Journal of Official Statistics, Vol. 11. No. 2, 1995, pp. 161–180

The Evolution and Development of Agricultural Statistics at the United States Department of Agriculture

Frederic A. Vogel[1]

Statistical information on the supply and location of agricultural foodstuffs is critical to any nation – but especially one in its developing stages. One of the first acts of the newly formed United States Department of Agriculture (USDA) in 1863 was to publish estimates of crop conditions. Estimation procedures used prior to the availability of modern sampling and estimation theories are contrasted with current methodology. Some modern day difficulties are similar to those encountered in the early years and point to the need for continuous research and development activities.

Key words: Forecasts; estimates; probability sampling; sampling frame.

1. Introduction

On July 10, 1863, the United States Department of Agriculture (USDA) initiated monthly crop reports on the condition of crops in 21 states loyal to the Union, plus the Nebraska Territory. The monthly crop reports and other agricultural statistics are issued now by the National Agricultural Statistics Service (NASS). The purpose of this paper is to trace the evolution of the development of statistical procedures in USDA during the past 130 years.

The early reports were based on collected data that were subject to various biases of judgement and sampling. The estimation process was crude and subjective. The ideas of probability sampling and accompanying estimation procedures had not yet been born. Benchmark estimates were provided first by the decennial and later by five-year censuses of agriculture conducted by the Bureau of the Census in the United States Department of Commerce. Forecasts and estimates for the years between census periods were based upon farmers indicating a percentage change from the preceding year which were averaged and applied against either the census base or the previous year to obtain the current estimate. In addition, administrative data such as shipments of fruit and vegetables, receipts at mills and elevators, and sales of livestock were used to revise the preliminary estimates based on farmer reports until new benchmark census data became available. Past comparisons between data reported by farmers and final revised estimates became an increasingly important basis for interpreting and converting current reports from farmers into estimates.

The early forecasts and estimates produced by NASS and predecessor agencies were primarily based on the art of subjectively evaluating survey data, interpreting

[1] National Agricultural Statistics Service, U.S. Department of Agriculture, Estimates Division, Washington, DC 20250-2000, U.S.A.

how survey data fit with knowledge of current weather and marketing trends, and anticipating how the survey data might later match up to administrative data, and finally, the benchmark census data. These early estimates, however, quickly became known for their general accuracy and had an influence on the markets. As the markets became more sensitive to the reports, there was an increasing need to make the estimates and forecasts more accurate. Therefore, from the beginning, improvements in statistical methodology were being continually sought.

Since 1940, the continued and increasing use of probability sampling and survey methods has improved the accuracy of official estimates and reduced the reliance on administrative data and census benchmark data. The improvements in sampling and modern estimation procedures have essentially removed the need for periodic census benchmarks to adjust the official estimates for major commodities. However, the administrative data, for example, on exports, millings, and livestock slaughter, are still used extensively with survey data to produce the current official statistics.

Allen (1992) provides a very detailed account of the statistical defensibility of the joint use of administrative and survey data. The use of administrative data along with survey data to produce official statistics in the USDA has been debated and studied throughout the years. Attempts to improve survey and estimation methodology to reduce reliance on administrative data have significantly improved the quality of the official estimates. However, administrative data continue to be used to improve the accuracy of the official estimates.

The following sections trace the evolution of the estimating procedures used by NASS in the U.S. Department of Agriculture.

2. In the Beginning

As the nation grew and developed in the 1800s, farming activities were pushing westward beyond the Mississippi river and transportation systems were developed along trade routes. There were sectional surpluses of farm products – grain and livestock in the Midwest – cotton and tobacco in the South. The services of merchandising specialists were required to buy, store, ship, and sell the products. As this market economy developed, there was an increasing need for statistics for those in the market to understand price changes and to plan production and distribution activities. The farmers especially felt a need for statistical information because they were generally at the mercy of the merchants who, because of the nature of their business, had more information about supply and prices than did the farmers.

The Census Act of 1839 established the Census of Agriculture (Benedict 1939). The Census of Agriculture, conducted continuously since 1840, was and still is the responsibility of the Bureau of the Census. The Census of Agriculture was conducted every 10 years until 1920 and every five years thereafter. Annual estimates of post-harvest agricultural production were prepared in intervening years by the U.S. Patent Office from 1839 to 1862 when the newly formed U.S. Department of Agriculture assumed responsibilities for all agriculture statistics except the Census of Agriculture.

There was a growing demand during 1839–1862 for pre-harvest measures of the quantities of the crops to be harvested at the end of the year. The most notable

attempt to furnish pre-harvest estimates of crop conditions was carried out by the editor of the *American Agriculturist*, a monthly magazine for farmers (Lee 1952). The editor first published comments about growing conditions based on reports from farmers receiving the magazine. The editor later recruited voluntary reporters from each county to respond to a mail inquiry about conditions in their county. These became the basis for monthly reports during the growing season published in the magazine. This basic concept was adopted by the Department of Agriculture after it was established.

The U.S. Department of Agriculture was established May 15, 1862. A reporting program fashioned on the work of the *American Agriculturist* was initiated. The first monthly crop report was published in July 1863.

On the tenth day of each month, in 1863, (May through October) a circular was mailed to a corps of 2,000 crop correspondents in the 21 states and the Nebraska Territory, whose names came from members of Congress and were generally distributed so that all counties were represented. The questions related to two matters: The average amount sown in 1863 compared with 1862 and the current appearance of the crop. The correspondents were asked to report for their locality rather than their own farms to ensure a greater geographic coverage. Locality was loosely defined and was assumed would represent as large an area as possible that was similar to the respondent's farm. For each crop, numerical answers were given with 10 representing an average of the amount of area sown making each number above or below 10 represent

Table 1. U.S. Department of Agriculture report of corn acreage and general condition July 1863

State	Average amount of corn sown compared with 1862	Appearance of crop at this date
Delaware	12	9
Illinois	11	9
Indiana	10	10
Iowa	12	11
Kansas	10	11
Kentucky	8	10
Maine	9	10
Maryland	10	8
Massachusetts	10	9
Michigan	10	10
Minnesota	13	10
Missouri	11	10
New Hampshire	9	10
New Jersey	11	10
New York	10	10
Ohio	11	10
Pennsylvania	11	9
Rhode Island	10	10
Vermont	10	11
Wisconsin	11	10
Nebraska Territory	8	10
General average	10 1/9	9 1/2

one-tenth of an increase or decrease. The number 10 was also used to represent an average appearance or condition of the crop. The assumption was that farmers would be knowledgeable about their locality and could report whether acreage was increasing or decreasing over the previous year and whether current crop conditions as affected by weather, insects, disease, etc., were above or below average. Table 1 was extracted from the July 1863 report. The full report contained similar data for 10 crops.

These averages were basically simple straight averages. Some early analysis discussed the merits of using the weighted average taking account of area under corn in each state versus the straight average over all states. As the growing season progressed, correspondents were asked to provide updates on the information furnished earlier. At the end of the year they were asked to estimate for their locality the production, yield, and price. The production estimates were determined by the correspondents reporting current year production in tenths of the previous year. Acreage estimates were determined by dividing estimated production by average yields.

By 1866, annual reports were initiated that included estimates of acreage, yield per acre and production of important crops, and numbers of livestock on farms on January 1. The use of percentage estimates began in 1876. The farmers in the "sample" were asked to estimate acreage, total production, and livestock numbers in their locality as a percentage of the previous year rather than in tenths of the previous year.

Prior to the 1880 agricultural census, only information about total crop production and livestock inventories was obtained. The 1880 census also obtained information about crop acreages. These census enumerations of acreage provided benchmarks for estimating crop acreages for years between census years. This was the beginning of the procedure still used in the 1990s to forecast and estimate crop production. The basic procedure is to calculate crop production as the product of the two separate estimates of acreage and yield per acre. Crop acreages once planted usually do not change very much between planting and harvest. There is also less year to year variability between crop acres than there is between yield per acre. In general, the estimates through the 19th century continued to be linked to the decennial Census of Agriculture conducted by the Bureau of the Census. The USDA relied upon correspondents reporting their assessment of year-to-year changes in their locality to make the annual estimates. As might be suspected, small year-to-year biases in the measures of change linked to a census could grow to a widening gap over the years between the USDA estimates and the next census benchmark level. This problem led to improved methodology. Becker and Harlan (1939) provide a detailed review of these early methods.

3. The 20th Century Before Probability Sampling

During the latter part of the 19th century, primary estimation efforts went into increasing the number of voluntary crop reporters. In 1882, the USDA appointed agents in each state to work on a part time basis and to build up the list of crop

correspondents who would report directly to Washington, D.C. The list of county crop reporters had grown to 10,000 people who along with an additional 28,000 township reporters were responding to inquiries from Washington, D.C.

By the late 1800s, the USDA was establishing federal statistics offices in each state with the agents becoming full time employees. These offices began developing their own lists of farm reporters who reported to the state office. Meanwhile, the Washington, D.C. lists of correspondents were also maintained. As the same inquiry was used in each survey, the result from one could serve as a check against the other. Statisticians in the headquarters office had both sources of information to use to establish the official estimates.

The process of reconciling all information into official estimates led to the creation of the Crop Reporting Board in 1905. This is now known as the Agricultural Statistics Board (Allen 1992). The chief statistician (later referred to the Chairperson) would invite two headquarters statisticians and two State Statisticians (originally called agents) to work as a committee to review the data and make the final estimates. This was a subjective process even though it required thorough knowledge of the items being estimated and how the survey data related to other administrative data and census benchmarks. The board "set" estimates that represented a compromise between the different survey results and interpretations.

During this same period, several individual states were independently developing their own Crop Reporting Services. As separate state and federal statistical systems developed, two problems occurred. One was the duplication of effort that occurred when two separate organizations maintained lists of farms and conducted their own surveys. More serious was that the separate state and federal reports did not always agree and caused confusion among those the statistics were meant to serve. As a result, the federal and state governments agreed to combine resources and have one office in each state prepare the agricultural statistics for both the USDA and state governments. These agreements were first initiated in 1917 and have been the basic system used throughout the 20th century.

The most important statistics produced by the department in the early days were the forecasts of the production of crops such as wheat, corn, and cotton followed by end of season estimates of actual production. For reasons given above, the forecasts and estimates of production were determined by separately estimating or forecasting acreage planted and average yields per acre. This procedure is still being used. There was no "sampling frame" of farms; there were only lists of correspondents who would voluntarily respond to a mailed inquiry.

In the absence of probability sampling theory much effort went into improving estimating procedures to measure crop acreages and to forecast crop yields. These procedures are discussed below in chronological order and are described more thoroughly by Becker and Harlan (1939).

3.1. Methods for forecasting probable yield per acre

3.1.1. Par method for estimating probable yield per acre

In 1912, the Par Method was adopted to translate farmer reported crop condition

values early in the crop season into a probable yield per acre that would be realized at harvest. The par method to forecast yield (y) consisted of the following components

$$\bar{y} = \frac{CY_m}{C_m} \quad \text{where}$$

C_m = The previous 10-year average condition for the given month.
Y_m = The previous 10-year average yield per acre realized at the end of the season.
C = Current condition for the given month.

The forecasting model was simply a line passing through the origin and $(C, \bar{y})$. A separate par yield ($\bar{y}$) was established for each state, crop, and month. In actual practice, subjective modification of the means was considered necessary to remove the effects of atypical conditions. For example, a drought which may occur only once every 10–15 years would greatly affect the 10-year average conditions and yield. To aid in these adjustments, 5- and 10-year moving averages were computed to identify unusual situations or trends, and if necessary, exclude the atypical observations.

3.1.2. Regression techniques for estimating probable yields

The development of simple graphic solutions prior to the use of regression and correlation theory was a major breakthrough as a practical means to forecast crop yields and was implemented in the late 1920s. Data for a sufficient number of years had been accumulated so final revised estimates of yields could be plotted against averages of condition reports from farmers for each crop in each state.

$\bar{y}$ = Final revised yield.

C = Crop condition for the given month.

$\hat{\bar{y}} = f(c)$, if the graphical regression of $\bar{y}$ against C happens to be linear (i.e., $f(c) = a + bC$).

Graphical regression techniques provided a consistent method to translate survey data into estimates which in effect adjusted for persistent bias in the data caused by the purposive sampling procedures. This method quickly replaced the par method and was adopted rapidly.

Mathematical methods were not used to fit the regression lines. Instead, lines were fit freehand because the method was not limited to linear relationships and years that fell "off the line" could be studied separately. There was still considerable subjectivity involved in interpreting these survey results to arrive at the yield forecasts.

Beginning in 1926, farmers were also asked to report a probable yield on their farms as well as the conditions in their locality on the inquiry used for the last forecast of the season prior to harvest. These probable yields were also plotted graphically by crop and by state to arrive at the official estimates. After harvest, farmers were asked to report actual average yields harvested.

3.1.3. Objective measurement of yield

Some early work was done to use objective methods to replace the practice of relying

on grower reported locality condition or probable yields. In 1925, a North Carolina statistician submitted a plan for counting the number of cotton plants, bolls, etc., in field plots consisting of 15 feet in a row of cotton. One aspect missing from this early work was an objective random method of sampling fields to remove the selectivity bias. A significant attempt in 1939 and 1940 to remove this bias was to select wheat fields at random along a specified route. From Texas to North Dakota, samples of grain from the selected fields were obtained for computing yield and quality estimates. King, McCarty, and McPeek (1942) documented this methodology.

The following discussion describes early attempts to estimate the acreage to be harvested.

3.2. Methods for estimating acres to be harvested

3.2.1. Ratio estimates

Because the ideas of probability sampling had not yet been formed, procedures used to estimate acres for harvest were more difficult than those to estimate average yields. Although the network of state offices continued to enlarge their lists of farm operators, there was no complete list of farms that could be used for survey purposes. Therefore, the estimating procedures relied upon establishing a base from the most recent Census of Agriculture and estimating the percentage change from year to year. A common procedure during that time was to include two columns in the questionnaire when asking the farmer questions about acreage planted to each crop. During the current survey, the farmer was asked to report the number of acres planted this year and the number of acres planted the previous year in each crop. This method was subject to several reporting biases including memory bias and led to matching "identical" reports from year to year to remove the memory bias. The matching of identical reports did improve the estimates, but was considerably more labor intensive because the name matching had to be done by hand. The process was also complicated by problems with operations changing in size, and was inherently biased because it did not account for new entrants to agriculture.

In the surveys initiated in the 1860s, farmers on the headquarters' lists were asked to report their judgement of the annual percentage change in crop acreages in their locality. Starting in 1888, farmers were asked to report acreages on their individual farms. By 1912, this method had completely replaced the locality questions about acreages. The average change in acreage over the reporting farms computed as a percentage of the previous year was multiplied by the previous year's estimate of acres to obtain the current estimate.

While the use of individual farm acres instead of general locality questions was considered to be a significant improvement, this method was subject to a potentially serious bias caused by the selective or purposive nature of the sample. In an effort to make an allowance for this bias, a relative indication of acreage was developed in 1922. This indicator became known as the ratio relative and contained the following components:

R_1 = Sample ratio of the acreage of a given crop to the acreage of all land in farms (or crops) for the current year.

R_2 = Same as R_1 but for the previous year.

$\hat{y}$ = $(R_1/R_2)*$ Previous year's acres in given crop.

The belief was that this ratio held the bias resulting from the purposive sampling constant from one year to the next. A reported limitation was the extreme variability in the acreage ratios between the sample units. This was countered by increasing the "number" of farms surveyed and weighting the results by size of farm.

In 1928, matched farming units reporting in both years were used to compute the ratio relative. This reduced the influence of the variability between sample units. When looking back at the ratio relative estimator from a current perspective, one may examine the estimate of Rel variance (also assuming probability sampling).

$$CV^2(\hat{y}) = CV^2(R_1) + CV^2(R_2) - 2\,COV(R_1 R_2)$$

This shows why using matching reports improved the ratio relative estimator. However, this did not solve the problem because by using matching reports, farms going into or out of production of a particular crop were not properly represented. Therefore, statisticians continued their efforts to develop a more objective method of gathering and summarizing survey data.

3.2.2. Pole count estimates of acreage

Some statisticians in the early 1920s would travel a defined route on the rural roads or via railway routes and record the number of telephone or telegraph poles opposite fields planted to each crop. The relative change in the pole count for each crop from year-to-year provided a measure of the average change in crop acreage. This method was generally unsatisfactory because large portions of the U.S. still did not have telephone service and the pole count method was therefore not widely used.

3.2.3. Crop meter estimates of acreages

A more refined method of estimating acreage was developed in the mid-1920s. A "crop meter" was developed and attached to an automobile speedometer to measure the linear frontage of crops along a specified route. The same routes were covered each year. This made possible a direct comparison of the number of feet in various crops along identical routes from the current year and the previous year. Hendricks (1942) described some of the properties of this estimator.

3.3. Dilemma of nonprobability surveys

Because of the selective/purposive nature of the samples for the surveys, the determination of the "official" estimates relied heavily upon a subjective appraisal of the survey data as plotted on charts and a reconciliation with administrative data when available.

In the 1930s, demands for more accurate data rapidly increased. The economic depression, the disastrous drought which created a virtual "dust bowl," in many states, Agricultural Adjustment Act programs, and a rapid change in farming

practices challenged the traditional estimating procedures. In 1938, a cooperative research program was initiated with the Statistical Laboratory at Iowa State University to develop sampling and estimation theory to deal with these challenges. Reliable methods not solely dependent on historical relationships as bases were needed for estimation – especially for single-time surveys or periodic surveys.

4. The 20th Century After Probability Sampling

A milestone in the evolution of statistical methodology for agriculture was the development of the master sample of agriculture as described by King and Simpson (1940) and King and Jessen (1945). This was a cooperative project involving Iowa State University, the U.S. Department of Agriculture, and the U.S. Bureau of the Census. This area sampling frame demonstrated the advantages of probability sampling. The entire land mass of the United States was subdivided into area sampling units using maps and aerial photographs. The sampling units had identifiable boundaries for enumeration purposes. The area sampling frame had several features that were extremely powerful for agricultural surveys.

By design, it was complete in that every acre of land had a known probability of being selected. Using rules of association to be described below, crops and livestock associated with the land could also be measured with known probabilities. The Master Sample of Agriculture was based on a stratified design – the strata defined to reflect the frequency of occurrence of farmsteads. Area sampling units varied in size in different areas of the country to roughly equalize the number of farm households in each area sampling unit.

The master sample was used for many probability surveys, but not on a recurring basis because of the added costs since the area samples had to be enumerated in person. The panel surveys of farm operators, while not selected using probability theory, were very much cheaper to conduct because the collection was done by mail. It was not until 1961 that pressures to improve the precision of the official estimates resulted in the U.S. Congress appropriating funds for a national level area frame survey on an annual recurring basis. During the early 1960s, the Master Sample of Agriculture was being replaced by a new area frame that was stratified using land use categories based on the intensity of cultivation of crops. This methodology is still used in the 1990s. The process of developing the area frame is now much more sophisticated relying upon satellite imagery and computer aided stratification as described by Tortora and Hanuschak (1988). The use of the area frame led to the development of some new estimators described below.

4.1. Area frame estimators

The sampling unit for the area sample frame is a segment of land – usually identified on an aerial photograph for enumeration. The segment size generally ranged from 0.5 to 2 square miles depending upon the availability of suitable boundaries for enumeration and the density of the farms. The basic area frame estimator was the design based

unbiased estimate of the total

$$\hat{y}_a = \sum_h \sum_i e_{hi} \cdot y'_{hi} \quad \text{where}$$

y'_{hi} was the ith segment total for an item in the hth stratum and e_{hi} was the reciprocal of the probability of selecting the ith segment in the hth stratum.

During the frame development process, the segment boundaries are determined without knowledge of farm or field boundaries. Therefore, an early (and continuing) difficulty was how to associate farms with sample segments during data collection. Three methods have evolved which are both referred to as methods of association and as estimators. Let y_{hil} be the value of the survey item on the lth farm with all or a portion of its land in the ith sample segment. Then, the different estimators depend on how survey items on farms are associated with the sample segments and follow:

Farm (Open): The criteria for determining whether a farm is in the sample or not is whether its headquarters are located within the boundaries of the sample segment. This was the method used at the inception of the use of the master sample and used until 1992. This estimator was most practicable when farm operations were generally homogeneous, that is, they produced a wide variety of items, some of which may not appear in the segment. This estimator was also useful for items such as number of hired workers and animals born that are difficult to associate with a parcel of land. The extreme variation in size of farms and the complex rules needed to determine whether the farm headquarters were in the segment led to the demise of the farm estimator.

$$y'_{hi} = \sum_l F_{hil} y_{hil}$$

where $F_{hil} = 1$, if the operator of farm l lives in the segment; 0 otherwise.

Tract (Closed): This concept was first tried in 1954. The tract estimator is based on a rigorous accounting of all land, livestock, crops, etc., within the segment boundaries regardless of what part of a farm may be located within the boundaries of the segment. The method offered a significant reduction in both sampling and nonsampling errors over the farm method, because reported acreages could be verified by the map or photograph. The estimator is robust in that the maximum amount that can be reported for a segment is limited by its size. This estimator is especially useful for measuring acres in specific crops.

$$y'_{hi} = \sum_l T_{hil} y_{hil} \quad \text{where } T_{hil} = \frac{\text{Amount of item on farm } l \text{ in segment } i}{\text{Total amount of item on farm } l}$$

Weighted: The difficulty with the tract estimate was that some types of information, such as economic, could only be reported on a whole-farm basis. This led to the development of the weighted procedure in the late 1960s. In this approach, data are obtained on a whole-farm basis for each farm with a portion of its land inside a sample segment. The whole farm data are prorated to the segment based on the proportion of each farm's land that is inside the segment. This estimator provided

the advantage of a smaller sampling error than either the farm or tract procedures. On the minus side, data collection costs increased 15–20 percent because of increased interviewing times, and intractable nonsampling errors are associated with determining the weights. This estimator is used to estimate livestock inventories, number of farm workers, and production expenditures.

$$y'_{hi} = \sum_{l} W_{hil} y_{hil} \quad \text{where } W_{hil} = \frac{\text{Acres of farm } l \text{ in segment } i}{\text{Total acres in farm } l}$$

Ratio: The area frame sample was designed so that 50%–80% of the segments were in the sample from year to year. This allowed the computation of the usual ratio estimators.

4.2. Multiple frame estimator

While the area frame sampling and estimating procedures were being refined in the 1950s, this period also saw a rapid change in the structure of agriculture. Farms became more specialized and much larger. This introduced more variability that required much larger area frame sample sizes.

The proportion of farms having livestock was decreasing rapidly during this period. The variation in numbers of livestock on farms with livestock also had increased dramatically.

The combination of these two factors meant that either resources for an extremely large area frame sample would be needed or alternative sampling frames were needed. In the early 1960s, H.O. Hartley at Iowa State University was approached about this problem. The result was his 1962 paper laying out the basic theory of multiple frame sampling and estimation and summarized in Cochran (1977, pp. 145–146). Cochran (1965) more fully developed the concepts of multiple frame sampling and estimation methodology. Fuller and Burmeister (1972) and Bosecker and Ford (1976) also developed multiple frame estimators with reduced variances.

As implied by its name, multiple frame sampling involves the use of two or more sampling frames. If there are two frames, there are three possible post-strata or domains – sample units belonging only to frame A, sample units belonging only to frame B, and finally the domain containing sample units belonging to both frames A and B. As pointed out by Hartley, the sampling and estimation theory to be used depended on knowing in advance of sampling whether the domain and frame sizes were known. This determined whether theories applying to post-stratification or domain estimation were to be used.

In the agriculture situation, the area sampling frame provided 100% coverage of the farm population. There was also a partial list of farms which could be stratified by size or item characteristic before sampling. Domain membership and sizes are unknown prior to sampling, thus sample allocation is by frame and domain estimation theories apply. The theory requires that after sampling, it is necessary to separate the sampled units into their proper domain. This meant area sample units had to be divided into two domains:

- Farms not on the list.
- Farms on the list.

By definition, all farms represented by the list were also in the area frame.

The Hartley estimator for this situation was

$\hat{Y}_H = N_a(\bar{y}_a + P\bar{y}'_{ab}) + N_b Q\bar{y}''_{ab}$ where

$\bar{y}_a$ represents area sample units not on the list, $\bar{y}'_{ab}$ represents area sample units overlapping the list frame and $\bar{y}''_{ab}$ representing the list frame and $P + Q = 1$.

The weights, (P and Q), were to be determined to minimize var ($\hat{y}_H$). This sampling and estimation theory was used in surveys to measure farm labor numbers and wage rates, and also to estimate farm production expenditure costs. Both involved lengthy questionnaires requiring personal interviews and were annual surveys. Because of the considerable variation in the sizes of farms and the sampling efficiencies that occurred from the stratification in the list frame, the majority of the weight went to the list frame portion of the estimator, that is, P was small and Q was large.

Hartley (1962) suggested an alternative to his estimator shown above. With the alternative estimator, units on the list frame that are in the area frame sample are screened out of the area frame portion of the survey. In other words, $P = 0$ and

$\hat{Y}_H = N_a\bar{y}_a + N_b\bar{y}''_{ab}$.

Additional analysis by Cochran (1965) suggested that for a fixed cost, the screening estimator would have the lower variance whenever the cost of sampling from the list frame is less than the difference between the cost of sampling from the area frame and the cost of screening the area frame sample to identify those also in the list frame.

For those reasons, the screening estimator is used exclusively today. The increased use of telephone enumeration for the list sample reflects personal to telephone enumeration cost ratios of 1 to 15 in some cases. The area frame sample is surveyed in its entirety in June each year. Farms that overlap the list frame are screened out and the area domain representing the list incompleteness is defined. During the next 12-month period, NASS conducts a series of multiple frame quarterly surveys to measure livestock inventories, crop acreages and production, and grain in storage. Other multiple frame surveys during the year cover farm labor and production expenditures. Each survey relies upon the multiple frame screening estimator.

This basic methodology has stood the test of time. Considerable changes have been made in sampling methodologies within sampling frames and the content of the surveys, but the fundamental Hartley estimators still form the backbone of the estimating procedures for the agricultural statistics program.

The domain determination as discussed by Vogel (1975) has been the most difficult operational aspect to tackle in developing, implementing, and using multiple frame methodology. As the structure of farms becomes more complicated with complex corporate and partnership arrangements, the survey procedures require a substantial effort to minimize nonsampling errors associated with domain determination.

4.3. Crop yield forecasts after probability sampling

Ever since the first crop report was issued in 1863, the early season forecasts of crop

production continued to be some of the most critical and market sensitive information prepared by the USDA. The development of probability sampling theory and the area sampling frame provided a foundation upon which to replace judgement based estimates of locality conditions to forecast yields per acre. In 1954, research was initiated to develop forecasting techniques based on objective counts and measurements that would be independent of judgement based estimates. The use of nonrepresentative samples of farmers continued to be used to report conditions in their locality and individual farms during this period, however.

Research on the use of corn and cotton objective methods began in 1954 followed by work on wheat and soybeans in 1955 and sorghum in 1958. Early results showed that a crop cutting survey at harvest time based on a probability sample of fields would provide estimates of yield per acre with good precision. There were two difficulties. One difficulty is to forecast yield before the crop is mature, and even more difficult before the plants have set fruit. The basic procedures that follow were developed and are largely still in place in the 1990s.

A two-step sampling procedure is used. First, a sample of fields is selected from those identified during the annual area frame survey as having the crop of interest. Self-weighting samples are selected. Observations within fields are made in two randomly located plots with each selected field. Selected plots for most crops include two adjacent rows of predetermined length. The probable yield per acre is a function of the number of plants, the number of fruit per plant, and the size or weight of the fruit. Early in the crop season, the number of plants are used to forecast the number of fruit, with historical averages used for fruit weights. After fruit are present, several measurements are obtained to project final fruit weight. For example, the length and diameter of corn ears are obtained from ears within the sample plots. When the crop is mature, the sample plots are harvested and the fruit counted and weighed for the final yield estimate. The early season counts and measurements from within the sample plots are combined with the data from the harvested fruit and become part of a data base that is used to develop forecasting models in subsequent years. After the farmer harvests the sample field, another set of sample plots is located and grain left on the ground is gleaned and sent to a laboratory where it is weighed and used to measure harvest loss. During the forecast season, historical averages are used to estimate harvest losses.

Simple linear and multiple regression models are used to describe past relationships between the prediction variables and the final observations at maturity. Typically, early season counts and end of season harvest weights and counts from within each unit are used. They are first screened statistically for outlier and leverage points as described by Beckman and Cook (1983). Once these atypical data are identified and removed, the remaining data are used to create current forecast equations.

The basic forecast models for all crops are essentially the same in that they consist of three components: the number of fruit, average fruit weight, and harvest loss.

The net yield per acre as forecast for each sample plot is computed as follows:

$$\bar{y}_i = (F_i \times C_i \times W_i) - L_i$$

Where:

$F_i =$ Number of fruit harvested or forecast to be harvested in the ith sample plot.
$C_i =$ Conversion factor using the row space measurement to inflate the plot counts to a per acre basis.
$W_i =$ Average weight of fruit harvested or forecast to be harvested.
$L_i =$ Harvest loss as measured from post-harvest gleanings (the historical average is used during the forecast season).
$\hat{\bar{Y}} = \sum(y_{i/n})$ for the n sample fields.

Separate models are used to forecast the number of fruit (F_i) to be harvested and the final weight (W_i). The variables used in each model vary over the season depending upon the growth stage at the time of each survey. At the end of the crop season, F_i and W_i are actual counts and weights of fruit for harvest.

The major contributor to the forecast error is the difficulty of forecasting fruit weight early in the season. Many factors such as planting date, soil moisture, temperatures at pollination time, etc., crucially affect a plant's potential to produce fruit. While the number of fruit can be counted early in the season, the plant does not always display characteristics that provide an indication of final fruit weight. While each plant's potential to produce fruit is affected by previous circumstances, that information is locked inside the plant – often until fruit maturity.

Over the years, the USDA conducted extensive research to improve the basic yield forecast models. Examples of this work appear in Arkin, Vanderlip, and Ritchie (1976). Models using weather data were continuously being developed and compared against the traditional objective yield models, but have always fallen short. The plant measurements reflected the effects of weather and the use of weather data does not add to the precision. Another effort involved an attempt to model the plant growth and to use these models for yield forecasting. These models, known as plant process models, did not prove to be feasible to use in a sample survey environment (Gleason 1982).

The basic objective yield surveys and forecasting models as developed in the 1950s are still being used in the 1990s. The use of a sample of farmers to report also has continued. There have only been two significant changes in the farmer survey since 1954. Starting in the late 1980s a probability sample of farms was selected from the large multiple frame survey conducted in June to estimate acres planted.

The other significant change is that the farms report expected yields only for their farm. The farm survey is still used each month along with the objective yield survey to forecast yields. Because of cost considerations, the objective yield surveys are only conducted in states producing a major portion of the crop. For example, the corn objective yield survey is only conducted in 10 states which collectively account for about 85% of the U.S. corn production. The farm survey is less expensive because data are collected by mail and telephone and is conducted in every state each month during the crop season.

The biggest problem facing both the statistician and the farmer in projecting yields is the uncertainty about future weather. That was true in 1863 and is still true in the

1990s. Any significant improvements in crop yield forecasting methodology will probably be closely connected to improved weather forecasts.

5. Current Issues

As farms become larger and more specialized, two estimation problems became critical. The primary reason these are troublesome is that most surveys were designed to produce totals rather than means. These involve imputation for missing data and adjustments for outliers. A third problem involves variance estimation for the complex sample designs being used. The imputation problem has received more attention and will be discussed first.

5.1. Imputation

Research by Bosecker (1977), Ford (1976, 1978) along with papers by Fellegi and Holt (1976), Platek and Gray (1985), and Kovar and Whitridge (1995) provide much useful information about the imputation problem. In the early 1970s, the "hot deck" procedure was developed and implemented into the Quarterly Agricultural Labor Survey. This survey provided quarterly estimates of numbers of farm workers by type of work, method of payment, and wages paid. The "hot deck" in this case did not make use of a single randomly selected closely matched donor. It was basically a large matrix consisting of moving averages of number of workers and wages paid from previous reports. The matrix had separate cells for type of work and method of payment.

The most obvious weakness of this method was that the sampling errors of the resulting estimates were understated because imputation was for individual farms which were further processed assuming the data had been actually reported. Also, the imputation method did not take into account the complex multiple frame design. The largest farm (if a nonrespondent) could receive the average of the most recent three reports regardless of their sizes or types.

The next imputation procedure used was developed by Crank (1979). Imputation was not on an individual farm basis as estimates for nonrespondents were obtained by treating them as a group or domain. The estimator for the nonresponse domain was based on two assumptions:

1. It is possible to determine for nonrespondents whether or not they have the item of interest.
2. The distribution for respondents *with* the item of interest will also represent the nonrespondents known to have the item of interest.

$$\hat{y}_h = \frac{N_h}{n_h}\left[(n_h^p + n_h^{rk})\bar{y}^p + n_h^u \bar{y}^r\right]$$

where $\bar{y}^p$ is the mean of positive sample units and $\bar{y}^r$ is the mean of all reporting sample units including those that do not have the item of interest.

Also, n_h^p, n_h^{rk}, and n_h^u are sample counts in the hth stratum of the number of sample units that, respectively, had the item of interest (n_h^p); that were nonrespondents but

were known to have the item of interest (n_h^{rk}); or were nonrespondents and it was not known whether they had the item of interest n_h^{ru}. The sample mean of respondents with positive data ($\bar{y}^p$) is weighted separately from the sample mean of all respondents ($\bar{y}^r$). In other words, nonrespondents known to have the item of interest are in effect given imputed values equalling the mean of respondents with the item. Nonrespondents whose status is unknown receive the overall average of all respondents.

One can see, after careful examination of the components, that the overall estimate is sensitive to the breakdown between nonrespondents whose status is known and those whose status is unknown in addition to the values used to estimate for them. The use of a new sample or a change in survey procedures can change the number of nonrespondents and also the number identified to have the item of interest.

A refinement of the Crank estimator has been developed which, similar to the "hot deck" procedure, in effect, imputes means for missing farms. It relies on the assumption underlying the Crank estimator that it is possible to determine a minimum amount of information for the missing records, i.e., whether or not they have the item of interest. Reported data within each sampled stratum are poststratified by geographic subregions which are contiguous groupings of homogeneous counties. A typical state will have seven to nine such regions. Means for positive reports $\bar{y}^p$ and all reporting operations y^r are computed as before, but by each separate region. The appropriate mean is then used to impute for a missing record.

For example, a missing record known to have an item of interest receives the mean for all positive reports lying in the same stratum and subregion as the missing record. Variance estimates are computed using reported and missing records alike. This understates the variance, but at a minimal level because the poststratified means introduce variability.

A closely related problem, but also one becoming more critical as farms become larger and more diverse, is the problem of outliers or extreme observations.

5.2. *Outliers*

Outliers are observations that have an undue influence on the survey estimate and sampling error. In agricultural surveys, outliers generally occur several ways:

- An operation that greatly increased in size between the time the sample frame was developed and the survey was conducted.
- An extremely large operation that was incorrectly classified in the sample design process and thus assigned to a sampled stratum.
- An ordinary operation that is assigned or falls into a stratum or Primary Sampling Unit that has an extremely small probability of selection (large expansion factor). A typical example is an urban segment in the area sample that unexpectedly contains an agricultural operation.

Several procedures are used in NASS to deal with outliers. Although survey data are subjected to thorough computer aided edits, a final step is a specific outlier review. First, the reported data are multiplied by the reciprocal of the probability of selection and the largest 44 data values are listed. For repetitive surveys, another listing shows the largest differences between the current and previous survey. The purpose of these

reviews is to search for errors that had not been previously corrected and to make corrections. Procedures used to identify outliers in the crop objective yield survey differ from those used in the multiple frame surveys.

Crop counts and measurement data from the objective yield surveys are analyzed using procedures described by Beckman and Cook (1983). Data points that individually have a significant effect on model coefficients are identified as outlier points or as leverage points and are deleted.

A form of a trimmed estimator is used to adjust for outliers found in multiple frame surveys. Distributions of historical reported data (10 years or more) expanded by the sampling weight are examined to identify cut-off values. These cut-off values are determined such that there is a less than 1% probability of its being exceeded considering the historical distributions. These cut-off values remain fixed over time. Current survey values exceeding these cut-off values are assigned the cut-off value. Data points from a sampling unit selected with certainty are exempted from this test. This procedure is an adaptation of the method described in Searls (1966).

5.3. Variance estimation

The sample designs used for the multiple frame surveys and objective yield surveys use stratified, multiple stage sampling within sample frame. The survey design involves a combination of cluster sampling, poststratification, and subsampling. These designs lead to unbiased and relatively efficient estimators. The variances of these estimators are difficult to estimate – in some cases design unbiased estimation of the variances is impossible.

The first attempts at variance estimation for the agricultural surveys assumed simple random sampling with no replacement. Some early work on variance estimation was done by Cochran and Huddleston (1969). These estimators were appropriate for the sample designs used at that time which were more single frame oriented. Kott and Johnston (1988) showed that these underestimated variances for current sample designs, and suggested new estimators.

Recent contributions by Francisco and Fuller (1986) also show that the variances used for the objective yield estimates are understated. They suggested an improved estimator and also suggested changing the sample design to permit unbiased estimation of the variance.

5.4. Other issues

During the period following the development of the area frame and the implementation of multiple frame sampling and estimation, the goal was to produce agricultural estimates that were statistically defensible and with less reliance upon administrative data. Considerable effort went into improving the area frame design – improving stratification using Landsat satellite imagery Hanuschak, Allen, and Wigton (1979, 1982) and implementing replicated sampling Fecso, Tortora, and Vogel (1986) are examples. The coverage of the list frame was expanded, record linkage routines to detect duplication were developed, and integrated surveys were implemented for cost and sampling efficiencies. A program to identify and minimize nonsampling errors was initiated.

Despite these efforts, the final determination of the official estimates is still based on the informed appraisal of the probability estimates and administrative data. For example, administrative data on wheat exports and millings are used along with survey data to produce the official estimates of wheat in storage periodically during the marketing season. The approach is to start with the total supply which is the previous year's production estimate and subtract the amounts exported and milled. Since the previous year's production estimate and the current survey estimate of quantities in storage are both subject to sampling and nonsampling errors, the measure of the quantity of wheat in storage derived by subtracting administrative data from production will differ from the current survey estimate of wheat in storage. At the end of a crop marketing season, the administrative data are used to evaluate the production estimate and revise it if necessary. Revisions beyond the range of sampling variability are still the reason for increasing research efforts to improve the estimators.

6. Look to the Future

This paper has traced the history of estimation methodology at NASS. Since the early 1960s, significant improvements have been made in sampling and survey methodology. Despite significant developments in statistical methodology, basic procedures to determine official estimates have remained essentially unchanged. Much reliance is still placed on the use of administrative data and balance sheets to evaluate and modify survey estimators.

This problem will not go away. First, the demand for current, accurate statistical information is insatiable. However, resources for federal statistical programs are not keeping up with inflation; this hampers efforts to produce "stand-alone" estimates. Therefore, reliance will continue to be placed on administrative data. More effort will go into developing data analysis procedures to extract information from both survey and administrative data to better explain the causes and sources of the ups and downs of livestock and crop production from one period to the next. For example, was an increase in livestock inventories caused by new producers or existing producers increasing herd sizes? Each has a possible different implication about future inventory levels or the length of the production cycle.

Estimators that remain stable in the presence of outliers are needed. Agricultural operations will continue to become larger, more complex and more specialized. Structure will change faster than sample frames can be updated.

Historically, currently, and in the future, the most market sensitive statistics will be the crop production forecasts. As satellite weather data produce better weather forecasts and more timely weather data, forecast models to improve the accuracy of the forecasts will be needed. The probability sample designs, survey, and estimating procedures have been developed to produce state and national estimates. County and local area estimates that are made available are still based upon large scale non-probability survey data. A bridge between these two data sources is needed to produce improved county estimates. As the "information float" shortens the time span in which data are most useful and as markets continue to become even more

time sensitive, there will be an increasing need to shorten the time span between data collection and dissemination of the results.

From a statistical estimation standpoint, agriculture involves many challenges. It has very diverse content and size distributions. Farms change size on a seasonal basis. Many of the commodities that are produced are perishable which presents difficulties in tracking the flow through the marketing system. Because of spoilage, grading, etc., amounts finally processed or marketed will differ considerably from the amount actually produced. The next decade and the next century will continue to offer challenges.

7. References

Allen, R. (1992). Statistical Defensibility as used by U.S. Department of Agriculture, National Agricultural Statistics Service. Journal of Official Statistics, 8, 481–498.

Arkin, G.F., Vanderlip, R.L., and Ritchie, J.T. (1976). A Dynamic Grain Sorghum Growth Model. Transactions of the American Society of Agricultural Engineers, 19, 622–630.

Becker, J.A. and Harlan, C.L. (1939). Developments in the Crop and Livestock Reporting Service Since 1920. Journal of Farm Economics, 21, 799–827.

Beckman, R.J. and Cook, R.D. (1983). Outliers. Technometrics, 25, 119–149.

Benedict, M.R. (1939). Development of Agricultural Statistics in the Bureau of the Census. Journal of Farm Economics, 21, 35–60.

Bosecker, R.R. (October 1977). Data Imputation Study on Oklahoma DES. U.S. Department of Agriculture, Statistical Reporting Service.

Bosecker, R.R. and Ford, B.L. (1976). Multiple Frame Estimation with Stratified Overlap Domain. Proceedings of the Social Statistics Section, American Statistical Association, 219–224.

Cochran, R.S. (1965). Theory and Application of Multiple Frame Sampling. Ph.D. Dissertation, Iowa State University, Ames, Iowa.

Cochran, R. and Huddleston, H. (1969). Unbiased Estimates of Agriculture. Statistical Reporting Service, U.S. Department of Agriculture.

Cochran, W.G. (1977). Sampling Techniques, Third Edition, New York: John Wiley.

Crank, K.N. (1979). The Use of Current Partial Information to Adjust for Nonrespondents. U.S. Department of Agriculture, Statistical Reporting Service.

Fecso, R., Tortora, R.D., and Vogel, F.A. (1986). Sampling Frames for Agriculture in the United States. Journal of Official Statistics, 2, 279–292.

Fellegi, I.P. and Holt, D. (1976). A Systematic Approach to Automatic Edit and Imputation. Journal of the American Statistical Association, 71, 17–35.

Ford, B.L. (1976). Missing Data Procedures: A Comparative Study. U.S. Department of Agriculture, Statistical Reporting Service.

Ford, B.L. (1978). Nonresponse to the June Enumerative Survey. U.S. Department of Agriculture, Statistical Reporting Service.

Francisco, C.A. and Fuller, W.A. (1986). Statistical Properties of Crop Production Estimators. Report on Cooperative Research with the U.S. Department of Agriculture, Statistical Reporting Service.

Fuller, W.A. and Burmeister, L.F. (1972). Estimators for Samples Selected from Two Overlapping Frames. Proceedings of the Social Statistics Section, American Statistical Association, 245–249.

Gleason, C.P. (1982). Large Area Yield Estimation/Forecasting using Plant Process Models. Paper presented at the Winter Meeting of the American Society of Agricultural Engineers.

Hartley, H.O. (1962). Multiple Frame Surveys. Proceedings of the Social Statistics Section, American Statistical Association, 203–206.

Hanuschak, G.A., Allen, R.D., and Wigton, W.H. (1982). Integration of Landsat Data into the Crop Estimation Program. International Symposium on Machine Processing of Remotely Sensed Data, Purdue University.

Hanuschak, G., Sigman, R., Craig, M., Ozga, M., Luebbe, R., Cook, P., Kleweno, D., and Miller, C. (1979). Obtaining Timely Crop Area Estimates Using Ground-Gathered and Landsat Data. Technical Bulletin No. 1609, USDA, Washington, D.C.

Hendricks, W.A. (February 1942). Theoretical Aspects of the Use of the Crop Meter. Agricultural Marketing Service, USDA.

King, A.J. and Jessen, R.J. (1945). The Master Sample of Agriculture. Journal of the American Statistical Association, 40, 38–56.

King, A.J., McCarty, D.E., and McPeek, M. (1942). An Objective Method of Sampling Wheat Fields to Estimate Production and Quality of Wheat. U.S. Department of Agriculture, Technical Bulletin No. 814.

King, A.J. and Simpson, G.D. (1940). New Developments in Agricultural Sampling. Journal of Farm Economics, 22, 341–349.

Kott, P.S. and Johnston, R. (1988). Estimating the Nonoverlap Variance Component for Multiple Frame Agricultural Surveys. National Agricultural Statistics Service, U.S. Department of Agriculture Staff Report SRB-88-05.

Kovar, J.G., and Whitridge, P.J. (1995). Imputation of Establishment Survey Data, 403–423. In Business Survey Methods, Cox, B., Binder, D.A., Chinnappa, N., Christiansson, A., Colledge, M., and Kott, P. (eds.). New York: John Wiley.

Lee, I.M. (1952). A Critical Evaluation of Available Agricultural Statistics. Journal of the American Statistical Association, 47, 269–280.

Platek, R. and Gray, G.B. (1985). Some Aspects of Nonresponse Adjustments. Survey Methodology, 11, 1–14.

Searls, D.T. (1966). An Estimator for a Population Mean which Reduces the Effect of Large True Observations. Journal of the American Statistical Association, 61, 1200–1204.

Tortora, R.D. and Hanuschak, G.A. (1988). Agricultural Surveys and Technology. Proceedings of the Survey Research Methods Section, American Statistical Association, 63–68.

Vogel, F.A. (1975). Surveys with Overlapping Frames – Problems in Application. Proceedings of the Social Statistics Section, American Statistical Association, 694–699.

Received April 1989
Revised December 1994