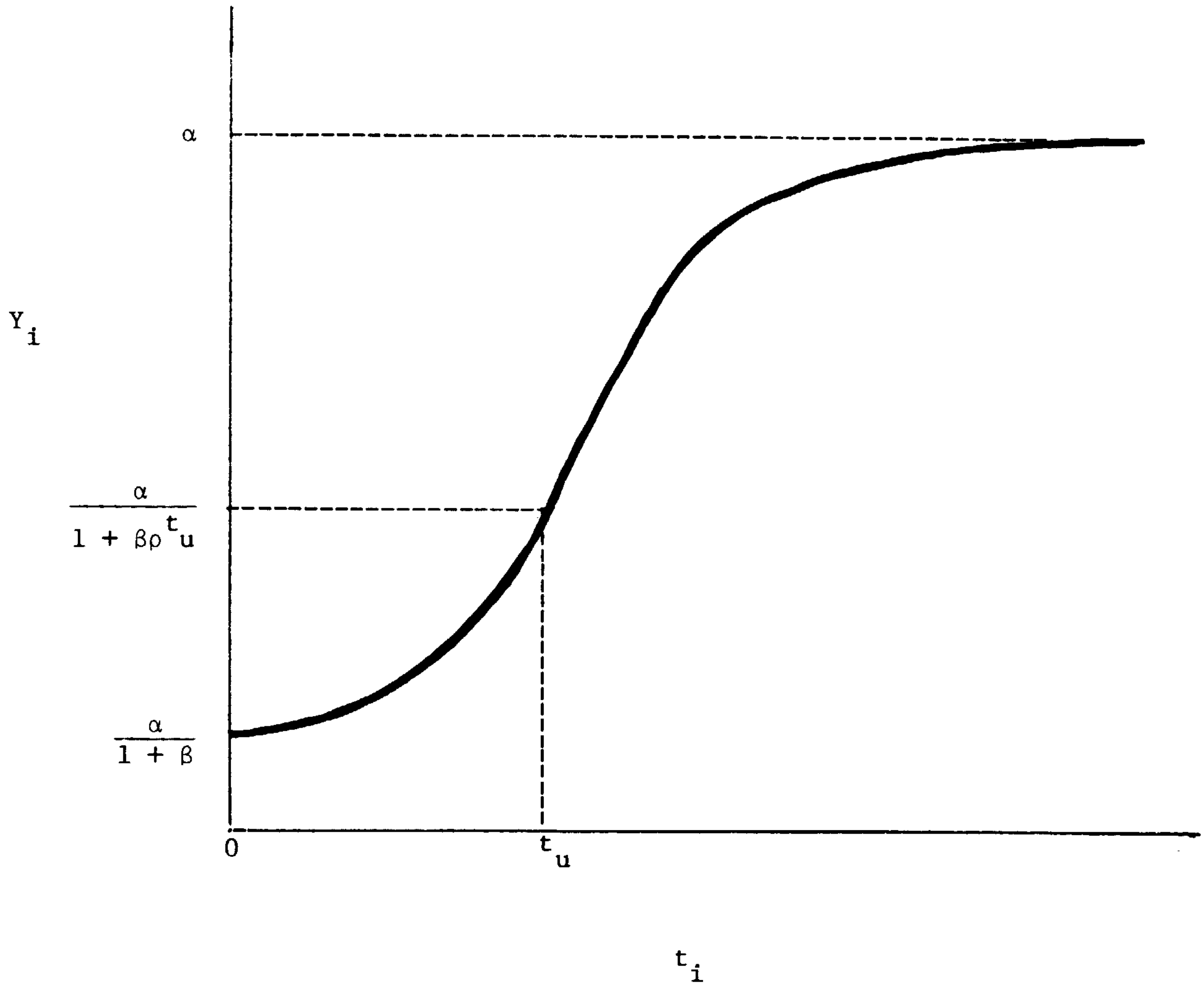
The disturbance term ε_i has been observed to have a functional relationship with the independent variable. Thus in reality, $\varepsilon_i = \varepsilon_i(t_i)$. The actual form of this disturbance will be discussed in more depth later in this report.

When used to explain corn growth, the model gives the relationship of kernel weight and development to the length of time the kernels have been growing. In previous research efforts, various combinations of independent and dependent variables were experimented with to serve as proxy for growth and growing time. These included "tassel emerged," "silked emerged," and "silk starting to dry" as the zero point for the time variable, and "dry ear weight," "dry kernel weight," and "wet kernel weight" as the growth variable. The optimal combination of those tested was "time since silking" and "dry kernel weight per plant." (XII)

The logistic growth model hypothesizes that dry kernel weight accumulates slowly in a plant during the earliest stages of ear development, increases at an increasing rate for a period of time, and then increases at a decreasing rate approaching an asymptotic maximum value. This asymptotic value represents the total dry weight per plant at maturity. It is this parameter that is expanded by estimates of plants per acre to produce a yield forecast.

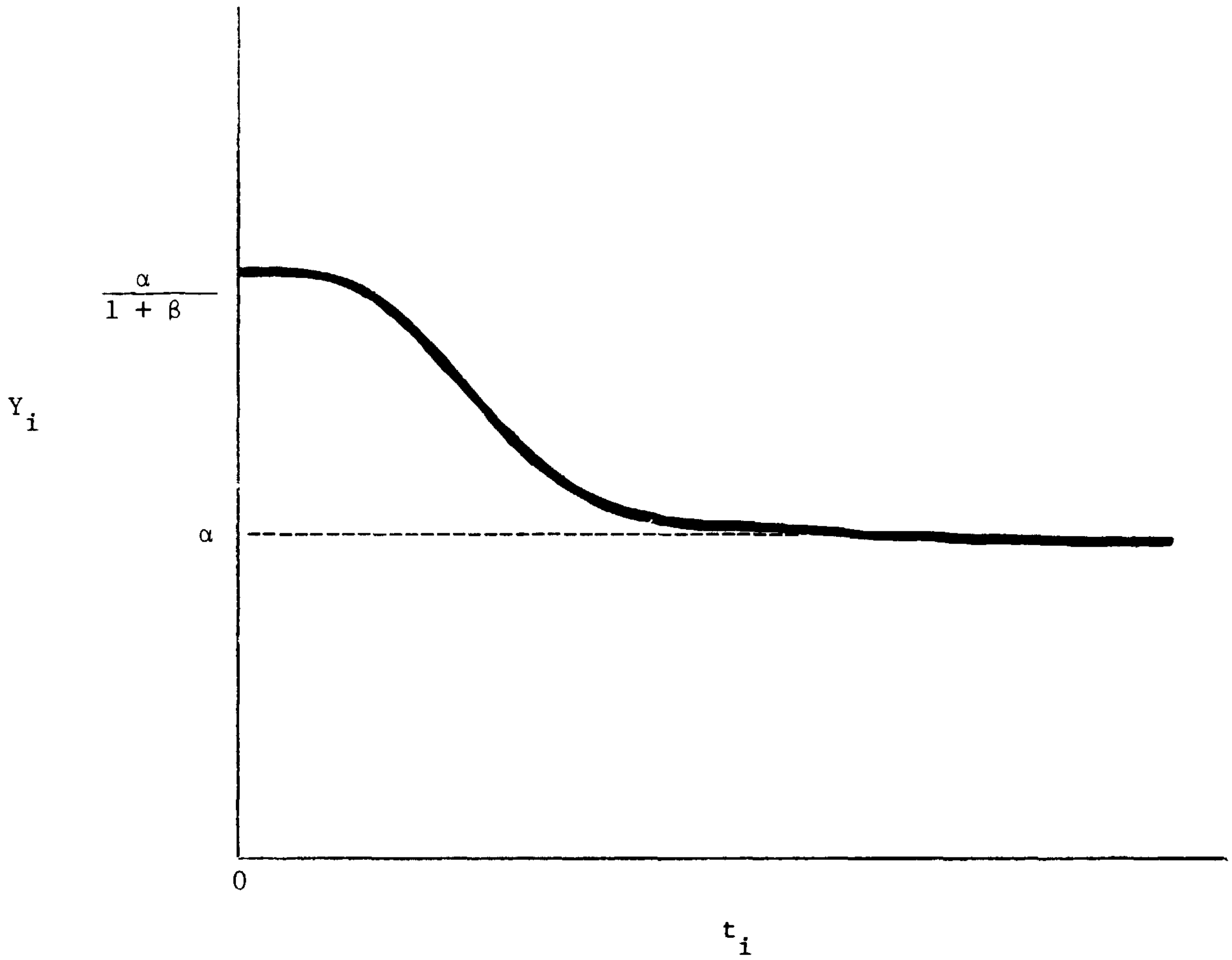
1/ References to sources will be indicated by parenthesized Roman numerals which are associated with entries on the reference page at the end of the report.

The logistic growth model is shown graphically below.



Plant survival has also been shown to be a phenomenon whose change over time can be expressed by a logistic equation. Unlike a growth function, a function modeling a survival process must have a nonpositive slope. This is what happens when the β parameter in the logistic function is negative.

The logistic survival model is shown graphically below.



SAMPLE DESIGN AND DATA COLLECTION

Subsamples of the June Enumerative Survey (JES) corn fields were randomly chosen in Texas and Iowa to serve as the corn growth research sample of fields. The JES was based on an area frame sample of segments where corn acreage and the acreages of other crops were obtained for each field in the segment. Because estimates of "yield per acre" and not "per field" were desired, fields were sampled with probabilities proportional to the JES expanded corn acreage. This procedure produced a sample which was self-weighted by acres per field.

Ninety sample fields were selected throughout Iowa while 30 fields were selected in two crop reporting districts in western Texas. A number of fields were lost due to farmer refusal or other circumstances resulting in a final sample size of 24 in Texas and 78 in Iowa.

Within each field two units were located randomly using procedures outlined in the objective yield "Corn, Cotton and Soybeans" Enumerator Manual. (VI) Each unit consisted of a row length of 100 corn plants. A plant was defined as all growth from a single seed including the main stalk and all tillers.

Data collection was carried out in four phases: plant population count, phenological event observations, ear removal and field measurements, and the laboratory determination of dry matter. A more complete description of all data collection procedures can be found in the enumerator's and laboratory manuals. (VII, VIII)

PLANT POPULATION COUNTS

During the first visit to each sample field, plant population determinations were made. Plants were counted in a 45 foot row segment in each randomly selected unit. Row widths were measured at three places in each unit. These measurements allowed plant population estimates to be aggregated to the field, state or area level.

PHENOLOGICAL EVENT OBSERVATIONS

The purpose of the second phase of the data collection, the phenological event observations, was to obtain the time of silk emergence for each plant in each unit. Silk emergence for a plant was defined to have occurred when silk was first observed on any ear of the plant or its tillers. The silking data for a plant was set as the data midway between the date of the visit when silk emergence was first observed and the date of the previous visit. Observations were made for silk emergence every 3-4 days during peak silking periods, and weekly during less active silking days.

At a time a plant had been observed to have silked, a yellow tag was placed on it. Only plants with such tags were included in the remainder of the sampling process as plants not silking were assumed to have no grain producing capability.

EAR REMOVAL AND FIELD MEASUREMENTS

Beginning the last week of July, weekly visits to the sample fields were made to carry out the third phase of data collection. These visits continued for 10 weeks or until harvesting of the field, whichever occurred first.

On each visit, four plants showing kernel formation were selected from each unit in the following manner. Blocks of ten plants were independently ordered for sampling on the first through tenth sampling visit. Within each block, the ten plants were randomly ordered. On any given visit, plants in the selected block were observed in the random order until four plants showing kernel formation had been sampled. A plant was considered to have kernel formation if any of its ears had such formation. Plants that had not silked (did not have a yellow tag) were excluded from the population being sampled.

The following "in-field" observation were made on each ear of each plant being sampled. First a determination was made as to whether kernel formation was present. Then ear length and circumference measurements were made without removing the ear from the stalk or disturbing the husk. Finally the ear was picked, husked and weighed. Each ear was carefully labeled as to whether it was a primary ear, secondary ear, etc., sealed in a plastic bag and sent to the laboratory.

The in-field measurements and weighing of ears were made to explore the usefulness of such measurements in doubling sampling procedures. Analysis of this portion of the data will be presented in a seperate research report.

LABORATORY DETERMINATION OF DRY MATTER

Sampled ears were sent to the laboratory for determination of dry matter content. The following procedures, performed on each ear arriving at the laboratory, made up the fourth phase of data collection.

Two kernel rows were chosen randomly from each ear. The kernels in each selected row were carefully removed from the cob to prevent damage or puncturing and to prevent removal of cob parts with them. Kernels from each individual row were weighed after removal from the cob, and dried in an oven for 72 hours at a temperature of 150^0F to standardize moisture content. This temperature and drying period were chosen because they were found to reduce moisture in grain at maturity to less than two percent, while not burning the immature grain coming into the laboratory early in the growing season. Kernels were weighed after this drying process to determine dry matter content. Determinations from each of the two sampled kernel rows were averaged and expanded by the number of kernel rows to compute a mean dry weight of grain for the ear.

DATA ANALYSIS

The analysis of data was performed seperately for each state. Logistic growth models will be discussed first.

LOGISTIC GROWTH MODELS

Data collected in the fields and laboratory during the summer of 1976 were prepared so that a logistic growth model could be fitted to such data via standard nonlinear regression techniques.

"Time since silking," the proposed independent variable, was defined to be the assigned Julian date of silk emergence (see section on Phenological Event Observations) subtracted from the date when the plant was sampled.

The estimated dry kernel weights (in grams) for each ear of a plant processed in the laboratory was summed to provide an estimate of dry kernel weight per plant. If a plant had zero dry kernel weight, it was treated as a "non-survivor" and deleted from data used to fit the growth model. Instead, non-survivors were used to formulate a logistic survival model which will be discussed in more detail later in this report.

Mean dry kernel weight and time since silking were estimated for the growing plants sampled from a field on each visit. Such means incorporated data from all plants in the same field that were sampled on the same date. Thus, for each field and sampling date for that field, there was one value for the independent variable (the mean time since silking) and one value for the dependent variable (mean dry weight per plant). These means were used in the regression procedures instead of individual plant data to insure independence of data points.

Estimates of the number of silked plants per acre were obtained for each unit by adjusting the plant population estimate by the proportion of tagged stalks that had silked. A silked plant population estimate for each field was obtained by averaging the estimates derived from the two units. Data entering the model from fields with higher silked plant populations were representative of a larger proportion of the total population of silked plants than data from fields with the low silked plant populations. Thus, such data were given more weight in fitting the model.

The weighted Marquardt nonlinear procedure in SAS (Statistical Analysis System) was used to regress the dependent growth variable on the time variable. Weights were developed from expanded silked plant populations, causing more weight to be given to fields with denser silked plant populations. For each field, the following weight was calculated.

$$w_i = \frac{\text{\# of silked plants per acre}}{10{,}000} \qquad i = 1,2,\ldots, n \qquad (2)$$

The weighted regression equation was the following:

$$(w_i)\ (y_i) = (w_i) \left[\frac{\alpha}{1 + \beta\rho^{t_i}} + \varepsilon_i \right] \qquad i = 1,2,\ldots, n \qquad (3)$$

In general least squares theory, several basic model assumptions are made:

ε_i is normally distributed (4)

$$E\ (\varepsilon_i) = 0 \qquad (5)$$

$$\text{Var}\ (\varepsilon_i) = \sigma_\varepsilon^2 \qquad (6)$$

$$\text{Cov}\ (\varepsilon_i, \varepsilon_j) = 0 \qquad (7)$$

Thus, the assumption was made that the residuals obtained from fitting the logistic growth model to the corn data were independently distributed with mean zero and a constant variance σ_ε^2. (As indicated earlier in this report, research conducted in 1974 and 1975 has indicated that the assumption of constant variance (6) is not valid. Variations of the regression equation will be discussed below whose residuals when fitted to the corn growth data do not exhibit a significant correlation with time). Figures A-1 and A-2 in the appendix have plots for both Texas and Iowa of the data from the entire growing season with the logistic model fitted to these data. Using data collected throughout the growing season, the SAS procedures estimated the model parameters as follows:

Iowa	Texas
$\hat{\alpha} = 141.2$	$\hat{\alpha} = 166.0$
$\hat{\beta} = 38.4$	$\hat{\beta} = 43.0$
$\hat{\rho} = .892$	$\hat{\rho}$ .897

In each case, $\hat{\alpha}$ would give the estimated mean dry kernel weight (in grams) per plant at maturity.

Since the growth model was being researched to determine its ability to "forecast" kernel weight at maturity, it was important to examine the parameter estimates from a regression in which only early season data were present. To this end, the above model was fitted to cumulative growth data each week as new data became available. For example, when the model was fitted at the end of four weeks of data collection, it was fitted to data from plants sampled on any of the first four visits to the field but excluded data from any plant sampled on subsequent visits. Figures A-4 and A-7 in the appendix give the values of the estimated parameters for each cutoff data for this unadjusted model. Notice that these estimates change as additional data are made available. By examination of these changes, it was hoped that some of the following questions may be answered.

How early in the growing season can a reasonable forecast be made?
How will the addition of later data affect this forecast?
What happens to the forecast errors as more data are added?

Before these and other questions are pursued, possible violations of model assumptions will be examined.

HETEROSCEDASTICITY

After fitting the model through all ten weeks of data, an examination was made to determine if the underlying assumptions of the model had been met. Earlier research into this growth model indicated that the assumption of constant variance did not hold and that the residuals from the regression have a significant positive correlation with time. The sample correlation coefficient, R, between the residuals and time was R = .49 in Texas, and R = .46 in Iowa for all ten weeks of data. Graphs of the residuals versus time for each state are found in the appendix figures A-13 and A-15.

This condition is commonly referred to as heteroscedasticity. As reported in Goldfeld and Quandt, ". . . one generally obtains inferior parameter estimates if ordinary least squares is applied to a model with heteroscedastic disturbances. Furthermore, the presence of heteroscedasticity may invalidate standard tests of statistical significance." (I, p.78).

This heteroscedastic problem was recognized during the data analysis of the 1974 Corn Growth Research Project and various adjustments to the model have been tried to eliminate or minimize this error. (XI)

At the present time, the following ones seem the most valid. A model more closely representing what has been observed is:

$$(w_i)\ (y_i) = (w_i) \left[\frac{\alpha}{1 + \beta\rho^{t_i}} + \varepsilon_i(t_i) \right] \qquad i = 1,2,\ldots, n \qquad (8)$$

where $\varepsilon(t)$ is a random variable with

$$E(\varepsilon\ (t)\ ) = 0$$

$\sigma_{\varepsilon(t)}$ is some increasing function of the independent variable t.

Under the above assumptions consider the following model:

$$z_i = \frac{w_i}{\hat{\sigma}_{u_i}\ (t_i)} \cdot \frac{\alpha}{1 + \beta\rho^{t_i}} + \frac{w_i}{\hat{\sigma}_{u_i}\ (t_i)} \cdot \varepsilon_i(t_i) \qquad (9)$$

where

$$z_i = \frac{(w_i)\ (y_i)}{\hat{\sigma}_{u_i}(t_i)}$$

and $\hat{\sigma}_{u_i}(t_i)$ is an estimate for the standard deviation of (10)

$$u_i(t_i) = (w_i) \left[\varepsilon_i(t_i) \right] \qquad (11)$$

The disturbance term in this model is:

$$\frac{u_i(t_i)}{\hat{\sigma}_{u_i}(t_i)} \qquad (12)$$

with variance

$$\text{Var}\ \frac{u_i(t_i)}{\hat{\sigma}_{u_i(t_i)}} = \frac{\sigma^2_{u_i(t_i)}}{\hat{\sigma}^2_{u_i(t_i)}} \tag{13}$$

The ratio is close to "one" for good estimators of $\sigma^2_{u_i(t_i)}$, and the model will be homoscedastic as long as the ratio remains constant over time.

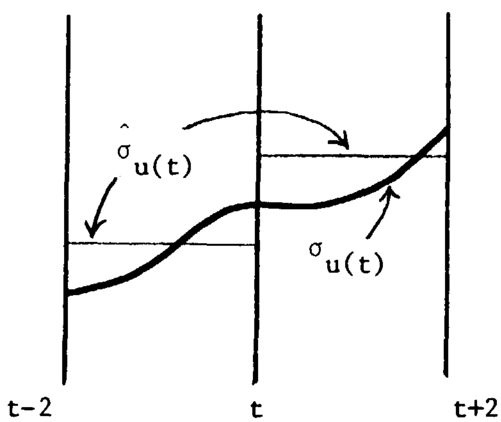
Two methods have proved effective in obtaining an estimate of

$\sigma_{u_i(t_i)}$

(a) Standard Error Adjustment

The first method of estimating $\sigma_{u_i(t_i)}$ was suggested in the analysis of a within year logistic wheat growth model. (III) The relevant range of the independent time variable was broken into two-day intervals and a sample standard deviation from the predicted value was computed for each interval. A small variation has been made in the procedure described in the wheat research report. The unadjusted model was first fitted to the data and the predicted y value for the mean time in each interval was determined. The sample deviation of the observations in each interval from this "expected y value" was used as $\hat{\sigma}_{u(t)}$, instead of the standard deviation about the sample mean as was done with the spring wheat model. The assumption was made that within such a small time period, $\sigma_{u(t)}$ changes so little that it could be assumed constant.

How good an estimate of $\sigma_{u(t)}$ does this method provide?



A step function was derived to approximate an unknown function that is most probably continuous. The classical way to improve such an approximation is to decrease the width of the intervals. Unfortunately, in this case, decreasing the size of the time intervals also decreases the number of observations available in each interval for estimating the standard deviation.

As stated above, it was decided to use an interval width of two days. Such an interval was small enough to allow for a reasonable approximation by a step function. In the Iowa data, the number of observations within a single interval ranged from a low of 2 to a high of around 30. The lower numbers were found in intervals located at the extremes of the time span. If all intervals contained a sufficient number of observations, this method of obtaining an estimate of $\sigma_{u(t)}$ could be satisfactory. Since this was not the case for all intervals, another methods of estimation was also considered.

(b) Logistic Adjustment

Because of the problems that could occur with the above method of adjustment, an attempt was made to determine a functional form for $\sigma_{u(t)}$.

An examination of the residuals from the regression of the unadjusted model indicated that the absolute value of the residuals of u(t) increased slowly at first, then increased more rapidly, finally leveling off to a broad horizontal band (See Figure A-17 in the appendix). This behavior suggested that $\sigma_{u(t)}$ itself could have a logistic structure.

If $\sigma_{u(t)}$ did have a logistic functional form, it would still be necessary to estimate the parameters for this function. To do this, the absolute value of the residuals from the regression with the unadjusted model were used in a non-linear regression to fit the model.

$$Q_i = \frac{E}{1 + FG^{t_i}} + V_i \qquad i = 1, \ldots, n \qquad (14)$$

where V_i has mean zero and small constant variance. Once these regression parameters E,F,G were estimated, the fitted equation

$$\hat{\sigma}_{u_i}(t_i) = \frac{\hat{E}}{1 + \hat{F}\,\hat{G}^{t_i}} \qquad i = 1, \ldots, n \qquad (15)$$

was used as an approximation for $\sigma_{u(t)}$. The homoscedastic model (9) was then run. This method of adjustment overcame one of the major problems present with the use of the standard error adjustment. It did not require a small interval to contain a certain minimum number of data points. The major disadvantage of this method of adjusting for heteroscedasticity is apparent. Its value as a supplier of an estimate of $\sigma_{u(t)}$ depends upon the ability of the logistic model to explain the pattern of regression residuals. Later in this report, a set of simulated data points will be discussed. These points were generated to fit a logistic equation with a disturbance term which was normally distributed with zero mean and logistic standard deviation. Examination of these simulated data gives some additional credibility to the assumption of a logistic standard deviation.

COMPARISON OF THREE MODELS

Comparisons using 1976 corn research data were made of the three logistic models discussed above: The unadjusted model, the standard error adjusted model, and the logistically adjusted model. Attention was concentrated on a comparison of the two homoscedastic models.

Because the purpose of this research was to develop forecasting techniques, it was important to consider how each of these models behaved at various cutoff dates throughout the growing season. Several of the cutoff dates were particularly important. Four weeks of data were available for the September 1 crop forecast, 8 weeks of data could have been available for the October 1 forecast, and all ten weeks had been processed by the November 1 estimate. Which of the three models does the best job of predicting corn growth? What properties should a model demonstrate to show that it can give a reliable grain forecast?

Two possible methods of evaluating these models were suggested in an earlier research report on corn. (XII) These are:

(a) The magnitude and sign of the departure of the forecast from the actual mean dry weight at maturity.

(b) The magnitude of the relative standard deviation of the primary parameter, α.

Mean dry weight at maturity was estimated from the subsample of plants that were sampled after reaching maturity. Maturity was defined to have been reached 60 days after silking. The mean was weighted according to silked plants per acre in the same manner as the nonlinear regression.

Small deviations of $\hat{\alpha}$ from the mean dry weight at maturity at early cutoff dates could indicate a reliable early forecast.

The relative standard deviation is the estimated standard deviation divided by the estimate of the primary parameter:

$$\text{Relative Standard Deviation} = \frac{\hat{\sigma}_\alpha}{\hat{\alpha}}$$

Other important considerations in comparing the three models include:

(c) The cutoff date on which the primary parameter α begins to stabilize.

(d) The presence of heteroscedastic error.

Graphs in the appendix of this report show the following for each of the three models in each state:

1. Absolute value of relative deviation of $\hat{\alpha}$ from the mean dry weight at maturity, over time (Figures A-19 through A-22).

2. The relative standard deviation of the primary parameter over time (Figures A-4, A-7,A-6, A-9).

3. The estimated primary parameter over time (Figures A-4, A-7, A-5, A-8).

In addition to these graphs, there are accompanying tables which give exact values of the variables which appear in the graphs. Below are some generalizations that can be made by reviewing the results presented in the graphs and charts.

In going over the data, it was immediately clear that considerably more variability exists in the results from the Texas data. This appears to be the result of the sample size being about one-third of that in Iowa. For either state, it is important to give special attention to what the data are doing at the end of four weeks and then at the end of eight weeks.

The relative error of the primary parameter is one of the more useful measures of the reliability of the forecast estimate and should be helpful in contrasting the two adjusted models. In Iowa, the standard error adjusted model has a relative error which drops below the 5 percent level by the fourth week, which is in time for the September 1 forecast. By the fifth week the logistically adjusted model's relative error is also below 5 percent. In Texas, where the sample size was considerably smaller, the relative error of the standard error adjusted model reached 5 percent on the fifth visit with both of the other models falling to about 5 percent level by the sixth week. The relative error of the standard error adjusted model remained consistently below that of the other two models, but deviation between the errors of any two of the models was relatively small after four weeks of data was available.

One important characteristic of a forecasting technique is stability. After a certain amount of data have been made available, inclusion of additional data should not cause great shifts in forecast levels.(More variability will of course occur in years when late season data represent growing conditions very different from that represented by earlier data). The earlier in the growing season that a model's forecast level begins to stabilize, the earlier such a forecast can be reliably used in making official yield forecasts.

Due to the lack of data in early weeks of a within-year growth model, one would expect considerable fluctuation from one cutoff date to another during these early weeks. What happened after four weeks when the September 1 forecast must be made? Texas with its small sample size was still fluctuating considerably at this stage. The Iowa forecast level at four weeks, using either of the two adjusted models, did not vary much from the levels found after six weeks, but there was still a jump in dry weight (a little under ten grams per plant) between the 4 week forecast and the 8 week forecast (October 1). It should be noticed that in the relevant range of the growing season (four weeks and after) the forecast levels obtained by using the standard error adjustment model and the logistically adjusted model were consistent.

An upward trend of the $\hat{\alpha}$ forecast levels over time in both states and all three models was observed. This condition was more noticeable in Iowa where each week the forecast level was several grams larger than the level the week before. This could indicate a serious problem in the forecasting technique. Could this condition be intrinsic to this forecasting process or is it a phenomenon peculiar to the 1976 growth data? This question will be addressed in more detail in the section on the simulation model.

One final evaluation tool that was used to compare the various models involved looking at what could be called maturity deviation. The mean dry weight at maturity was calculated and the percent relative deviation from mean dry weight at maturity was determined for each cutoff date.

$$\text{maturity deviation} = \left| \frac{\hat{\alpha} - \alpha_m}{\alpha_m} \right| \times 100 \qquad (13)$$

Where $\hat{\alpha}$ is the forecast estimate and α_m is the mean dry weight at maturity.

In the Texas data there seemed to be a good bit of variability between models and from week to week for the same model. No one model appeared to perform appreciably better than any other when evaluated by this criterion. Using the Iowa data, the unadjusted model had a maturity deviation which fell under the 5 percent mark in four weeks (the September 1 forecast). The standard error adjusted model used with the Iowa data did not reach the 5 percent level until nine weeks of data were included while the logistically adjusted model never reached it. This relatively poor showing by the two adjusted models resulted from the upward trend of the $\hat{\alpha}$ levels already discussed. Again, it becomes important to know if this trend is intrinsic to the procedures used.

When heteroscedasticity was discussed above, it was stated that the violation of this model assumption results in inferior parameter estimates. One should examine how well the adjusted models performed in removing this violation. The figures labeled A-14 and A-16 in the appendix give the sample correlation coefficients of the residuals with time. Under the hypothesis that the correlation is zero, the charts also give the probability that the random variable (the random sample correlation coefficient) will be greater than the computed value of R for this set of data. This probability is computed under the null hypothesis. Thus low values of this probability would correspond to rejection of the null hypothesis while large values would lead to a failure to reject. Using a test with a rejection level of .05, the unadjusted model in Iowa would lead to rejection of the null hypothesis at each cutoff date. The logistically adjusted model would lead to rejection only for the model fit based on six weeks of data. The standard error adjusted model in Iowa did not remove the heteroscedastic error as well as the logistic model. The null hypothesis would be rejected using this model for all cutoff dates greater than five weeks. This would even be the case if the rejection level was lowered to .01. In the Texas data, the two adjusted models performed similarly, each failing to reject the null hypothesis of zero correlation at every cutoff date, with the exception of the fifth week using the logistic adjustment. Here there was a significant negative correlation with time, indicating that the logistic adjustment overcompensated for the changing variance.

ANALYSIS OF DATA SIMULATION

To answer a number of questions raised earlier in this report, it was important to obtain data that was free from characteristics peculiar to a single growing season. Since similar data from distinct growing seasons were limited, the alternative was to construct a set of simulated points from the mathematical model that had been hypothesized to express the accumulation of dry kernel weight over time. This approach has the added advantage of providing a check on the logistic functional form as an appropriate model for dry matter accumulation.

Simulated data points showing dry kernel weight versus time were constructed from the basic logistic model to represent a typical growing season for corn. The data from ten weeks of visits in Iowa were taken as a base for the simulation. The estimated parameters $\hat{\alpha}$, $\hat{\beta}$, $\hat{\rho}$, from the unadjusted non-linear regression on the growth data were chosen for the simulation model. $\hat{E},\hat{F},\hat{G}$, the estimated parameters from the non-linear regression on the absolute value of the residuals were also used to express the dependence of the residuals on time. For each observation in the Iowa data, the independent time variable was plugged in to produce the following simulated dry kernel weight:

$$y_i = \frac{\hat{\alpha}}{1 + \hat{\beta}\hat{\rho}^{t_i}} + v_i(t_i) \qquad \begin{aligned} \hat{\alpha} &= 141.174 \\ \hat{\beta} &= 38.356 \\ \hat{\rho} &= .8924 \end{aligned} \qquad (14)$$

where $v_i(t_i)$ was a normal random variable generated by the SAS random variable generating procedure. This random variable had mean zero and standard deviation:

$$\sigma_{v_i(t_i)} = \frac{\hat{E}}{1 + \hat{F}\hat{G}^{t_i}} \qquad \begin{aligned} \hat{E} &= 33.977 \\ \hat{F} &= 44.777 \\ \hat{G} &= .8904 \end{aligned} \qquad (15)$$

Thus, these simulated data points were generated following the assumptions that were made when applying the logistically adjusted model. A graph of these simulated data points can be found in Figure A-3 in the appendix. A comparison of this data with the 1976 Iowa growth data on Figure A-2 indicates that a "good copy" has been created. This in itself gives additional credibility to the use of the logistically adjusted model to eliminate heteroscedasticity.

The unadjusted logistic model and the two adjusted variations were fitted to the simulated data in the same manner as they were fitted to the original sets of data. The only difference was that the weighting factor determined by silked plants per acre was eliminated because of the nonexistence of between-field differences in the simulated data. An examination of the parameter estimates and errors obtained from the regressions was made.

An upward trend of the $\hat{\alpha}$ forecasts levels over time was observed in 1976 growing season data from both states. Construction of simulated data was first suggested to indicate whether this trend represented only 1976 growing conditions or whether it was a phenomenon likely to occur whenever the logistic model was fitted to growth data at various cutoff dates. If this trend was observed when fitting simulated growth data, it could easily be concluded that the upward trend was intrinsic to the forecasting process. Figure A-11 in the appendix is a graph showing the $\hat{\alpha}$ values for the ten cutoffs. An upward trend is not exhibited by these simulated data points. In fact, the standard error adjusted model has a stable estimate by four weeks with the logistic model giving a very stable estimate by five weeks of data.

When the three models were fitted to the simulated data, the standard relative error of the primary parameter behaved in a similar manner to its behavior when the models were fitted to the 1976 Iowa data. The relative error of the standard error adjusted model remained below that of the other models, but the differences were not great. The error terms came close to the 5 percent level at the fifth week, and then steadily declined (See Figures A-10 and A-12).

The "mature dry kernel weight" of data in the simulation was the $\hat{\alpha}_s = 141.17$ used in the construction of the data points. Maturity deviation for the simulated points was defined to be:

$$\text{Maturity Deviation} = \left| \frac{\hat{\alpha} - \hat{\alpha}_s}{\hat{\alpha}_s} \right| \times 100\% \qquad (16)$$

where $\hat{\alpha}$ is the forecasted "dry kernel weight at maturity" for the cutoff date in question. A table and graph of such deviations appears in the appendix on Figures A-23 and A-24.

These maturity deviations were then used as a method of evaluating the three models when fitted to the simulated data. The results were encouraging. Both adjusted models reached a maturity deviation of about 5 percent or lower after only three weeks of data. The deviation of the logistic model rose to over 10 percent at the four week cutoff, but the maturity deviations of both adjusted models remained below 4 percent for the remaining cutoff dates.

Figure A-18 shows the effect that the two adjusted models had on removing the heteroscedastic error built into the simulated data points. Since the standard deviation of these points was constructed to be a logistic function of time, it was expected that the logistically adjusted model would do the better job of removing the dependence on time. This was not the case. The logistically adjusted model produced residuals significantly (at the 5 percent level) correlated with time at cutoff dates of six weeks, nine weeks and ten weeks. The residuals from the regression involving the standard error model were never significantly different from zero (at the 5 percent level).

LOGISTIC SURVIVAL MODEL

During the process of fitting data for the logistic growth model, the data points with zero growth were deleted. Most of these were termed "non-survivors" and were used to forecast a survival ratio to help in the determination of surviving plants per acre.

If a plant was sampled more than six days after silking and no kernel formation was discerned, then the assumption was made that growth would not occur--the plant had failed to survive. The survival variable for such plants was set to zero, while plants with at least one ear showing growth had the survival variable equal to one.

If a plant was sampled within six days after its silking date and no growth was then present, it was not possible to accurately determine whether no growth would occur in the future or whether normal growth had not yet commenced. For this reason, data from plants sampled six or fewer days after silking were deleted from the survival analysis.

As in the logistic growth regression procedures, individual plant data were replaced by means taken over plants in the same field that were sampled on the same date.

The logistic survival model is of the form

$$s_i = \frac{\gamma}{1 + \delta Q^{t_i}} + \mu_i \qquad i = 1,2, \ldots, n \qquad (17)$$

where γ, δ, Q are parameters

$0<\gamma,Q<1$

$\delta<0$

μ_i = disturbance term

t_i = independent variable

s_i = dependent variable

The independent time variable is still time since silking, but in the survival model, t is referred to as "survival time". "Survival time" was often not equal to "time since silking" in the growth model since plants included in the each mean could change. The dependent variable, s_i, is the survival ratio, a non-negative number less than one. The weighted Gradient nonlinear regression procedure in SAS was used with weights as in (2). An examination of the residuals failed to discover a heteroscedastic error present so no adjustment was necessary.

The primary parameter, γ, is the asymptotic value of s_i. It is this survival ratio at maturity that was used to forecast "surviving plants per acre at maturity" from early season estimates of "silked plants per acre".

The survival model parameters estimated for Iowa and Texas using all ten weeks data are given below. The survival model was fitted to the data at various cutoff dates with little change from cutoff to cutoff.

	IOWA	TEXAS
γ	.980	.990
$\hat{\sigma}_{\gamma} / \hat{\gamma}$	.002	.003
δ	-.040	-.040
$\hat{\sigma}_{\delta} / \hat{\delta}$	44.723	35.125
Q	.700	.700
$\hat{\sigma}_{Q} / \hat{Q}$	4.604	4.279

Plots of the survival data for each state after ten weeks are contained in Figures A-25 and A-26.

COMPLETING THE FORECAST

This section will discuss the way the information generated in the logistic growth and survival models was used to forecast corn yields. A copy of the worksheet used is found in the appendix in Table A-27.

Plant population counts made early in the season were expanded to estimate plants per acre. The phenological observation data were used to determine what percentage of these plants had actually silked, so an estimate of silked plants per acre could be made. The survival ratio generated by the logistic survival model is multiplied by this estimate to produce a forecast of plants per acre with grain at maturity. The primary parameter estimate in the logistic growth model, which is the forecast "dry kernel weight per plant at maturity" is adjusted to the standard 15.5 percent moisture weight (The factor .982 was based on a previous study involving paired sampled of mature kernels. One sample underwent the research drying method while the other was dried using the warmer official ovendrying method for determining moisture content.) (IX) The forecast 15.5 percent moisture grain weight per plant at maturity was multiplied by the forecast number of plants per acre with grain at maturity to produce the forecast biological yield. This biological yield was then adjusted by a historically based ratio to account for harvesting loss in the field, producing a forecast of harvested yield.

CONCLUSIONS AND RECOMMENDATIONS

The 1976 Corn Research Project has demonstrated that within-year regression modeling can provide a good corn forecast by October 1 and a reasonable forecast by September 1. A forecast by this method prior to September 1 would not be feasible.

Further research efforts should be made to improve the stability of the September 1 forecast. Results from the regression of the simulated data for a forecast at this point in the growing season indicated that under ideal conditions (no extreme between field differences or major weather deviations during the growing season) a very stable September 1 forecast is possible. Recall that with four weeks of data, the forecast from the simulated data had a deviation from mature dry weight under 5 percent.

Methods to deal with less than ideal situations should be studied. Possibilities include the following:

1. Incorporate weather variables into the model. One way that this could be done is to use "stress-free days since silking" as the independent variable instead of simply "days since silking." This would give a more accurate count of the number of days when actual growth had taken place, but would not, of course, take into account future weather conditions.

2. An attempt should be made to relate growing conditions in the field to changes in dry kernel weight in the laboratory. If such relationships and the associated lag times can be estimated, they could be used to improve the earlier forecast indications.

3. The possibility of between field differences has not been addressed by this research. Instead, data from all fields have been put together in a single model. The possibility of running more than one model per state should be explored. Field data could be aggregated based on criteria using weather conditions, agricultural practices, or possible combinations of these.

A great deal of attention in this research report has been given to developing and comparing variations of the logistic model that would alleviate the heteroscedastic error. The standard error model performed slightly better than the logistically adjusted model when evaluated by several of the criteria. However, the difference between the performances was not great enough to justify abandonment of the logistically adjusted model. It is recommended that both models be applied to data from the coming season, reevaluated, and a weighted average between the two be found based on their performances over the 1975, 1976 and 1977 growing seasons. In developing forecast indications for the 1977 growing season for use by the State Statistical Offices of involved corn states, a single model should be used for every forecast during the season. In this way, changes in forecast levels would indicate only changes occurring in the fields and not differences between models. If a combination model has not been developed for use by September 1, it is recommended that the standard adjusted model be used.

Because of the labor intensive nature of the data collection procedures, efforts should continue to look into modifications of these procedures that could reduce data collection and processing manhours. Possibilities include sampling fewer plants per visit or fewer kernel rows per ear. An evaluation of double sampling techniques using in-field measurements is already underway.

REFERENCES

(I) Goldfeld, Stephen M. and Quandt, Richard E. 1972 Nonlinear methods in econometrics. North-Holland Publishing Company. Amsterdam, London.

(II) Huddleston, H.F. and Wilson, W.W. 1975 Research in yield forecasting. Proceedings from the National Conference of the Statistical Reporting Service. U. S. Department of Agriculture, 74-88.

(III) Nealon, Jack. 1976 The development of within-year forecasting models for spring wheat (other than Durum). Research and Development Branch, Research Division, Statistical Reporting Service, U. S. Department of Agriculture.

(IV) Nealon, Jack. 1976 The development of within-year forecasting models for winter wheat. Research and Development Branch, Research Division, Statistical Reporting Service, U. S. Department of Agriculture.

(V) Nealon, Jack. 1976 Within-year spring wheat growth models. Research and Development Branch, Research Division, Statistical Reporting Service, U. S. Department of Agriculture.

(VI) 1976 Corn, Cotton and Soybeans Enumerator's Manual. Crop Reporting Board, Statistical Reporting Service, U. S. Department of Agriculture.

(VII) 1976 Corn Growth Research Project Enumerator's Manual, Research Division, Statistical Reporting Service, U. S. Department of Agriculture.

(VIII) 1976 Corn Growth Research Project Laboratory Manual. Research Division, Statistical Reporting Service, U. S. Department of Agriculture.

(IX) Oven Methods for Determining Moisture Content of Grain and Related Agricultural Commodities. Grain Division, Consumer and Marketing Service, U. S. Department of Agriculture. GR Instruction 916-16.

(X) Rockwell, Dwight A. 1975 Nonlinear estimation. Research and Development Branch, Research Division, Statistical Reporting Service, U. S. Department of Agriculture.

(XI) Rockwell, Dwight A. and Nealon, Jack. 1976 An application of nonlinear estimation. Presented at the Ninth International Biometric Conference. Research and Development Branch, Research Division, Statistical Reporting Service, U. S. Department of Agriculture.

(XII) Wilson, Wendell W. 1974 Preliminary report on the use of time related growth models in forecasting components of corn yield. Research and Development Branch, Research Division, Statistical Reporting Service, U. S. Department of Agriculture.

APPENDIX

MEAN DRY KERNEL WEIGHT VERSUS TIME
TEXAS

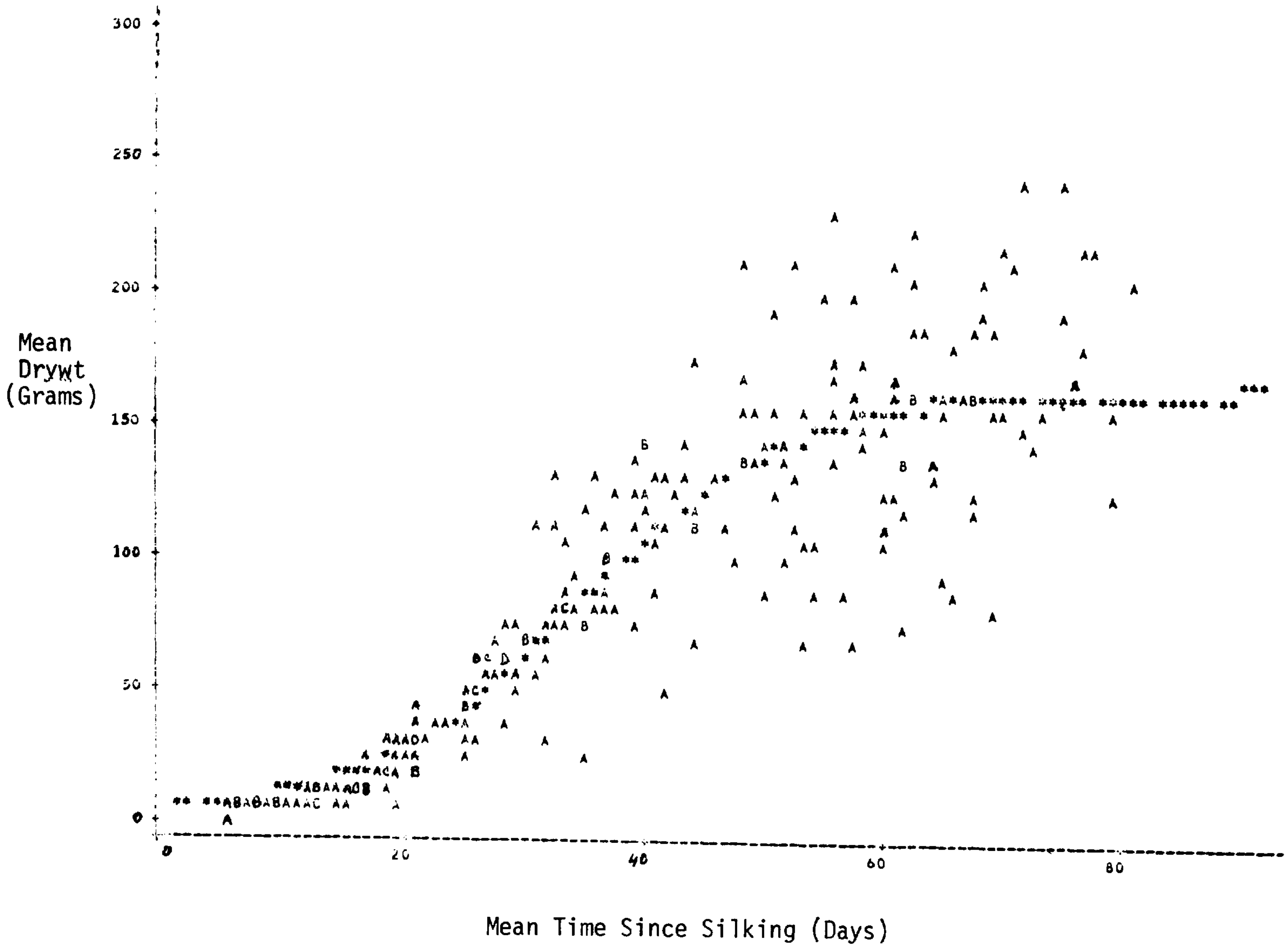


FIGURE A-1

MEAN DRY KERNEL WEIGHT VERSUS TIME

IOWA

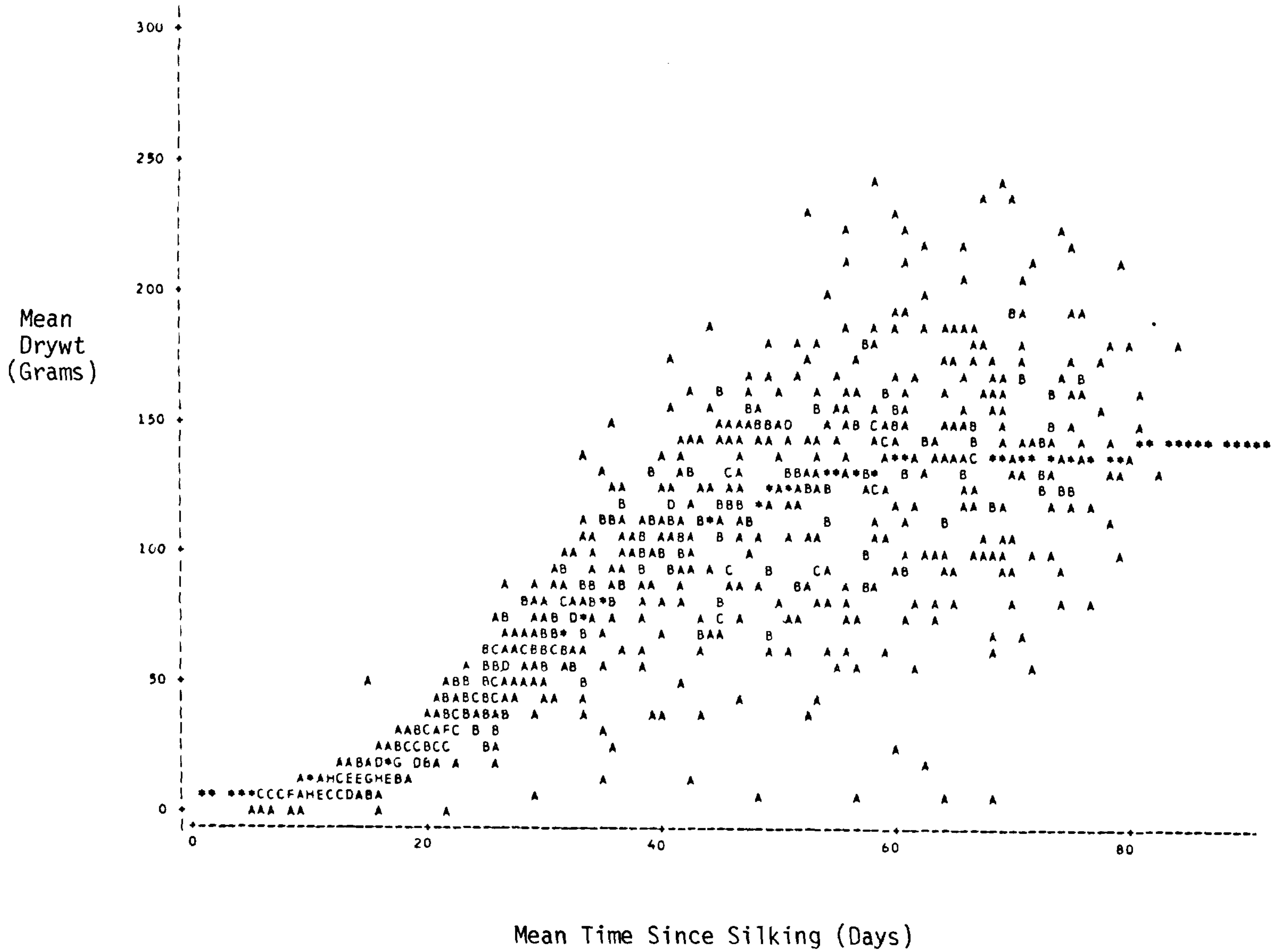


FIGURE A-2

MEAN DRY KERNEL WEIGHT VERSUS TIME
SIMULATED DATA

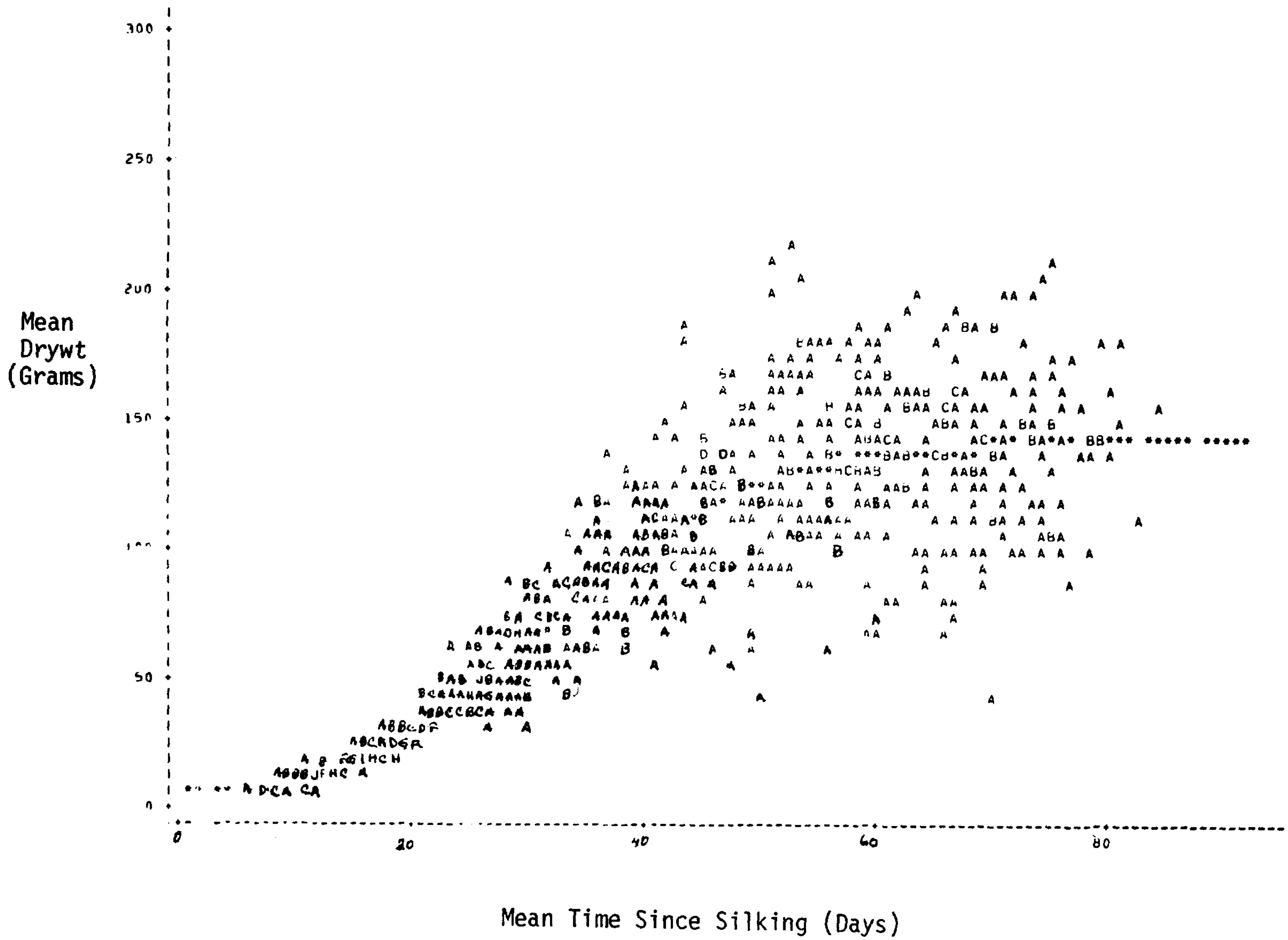


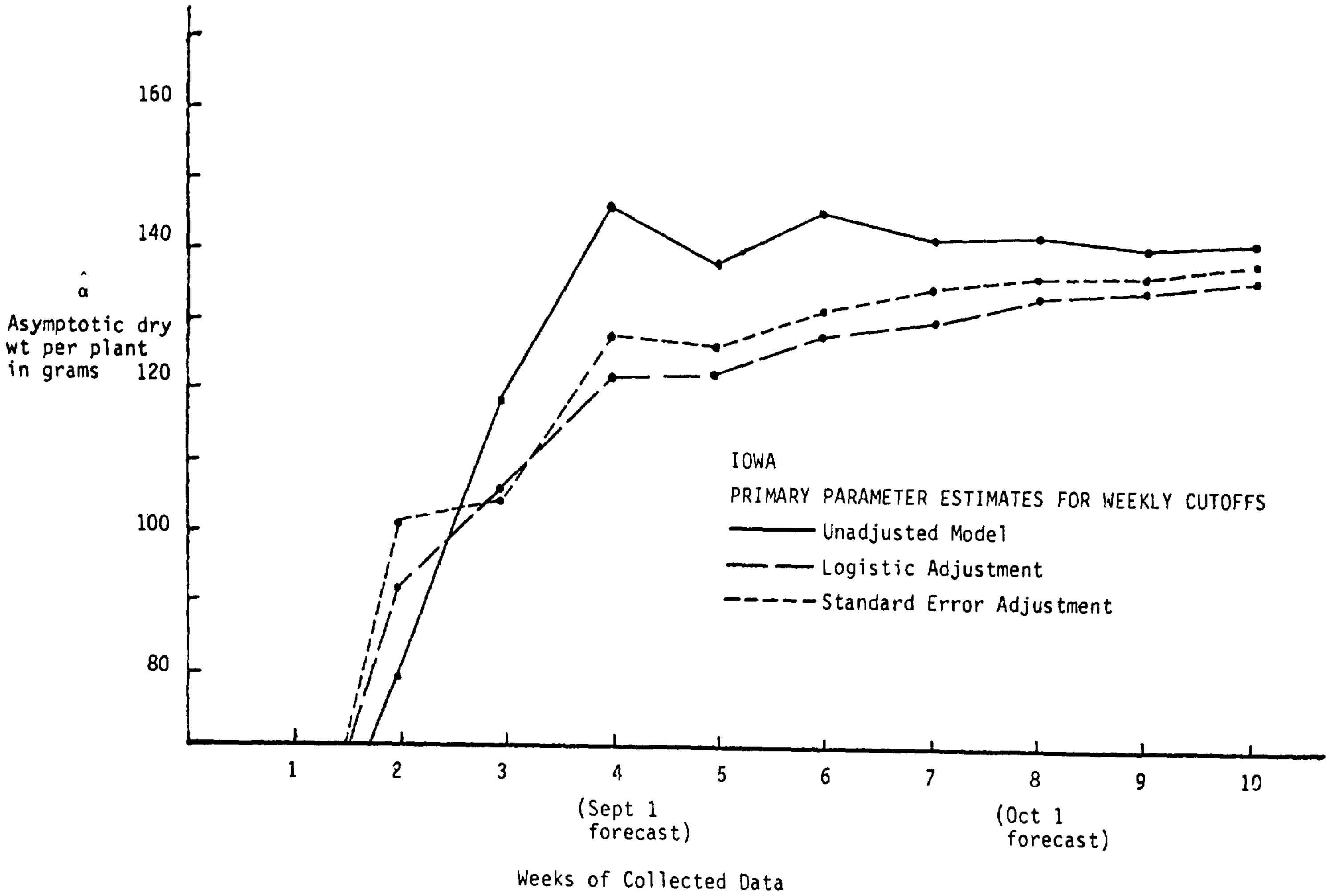
FIGURE A-3

FIGURE A-4

IOWA

PARAMETER ESTIMATES AND RELATIVE ERRORS

WEEKS OF DATA	MODEL	$\hat{\alpha}$	$\sigma_{\hat{\alpha}}/\hat{\alpha}$	$\hat{\beta}$	$\sigma_{\hat{\beta}}/\hat{\beta}$	$\hat{\rho}$	$\sigma_{\hat{\rho}}/\hat{\sigma}$
1 Week	Unadjust	41.2	35.0	44.3	23.1	.808	4.2
	Logistic	45.7	51.5	45.8	37.1	.813	3.1
	St Error	41.5	40.8	38.4	25.7	.817	3.1
2 Weeks	Unadjust	78.7	17.1	84.6	16.4	.826	1.9
	Logistic	91.6	22.9	75.2	16.3	.840	1.2
	St Error	100.2	24.1	77.7	16.4	.845	1.2
3 Weeks	Unadjust	117.8	12.2	73.2	13.6	.858	1.2
	Logistic	106.3	11.2	77.3	8.8	.849	.8
	St Error	104.0	10.0	73.1	7.6	.851	.7
4 Weeks	Unadjust	146.7	9.2	67.7	14.4	.874	>.1
	Logistic	122.9	7.4	81.0	7.2	.856	.6
	St Error	128.6	4.8	79.1	6.0	.861	.5
5 Weeks	Unadjust	137.6	5.3	57.5	18.1	.877	.9
	Logistic	122.1	4.5	78.3	7.2	.858	.5
	St Error	126.0	4.3	76.3	6.5	.861	.5
6 Weeks	Unadjust	145.0	4.5	46.8	18.5	.887	.8
	Logistic	127.7	3.3	75.9	7.4	.862	.5
	St Error	131.4	3.2	72.8	7.2	.866	.4
7 Weeks	Unadjust	141.0	3.0	43.8	19.5	.888	.8
	Logistic	129.6	2.6	75.3	7.1	.863	.4
	St Error	134.5	2.3	70.1	7.2	.869	.4
8 Weeks	Unadjust	142.4	2.5	39.7	20.2	.892	.8
	Logistic	132.6	2.2	73.9	7.1	.866	.4
	St Error	136.4	2.0	67.3	7.6	.872	.4
9 Weeks	Unadjust	140.3	2.0	40.2	21.7	.890	.8
	Logistic	133.3	2.0	73.3	7.1	.866	.4
	St Error	136.8	1.8	67.1	7.4	.872	.3
10 Weeks	Unadjust	141.2	1.8	38.4	21.5	.892	.7
	Logistic	135.0	1.8	73.4	6.9	.867	.4
	St Error	138.0	1.6	65.6	7.6	.873	.4



160
140
120
100
80
$\hat{\alpha}$
Asymptotic dry wt per plant in grams
1
2
3
4
5
6
7
8
9
10
(Sept 1 forecast)
(Oct 1 forecast)
Weeks of Collected Data
IOWA
PRIMARY PARAMETER ESTIMATES FOR WEEKLY CUTOFFS
Unadjusted Model
Logistic Adjustment
Standard Error Adjustment

FIGURE A-5

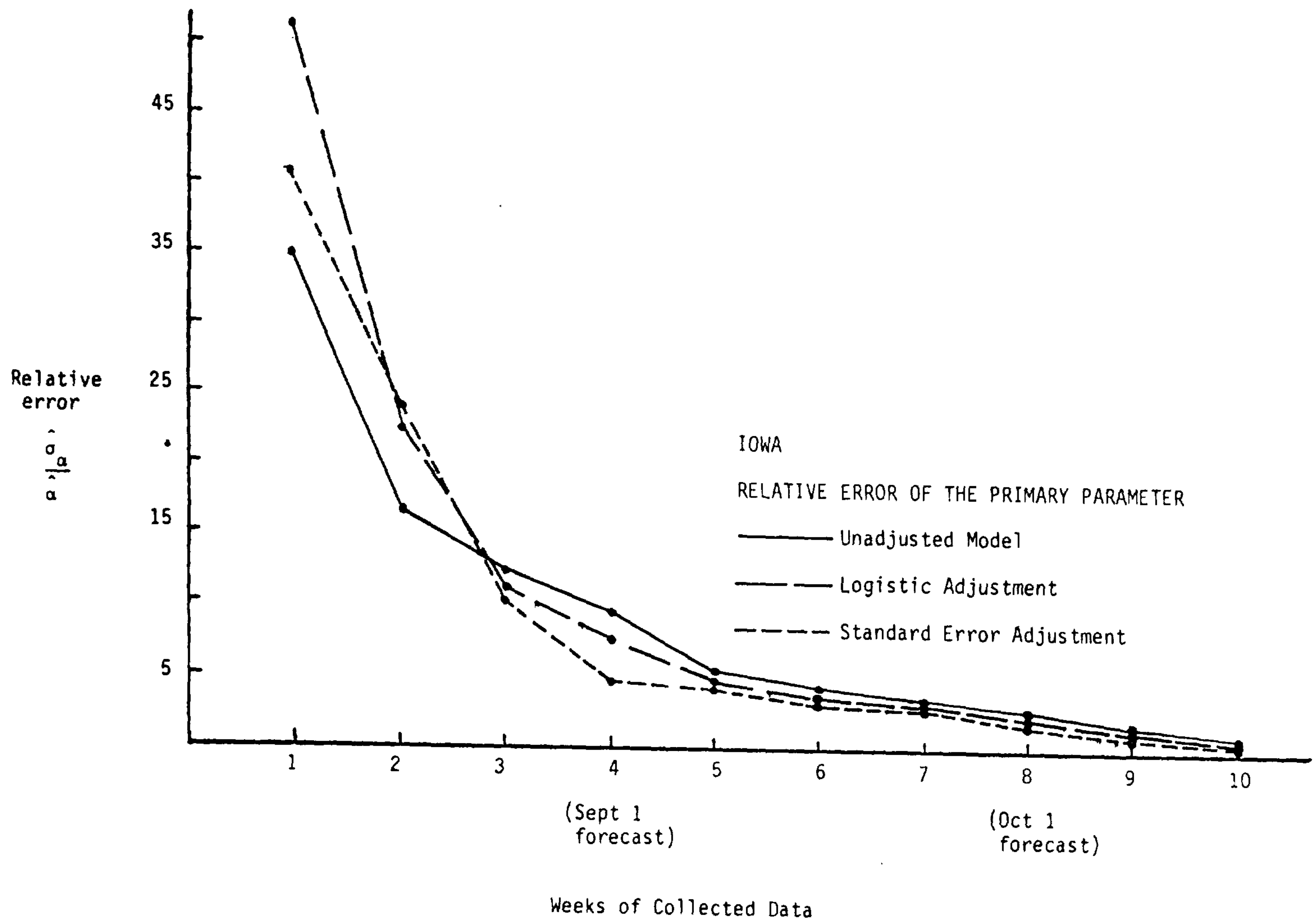

IOWA
RELATIVE ERROR OF THE PRIMARY PARAMETER
Unadjusted Model
Logistic Adjustment
Standard Error Adjustment
Relative error
$\hat{\sigma}_{\alpha}/\hat{\alpha}$
45
35
25
15
5
1
2
3
4
5
6
7
8
9
10
(Sept 1 forecast)
(Oct 1 forecast)
Weeks of Collected Data

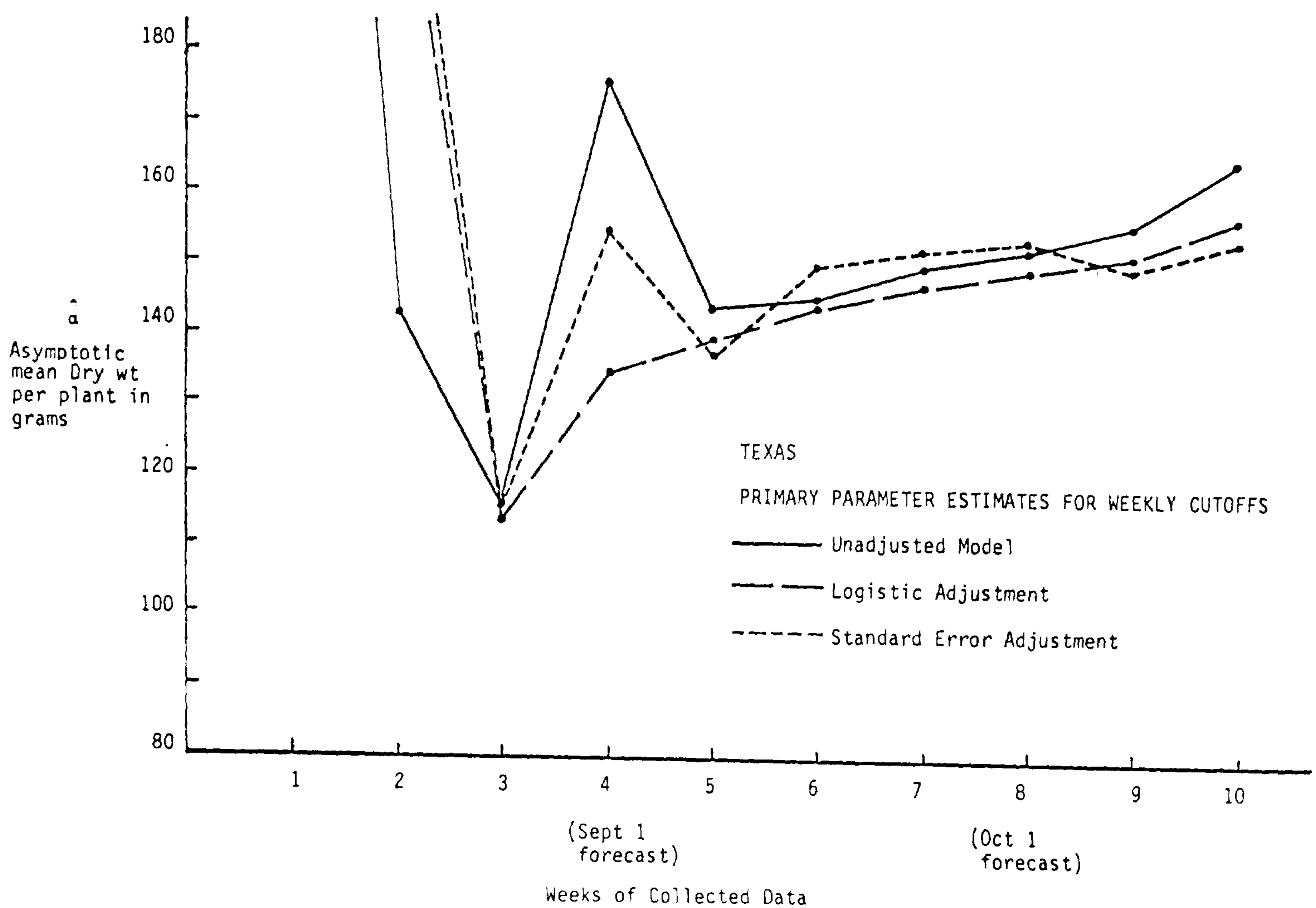
FIGURE A-6

FIGURE A-7

TEXAS

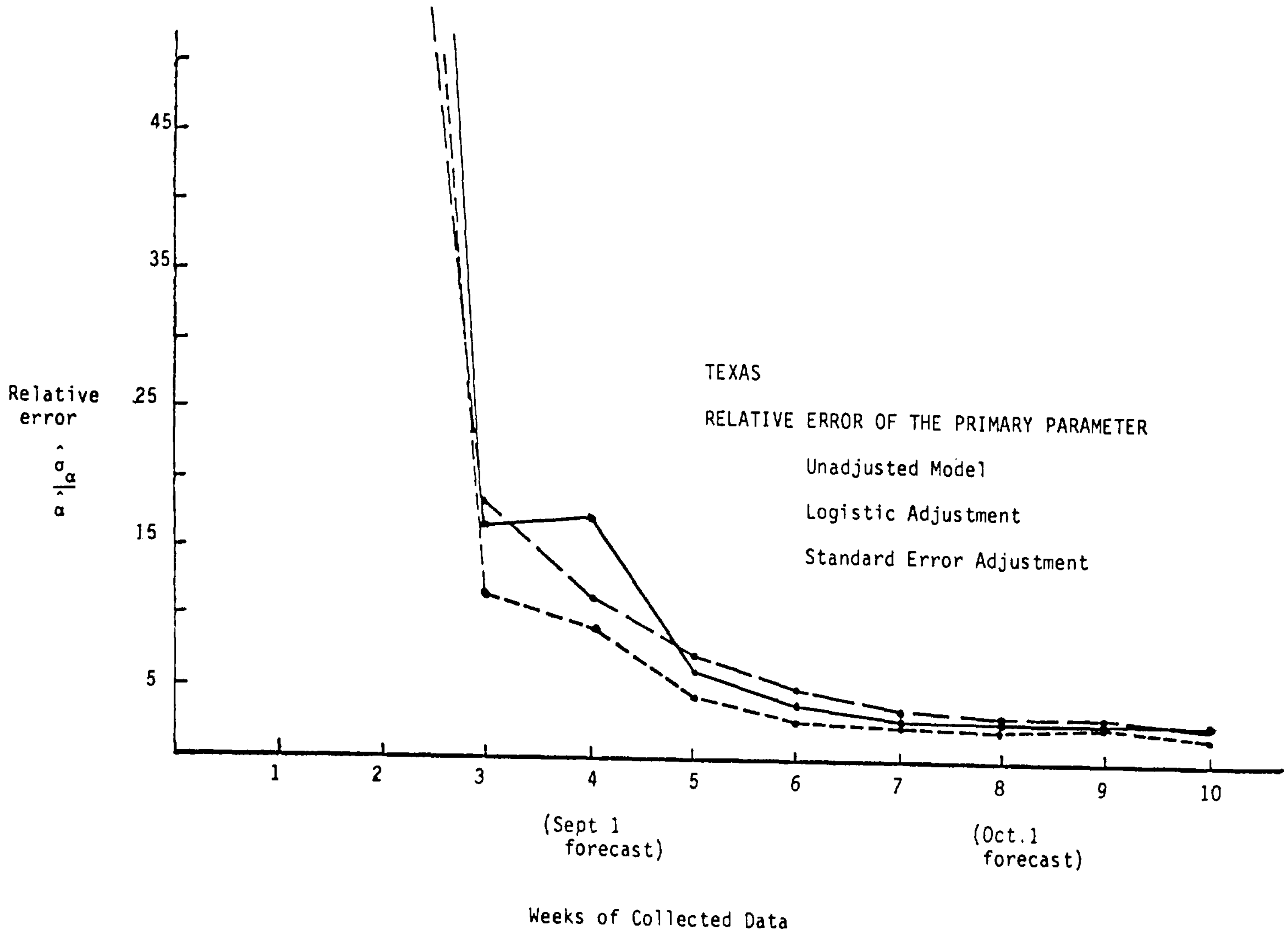
PARAMETER ESTIMATES AND RELATIVE ERRORS

WEEKS OF DATA	MODEL	$\hat{\alpha}$	$\sigma_{\hat{\alpha}}/\hat{\alpha}$	$\hat{\beta}$	$\sigma_{\hat{\beta}}/\hat{\beta}$	$\hat{\rho}$	$\sigma_{\hat{\rho}}/\hat{\rho}$
1 Week	Unadjust						
	Logistic	(Too little data for proper convergence)					
	St Error						
2 Weeks	Unadjust	143.3	90.6	130.3	47.1	.850	3.8
	Logistic	222.3	79.4	155.1	110.9	.864	1.9
	St Error	249.0	86.0	214.1	106.8	.857	1.6
3 Weeks	Unadjust	117.0	17.0	98.0	22.1	.850	1.8
	Logistic	114.8	18.3	91.5	17.0	.850	1.5
	St Error	116.1	11.8	103.3	10.3	.848	.9
4 Weeks	Unadjust	176.9	17.4	81.4	18.6	.879	1.4
	Logistic	136.1	11.4	98.1	12.0	.857	.9
	St Error	156.4	9.2	105.5	10.7	.863	.7
5 Weeks	Unadjust	144.3	6.9	85.3	26.8	.866	1.3
	Logistic	139.1	7.5	93.2	8.9	.859	.7
	St Error	137.6	4.4	101.0	10.9	.858	.6
6 Weeks	Unadjust	146.3	4.1	93.6	32.0	.864	1.3
	Logistic	145.5	5.1	94.8	12.5	.861	.7
	St Error	151.5	2.9	103.9	9.2	.862	.5
7 Weeks	Unadjust	150.6	3.2	82.7	32.4	.870	1.2
	Logistic	147.9	3.8	94.4	11.3	.863	.6
	St Error	153.2	2.7	94.5	9.3	.867	.5
8 Weeks	Unadjust	153.0	2.8	78.5	34.4	.873	1.2
	Logistic	150.0	3.3	93.1	11.6	.865	.6
	St Error	154.3	2.4	91.7	9.4	.869	.5
9 Weeks	Unadjust	156.7	2.8	62.7	35.9	.882	1.2
	Logistic	152.3	3.2	88.5	12.6	.868	.6
	St Error	150.3	2.6	86.9	10.5	.869	.5
10 Weeks	Unadjust	166.0	2.9	43.0	30.9	.897	1.0
	Logistic	157.7	2.9	83.9	12.8	.872	.6
	St Error	154.8	2.4	73.0	12.5	.876	.6



180
160
140
120
100
80
$\hat{a}$
Asymptotic mean Dry wt per plant in grams
1
2
3
4
5
6
7
8
9
10
(Sept 1 forecast)
(Oct 1 forecast)
Weeks of Collected Data
TEXAS
PRIMARY PARAMETER ESTIMATES FOR WEEKLY CUTOFFS
Unadjusted Model
Logistic Adjustment
Standard Error Adjustment

FIGURE A-8



TEXAS
RELATIVE ERROR OF THE PRIMARY PARAMETER
Unadjusted Model
Logistic Adjustment
Standard Error Adjustment
Relative error
$\hat{\sigma}_\alpha / \hat{\alpha}$
45
35
25
15
5
1
2
3
4
5
6
7
8
9
10
(Sept 1 forecast)
(Oct.1 forecast)
Weeks of Collected Data

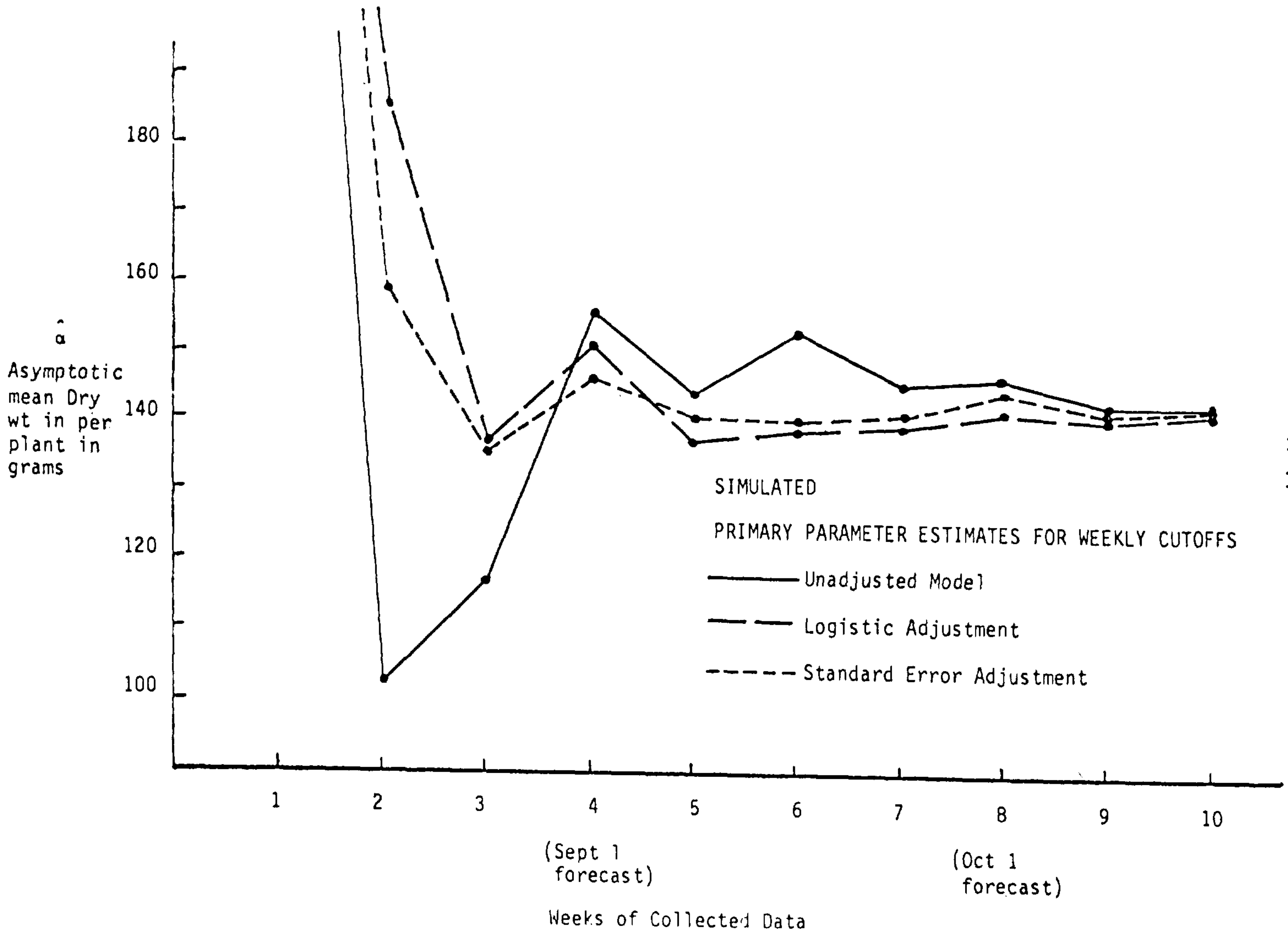
FIGURE A-9

FIGURE A-10

SIMULATED DATA

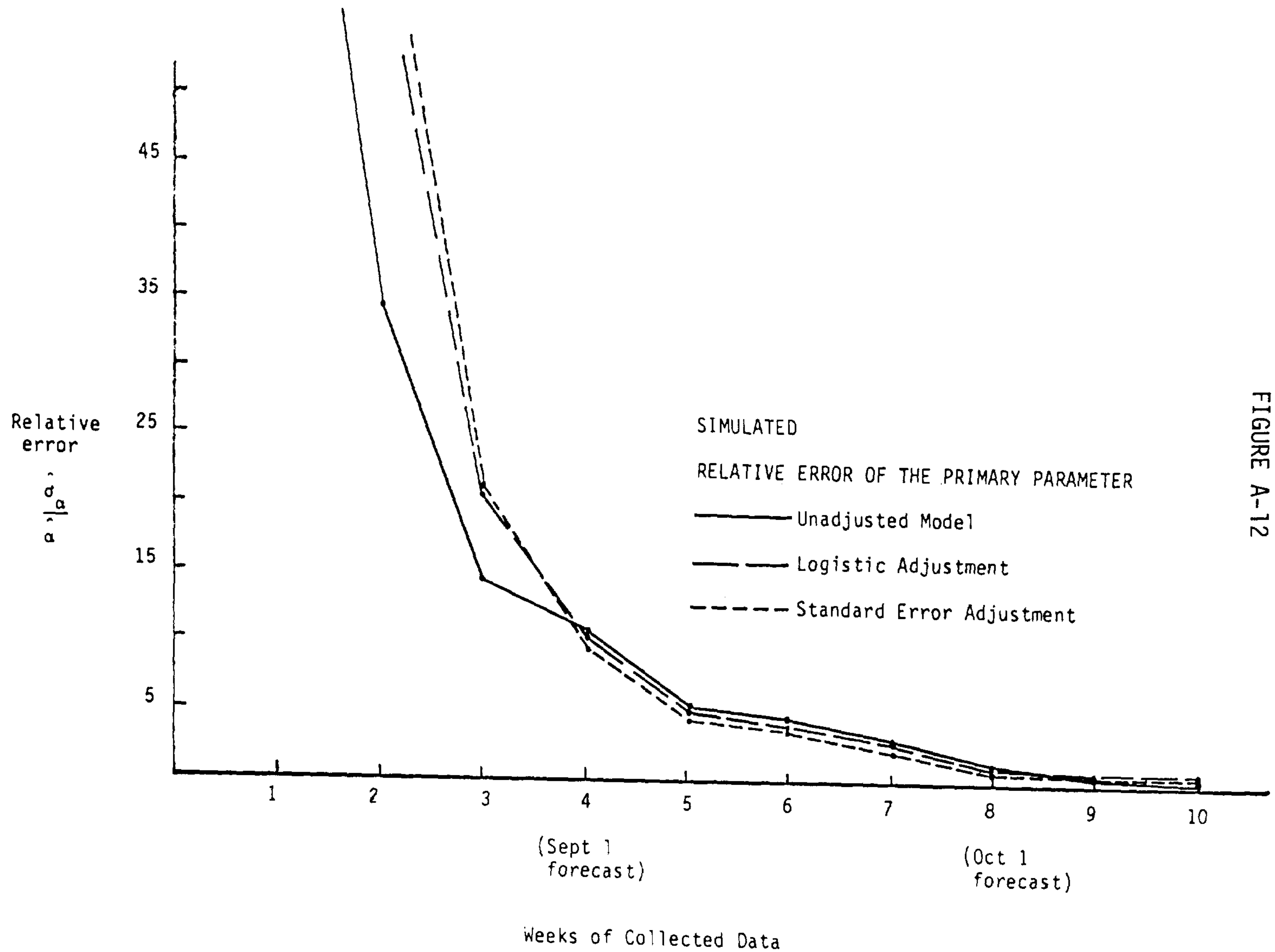
PARAMETER ESTIMATES AND RELATIVE ERRORS

WEEKS OF DATA	MODEL	$\hat{\alpha}$	$\sigma_{\hat{\alpha}}/\hat{\alpha}$	$\hat{\beta}$	$\sigma_{\hat{\beta}}/\hat{\beta}$	$\hat{\rho}$	$\sigma_{\hat{\rho}}/\hat{\rho}$
1 Week	Unadjust Logistic St Error	(too little data for proper convergence)					
2 Weeks	Unadjust	103.0	34.3	34.5	20.1	.876	1.8
	Logistic	186.6	85.1	53.4	77.5	.891	1.3
	St Error	160.6	68.3	47.1	59.7	.888	1.4
3 Weeks	Unadjust	117.0	14.4	38.4	10.5	.879	1.1
	Logistic	137.4	20.7	40.5	15.8	.867	.7
	St Error	136.0	21.1	40.9	15.9	.886	.7
4 Weeks	Unadjust	157.2	10.3	42.6	9.1	.893	.7
	Logistic	152.7	10.2	43.4	7.6	.890	.4
	St Error	148.0	9.2	42.5	6.8	.889	.4
5 Weeks	Unadjust	144.5	5.5	37.6	11.7	.893	.7
	Logistic	137.9	5.2	40.3	4.8	.888	.4
	St Error	141.4	4.8	40.6	4.5	.889	.3
6 Weeks	Unadjust	153.7	4.8	33.5	13.1	.901	.7
	Logistic	139.1	3.9	40.1	5.3	.888	.4
	St Error	141.3	3.6	40.0	4.4	.889	.3
7 Weeks	Unadjust	146.5	3.0	34.5	15.1	.897	.6
	Logistic	140.2	2.8	40.2	4.7	.889	.3
	St Error	142.1	2.7	39.9	4.2	.890	.2
8 Weeks	Unadjust	147.8	2.1	33.5	14.5	.898	.6
	Logistic	143.3	2.1	40.5	3.8	.890	.2
	St Error	144.7	1.8	39.8	4.2	.981	.2
9 Weeks	Unadjust	142.8	1.6	36.9	16.0	.893	.6
	Logistic	141.3	1.7	40.4	3.7	.889	.2
	St Error	141.5	1.6	40.2	4.1	.889	.2
10 Weeks	Unadjust	142.9	1.4	36.9	16.0	.893	.6
	Logistic	141.7	1.5	40.5	3.6	.889	.2
	St Error	142.3	1.4	40.2	4.0	.890	.2



α̂
Asymptotic mean Dry wt in per plant in grams
180
160
140
120
100
SIMULATED
PRIMARY PARAMETER ESTIMATES FOR WEEKLY CUTOFFS
Unadjusted Model
Logistic Adjustment
Standard Error Adjustment
1
2
3
4
5
6
7
8
9
10
(Sept 1 forecast)
(Oct 1 forecast)
Weeks of Collected Data

FIGURE A-11



SIMULATED
RELATIVE ERROR OF THE PRIMARY PARAMETER
Unadjusted Model
Logistic Adjustment
Standard Error Adjustment
Relative error
$\hat{\sigma}_{\alpha} / \hat{\alpha}$
45
35
25
15
5
1
2
3
4
5
6
7
8
9
10
(Sept 1 forecast)
(Oct 1 forecast)
Weeks of Collected Data

FIGURE A-12

REGRESSION RESIDUALS VERSUS TIME

IOWA

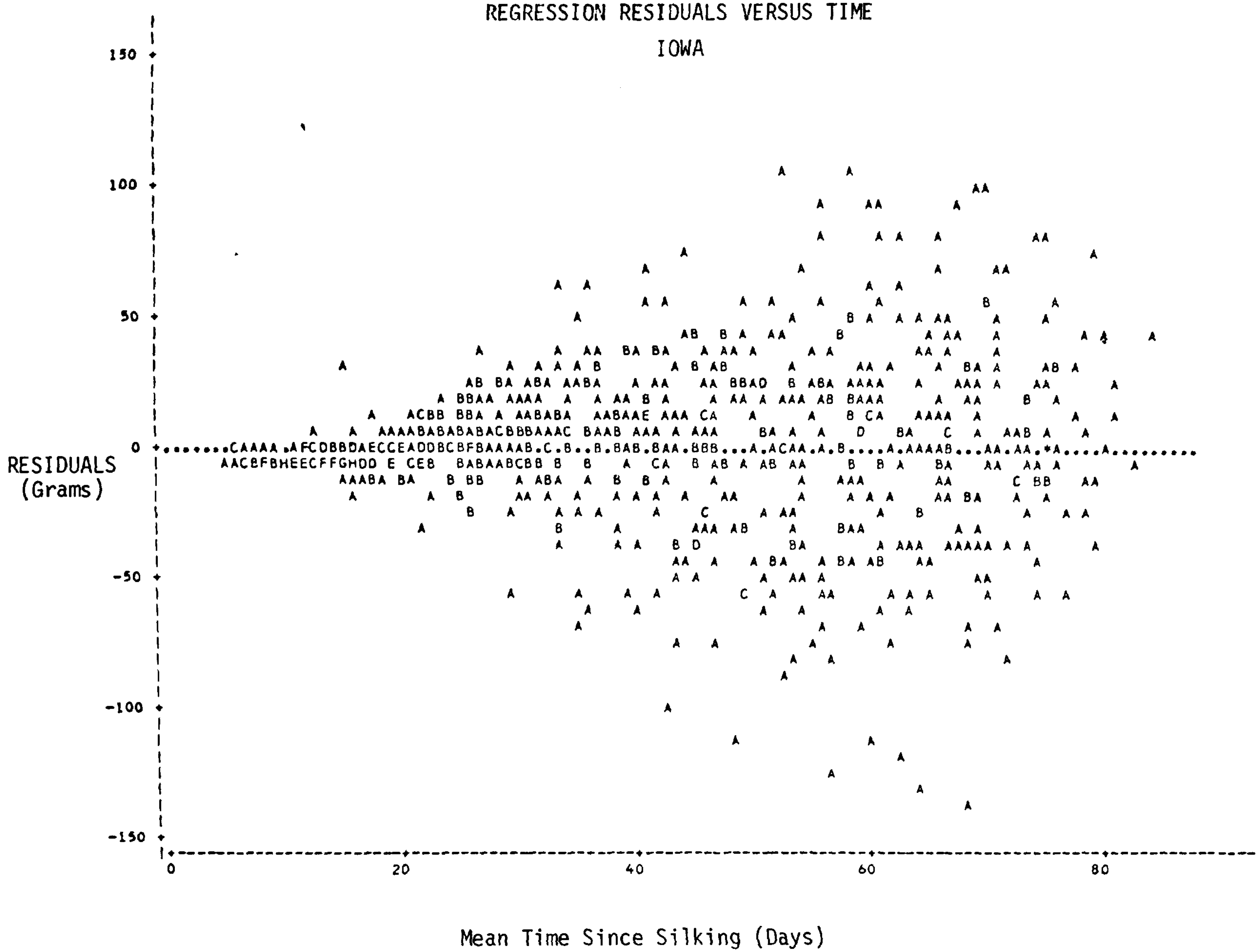


FIGURE A-13

FIGURE A-14

CORRELATION OF REGRESSION RESIDUALS WITH TIME

IOWA

Weeks Of Data	Unadjusted Model		Logistic Adjustment		St Error Adjustment	
	R	Prob>|R|	R	Prob>|R|	R	Prob>|R|
1	.59	.0001	-.10	.3654	.08	.5095
2	.52	.0001	-.03	.7481	.10	.2371
3	.47	.0001	.03	.6639	.03	.6239
4	.48	.0001	.01	.8479	.07	.2140
5	.51	.0001	.05	.3200	.08	.0936
6	.52	.0001	.09	.0411	.13	.0034
7	.51	.0001	.08	.0737	.13	.0021
8	.50	.0001	.07	.0523	.13	.0007
9	.49	.0001	.07	.0724	.12	.0014
10	.46	.0001	.05	.1879	.11	.0023

REGRESSION RESIDUALS VERSUS TIME

TEXAS

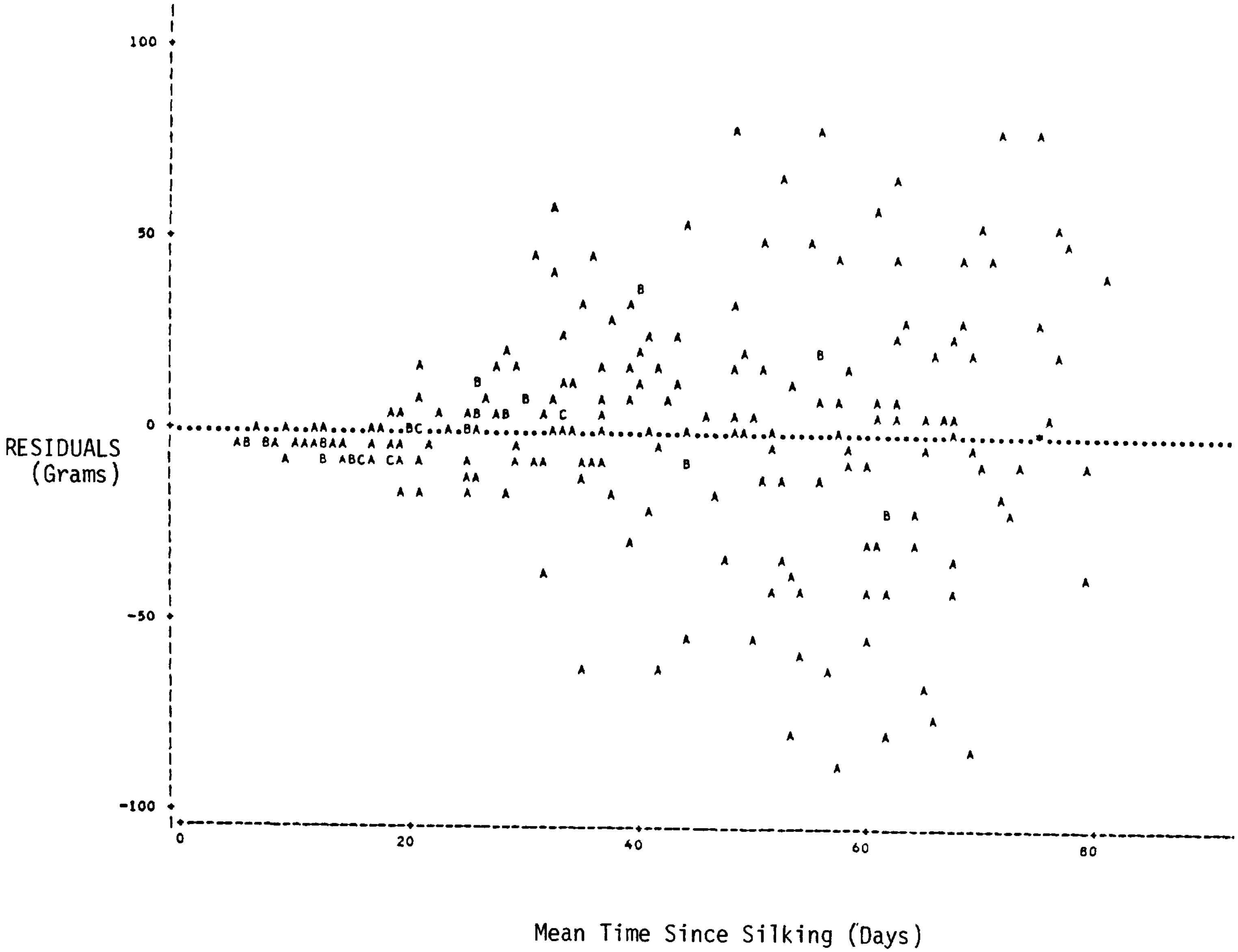


FIGURE A-15

FIGURE A-16

CORRELATION OF REGRESSION RESIDUALS WITH TIME

TEXAS

Weeks Of Data	Unadjusted Model		Logistic Adjustment		St Error Adjustment	
	R	Prob>|R|	R	Prob>|R|	R	Prob>|R|
1	.34	.1579	.41	.0767	-.13	.6036
2	.45	.0028	.03	.8703	-.05	.7443
3	.37	.0026	.17	.1749	-.06	.6269
4	.46	.0001	.09	.4146	.08	.4589
5	.44	.0001	-.22	.0190	.02	.8468
6	.50	.0001	.09	.3131	>.01	.9624
7	.50	.0001	.04	.6541	>.01	.9874
8	.50	.0001	.05	.5043	>.01	.9942
9	.53	.0001	.08	.2196	.05	.4885
10	.49	.0001	.12	.0665	.13	.0609

ABSOLUTE VALUE OF REGRESSION RESIDUALS VERSUS TIME

IOWA

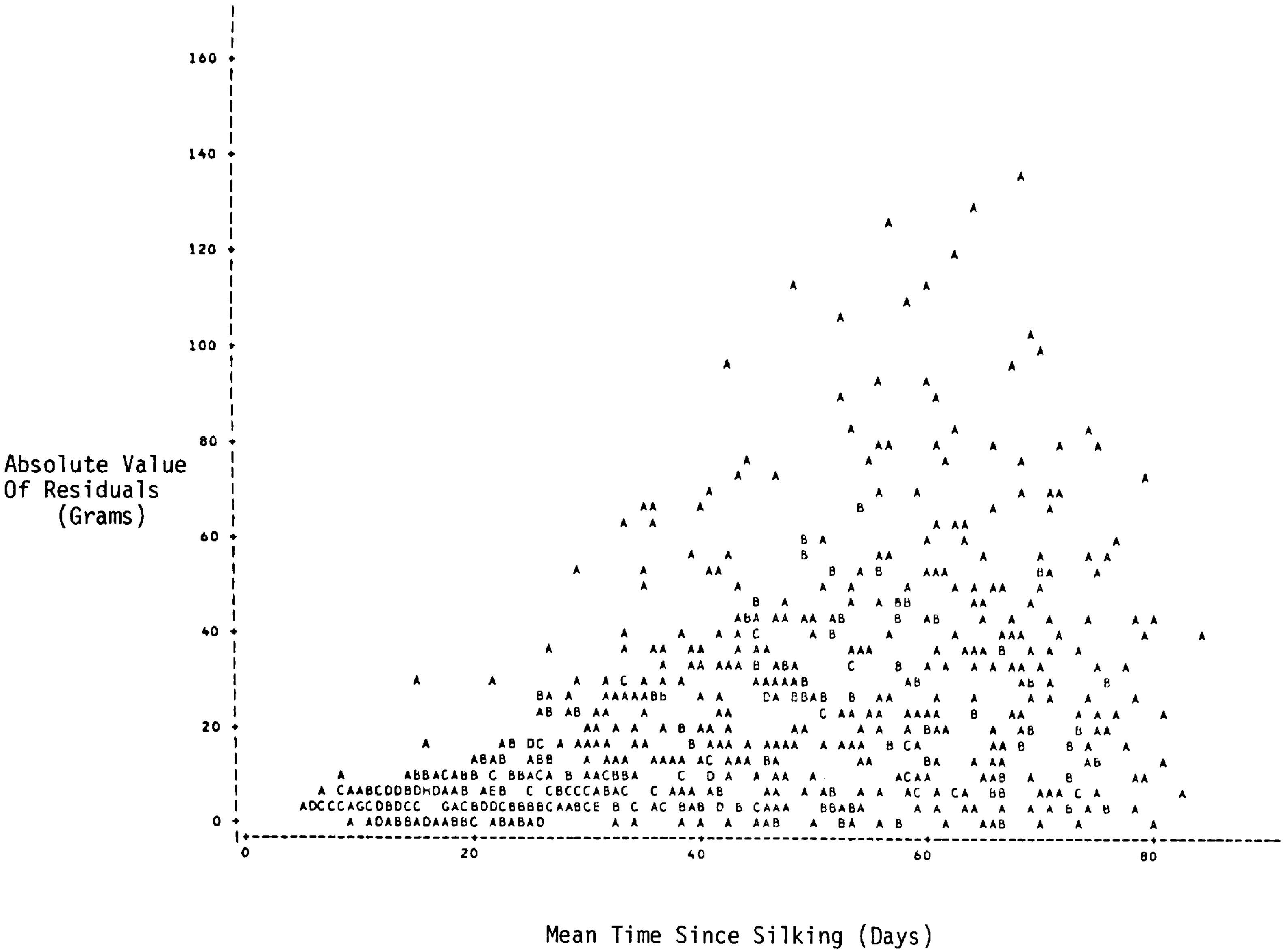


FIGURE A-17

FIGURE A-18

CORRELATION OF REGRESSION RESIDUALS WITH TIME

SIMULATED DATA

Weeks Of Data	Unadjusted Model		Logistic Adjustment		St Error Adjustment	
	R	Prob>|R|	R	Prob>|R|	R	Prob>|R|
1	.48	.0001	-.18	.1211	.09	.4400
2	.56	.0001	.02	.7683	.07	.3681
3	.54	.0001	.04	.5905	.03	.6156
4	.57	.0001	-.02	.6789	.03	.6475
5	.56	.0001	.05	.2745	.03	.6195
6	.57	.0001	.13	.0045	.06	.1649
7	.59	.0001	.06	.1576	.04	.3115
8	.52	.0001	-.07	.0816	.05	.2431
9	.50	.0001	.08	.0449	-.003	.9370
10	.48	.0001	-.09	.0096	-.002	.9612

FIGURE A-19

DEVIATION FROM MEAN DRY WEIGHT AT MATURITY

IOWA

A mature plant was defined as one where time since silking was greater than sixty days.

Mean Dry Weight at Maturity
144.15 grams per plant

Weeks Of Data	Model	Deviation	% of Absolute Deviation
1	Unadjust	-102.95	71.42
	Logistic	- 98.45	68.30
	St Error	-102.65	71.21
2	Unadjust	- 65.45	45.40
	Logistic	- 52.55	36.46
	St Error	- 43.95	30.49
3	Unadjust	- 26.35	18.28
	Logistic	- 37.85	26.26
	St Error	- 40.15	27.85
4	Unadjust	2.55	1.77
	Logistic	- 21.25	14.74
	St Error	- 15.55	10.79
5	Unadjust	- 6.55	4.54
	Logistic	- 22.05	15.30
	St Error	- 18.15	12.59
6	Unadjust	.85	.59
	Logistic	- 16.45	11.41
	St Error	- 12.75	8.84
7	Unadjust	- 3.15	2.19
	Logistic	- 14.55	10.09
	St Error	- 9.65	6.69
8	Unadjust	- 1.75	1.21
	Logistic	- 11.55	8.01
	St Error	- 7.75	5.38
9	Unadjust	- 3.85	2.67
	Logistic	- 10.85	7.53
	St Error	- 7.35	5.10
10	Unadjust	- 2.95	2.05
	Logistic	- 9.15	6.35
	St Error	- 6.15	4.27

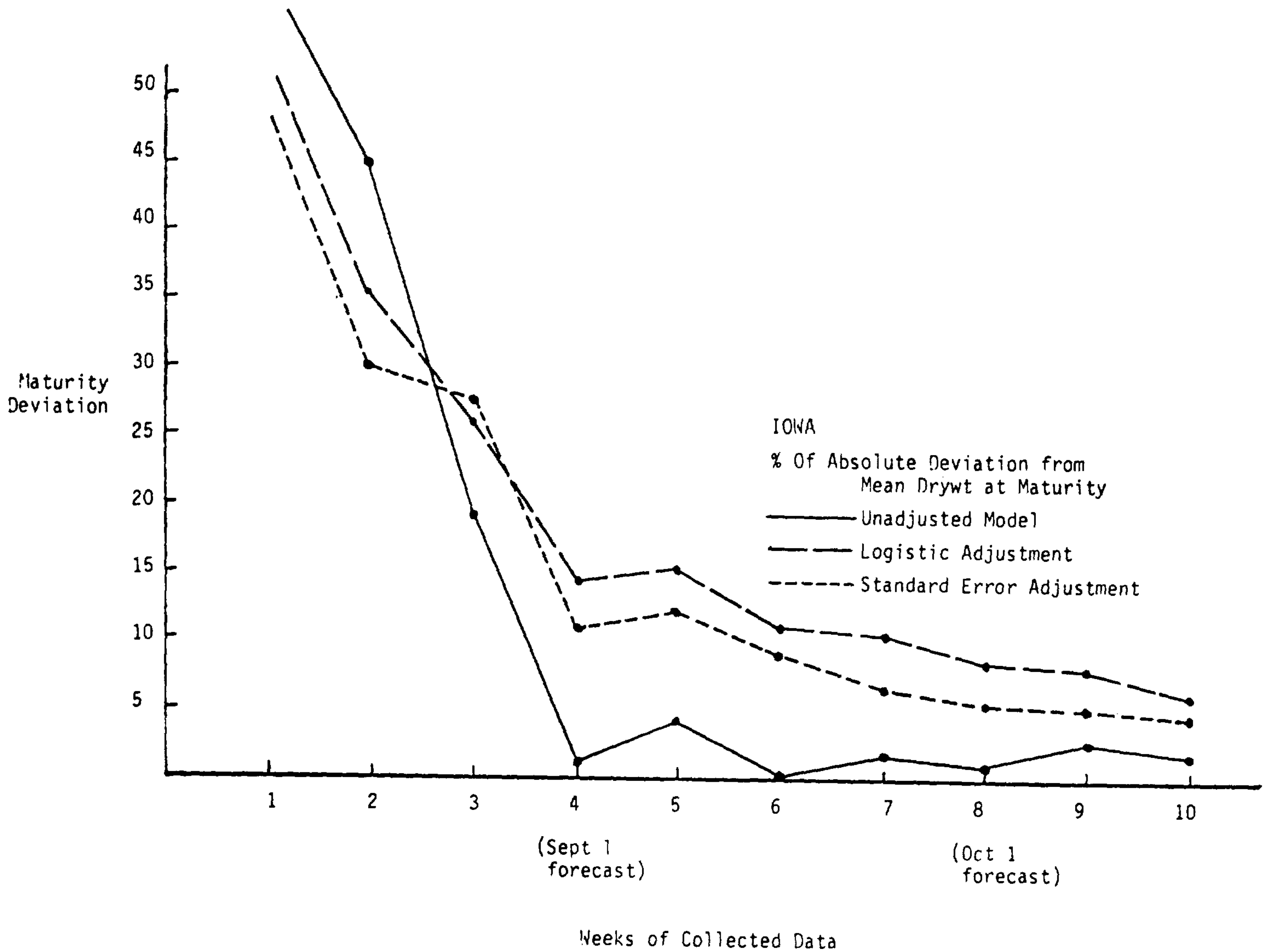

Maturity
Deviation
50
45
40
35
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9
10
(Sept 1
forecast)
(Oct 1
forecast)
Weeks of Collected Data
IOWA
% Of Absolute Deviation from
Mean Drywt at Maturity
Unadjusted Model
Logistic Adjustment
Standard Error Adjustment

FIGURE A-20

FIGURE A-21

DEVIATION FROM MEAN DRY WEIGHT AT MATURITY

TEXAS

A mature plant was defined as one where time since silking was greater than sixty days.

Mean Dry Weight at Maturity

166.95 grams per plant

Weeks Of Data	Model	Deviation	% of Absolute Deviation
1	Unadjust	-	-
	Logistic	-	-
	St Error	-	-
2	Unadjust	23.65	14.17
	Logistic	55.35	33.15
	St Error	82.05	49.14
3	Unadjust	-49.95	29.92
	Logistic	-52.15	31.24
	St Error	-50.85	30.46
4	Unadjust	9.95	5.96
	Logistic	-30.85	18.48
	St Error	-10.55	6.32
5	Unadjust	-22.65	13.57
	Logistic	-27.85	16.68
	St Error	-15.45	9.25
6	Unadjust	-20.65	12.37
	Logistic	-21.45	12.85
	St Error	-15.45	9.25
7	Unadjust	-16.35	9.79
	Logistic	-19.05	11.41
	St Error	-13.75	8.24
8	Unadjust	-13.95	8.36
	Logistic	-16.95	10.15
	St Error	-12.65	7.58
9	Unadjust	-10.25	6.14
	Logistic	-14.65	8.78
	St Error	-16.65	9.97
10	Unadjust	- .95	.57
	Logistic	- 9.25	5.54
	St Error	-12.15	7.28

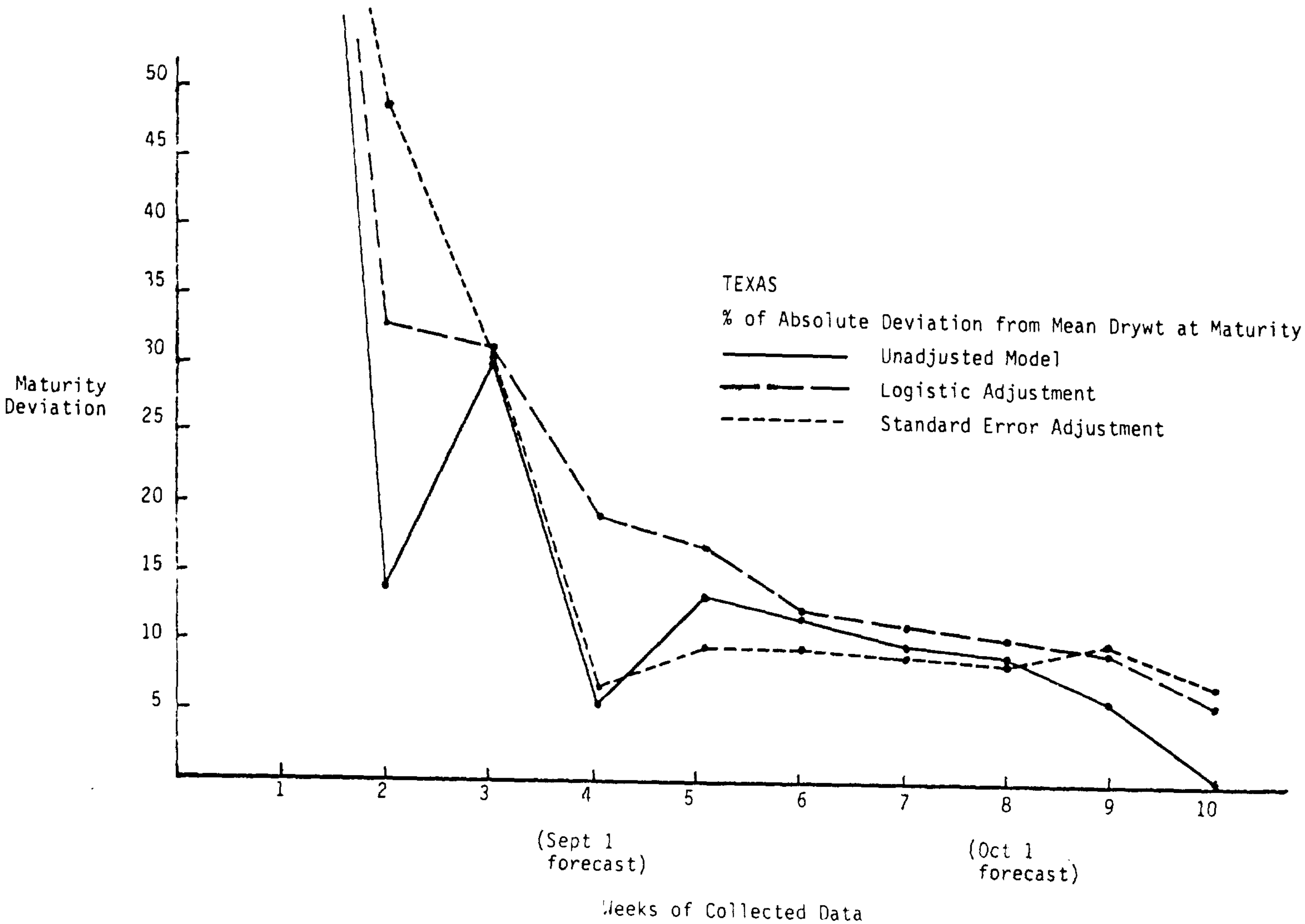

TEXAS
% of Absolute Deviation from Mean Drywt at Maturity
Unadjusted Model
Logistic Adjustment
Standard Error Adjustment
Maturity Deviation
50
45
40
35
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9
10
(Sept 1 forecast)
(Oct 1 forecast)
Weeks of Collected Data

FIGURE A-22

FIGURE A-23

DEVIATION FROM MEAN DRY WEIGHT AT MATURITY

SIMULATED DATA

Asymptotic dry weight at maturity was defined to be $\hat{\alpha}$ value used in constructing the simulated data.

Asymptotic Dry Weight at Maturity

141.17 grams per plant

Weeks Of Data	Model	Deviation	% of Absolute Deviation
1	Unadjust	-	-
	Logistic	-	-
	St Error	-	-
2	Unadjust	-38.17	27.04
	Logistic	45.63	32.32
	St Error	19.43	13.76
3	Unadjust	-24.17	17.12
	Logistic	- 3.77	2.67
	St Error	- 5.17	3.66
4	Unadjust	16.03	11.36
	Logistic	11.53	8.17
	St Error	6.83	4.84
5	Unadjust	3.33	2.36
	Logistic	- 3.27	2.32
	St Error	.23	.16
6	Unadjust	12.53	8.88
	Logistic	- 2.07	1.47
	St Error	.13	.09
7	Unadjust	5.33	3.78
	Logistic	- .97	.69
	St Error	.93	.66
8	Unadjust	6.63	4.70
	Logistic	2.13	1.51
	St Error	3.53	2.50
9	Unadjust	1.63	1.15
	Logistic	.13	.09
	St Error	.33	.23
10	Unadjust	1.73	1.23
	Logistic	.53	.38
	St Error	1.13	.80

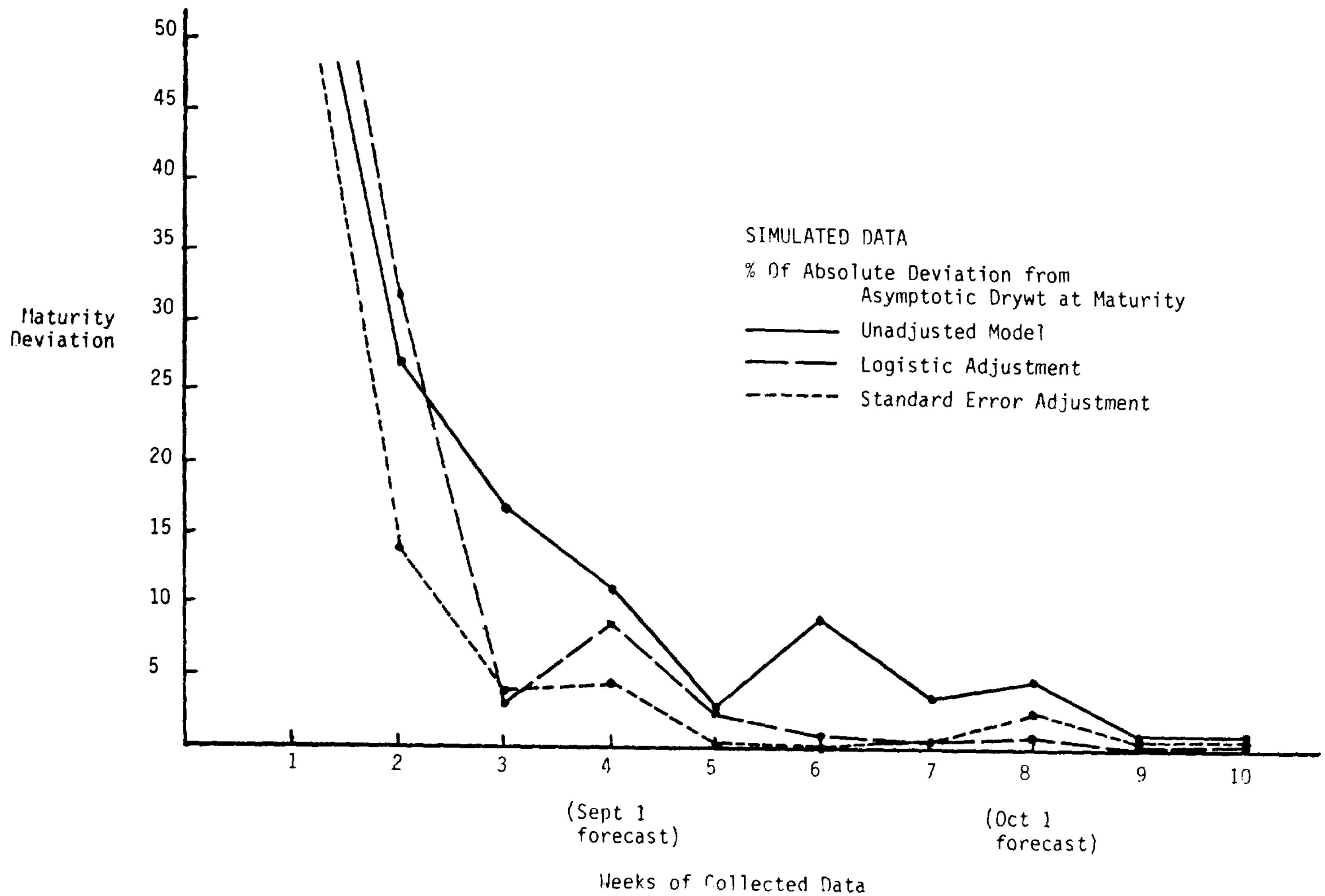

SIMULATED DATA
% Of Absolute Deviation from Asymptotic Drywt at Maturity
Unadjusted Model
Logistic Adjustment
Standard Error Adjustment
Maturity Deviation
50
45
40
35
30
25
20
15
10
5
1
2
3
4
5
6
7
8
9
10
(Sept 1 forecast)
(Oct 1 forecast)
Weeks of Collected Data

FIGURE A-24

SURVIVAL RATIO VERSUS TIME

IOWA

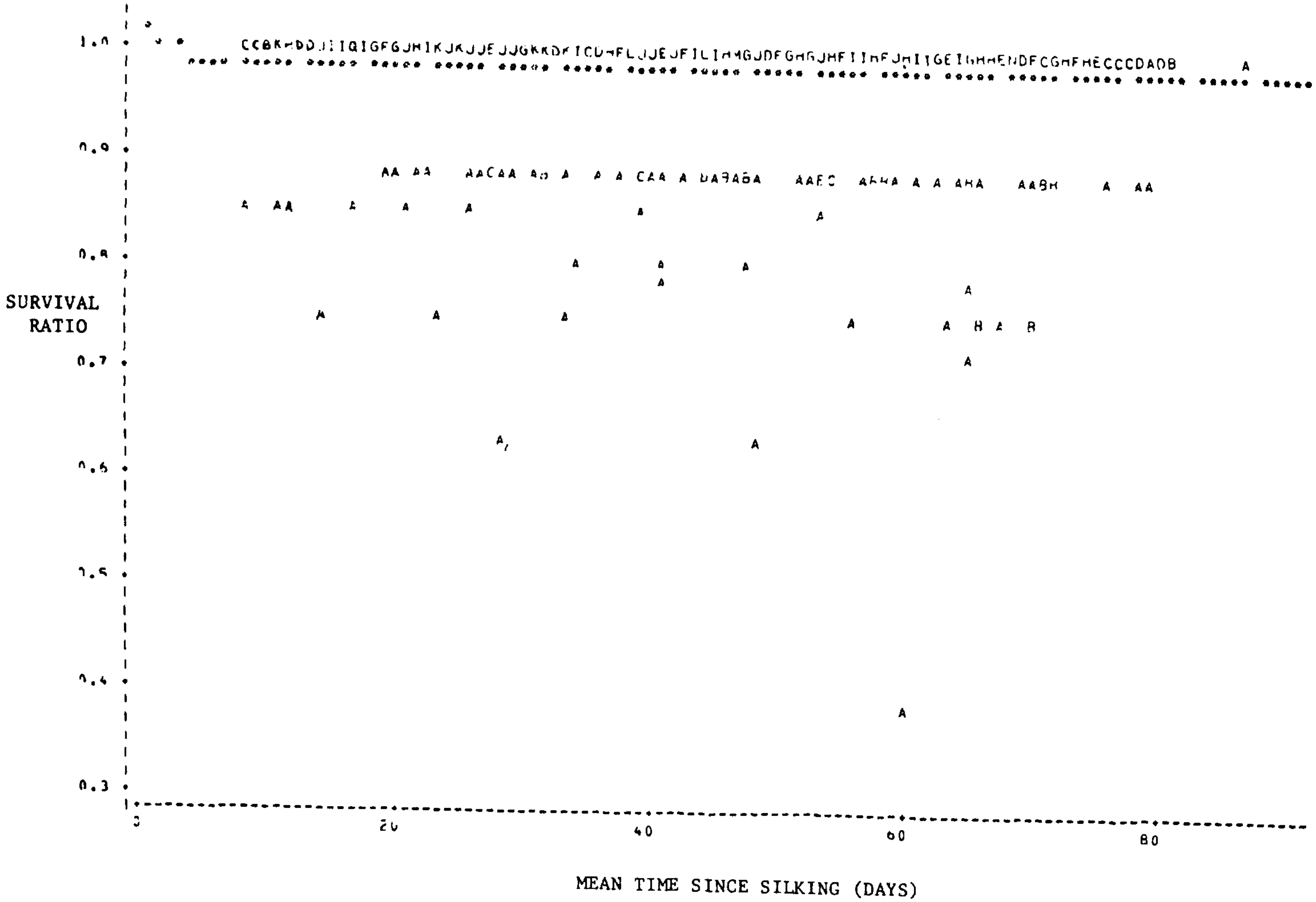


FIGURE A-25

SURVIVAL RATIO VERSUS TIME

TEXAS

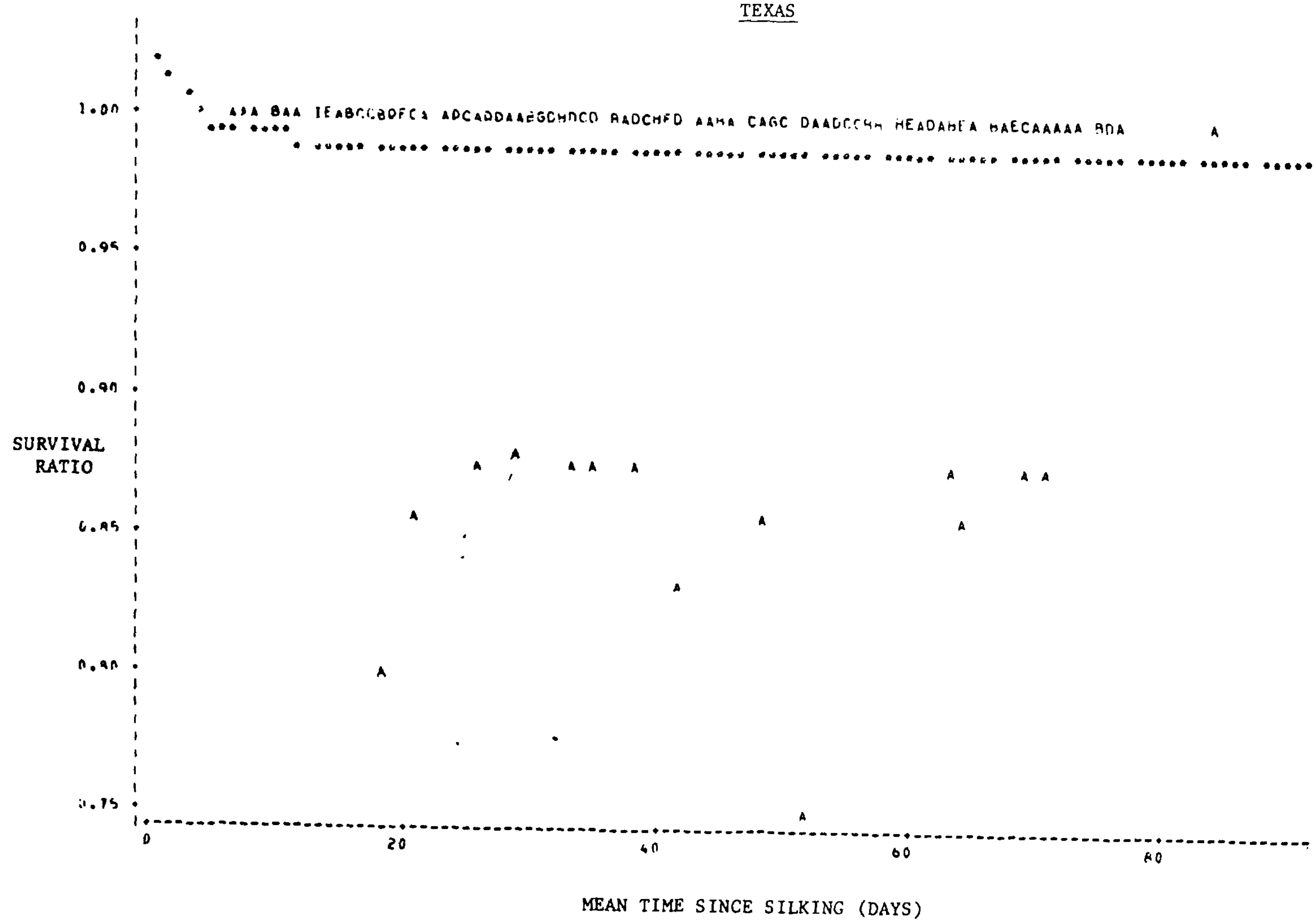


FIGURE A-26

FIGURE A-27

1976 CORN YIELD FORECASTING RESEARCH - WORKSHEET

State or Area ______________________________ Corn for Grain

Number of Sample Fields (for Grain) __________ Forecast Date ________

Plants With Potential to Produce Grain

Item 1 - Total number of plants per acre........................ __________

Item 2 - Relative standard error of plants per acre............. ________ %

Item 3 - Total number of silked plants per acre................. __________

Item 4 - Relative standard error of silked plants per acre...... ________ %

Forecast Number of Plants Per Acre With Grain at Maturity

Item 5 - Forecast ratio of plants in Item 3 surviving with grain at maturity.. __________

Item 6 - Relative error of the primary survival parameter....... ________ %

Item 7 - Forecast number of plants per acre with grain at maturity (Item 3 x Item 5)................................... __________

Forecast Standard Moisture (15.5%) Grain Weight Per Plant At Maturity

Item 8 - Forecast "dry" grain weight per plant at maturity...... __________ grams

Item 9 - Relative error of the primary growth parameter......... ________ %

Item 10 - Adjusted 15.5% moisture grain weight per plant at maturity (Item 8 x .982)/.845................................ __________ grams

(________ lbs.)1/

Forecast Yield Per Acre

Item 11 - Biological Yield per acre (Item 7 x Item 10)/453.59... ________ lbs.

(________ bu.)2/

Item 12 - Forecast harvested yield per acre (Item 11 x   )..... ________ lbs.

(________ bu.)2/

1/ Pounds equivalent is based upon one pound being equal to 453.59 grams.

2/ Pounds are converted to bushels based upon 56 lbs. of 15.5% moisture corn grain equaling one bushel.