A SURVEY OF AGRICULTURAL ECONOMICS LITERATURE

VOLUME 2

Quantitative Methods in Agricultural Economics, 1940s to 1970s


Published by the University of Minnesota Press, Minneapolis, for the American Agricultural Economics Association
Technical Developments in Agricultural Estimates Methodology

Harry C. Trelogan, C. E. Caudill, Harold F. Huddleston, William E. Kibler, and Emerson Brooks
Statistical Reporting Service, United States Department of Agriculture

History of Methodology for Agricultural Statistics before 1940

It is only in recent years that farm data, although found in ancient records (for example, in Egyptian scrolls), have been developed to represent large areas or nations in quantitative terms. Even in the United States, now in the forefront in applying farm data to economic and political problems, the development of an information system for the collection, processing, dissemination, and interpretation of agricultural statistics has been a slow process. Here we will characterize briefly the period before World War II, noting the progress in crop and livestock estimates that did occur, before dealing more comprehensively with the postwar period when the application of modern statistical technology was accelerated.

The statistical base for an agricultural information system evolved in response to national development needs. Its growth and effective contributions to development were favored by recognition of its value by political leaders, accompanied by scientific progress that facilitated the collection and analysis of the necessary data. Farmer George Washington, with a strong aversion to urban merchants who obtained market advantage by acquiring and holding supply information, left a legacy of beneficial influence. Government support for making market information public began in 1839. Progress came more rapidly during periods of national adversity when the availability of farm products was of crucial importance.
The Civil War created an early opportunity in this direction. Abraham Lincoln sought to persuade the British to stay out of the conflict by calling their attention to the prospective food supplies from the North, which he was confident would compare favorably with the more obvious supply of cotton from the South. Arrangements for crop estimates were made quickly in 1862 in the newly created office of Commissioner of Agriculture. These estimates proved their worth, laying the foundation for a strong farm statistical program that was founded in the tax system with a specific appropriation in 1865. Subsequent crisis periods—panics caused by needs for foreign exchange, business depressions, droughts and floods, and two world wars—created demands for better supply information, based on reliable statistics. Only the highlights of the technological revolution in statistical methods that brought the United States to its present preeminent position in this work can be treated here.

Originally the problems consisted mainly of obtaining nationwide data on the acreage and production of the major grain crops and numbers of livestock. Census data were adequate for the development of supply weights by production areas. Since the farms within a given production area were similar in many respects, an alert observer could develop acceptable production estimates, and the inexpensive and dependable mail service could be used for data collection from both farmers and county agents.

Data were obtained from farms by crop reporters in every county, usually in terms of percentages of the previous year. These data were brought together for state estimates that in turn were combined in Washington to make national estimates. Statistical reporters in each state were expected to be familiar enough with growing areas and conditions to be able to exercise judgment regarding the reasonableness of the resulting estimates. Examination of the data by experienced statisticians led to national estimates that were essentially summarizations of state estimates.

For several decades great dependence was placed on human judgments in arriving at the single best estimate for each item, for each state, for each growing season, and for annual national estimates. Crops and livestock were followed through the markets to export or consumption as a means of verifying each estimate. As more and more evidence was accumulated, year-end estimates were revised. Unless substantial revisions were clearly needed beforehand, the policy was to examine all estimates at the end of the following year for possible revisions so that the annual estimates became the bases for comparisons in the next year. Each item was examined for compatibility with the next Census of Agriculture when those data became available.

The Agricultural Census, which was developed outside the USDA in 1840, was designed to provide a full count of American farms taken at ten-year in-
tervals (later changed to five-year intervals). Even though the completeness of the censuses—which were taken by temporary political employees until 1960—varied greatly, they provided more detailed farm data than any other source, and they were used to true up crop and livestock estimates based on sample data. Experience showed that errors in a low census estimate tended to be compounded in subsequent years because of a cumulative downward bias in the annual estimates. With a ten-year interval between censuses the error could become substantial at times. For example, the 1899 USDA estimate of corn acreage was 82,109,000 acres compared with 94,914,000, or nearly 16 percent more, reported two years later by the census for 1899. The USDA estimate of wheat acreage was 44,055,000, whereas the census figure was 52,589,000 or 19 percent more. Consequently, a corps of special field agents was created at the turn of the century, in part to overcome this known bias of farm crop reporters.

In 1905 the Crop Reporting Board was established, with substantial representation from the field workers, to apply group judgment in making the best estimates from all sources of data. The statistical methods developed out of the experience of men with little formal training during these early decades of crop estimates were generally creditable.

Since 1900 when the Census used electric machines for the first time to tabulate part of the agricultural data, the Census Bureau has held a position of leadership in developing mechanical and electronic data handling capabilities. It played a pioneering role in the development of Hollerith (punch card) machines and eventually electronic computers.

USDA crop estimators combined returns from township reports, county reporters, state agents, and field agents (each covering several states) but changed methods little despite an influx of suggestions during World War I. Many ideas were explored but rejected, including the possibility of expressing condition reports in terms other than percentages of “normal,” the recruitment of county agents as reporters, and the use of threshers’ returns for grain. Other ideas were adopted, such as the use of round numbers for weights by counties, the development of quantitative estimates as the basis for forecasts, the reporting of intentions to plant and to breed—all of which gradually gained wide acceptance. Two innovations of lasting benefit were the introduction of Civil Service appointments for statisticians in 1914 to assure the availability of qualified staff and state cooperative arrangements in 1917 to reduce confusion and to strengthen the estimates especially in states that had annual state farm censuses. The number of reporters and the number of crops reported both continued to increase, and many special surveys were undertaken during World War I.

In the postwar period expansion of the information system continued with
greater stress on prices; a gradual transition from reporting "judgment data" to "individual farm data" for acreage began and was in full use by 1926. In 1924 a Rural Mail Carrier Survey for acreages began, to be taken in September of each year. Data on intentions to plant, collected earlier in each year, were converted into "Prospective Plantings" by the removal of bias.

A movement to give the staff members statistical training which emphasized sampling and correlation took hold. Several types of graphic analyses were developed in the late 1920s, based on ratios of currently reported acreage to the acreage reported for the same farm the previous year, the acreage reported in the census year, and the total acreage in the farm. These analyses revealed different types of bias occurring for such items as cash crops—biases that needed to be taken into account in forecasting. During the same period several researchers in other parts of the Bureau of Agricultural Economics developed new statistical techniques that proved to be very useful. The multiple curvilinear correlation technique developed by Ezekiel [1924] was one of these. C. R. Sarle simultaneously developed graphic methods to reduce the laborious process for crop estimates, and Bean [1930] commented on the graphic method.

Objective measures for estimating acreages were explored including counting the fields planted to particular crops from train windows along specified stretches of track. The number of telephone poles bordering each field indicated size. In 1923 this approach was adapted by installing meters on automobiles to measure road frontage bordering fields of specified crops. Later yield forecasts were checked by counting plants or cotton bolls to get objective checks on judgment data. Condition reports gave way to forecast yields per acre read from regression charts indicating past relationships. Allowances had to be made for such observed phenomena as potential boll weevil damage in cotton, weather conditions, and acreage abandonment during the growing period. Similar innovations were introduced for fruits and vegetables to respond to needs for quantitative data by specialized areas of production.

Livestock estimation came in for special attention with the postwar price decline. The Rural Mail Carrier Survey, in which the carrier put an inquiry card in the mailboxes of ten growers on his route, was initiated in 1922 for a pig survey. This proved to be a practical means of getting an acceptable sample, and it was extended to other livestock and livestock products (for example, milk produced yesterday) for many years afterward. Historic questions (number this year and last year) and questions on intentions to breed were perfected and widely adopted. Data on livestock movements were gathered from numerous sources, along with births and deaths, enabling estimators to maintain balance sheets by states.
Beginning in 1925 the Census of Agriculture has been taken every five years, thus reducing cumulative error problems in crop estimating. Several decades passed before the advocates of getting at least benchmark data by sampling methods were heard. These demands were heard after mail censuses failed to achieve close enough approximations to full counts and after annual sample-estimate benchmarks had been adopted by the USDA for crop estimates.

Annual prices received for agricultural products in local markets have been reported since 1866. Data on monthly prices for crops were first collected in 1908, on monthly prices for meat animals and livestock in 1910. Needs arising during World War I stimulated the collection of data on prices paid by farmers, first from the farmers but later from suppliers, and also stimulated reports on the supply and demand for farm labor. Later refinements included obtaining local farm product prices from elevators, mills, and dealers, abandoning the mid-month price for milk in favor of averages for the month as a whole, substituting market season averages for December 1 prices for crops, and realigning crop reporting districts to coincide more closely with marketing areas.

Index numbers have been used to express changes in prices for crops, for livestock, and for all farm products. Growth of the parity concept stimulated interest as the concept became embodied in farm policy legislation. Such developments called for close attention to the maintenance of adequate samples and to index number construction with minimum bias. In general, the Laspeyres formula was used.

The twin emergencies of depression and drought during the 1930s presented many new demands and problems to crop estimators and diverted attention from recognized technical research needs. Nevertheless, the quest for data on weather and weather forecasting with respect to crop forecasting continued. By the late thirties, however, researchers were sensing the futility of weather analysis and recognizing other means of predicting crops. New approaches were sought, but World War II intervened, disrupting their development.

**Technical Developments in Agricultural Estimates**

*Methodology, 1940-50*

During the 1940s technical developments in the Division of Agricultural Statistics occurred in all aspects of its work—data collection, tabulation, analysis, publication, and dissemination. As American industry expanded its operations to provide the vast quantities of war matériel required by Britain and France, the number of workers available to American farm operators sharply
decreased. Consequently Agricultural Estimates services were requested to expand and improve farm employment information. By the fall of 1941 the decision had been made to set up large-scale farm labor enumerative surveys, using a probability sample based on areas of land throughout the country.

The principal technical problem was how to associate farms with selected sample areas. If data were obtained for all farms with any land inside the sample segment, large farms would be overrepresented. If only farms with all their land inside the segment were enumerated, small farms would be overrepresented. After considerable study it was decided to designate a “headquarters” for each farm with land inside the sample segment; if the headquarters was inside the segment, data for that farm would be used in the analysis. Enumerators were given detailed criteria for determining farm headquarters:

1. With only one occupied or unoccupied dwelling on the farm, the dwelling is the headquarters.
2. With two or more dwellings and the operator living on the farm, the operator’s dwelling is the headquarters.
3. With two or more dwellings and the operator living off the farm, the dwelling of greatest value is the headquarters.
4. With no dwelling but with a building on the farm, the building is the headquarters. With two or more buildings, the one of greatest value is the headquarters.
5. With no buildings on the farm, the main entrance is the headquarters. The main entrance is the point where the farm operator usually turns off a public road, private road, trail, or path to the farm he operates. If a farm with no buildings is composed of two or more separate tracts of land, the headquarters for the farm is the main entrance to the tract with the greatest value.
6. With no buildings on the farm and no point regarded as the main entrance, then the farm headquarters is the northwest corner of the tract with the greatest value.

Later, when segments had been delineated in towns and cities, the operator’s residence served as the headquarters and the other categories were no longer needed.

As demands increased for additional data on a wider range of subjects, a plan evolved for the creation of a nationwide sample from which subsamples could be drawn for probability area surveys in order to collect data on almost any phase of American agriculture. The result was the establishment in 1944
at Iowa State University of the "Master Sample of Agriculture" as a joint venture with the Division of Agricultural Statistics and the Bureau of the Census. As finally constituted, the Master Sample was a scientifically drawn sample of about 67,000 areas of land, each area having natural boundaries (as far as possible) and each containing on the average the headquarters of five farms, making a total of approximately 300,000 farms or roughly 5 percent of the estimated 6,000,000 farms in the United States at that time. The sample segments varied in size from less than a square mile in Indiana to over a hundred square miles in Nevada. The sample (1/18 of the segments) was drawn in a systematic fashion with a random starting point. Thus, every acre of land in each of the 3,000 counties in the United States had a known chance of being included in the Master Sample, which was subdivided into three major divisions: (1) incorporated towns and cities, (2) unincorporated places, and (3) open country. The sampling rate of 1/18 rather than 1/20 (5%) resulted from an early plan to draw one of each 18 sections or one-half of a township. A 5-percent sample was considered necessary to provide state acreage estimates of major crops with acceptable sampling errors of around 6 percent.

Master Sample segments were used by the Census Bureau in connection with the 1945 Census of Agriculture. Probably the first subsample drawn for actual use was the so-called 101 County General Purpose Sample. The following procedure was used in drawing the sample:

1. All counties in the United States were classified into twenty groups on the basis of major type of farming region (corn, cotton, dairy, general and self-sufficient, range livestock, western specialty wheat, and residual) and major geographic region (Northeast, North Central, South, and West).

2. The result was twenty groups of counties, with each group wholly within one major type of farming region and within one major geographic region.

3. The twenty groups were subdivided into 101 strata (with about 60,000 farms in each stratum) by using component indexes; approximately twelve Census of Population and Census of Agriculture items (such as demographic farm characteristics, farm labor force, county characteristics, and household welfare items) were combined, by using two or three different weighting systems on the twelve variables, into two or three component indexes (Haygood [1945]; Haygood and Bernert [1945]; Hansen, Hurwitz, and Madow [1953, vol. 1, pp. 387-390]).
4. From each of the 101 strata one county was selected by the use of random numbers, so that the probability of a county being chosen was proportional to the number of farms in that county.

The 101 County General Purpose Sample was designed to provide an efficient national sample that would be representative of major type of farming regions. Within these regions it would be representative of certain important socioeconomic variables of the region. It was not designed to be representative of individual states.

For the special farm labor enumerative surveys made in March, May, and September 1945, the 101 County General Purpose Sample was augmented by an additional 57 counties selected by the same procedure. Special segments were developed in the 158 counties for the wage surveys with each containing about five farm headquarters.

In an attempt to provide some of the economic data needed on agriculture in the United States, surveys of agriculture were made in April, July, and October 1945 and in January 1946. Interviews were obtained from about 2,800 farmers each quarter. The 101 County General Purpose Sample was used for these surveys. Although the results were of limited direct use, the surveys clearly represented the type needed to provide answers to the multitude of economic questions facing USDA officials and other decision makers and analysts. Not until the 1970s would there be a definite program of early quarterly surveys to lay the groundwork and to provide answers of sufficient scale to cope with the technical problems involved.

In 1946 the Division of Special Farm Statistics was established in Agricultural Estimates specifically to inaugurate a program of periodic enumerative surveys. The first of these (January, 1947) was an enumerative survey of 15,000 farms associated with a selection of Master Sample segments in 800 counties. This survey initiated the economical practice of using a short questionnaire for basic items and a long questionnaire for data on additional questions. Other innovations were made after careful pretesting in the field. The system of state supervisors, part-time district supervisors, and local enumerators—a system that had been largely developed in the course of farm employment surveys during World War II—was expanded and improved for the January 1947 enumerative survey and other surveys.

The varied techniques and procedures used by Agricultural Estimates developed out of the willingness of the staff to undertake unusual surveys such as the series of enumerative and objective yield surveys made in 1949 and 1950. The governors of Virginia and North Carolina got into a friendly argument over which state had recorded the greatest gains in corn yields in the previous ten years. The Agricultural Estimates offices in Washington and in
the two states, in cooperation with their respective state colleges of agriculture, conducted special enumerations and field counts to decide the matter. The results indicated that Virginia was the winner in both years. A probability sample of areas of land in all but a few urban counties in each state resulted in an enumeration of 2,400 of the 173,000 farms in Virginia and 3,300 of the 287,000 farms in North Carolina. This was one of the earliest attempts to make objective counts of corn on farms selected by a probability sample of land areas on a state basis. In the first year the responsibility for field operations in Virginia was divided between two agencies and the results demonstrated the advisability of avoiding such arrangements.

Another survey that added to the accumulating technical skills of Agricultural Estimates was the survey made of farm housing in 1950. Observations, measurements, and interview information were obtained on some 20,000 farms in 382 counties in 45 states. In the South pictures were taken of each house in the sample segments. Because of the “closed season” on field surveys between March 15 and May 15 when the federal Census was being taken in the field, the housing survey was started in the southern tier of states from New Jersey to California on February 20 and in the remaining states on May 15.

Beginning in about 1944 it became necessary to obtain approval of the Division of Statistical Standards of the Bureau of the Budget for the purpose and procedures of any survey made by a federal agency in which nine or more people would be contacted. This regulation forced Agricultural Estimates to give careful consideration to the purpose, subject matter, questionnaire design, sampling plan, field procedures, and analytical methodology for its projects, especially for enumerative and objective yield surveys. Although obtaining clearance from the Bureau of the Budget was often frustrating, the result was an improved product. It also resulted in the upgrading of the technical competence of the staff in survey planning, questionnaire design, sampling techniques, operating procedures, and statistical analysis. In 1946, for the first time, individuals were hired because of their special talents in formulating survey inquiries. The Division of Agricultural Statistics sent out annually about ten million questionnaires, and after 1944 additional attention was given to their construction and use with an undoubted improvement in their effectiveness.

**Technical Advances in Agricultural Statistical Methodology, 1950-70**

Building primarily on concepts and methodology developed and used for special surveys on an ad hoc basis between 1930 and 1950, the statistical meth-
odology used for crop and livestock estimates underwent dramatic changes during the two decades following 1950. Pressure for these changes came from several sources, among them the congressional investigation of the 1951 cotton estimates and new developments in automatic data processing.

The congressional investigation highlighted the need for updating the methodology used for crop and livestock estimates and led to increased funding to implement new procedures as they were developed. Following the investigation a long-range plan was prepared for the agricultural estimates program of the USDA. The plan had four objectives, commonly referred to as Project A (on acreage, yield, and production and livestock inventories), Project B (on farm prices), Project C (on data handling methods), and Project D (on new or additional types of estimates).

Project A contained the plan for meeting the most fundamental needs of the estimating program, and at the same time it provided the framework for implementing the other projects. Consequently this part of the overall plan received the most attention during the 1950s and 1960s. Project A was most concerned with the implementation of a nationwide probability system of surveys which would provide independent unbiased estimates with known precision. The system was designed to use the Master Sample area frame developed during the 1940s at Iowa State University (King and Jessen [1945]). A probability sample of about 17,000 area sampling units averaging about one square mile in size was designed to strengthen state and national crop and livestock estimates. This sample was fully implemented for the forty-eight contiguous states by 1965. The results of studies conducted during the 1950s were used to modify and improve procedures as the implementation process evolved. These studies revealed serious deficiencies in the available area sampling frame for the western and more urbanized eastern states, leading to the construction of a new land use area frame in these states (Hudleston [1965]). The work was completed by 1965 and provided a major resource for sampling United States agriculture. By 1978 new land use area frames will be operational in each of the forty-eight contiguous states.

Most of the major technical advances in agricultural estimating during the 1950s were directly related to progress on the implementation of probability sampling. The new system provided for (1) the use of area sampling in continuing operational surveys, (2) the combined use of list and area sampling for agricultural surveys, (3) crop yield forecasting based on probability samples of fields, (4) the use of cross-sectional surveys of plant characteristics in developing yield models, and (5) the refinement of "crop cutting" techniques for estimating crop yields.

Area sampling, although used in a limited way before 1950, was viewed by many as being too expensive for a continuing system of surveys. Arnold King,
one of the developers of the Master Sample Frame at Iowa State during the 1940s, once stated that he went into private survey work because he could see no possibility that area frame methodology would ever be used in the USDA crop and livestock estimates program. Many improvements and innovations were made in this methodology as it was implemented. We have already mentioned the development of a new frame based on land use strata. Other significant changes included the use of smaller sampling units than those visualized when the Master Sample Frame was constructed. This change, combined with use of the "closed segment" estimator, reduced survey costs with only minimal increases in sampling variance (Hendricks, Searls, and Horvitz [1964]).

Earlier surveys revealed the susceptibility of area sampling to the bias introduced by the "extreme value" or "outlier" problem, particularly for livestock characteristics. This problem was especially troublesome in state estimates because of the relatively large changes in level of estimates it caused from survey to survey (Searls [1963]). Use of "censored" or "truncated" estimators partially solved this problem, but the primary solution came from the use of a list frame in combination with the area sample. The list frame contained the relatively few livestock and poultry operators within each state who were classified as "large." This procedure enabled the sample to cover, with a very small sample of operators, a large part of the population to be estimated, and it stabilized the level of estimates by reducing sample variance at a relatively low cost.

**Crop Yield Estimates and Forecasts**

The area sample provided for the first time a probability sample of fields for which data could be collected from farmers to generate independent estimates of crop yields (Hendricks [1963]). With the development and introduction of objective yield models for forecasting this sample of fields is now used only for estimates at or after harvest.

Project A plans heavily emphasized the improvement of early season yield forecasts for major crops such as corn, cotton, soybeans, and wheat. The greatest contributions to yield estimating during the two decades following 1950 were the development and operational use of objective forecast models based on actual plant measurements and the use of very small sample plots for preharvest observations and eventual harvest. Objective yield forecast models based on cross-sectional surveys of plant characteristics (number of plants per acre, number, size, and weight of fruit, number of nodes) were developed between 1950 and 1970. Data for running these models are obtained from a probability sample of fields enumerated in the area sample. Within
each field small randomly located plots are estimated early in the growing season. These plots are visited once each month to obtain plant counts and measurements until the crop is mature. At this time the plots are harvested and the production is weighed to obtain an estimate of biological yield. To estimate the yield actually harvested by the farmer, a gleaning survey is conducted in the same fields to estimate harvest loss. The net yield is obtained by subtracting the harvest loss from the estimate of biological yield.

The use of very small plots was a major advance in estimating crop yields based on "crop-cutting" techniques. "Crop-cutting" surveys for estimating yields were in use before 1950, but the early surveys usually involved the harvest of entire fields. The procedure was generally applied only in areas where an ample supply of cheap labor was available, which was not the case in the United States. Small plots, on which data collection is relatively inexpensive, can lead to biased estimates unless very precise, well-defined procedures are followed regarding what to include. With wheat plots of about .0001 acres, the inclusion of one additional wheat plant from outside the plot will cause an upward bias of 10,000 in plant population per acre. A major training effort was required to assure that counts and measurements were as precise as possible. In fact, all of the changes in methodology following 1950 have required significant increases in training. Today, approximately one-fourth of the survey budget is devoted to training.

Although most of the early advances in techniques in the United States were associated with the area sample approach which required data collection by personal interviews, some attention was devoted to probability mail surveys, particularly with regard to minimizing bias as a result of nonresponse. Procedures developed for handling this problem are discussed by Hendricks [1949].

**Multiple Frame Estimation Theory**

Surveys involving more than one sampling frame had been used before 1950, but little theory had been developed for two or more frames in a single survey design. Cooperative agreements between the Statistical Reporting Service and H. O. Hartley led to the development of theory for multiple frame sampling (Hartley [1962]). Beginning with the development of "large operator" lists in the Project A area sample, this methodology based on relatively complete lists of all livestock producers is now used as a second major system of probability surveys for livestock and poultry estimates. Also, quarterly probability surveys to obtain farm labor and wage rate data have replaced the monthly nonprobability surveys that previously served this purpose. These surveys include a list of known employers of agricultural labor in conjunc-
tion with an area sample. The area sample estimates that proportion of the population not covered by the list frame. Since the lists used for these surveys are not constructed to contain all farms, extension of this methodology to other surveys has been limited; however, a major effort is currently under way (to be completed by 1978) to build and maintain a list of all farm operators. This will enable the use of multiple frame sampling for practically all major surveys included in the agricultural estimates work of the USDA.

**Advances in Automatic Data Processing**

Paralleling the development in statistical theory and methodology was the rapid change after 1950 in the methods for processing data. All of the changes in agricultural estimating methodology since 1950 are highly dependent on the capabilities of the modern computer for data analysis and reduction. Beginning with a small first-generation computer in 1958, statistical work has employed each new generation of computers as they have evolved. Today, data processing for agricultural estimates involves a network, with each state statistical office and the Washington office tied directly to the same computer. This network enables rapid data transmission between the field and Washington and also provides large-scale computing capabilities to each state statistical office.

**Future Possibilities for Technological Advances**

Work is now under way in two areas which will have a major impact on agricultural statistical methodology by 1990. One area is the collection of data by satellite (remote sensing). The other area involves the development of techniques for physiological crop modeling, employing environmental as well as plant characteristic variables. Complete details on how the research in these areas will be used are not now available, but it is already clear that these methods are potentially of great value in improving crop acreage and production estimates (Ray and Huddleston [1976], Arkin, Vanderlip, and Ritchie [1976]).