THE REDESIGN OF THE AGRICULTURAL LAND VALUES SURVEY

Ron Fecso, National Agricultural Statistic Service, USDA
3251 Old Lee Hwy., Room 305, Fairfax, Va. 22030-1501

Key Words: area frame, spatial correlation

INTRODUCTION

Statistics on agricultural land values are made annually by the U.S. Department of Agriculture's Economic Research Service. A benchmarked estimator is used which is a composite of data from several sources. The Agricultural Census provides the benchmark every five years, with estimates of the average value per acre of all land and buildings. This benchmark data is adjusted annually using a year-to-year change ratio. The change ratio used until 1995 was based on data collected by the National Agricultural Statistics Service in the Agricultural Land Values Survey (ALVS), a list frame survey conducted in January. The ALVS was designed to produce survey estimates of certain land values and rents for farmland in the 48 conterminous states in the U.S. Specific estimates include: the average market value of cropland, pasture or grazing land, and on-farm woodland; and rents for cropland and pasture.

Operationally, the ALVS did not collect data on land and building values nor the acres or value of other fields such as ponds, feedlots, and airstrips. It focused only on the collection of data needed to estimate year-to-year ratios for the land-use items. The missing components are needed to arrive at the total value of agricultural holdings. Thus, benchmarking estimators are used because a direct estimate of total and per acre real estate value from the ALVS alone was not possible.

The benchmark estimation strategy begins with estimates of the change in value per acre for cropland, pasture and woodland. These estimates are derived from year-to-year ratios of the average value estimates from the ALVS. Land-use items were weighted by Census acres in the type of land to produce a change ratio for these three major agricultural land uses combined. These ratios were then applied to the corresponding prior year estimate (or the benchmark) to produce current year estimates.

Despite many changes to the survey over the years, data users and collectors came to a consensus that the underlying list-frame survey methodology that produced these statistics needed improvement. The potential design problems included undercoverage, weakly specified concepts of value, uncertain association rules, and high nonresponse rates. The complex nature of the estimate and limited study of the errors in the associated data made assessment of the quality of the data series difficult at best.

To improve the quality of official statistics about agricultural land value and rents received for agricultural land, an interagency team was formed to discuss methodology to improve quality. The redesign presented in this paper was proposed as an alternative to the Agricultural Land Values Survey (ALVS) design. Data was first collected using this alternative design in 1994. Based on the success of the initial experience with the redesigned survey, collection of ALVS data was switched to the alternative design in 1995 and data collection with the prior methodology ended.

This paper compares the survey designs and how the redesign is an improvement over the prior design. Criteria considered in the decision to redesign, such as coverage, specification of concepts, sampling frame choice, response rates, and the editing and imputation of the data, are briefly discussed.

WHY CONSIDER AREA FRAME SAMPLING?

The ALVS was a standalone mail survey using a list frame sample and some telephone follow-up. The redesign incorporated the ALVS data collection as an add-on to the June Quarterly Agricultural Survey (QAS), an area frame sample with an in-person interview. The redesign aims to improve data quality by improving the measurement of quality issues such as sampling error and through increased control of nonsampling error sources. For example, there were no formal quality control activities during list data collection, but supervisors at all levels were instructed to assure that the appropriate survey procedures were being carried out. In contrast, area sampling has a formal quality control process for enumeration.

AREA FRAME SAMPLING

Area frame construction and sampling are complex processes, but NASS's area frame can be conceptualized easily. First, consider the 48 conterminous states as subdivided into areas of land (called segments) from one-quarter to one mile square. Also, any area of land must be found in some segment and no area of land can be in more than one segment. That is, a one-to-one relationship exists between any point on the map and a segment. Thus, any measurable characteristic that can be uniquely associated with the land areas inside the segments can be estimated with an appropriate sample from this frame and an appropriate estimator. For example, estimates of the total area of land planted to soybeans, the total or average value of that land, or the number of hogs on that land at a specified point in time are possible.

Area frame segments are grouped into a hierarchical stratification. The highest level of stratification, a state, contains all the segments within a state. Within each state, segments are subgrouped into land-use strata based on classifications such as intensive cultivation, ag-urban, and pasture/range. The lowest level of stratification, substratum, groups the segments within a land-use stratum into geographically adjacent or nearly adjacent segments.

This hierarchical stratification serves three purposes. First, required state level estimates are facilitated. Secondly, frame development is partitioned into manageable blocks. Thirdly, many desired parameters associated with agriculture are more efficiently estimated because the activity often varies by state, land use, and geographic area. Thus, while constructed to be a multipurpose frame, the design is quite efficient for many items. For example, land values vary by state, the type of crop the land can support, and proximity to large cities or recreational areas. These characteristics have very strong spatial correlation. Thus, an area frame would be expected to provide design efficiencies when estimating land values. Further, the geographic control of the sample should prove useful in studying nonresponse and other data quality issues. Finally, approximately 80 percent of the segments are retained from one year to the next in a rotating sample scheme. This makes the design attractive for temporally correlated data such as land value.[1]

The change to an area sampling frame was made based on the potential for data quality improvement. Besides having the ability to compute sampling errors by using the new design, other nonsampling errors could be dealt with more effectively. Anticipated benefits from using an area sampling frame included: improved coverage of the target population, improved spatial distribution of samples, reduced nonresponse, use of spatial information to improve editing and corrections for nonresponse, reduced between-year variance using panel aspects of NASS's area design, and reduced response variance and/or bias by better specifying the reporting unit land.

Table 1 -- DESIGN CHANGE COMPARISON

	OLD	NEW
FRAME	LIST	AREA
MODE	MAIL/PHONE	FACE-TO-FACE
TYPE	STANDALONE	ADD-ON
TIMING	JANUARY	JUNE
ENUMERATION QUALITY CONTROL	INFORMAL	FORMAL

[1]For additional detail about NASS frames, see Fecso, et al., 1986 and Cotter and Tomczak, 1994.

AREA FRAME DATA

The agricultural value questions were added to an ongoing survey, the area frame portion of the June Agricultural Survey (JAS). The additional survey questions were confined to one page that contained two main parts. The parts were:

Tract Acres (Land and Buildings Values) and
Cash Rents (For Land Inside the Tract).

The current questionnaire version includes opinion questions on how much the land has changed in value over the year (for editing across years) and excess development value potential (for identifying valid extreme reports).

The second part of the agricultural land values section screens for land rented for cash inside the tract. The number of acres being rented strictly for cash is recorded for cropland and for pasture. The cash dollar amount per acre (paid by tract operator on a calendar year or crop year basis) is requested. For cash rent items, the market values reported for the tract acres are assumed to apply to the rented acres.[2] Thus, the data items are available to create estimates for rent per acre and ratios of rent to land market values.

The tract land and buildings value part focuses on the collection of data needed to develop an alternative to the year-to-year ratio-benchmarked estimator. A direct estimate of total real estate value and value per acre from this one data source is a future alternative to the current benchmark-ratioed estimate.

Beginning in 1995, the change in value per acre for cropland, pasture and woodland was estimated using a ratio estimator with data from matched segments retained from the prior year JAS. This data replaces the land value data collected from list frame samples until

[2] In 1994, values were asked for both tract acres and rented acres. Analysis of the reports revealed no differences between the values. This is reasonable to expect since there should be a strong spatial correlation. Thus, the working group agreed to remove the rented acres value questions as a burden reduction effort.

1994 for use in estimating the change in value for cropland, woodland, and pasture land.

Due to question wording changes, matched sample data for the total value reported for dwellings, buildings, other structures in the tract will not be available until the 1996 survey. Thus, this paper focuses only on the three land-use values: cropland, pasture, and woodland.

After 1996, a ratio incorporating data on structures can be evaluated to replace the current ratio based on land value only. Further, the acres in other fields (such as ponds, feedlots, and airstrips) and the value of these acres will be available and might be used to derive direct estimates of the total value of agricultural holdings. If the anticipated quality improvements are achieved, the survey might be used to set the level of agricultural land values independently of the complex estimator currently used or as a composite with the Census benchmark data.

FRAME COVERAGE ERROR

Coverage error attributable to the sampling frame is the deviation of the sampling frame from an unduplicated and complete "listing" of the target population. The "listing" may be of primary sampling units or clusters of population units rather than an individual listing. Yet, the idea remains the same--if a census of the sampling units using the survey data collection methodology would provide information from each unit in the population once and only once, then there is no frame coverage error.[3]

When the sampling frame is not a simple listing of the desired population units, rules are necessary to associate the population of interest to the reporting units on the sampling frame. The ALVS is such a case. The population is agricultural land while the reporting units on the frame are people associated with agriculture near the target land area. When there is a well-defined relationship between frame units and units in the population of interest, usually a one-to-one or cluster sampling relationship, estimators can be developed to account for the relationship, thus eliminating that frame problem. Unfortunately, land areas are not clearly associated with ALVS reporters from the list frame and potential biases occur. Essentially, one cannot show that the list frame estimator is consistent, while the area frame sampling estimator is consistent.

The association of land with reporters is well defined. Survey respondents in the June Agricultural Survey were asked opinions of land values for land operated in the agricultural tract they operate. With area frame sampling, all land areas have a selection opportunity that is positive and unique. Thus, frame coverage error for the main land types or for the values for "other land" and "buildings and capital improvements" are virtually nonexistent since they too are identified during data collection.

RESPONSE RATES

While the list design asked respondents about values in their "locality," the area design gets more personal by asking specifically about their operation's value. During the design stage, some team members expressed concern that asking for information representing a significant portion of a person's gross worth could be nonresponse inducing and thus negatively affect NASS's major survey, the JAS. Yet, other farm surveys request much more detail about assets. Thus, a major design decision criterion was response rate. The list frame surveys had about a 40 percent response rate, while about 70 percent of the land value sections in the area survey have useable data. While a few respondents voiced dislike of the new questions, there was no evidence that these questions were more disliked by respondents than other questions in the interview. Response rates for the survey did not change because of the added questions.

More important than the rate itself, nonresponse bias is better understood. With the list frame survey, estimation was based on an implicit missing-at-random assumption. Nothing supported such an assumption. With the area survey, spatial and temporal data are available to make various imputation cell strategies possible.

REPORTING ERRORS

Two reporting unit concepts associating land with people reporting by phone or mail have been used in the ALVS: (1) survey respondents providing opinions on "average" farmland values in their locality and (2) survey respondents reporting opinions of "average" farmland values in their county. Both association concepts present opportunities for coverage as well as reporting errors.

The LOCALITY REPORTING CONCEPT relies on the reporters knowledge of the presence of various types of land in their locality. Thus, the respondent's ability to formulate a response contributes to response error.[4] The major deficiency with the locality

[3]The existence of extraneous units on the sampling frame also is considered a coverage error. This problem creates inefficiency in the design but not necessarily the biases that under- or over coverage create. An area frame generally has extraneous units which are screened out during enumeration. Since the JAS is conducted and the agricultural tracts are already screened out for purposes other than the land value data, the screened tracts provide an efficient frame for this add-on data.

[4]For a discussion of locality reporting issues in crop yield reports, see Fecso (1991).

concept is the vagueness of the reporting unit. Specifically, for which agricultural land area should the individual report? If one respondent references an individual operation as "locality" while another respondent in the township references the entire township, duplicative reporting occurs. Also, despite the stratification used in the ALVS (grouping urban and agri-urban counties into separate strata), respondents do not know the "locality" boundary and duplication can occur across strata. Some farmed land areas may have a disproportionately low or no chance of being reported. For example, flood plains may have fewer, if any, residences. Thus, the "locality" reporting concept used with list sampling is prone to specification error and is not likely to be a proportional representation of the various land values in the population.

With the COUNTY REPORTING CONCEPT, respondents within a county are asked to report a value on the same "reporting unit," the county. This approach, while technically defining complete coverage and not vague like the locality reporting concept, was not successful. Respondents do not know the amount of land within the county at various value levels and, thus could formulate different responses for the same fixed value.

For area frame respondents, the specification of the land area to be valued is clear. It is the operated parcel of land identified with the respondent and drawn on aerial photography during the personal interview.

EDITING AGRICULTURAL LAND VALUES

Data editing during the list frame survey was limited. Reported values were screened for reasonableness during a review of the questionnaire when it arrived at the NASS field office (called a State Statistical Office or SSO). Questionnaires were also reviewed for legibility to keyenter. Data was edited out if reported on the wrong basis, such as total rather than per acre value. Outlier printouts based on fixed high and low cutoff points were produced for review. Instructions specifically asked not to edit outliers unless they were shown to be incorrect, as the survey sponsor (ERS) wanted to review these values as well.

The sponsor used a statistical edit, "fencing" of reports from within each county, to control the impact of outliers. Applying range edits, made at a state or even a county level, to data with the intent to change extremely high or low values to missing values, presents problems. For example, an "outlier" data review in several states with a very wide range of values due to diverse agriculture and rural/urban mixtures found that most of the extreme values are believable. Reports of $500,000+ per acre land in a major urban area of the county are realistic, especially for specialty cropland such as nurseries. Yet, $500 per acre land could be found in remote areas of the same county. Unfortunately, good data could be lost as well as bad with a fencing approach, especially in counties with wide variation in land prices.

Area Data Edits

Reported values in the area survey are also screened for reasonableness during a review of the questionnaire when it arrives at the field office. At this stage, editing is a difficult task, but more is known on the specific characteristics of the tract than when editing list frame data. Enumerator notes in the margin of the questionnaire are a help when present. Also, field office staff are often familiar with land values in many areas of their state, especially when specific circumstances might create unusually high or low value.

Area frame sampling allows for more refined computer editing using spatial information. After key entry, computer edits compare the reported tract values against values from the smallest possible locality. For new segments, tract reports are compared to segment level averages, providing a spatial comparison among nearest neighbors. [5]

Editing Data From a Matched Segment Design

The 1995 survey presented a significant opportunity to improve the data quality. For matched samples, historic data was used to create edit limits as a check for outlier values. The limits are developed using land values available from past survey data. The best edit limits for a segment are based on the reported data (with past outliers removed) from tracts in the segment.

For segments without historic data, edit limits are based on reported data from the same stratum type within the county, also adjusted for outliers.[6] If there were no reports from segments in the stratum type in the county, the range is based on all segments in the county. If there are no historic reports for the county, district (a contiguous group of counties) limits are created.

Finally, the segment average check is the default. Thus, edit limits are tailored to reflect the "nearest" spatially available characteristics for the respective segment. All data for segments with data values failing the edit are printed for verification. The verification includes inspection of the questionnaire for enumerator notes or for key-entry error. Any prior survey data for the tract should be reviewed. If the reports are consistent and not challenged by enumerator notes, the data are kept. If the data are not consistent and the

---

[5]This is also true for old segments not having a historic range for comparison.

[6]Strata 10 to 19, the highly cultivated strata, are grouped into a type, as are strata 20 to 29, the under half-cultivated strata; strata 30 to 39, the ag-urban strata; strata 40 to 49, the low density of cultivation strata; etc.

current survey data is deemed incorrect, it is edited to missing (for machine imputation later).

Post-Enumeration Editing

For land values collected in the area survey, spatial and temporal correlations allow for more informative post-enumeration edits as compared to list based survey data. Outliers in the fit of simple models of cluster effects and type of cropland effectively point out problematic data. Graphical products found in many statistical software packages provide a convenient way to explore the between tract and between- year within-tract differences. Outliers found graphically are easily linked and found in the "spreadsheet." With appropriate sorting, outliers can be reviewed with other "nearby" responses and past responses for the tract and its neighbors easily viewed.

The nearby and past information help to choose any necessary corrections. Remaining keyentry errors, such as an extra or omitted zero, become obvious and can be corrected. When a value is thought to be incorrect but no sure decision can be made concerning an appropriate replacement value, the value is "set to missing."

This editing was done at the sponsor's headquarters. After the editing was completed, SSO's conducted a final review. Printouts of each data item, sorted by segment, tract, and year of interview, were produced and sent to each state. The printout identified records with central office changes and both the original data and the change. SSO's could correct the 1994, 1995 or 1995 data and send back the recommended changes. When differences between years for a tract are not resolved, all data items could be set to missing, allowing imputation to make the adjustments in a manner less influential on the time series of estimates.

IMPUTATION

Since the land area by type is always present for area frame data, only the nonresponses for the value items needed an adjustment strategy. Rather than adjust weights, direct expansions for land values and cash rents are made after an imputation process for missing land values. Here too, the strong spatial and temporal correlations provide the underpinnings for a missing data adjustment methodology.

The strategy follows the same "search the nearest neighbors first" approach used in creating edits. Rather than create local ranges as done in editing, local medians of reported data or the ratio of year-to-year reports were calculated and used to impute values. When data was missing for one year but available for another in the tract, a median-local ratio adjustment was made. With data missing both years, the most local median responses were imputed for both years.

CONCLUSION

The decision to switch survey designs from a list to an area frame sample was based on the general strengths of the area frame survey and hypothesized usefulness of spatial and temporal correlation of land values. Experiences to date have confirmed the anticipated methodological improvements. Yet, further research is needed to reduce manual intervention during editing and imputation and to develop better benchmark and direct estimators.

REFERENCES

Cotter, J. J. and Tomczak, C. M. (1994), "An Image Analysis System to Develop Area Sampling Frames for Agricultural Surveys," *Photogrammetric Engineering & Remote Sensing*, Vol. 60, No. 3, pp. 299-306.

Fecso, Ron (1991), "A Review of Errors of Direct Observation in Crop Yield Surveys," in Biemer, et al. (Eds.), *Measurement Errors in Surveys*, 1991, pp 327-346, New York, Wiley.

Fecso, R., Tortora, R. D., and Vogel, F. A. (1986), "Sampling Frames for Agriculture in the United States," *Journal of Official Statistics*, Vol. 2, No. 3, pp. 279-292.