WSS TALK 10/27/93

YIELD MODELS FOR CORN AND SOYBEANS BASED ON SURVEY DATA

Thomas R. Birkett, National Agricultural Statistics Service, USDA

Abstract

The National Agricultural Statistics Service uses survey data to forecast yields for major agricultural commodities, including corn and soybeans. The survey data contains variables that become the independent variables in linear forecasting models. This paper describes the forecasting models, showing what the key survey variables are and examining how they are related to final yield.

Introduction

The National Agricultural Statistics Service (NASS), an agency of the United States Department of Agriculture, conducts monthly field surveys in the late summer and fall to forecast corn and soybean yields. Summarized data from the survey forms the independent variables for a statistical model that predicts the current season final average yield. The survey data include variables correlated with the final average number of ears or pods that will be harvested, along with variables correlated with the final average grain weight per ear or weight per pod. This paper gives a short description of these variables and how they are used to forecast final average yield.

Description of the Objective Yield Surveys

In June, NASS conducts a very large survey of agricultural land use in the U.S. to estimate the current season's acreage planted to corn and soybeans. From the base generated by this survey, NASS draws a random sample of corn and soybean plots. This is done through a two stage process, in which fields are selected and then random locations are designated within each selected field. The procedure is carried out so that a simple random sample is obtained, and each planted acre of corn or soybeans has an equal chance of being included in the sample. This simple random sample property is an important assumption for the statistical models to be applied to the survey data.

The randomly located plots are a few square feet in area. Within the plots, enumerators count and measure variables that are positively correlated with final yield. Among the variables collected for soybeans are number of plants per acre, number of nodes per plant, number of lateral branches per plant, number of blooms, dried flowers and pods per plant, and number of pods with beans per plant. For corn the NASS enumerators count the number of stalks per acre, number of stalks with ears, number of ear shoots, and number of ears with kernels per acre. They also husk a random sample of ears near the plot and measure the length of a typical kernel row on each ear. Just prior to farmer harvest of the corn or soybean field in which the sample is located, the enumerator harvests the plot and obtains the final yield. The same sample plots are revisited each month starting in August until farmer harvest.

Samples are laid out in all the major corn and soybean producing states. Data are collected during the period from the 21st of the previous month until the first of the month. Starting in August and continuing through November, around the 10th of each month the USDA releases yield estimates for each state based on the survey.

Variables in the Regional Models

The best relationship between the survey data and final yield is found at the regional level, the region being the set of states in the survey. Consequently, the plot level data is summarized to the state and then to the region level, where it is modeled against the region yield. Each monthly regional model normally has one independent variable X.

The form of the regional linear model is either

$$Y = \alpha + \beta X + \epsilon$$

or

$$Y = \alpha + \beta_1 X + \beta_2 X^2 + \epsilon$$

where

Y = average regional yield and

α, β's are unknown model parameters.

X is the known independent variable, and

ϵ is the difference between Y and its expected value.

In the examples used in this paper, the soybean model has the quadratic term while the corn model is limited to the linear term.

The values for X for corn and soybeans are shown in the following tables.

SOYBEAN VARIABLES BY MONTH	
August	estimated number of lateral branches per acre
September	Estimated number of pods with beans per acre
October-December	(estimated number of pods per acre) X (net weight per pod)

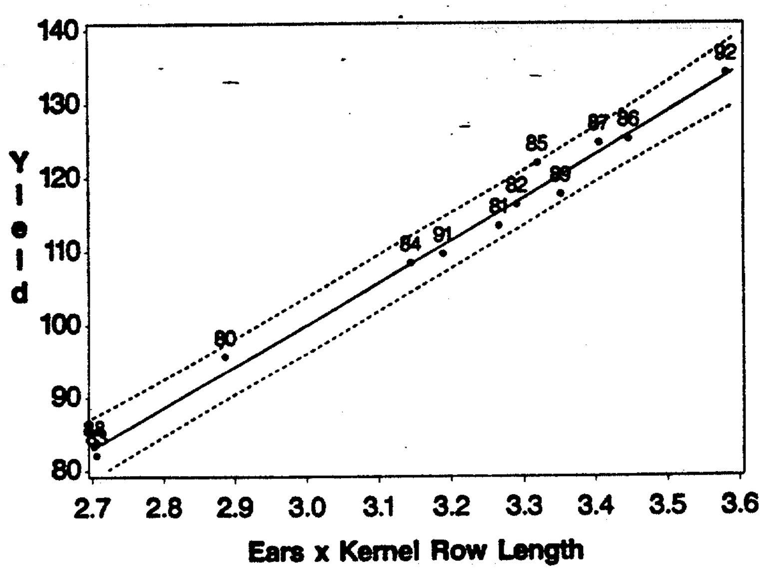
CORN VARIABLES BY MONTH	
August	(stalks with ears + ears with kernels per acre) X (average kernel row length per ear)
September	(Ears with kernels per acre) X (average kernel row length per ear)
October-December.	(Ears with kernels per acre) X (average grain weight per ear)

Maturity Adjustment

While NASS conducts the survey during the last ten days of each month, the overall maturity of the crop at that time will vary from year to year, depending on when it was planted, subsequent weather, etc. The forecasting power of the model is enhanced by classifying each plot by stage of maturity and limiting the independent variable calculations to data from pre-selected stages. This adjustment allows the independent variables to be more comparable across years. Variables not used directly in X (such as nodes and blooms, dried flowers and pods) are used for maturity classification. Consequently, the predictor variable is not a function of all the data, but only those plots in a stage that has exhibited good predictive power for final yield. This criteria normally means the exclusion of very immature samples in the first month of the survey. After that the vast majority of the samples are used directly in X.

A plot of the data in the September 1 corn regional model is shown below. (The digits plotted represent the years 1980-1992).

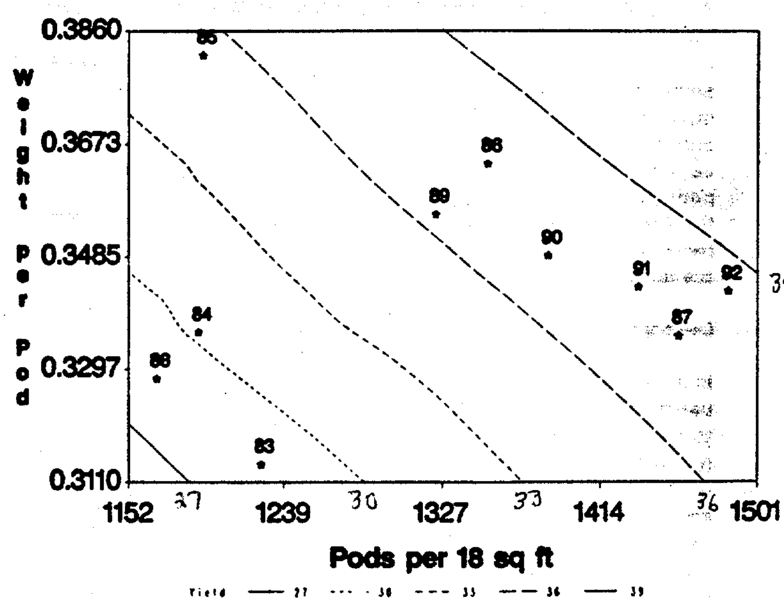
September 1 Corn Model



Relationship of the number and weight variables to final average yield

As mentioned above, the survey variables are selected to correlate with the components of final yield, which are number of ears or pods and weight per ear or pod. It is quite illuminating to view the 3-dimensional distribution of final yield and the factors of the independent variables to see how they explain the yield level. Since the independent variable in the model can usually be factored into the product of a variable correlated with final weight and one correlated with final number, we can plot the fitted model surface over the weight X number plane. The projection of selected levels of the fitted yield surface onto this plane is easier to analyze. An October example for soybeans and a November one for corn are shown below.

Soybeans - October 1, 1983-1992

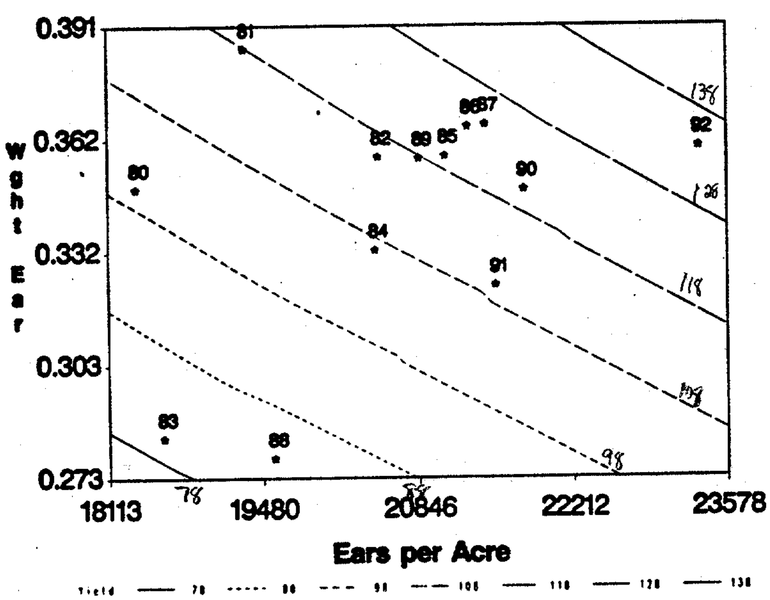


For soybeans, the weight per pod is in grams, and the yield contours projected from the fitted model surface onto the plane are 27, 30, 33, 36 and 39 bushels per acre. On October 1, usually about half the crop has been harvested, and the weight is for just those harvested samples. The pods per 18 square feet is for all samples as of October 1.

This graph contains a great deal of information about soybean yields. The years divide into two distinct groups, with 1983, 1984 and 1988 in the lower left corner, and the remaining years distributed along the 36 to 39 bushel contour region. The years 1983, 1984 and 1988 were severe drought years in the corn belt, and both pod counts and weight were depressed to the point that yields averaged around 30 bushels. In the remaining years, conditions were more normal, and average yields were generally around 36 to 39. So far there has not been a year where weight and numbers of pods were simultaneously near record levels. There is an obvious negative correlation between average weight and number. The two variables interact inversely with each other to produce approximately the same yield, even though the weight and number variables are varying quite widely. The heaviest average weight occurred in 1985, but it had drought-like numbers of pods. At the other extreme, 1987 had the lowest October 1 weight of the normal years, but its pod counts were the second highest. 1992, which has the record yield to date, had the highest number of pods on October 1.

Since the surface is based on a model with a quadratic term, one can see the spacing between the contours increases as the yield level increases. This implies that there are diminishing increases in yield as the average weight and numbers increase. Also, since the contours are at roughly 45 degree angles, one can deduce that increases in weight or numbers will increase yield. However, this is survey data, and numbers and weight do not vary independently (they vary inversely) so an increase in one will normally be associated with a decrease in the other and vice versa.

Corn - November 1, 1980 - 1992



For corn, usually about two-thirds of the crop is harvested by November 1. The grain weight, in pounds, is just for the harvested samples. The ear counts are for all samples as of November 1. The projected yield contours from the fitted surface are 78, 88, 98, 108, 118, 128 and 138 bushels per acre.

Here we see the two drought years, 1983 and 1988, in the lower left corner. There appears to be less dependence between the weight and number variables for corn than there was with soybeans. Some years, such as 1985, 1986 and 1987 are pushing the limit on both ears and weight. In 1992, ear density increased dramatically, while the ear weights maintained an average level for non-drought years. 1992 set a new record for yield by a large margin, driven by the large ear counts.

Since the corn model has no quadratic term, the spacing between the contour levels is constant. The 45 degree contours indicate both weight and numbers drive final average yield. If conditions are generally good, it is possible to have both large ear counts and above average weights in the same year, something that is not generally seen with soybeans.

Conclusion

Average corn and soybeans yields can be predicted by observing variables that are correlated with final numbers and weights. In corn both counts and weights can be high at the same time, producing record yields. With soybeans, however, final counts and weights are inversely related, producing relatively constant average yields in non-drought years.

References

Birkett, T.R. (1990), "The New Objective Yield Models for Corn and Soybeans", National Agricultural Statistics Service, SMB-90-02, Washington, DC, 20250.

SAS Institute Inc., SAS/GRAPH Software: Reference, Version 6, First Edition, Volume 1 and 2, Cary, NC: SAS Institute, Inc., 1990.

Searle, S.R. (1971), Linear Models, New York: John Wiley & Sons, Inc.

The author can be contacted at

NASS/USDA, Room 4813,
14th and Independence, SW,
Washington,DC, 20250

202-720-5359