YRB 68-08

July 31, 1968

A NEW SOURCE OF DATA FOR PRODUCTION ECONOMISTS

by

Fred H. Abel,* Donald D. Durost,*

and Harold F. Huddleston**

Not for publication. Presented in this form for review purposes only.

*Agricultural Economist, Farm Production Economics Division, Economic Research Service, USDA.

**Statistician, Statistical Reporting Service, USDA.

A NEW SOURCE OF DATA FOR PRODUCTION ECONOMISTS

Agricultural economists and other agricultural scientists have long investigated the relationship of input levels and other production practices to yields. These studies often are based on data from experiment station plots or on data obtained from a survey of farms. Data from large sample surveys of fields planted to particular crops, useful in studies like these, are now being collected annually.

The vehicle for collecting the data, the Objective Yield Survey of the Statistical Reporting Service has been operational for several years to improve agency yield per acre estimates. Thus, the detailed data have not been published and are not yet readily available. 1/

The vehicle for collecting the data, new uses for the data, and methods of making the data more readily available are examined.

The Objective Yield Survey

The objective yield survey (OYS) was instituted in the late 1950's to provide objective data on which to base forecasts of final yields. Projections of final yields for major crops are made by SRS several times during each growing season. The method of making these forecasts is not discussed because it is discussed adequately elsewhere (2, 3, 4, 5, 6, 10, 11, 12, 14, 17, 18) 2/. Rather the focal point will be the potential use of these plant characteristics and related data for other purposes.

Before discussing the data, a brief indication of how the sample fields and plots within fields are selected will be given.

Each June a probability area sample survey is conducted which provides information on acreages planted to various crops and on livestock and other items. All fields in each sampling unit are delineated on aerial photographs and the kind of crop and acreage in each field ascertained. For the OYS sample for each crop a subsample of fields is then selected with probabilities proportional to acreage.

In 1968, the number of fields included in the preharvest OYS is 2,450 for wheat; 3,930 for corn; 1,645 for soybeans; and 2,580 for cotton. The samples were selected from many states and represented over 90 percent of the production and acres of each of these crops. Within each sample field two small plots are marked by small stakes so the same plots may be visited from time to time during the growing season to obtain the data for making forecasts. The plots are harvested as soon as the crop is mature for purposes of estimating yield. For some crops the fruit is sent to laboratories for analysis. Immediately after harvest the fields are again visited, using another sample of plots to measure harvesting losses.

The yield forecasting models are designed to use measurements of plant phenological characteristics. Most of the visits to the field are to obtain objective plant measurements for this purpose. In this article, the use of these plant measurements for the projection of yield within years will not be considered. 3/

The Characteristics of the Data

The data obtained by the OYS can be grouped as objective and interview data. Objective data include such items as row spacing, plant population, yield, field loss, insect damage, and analysis of the fruit (oil and protein content of soybeans, etc.). Data, obtained in an interview with the farmer includes: Fertilizer applied, expected yield, variety, planting data, insecticides used, etc. The exact data obtained depends upon the crop. The potential value of data other than the plant characteristics and objective yield data has been recognized (16).

When considering new data it is instructive to compare it with types of data already available. Data from experimentation station plots and one time surveys will be considered. The data characteristics will be appraised from the point of view of the production economist.

Experiment station plot data have the advantages of being objective and measured quite accurately. Also, the experiment is controlled so that input levels and cultural practices not used on farms can be analyzed. The disadvantages are: (1) The data are typically from a small geographical area and therefore do not represent response farmers can expect, (2) there is no information obtained on what input levels or cultural practices farmers are using, and (3) there is no information on what response farmers are getting from their input levels and cultural practices.

Data obtained from a one time survey of farms have the advantage of reflecting what input levels and cultural practices farmers are using and indicating what response farmers are getting. They have the disadvantage of being subject to "interview" error. Interview errors result from failure of farmers to understand questions, their not being able to recall answers or not knowing the answer and having to guess. The data generally are not objective. Farm surveys are expensive and thus the size of the samples are usually small. The results generally do not represent a large geographical area.

The data from the OYS are from a large sample. They are representative of fields growing the crop and of farms growing the crop in the area sampled. Objective and interview data are or could be obtained that would indicate what input levels and cultural practices are being used by farmers and what yield response farmers obtained. The interview is conducted during the growing season so recall error should be small.

Potential Uses of the Data

In discussing the potential uses of the data, it is assumed that the individual plot data or any aggregate of it are readily available. Although there are many possible uses of the data, most can be grouped as either descriptive or analytical.

Descriptively, the OYS data provide very good information on what is happening in the production of the selected crops. These data can be used directly by farmers or persons advisng farmers to compare their practices and input levels with those prevailing in the area. They also can be used indirectly as inputs to budgets for linear programming or budgeting studies.

Analytically the data can be used to explain or predict yields. They also provide a basis upon which to analyze farmers response to new practices, innovations or varieties. Plant population data from the OYS, aggregated to the state level have been used in studies of corn yields (15).

The potential use of the data is, of course, much greater than outlined above. This is particularly true if additional data are collected. Additional data in the form of soil and topographic maps for the sample segments could be prepared. It would then be possible to relate yields as well as land use to soil and topographic factors. Also, data could be collected on cultural practices and varieties (16).

These data would afford an opportunity to evaluate the effects, under actual farming conditions, of new varieties and practices in relation to soil characteristics and other variables. It would be a powerful tool to overcome the limitations of results obtained by controlled experimentation in laboratories, in greenhouses, or on experimental farms (16).

Provided the justification is sufficient and necessary financial resources become available, many other extensions of the OYS are also possible.

It is perhaps useful to illustrate the nature and potential use of the data for analytical purposes. An analysis of yearly corn yields based on data from the OYS, weather indexes 4/ and SRS's published statistics on acres harvested and yield of corn is reported.

The model regressed acres harvested (AH), plant population (PP), nitrogen (N), Potassium (K), Phosphorus (P), time (T), June weather index (JUW), July weather index (JYW), August weather index (AW), Indiana (IND), Illinois (ILL), Missouri (MO), and Iowa (IA), on yield (Y). The data on PP, N, K, and P are from the OYS. The OYS estimates are based on samples that vary greatly in size between states and over time, but all samples are sufficiently large to produce reliable state estimates. Time series data are combined with cross section survey data for the years 1957 to 1966 for the five corn belt states (the four listed above and Ohio), giving 50 observations. The state variables are 0, 1 dummy variables and measure the consistent difference in yields between Ohio and the state in question after the effects of the other variables have been taken into account.

The results are presented below. The number in parenthesis under the coefficient is the estimated t value. The level of significance is indicated for the 0.10, 0.05, and 0.01 level by *, **, and *** respectively.

$$Y = 11.27 - \underset{(.4)}{.00048}\ AH$$

$$+ \underset{(.19)}{.0335}\ N$$

$$+ \underset{(1.5)}{.379}\ P^{*} - \underset{(1.3)}{.306}\ K^{*} + \underset{(2.59)}{.0034}\ PP^{***}$$

$$+ \underset{(4.1)}{27.3}\ JUW^{***} + \underset{(3.6)}{14.1}\ JYW^{***} + \underset{(2.11)}{4.3}\ AW^{**} + \underset{(.86)}{.741}\ T + \underset{(1.8)}{8.9}\ IND$$

$$+ \underset{(1.3)}{14.1}\ ILL^{*} - \underset{(.54)}{4.5}\ MO + \underset{(.99)}{11.6}\ IA$$

R^2 was .92 and the standard error of Y was 4 bushels.

The model fits the data quite well, as 92 percent of the variation in yield is explained by the variables in the model. The nonsignificance or low level of significance of the coefficients of such important variables as N, P, K, and T and the wrong sign for K, reflect the fact that all these variables tend to vary in the same way (multicollinearity). This makes it difficult to measure their independent effects accurately as shown by the low T values. If more disaggregated data, such as the individual plot data, were used, it is likely that these independent effects could be measured accurately. Multicollinearity does not affect the explanatory power or predictability of the model. That is, the joint effect of all the variables is measured accurately.

The individual plot data would also be useful in answering questions such as: Are farmers with the highest corn yields also the farmers using the greatest quantity of fertilizer per acre, using more plants per acre, and using narrow row spacing? Using this data, extension workers could make recommendations based on the results obtained by farmers under stated conditions. Also, if the practices of today's leading farmers are known, it would indicate what the average farmer in the future might do.

The OYS data also supplies agricultural scientists with a tool to determine the relationship between experimental plot and farm results. Thus, they would have more information to help estimate the consequences of adoption of new practices by farmers.

OYS data have been used to advantage in yield projections. Yield projections are important in the administration of federal farm programs and the best possible yield projections are needed. Historically yield projections have been yield trends modified by subjective judgments. An example of how additional input data such as the OYS data can modify such analysis is the recent experience with corn yield projections. In 1962 the ERS projected corn yields for 1967 was 70 bushels. When the projected corn yields were reviewed in 1963, available new data did not warrant changing the projected yield. In 1964 historical yield trends <u>and</u> OYS data for states on the quantity of nitrogen used and plant population per acre were available. With this data, the projected corn yield for 1967 was increased to 76 bushels. The actual yield for 1967 was 78 bushels, two bushels above the 1964 projection and 8 bushels above the initial projection in 1962. The six bushels different in the two projections can be largely attributed to the additional state input data.

These examples show only a few of the many possible uses of the OYS data. We believe they also show how more disaggregated data, such as the plot data obtained in the OYS, could be used to advantage.

Conclusion

The potential usefulness of the OYS data depends on the availability of the data and the level of aggregation of the available data. If accurate analysis and maximum use of the data are to be made, it is important that disaggregated (plot) data be available.

Ideally, the individual plot data would be published. The counties and crop reporting districts in which the plots are located should be specified. All data from the plots, relating to the plots and relating to the farms on which the plots are located should be published together. Some of the data about the farms could come from the June Enumerative Survey.

Currently, much of these data are not available from published sources. What data are available are highly aggregated, and published in several publications. Production economists and other agricultural scientists must make their data needs known and try to have the data made available at a level of aggregation and in a format most useful to them.

The OYS surveys provide a great deal of unique information and could be expanded to include additional useful data. It provides a very good means of collecting data from a large random sample for production economists and other agricultural scientists. The survey vehicle provides a method of attacking new kinds of problems as well as attacking old problems in new ways.

1/ Some data have been published in "Crop Production," published by the Statistical Reporting Service, USDA. It is published monthly and there is an annual summary. Objective yield data have been published in the December, January, and February monthly issues. The data are often identified as "obtained from a selective sample of fields."

2/ The number in parenthesis refers to the publication listed at the end of this article.

3/ Most of the potential use would be by agronomists or by firms that want to predict final yield during the growing season.

4/ The weather indexes are constructed to reflex the difference between the actual evapotranspiration and the normal evapotranspriation (13).

Selected References

1. Abel, M. E., and Waugh, F. V., "Relationships Between Group Averages and Individual Observations," Agricultural Economics Research, Vol. XVIII, No. 4, 1966, pp. 105-115.

2. Hendricks, W. A., "Forecasting Yields with Objective Measurements," J.F.E., Vol. 45, No. 5, Dec. 1963, pp. 1508-1513.

3. Hendricks, W. A., and Huddleston, H. F., "Objective Forecast of Cotton Yield," Agricultural Economics Research, 7: 108-111, 1955.

4. Hendricks, W. A., and Huddleston, H. F., "Objective Forecasts of Cotton Yields", A. E. R., Vol. IX, No. 4, 1957, pp. 20-25.

5. Hendricks, W. A., "Validity of Objective Estimates of Corn Yields," A. E. R., Vol. VII, No. 3, 1955, pp. 69-72

6. Houseman, E. E., and Huddleston, H. F., "Forecasting and Estimating Crop Yields from Plant Measurements," F. A. O. Monthly Bulletin of Agricultural Economics and Statistics, Vol. 15, No. 10, Oct., 1966.

7. Houseman, E. E., and Reed, T. J., "Application of Probability Area Sampling to Farm Survey," Agricultural Handbook 67, U.S.D.A., 1954.

8. Huddleston, H. F., "A New Area Sampling Frame and Its Use", J.F.E., Vol. 47, No. 5, Dec. 1965, pp. 1524-1533.

9. Huddleston, H. F., "Objective Methods in Forecasting Components of Corn Yield," A.E.R., Vol. X, No. 2, 1958, pp. 49-53.

10. Kelly, B. W., "Preliminary Report on Objective Procedures for Soybean Yield Forecasts," A.E.R., Vol. IX, No. 4, 1957, pp. 139-141.

11. Kelly, B. W., "Probability Sampling In Collecting Farm Data," J.F.E., Vol. 45, No. 5, Dec. 1963, pp. 1515-1520.

12. Nordquist, A. V., "Livestock Estimates and the Search for Further Improvement", A.E.R., Vol. VIII, No. 4, 1956, pp. 128-130.

13. Palmer, W. C., "Meterological Drought", Research Paper No. 45, U. S. Dept. of Commerce, Weather Bureau, 1965.

14. Parr, R. D. and Calvin, L. D., "Research on Objective Forecasts of Filbert Production," A.E.R., Vol. VIII, No. 3, 1956, pp. 92-94.

15. Shaw, L. H., and Durost, D. D., "The Effect of Weather and Technology In The Corn Belt, 1929-62" Agricultural Economic Report No. 80, U. S. Dept. of Agr., Economic Research Service, 1965.

16. Trelogan, H. C. and Houseman, E. E., "Assembly and Retrieval of Production Information for Agricultural Business", Speech presented at Annual Meeting of the Agricultural Research Institute, Washington, D. C., 1967.

17. Trelogan, H. C., "The Changing World for Agricultural Statistics", J.F.E., Vol 45, No. 5, Rec. 1963, pp. 1500-1506.

18. Vickery, R. E., "An Appraisal of Interview Procedures in Farm Surveys", A.E.R., Vol. VIII, No. 2, 1956, pp. 59-65.