Appendix C.
Statistical Methodology

THE CENSUS MAIL LIST AND SCREENER PHASE

The National Agricultural Statistics Service (NASS) maintains a list of farmers and ranchers from which the census mail list (CML) is compiled. The goal is to build as complete a list as possible of agricultural places that produce and sell, or would normally sell, $1,000 or more of agricultural products per year. This is the same list used to define sampling populations for NASS surveys conducted for the agricultural estimates program. Each record on the list includes name, address, and telephone number plus additional information used to efficiently sample and administer the NASS census of agriculture and its agricultural estimates programs.

NASS builds and improves the list on an ongoing basis by obtaining outside source lists. Sources include state and federal government lists, producer association lists, seed grower lists, pesticide applicator lists, veterinarian lists, marketing association lists, and a variety of other agriculture-related lists. NASS occasionally obtains special commodity lists to address specific list deficiencies. In 2000, NASS began an intensive push to increase list coverage in preparation for the census.

Most names on a newly acquired list are already on the list sampling frame. Those found on the list are set aside. Those not found are treated as potential farms until NASS can confirm their existence as a qualifying farm. Field offices routinely contact these potential farms to determine their status, however, the increased pre-census list building activity generated much more followup work.

Beginning in April 2002, NASS conducted the 2002 Farm Identification Survey to screen 591,288 potential farms before placing them on the CML. These records were mailed a one-page report form and a nonresponse followup mailing was made in May 2002. A second mailing to a group of 568,692 additional potential farm records was conducted in mid-July 2002. There was no followup mailing. The entire screener phase confirmed 349,664 qualifying farms that were added to the CML. A total of 282,901 names were confirmed as out of scope and were dropped from the list. Names returned as undeliverable-as-addressed totaled 92,203 and they were excluded from further census mailings. The remaining 435,212 names did not respond and were mailed census forms although they were not added to the CML as active farms.

During the spring and summer of 2002, measures were taken to improve name and address quality. Checks were made to detect and remove duplication both within states and across states. List addresses were processed through the National Change of Address registry and the Locatable Address Conversion System to ensure they were correct and complete. Records on the mail list with missing or invalid phone numbers were matched against a nationally available telephone database to obtain as many phone numbers as possible.

Records requiring special handling for census data collection or for analysis and summarization were identified. These were mostly farms considered unique because of their size or because they produced specialty commodities.

The official CML was established on September 1, 2002. The list contained 2,841,788 records. There were 1,839,533 records that were thought to meet the NASS farm definition and 1,002,255 potential farm records.

CENSUS SAMPLE DESIGN

All name and address records on the final CML received a 2002 Census of Agriculture report form. Two different types of census report forms, sample and

nonsample, were used to collect data. Sections 1 through 16 and 22 through 25 of the sample form were identical to sections on the nonsample census form. Sections 17 through 21 of the sample form contained additional questions on usage of fertilizers and chemicals, farm production expenditures, value of machinery and equipment, value of land and buildings, and hired workers. There were 12 regional versions of the nonsample form and 13 regional versions of the sample form with listings of crops varying by region.

The sample form was mailed to all mail list records in Alaska and Rhode Island and to a sample of records in other states. Mail list records were selected into the sample with certainty if they (1) were expected to have large total value of agricultural products sold or large acreage, (2) were in a county with less than 100 farms in 1997, or (3) had other special characteristics (e.g., abnormal farms such as institutional farms, experimental and research farms, Indian reservations, etc.). Mail list records in counties containing 100 to 199 farms in 1997 were systematically sampled at a rate of 1 in 2; counties containing 200 to 299 farms in 1997 were systematically sampled at a rate of 1 in 4; counties containing 300 to 399 farms in 1997 were systematically sampled at a rate of 1 in 6; and counties containing 400 or more farms in 1997 were systematically sampled at a rate of 1 in 8. The mail list records not chosen to receive the sample form received the nonsample form. This differential sampling scheme was used to provide reliable data for the sample sections of the report form for all counties.

The regional report form versions and the sampling scheme were used to provide reliable data for a large number of items/commodities at the county level, while reducing response burden.

EDITING DATA AND IMPUTING FOR ITEM NONRESPONSE

The mailing label on all forms returned to the National Processing Center (NPC) were scanned using bar code readers to capture identifiers and for check-in purposes. Forms determined to represent qualifying, in-scope farms were submitted for imaging. A snapshot was taken of each page of every report form and optical mark recognition (OMR) and intelligent character recognition (ICR) techniques were used to capture reported data from the images. The ICR engine determined a confidence level for every cell read. Any cell with a confidence level below a prescribed value was referred to analysts to review and correct from the image, when necessary. The images and the captured data were transferred to NASS on a flow basis. Data collected by telephone were captured using computer-assisted telephone interview software.

Captured data were processed through a format program. This program verified that record identifiers were valid and checked the basic integrity of the data fields. Rejected records were referred to analysts for correction. Accepted records were posted to the database.

All 2002 census data were passed through a complex computer edit. Data were batched by state for submission to the computer edit. The edit determined whether a reporting operation met the minimum criteria to be counted as a farm in the census. Operations failing to meet the minimum criteria were referred to analysts for verification. The edit examined each report for reasonableness and completeness and determined whether to accept, delete, impute (supply), or alter the reported value for each data record item.

Whenever possible, imputations, deletions, and changes made by the editing system were based on related data on the respondent's report form. For some items, such as operator characteristics, available data for that farm from the previous census were used. Values reported on previous NASS surveys were used, where applicable.

When these and similar methods were not available and values had to be supplied, the imputation process used information reported for another farm operation in the same state or in a neighboring state with characteristics similar to those of the farm operation with incomplete data. For example, a farm operation that reported acres of corn harvested, but did not report bushels of corn harvested, was assigned the same bushels of corn per acre harvested as that of another farm from that region having similar characteristics and reporting an acceptable yield. Assigned values for one operation could come from more than one respondent because imputation for missing items in each section of the report form was conducted separately.

Each execution of the computer edit consisted of records from only one state. Successfully edited records were made available as potential "donors," to supply values needed in subsequent imputations. These records were accumulated into pools of donors according to geographic location, so that each pool might be used during the computer edit of any reports from appropriate states. When imputation was required, a report's collective imputation needs for a section were used to identify a group of matching variables for the report which contained acceptable data relating to the missing items. For example, acres of corn harvested would be a matching variable for bushels of corn harvested, in consideration of the high correlation between the two items.

Similarity to the report being edited was evaluated for the matching variables for all farms in the appropriate donor pool. Values were imputed from the donor report considered most similar, referred to in this context as the "nearest neighbor" to the report being edited. Similarity between the edited record and a donor was calculated as the Euclidean distance between their selected matching variables. As part of the distance computation, the values of the matching variables were normalized to have the same variance within each donor pool. Latitude and longitude were consistently included in all imputation requests as matching variables, so that geographic proximity played a role in all donor selections.

Imputation conformed to logic provided by the complex edit. When appropriate, only donors able to contribute a nonzero imputed value were considered. For a farm reporting harvested corn acreage, for example, imputed bushels of corn harvested would be taken only from farms with harvested corn. In addition, imputed values were often adjusted. In some cases, acceptable data in another field of the edited report were used to establish a ratio between the edited report and the donor report. This proportion was applied to the imputed value as a scale factor. In the corn example, total bushels of corn from the donor would be scaled by the ratio of the acres of corn in the edited report to those in the donor report.

To maintain consistency with the complex edit, the imputed values in most sections of the report were tested to ensure they satisfied critical relationships among items within the section. If any of these constraints were not met, alternative donors were considered in order of their similarity to the edited report, until all the constraints for the module were satisfied.

In some cases, nearest-neighbor imputation was not possible. The requirement of a positive imputed value might rule out all available donors, resulting in an imputation failure. However, if some members of the donor pool were found to satisfy this requirement, then as many as 25 nearest neighbors were given further consideration. But if none of the candidate donors could provide qualifying data, the result was also noted as an imputation failure. Processing of records that encountered these imputation failures was suspended at the section where the failure occurred. These records were made available for analyst review and later reconsidered by the automated edit as a followup to corrective actions taken by the analyst.

The donor pool for each region was frequently updated with records from its area which had completed the editing process. As records were added to the donor pool, the records became available to donate values to incomplete reports subsequently edited for that region. Prior to editing, all donor pools were empty and no donors were available. Initial donor pools were created by giving special treatment to the first batches of data received from each state. Similar to the way that imputation failures were resolved through analyst review of the reports, early reports from initial batches were reviewed and adjusted manually by teams of analysts. This process was employed until each donor pool became self-sufficient in consistently providing imputed values for its region through the automated nearest-neighbor selection process.

To streamline editing once they had reached a mature stage in their growth, donor pools for some regions were not expanded in size beyond a chosen plateau. This provided assurance that computer edits would not exceed a reasonable processing time for nearest-neighbor searches. Although their size was limited, these donor pools did not become static. They were regularly recreated with representative samples of all records available from their regions. Within a given region, all successfully edited sample form records were included in the appropriate donor pool. Successfully edited nonsample form records were ordered by farm size and sales volume for a given

region, and then systematically sampled. Every "ith" record from the nonsample form list was joined to the complete list of sample forms for its region to form a refreshed donor pool. The steady renewal of donor pools for regions with large numbers of records assured a more diverse selection of donors over time.

All records with data changes were resubmitted to the edit to verify that acceptable corrections were made. Records with imputation failures were referred to an analyst for resolution. Corrected data were posted and the record was re-edited.

The complex edit ensured the full internal consistency of the record. Analysts were provided an additional set of tools to review record-level data across farms. These examinations detected extreme outliers or unique data distribution patterns that were possibly a result of reporting, recording, or handling errors. Potential problems were researched and, when necessary, corrections were made and the record re-edited.

NONRESPONSE AND SAMPLE ESTIMATION

Statistical estimation procedures were used to account for whole farm nonresponse and sample data collection. The procedures for nonresponse were necessary because some farm operators did not respond to the census despite numerous attempts to contact them. Statistical estimates for sample-form-only data items had to be calculated since, by design, the data were not collected from every farm. Nonresponse and sample estimation procedures were not applied in Alaska and Rhode Island because all farms received the sample form and data were collected from all farms.

Treatment of Farms Selected for the Screener Phase

The screener phase and followup strategies resulted in several possible outcomes depending on whether the screener name responded and was in or out of scope. Each of these outcomes was handled differently to adjust for nonresponse.

Names responding to the screener as out of scope (nonfarms) were excluded from the CML. If the respondent answered the screener as in scope, the respondent was added to the CML and received a census form. If this in-scope screener respondent answered the census form, the operation's report was eligible to be used to help account for nonrespondents to the census. If the in-scope screener respondent failed to respond to the census form, that operation's data were accounted for by census respondents.

Records for operations that did not respond to any of the three screener mailings were not considered to be part of the CML, but they were sent a census form. Screener nonrespondents that responded as in-scope operations on the census were assigned a fixed nonresponse weight of 1 for census tabulations. Screener nonrespondents that failed to respond to the census form were treated in summarization as if they never existed on a mail list.

Whole Farm Nonresponse Estimation

Whole farm nonresponse to the census occurred when no data were received from an operation on the CML. Records deemed to represent either a large farm, as defined by the total value of production or acreage, or a unique farm operation received intensive telephone or personal followup during census processing to obtain a response. If these attempts failed, data were imputed for the record. These large and/or unique records were designated as "Must" records and were assigned a fixed nonresponse weight of 1, meaning their data were not used for nonresponse adjustment. Screener respondents with reported sales above a certain level automatically became Must records.

During mail list development, the field offices, in an effort to reduce respondent burden, identified operations that participated in multiple NASS surveys, and those that had special reporting relationships with an enumerator. The records for these operations were "Tagged." The field offices assumed full responsibility for the data collection for any Tagged operations, including imputing data for them if a response was not obtained. Tagged records became Must records. They had a nonresponse weight of 1 and the reports were not used for nonresponse adjustments.

Whole farm nonresponse that occurred within the remaining universe of records, called non-Musts, was

accounted for by a statistical weighting procedure. All responding non-Musts in a state were put into mutually exclusive weighting groups based on their size and county as recorded on the CML database. Statistical models were used to estimate the number of nonresponse farms that were in scope for each weighting group. The weights of the responding farms in each weighting group were increased to account for nonresponding farms in that group.

Throughout the data collection period, changes and additions were made to the CML. Records added after the initial CML was created on September 1, 2002 were designated as new adds, treated like screener nonrespondents, and given a nonresponse weight of 1. New adds responding as in-scope records to the census were subsequently subtracted from the measurement of undercoverage. New adds linked to operations originally on the CML were not considered new adds. New adds occurred any time after the CML creation and before final weighting in February, 2004.

Some operators were sent more than one census form. These operators were required to fill out a separate form for each operation. Also, an operator may have had an operation for which a census form was not received, but the existence of which was noted on the form of the known operation. That operator was sent a new census form or enumerated by telephone to obtain data for that previously unknown operation. If a response was obtained for the previously unknown operation, the nonresponse weight for the new record was set equal to the nonresponse weight for the original operation reporting its existence. If no response was obtained for the previously unknown operation, it was treated as out of scope.

Some large farms operating in more than one county were treated as distinct county-specific operations to more accurately allocate data to counties. Similarly, large farms operating in more than one state were treated as distinct state-specific operations. Split add records were created for these operations and they were assigned the same nonresponse weight as the original CML operation. Controls ensured the calculated and nonresponse weights never exceeded 2. The nonresponse weights were systematically rounded to integers and an integerized weight of either 1 or 2 was assigned to each record. The integerization process eliminated any impact rounding would have had on census farm counts and totals in each county and in cross tabulations.

Tables A and C quantify the effect of the nonresponse estimation procedures on selected census data items. These tables contain percentages of the census aggregates that were contributed by nonresponse adjustments. As noted earlier, names included in the screener sample that never responded were treated as if they never existed on a mail list. Any in-scope farm in this group was missed and, consequently, "attributed" to the coverage adjustment. This is shown in Table C. For selected items, estimates of what was attributed were reallocated to nonresponse to obtain "corrected" values, which appear in Table A. This was possible at the state level only. The differences between state-level nonresponse adjustment numbers in the first line of Table C and their counterparts in Table A represent the amount reallocated.

There was no such reallocation in Hawaii because records in that state were not adjusted to account for coverage errors. No tables appear for Alaska because records were not adjusted for nonresponse or coverage.

The estimates provided in Tables A and C do not reflect the effect of item nonresponse on individual census data items. The effect of this item nonresponse is discussed in the ''Nonmeasurable Census Error'' section.

Sample Estimation

All Must records were preselected to receive the census sample form. Non-Must records were sampled to determine which would receive the sample form and which the nonsample form. All records in some small counties automatically received the census sample form but these records were not necessarily Must records. Nonresponse adjustment was allowed for the non-Musts.

Weights applied to the sample items appearing on the sample form only (Sections 17 through 21) were calculated by multiplying the farm's coverage-adjusted weight, which is described later, by the sample factor (e.g, 6 for a farm sampled with a 1-in-6 rate, 1 for a Must). An adjustment was made that ensured the number of farms operating in a county as estimated from the sample matched the number estimated from

the full census. Before computing published tabulations based on the sample, each record's sample weight was integerized to eliminate the impact rounding would have had on census farm counts and totals.

Operators with more than one operation were sampled as one record and received the same census form for each operation. Operations added after sampling were treated differently depending on whether or not the record was linked to a record on the original CML. Added operations that linked to a record on the original CML were mailed the same census form as the original CML operation. Added operations that were **not** linked to a record on the original CML were mailed the sample form.

MEASURABLE CENSUS ERROR

The root mean squared error of an estimated data item from the census provides a measure of the error a field office associated with completing a census. It measures the variation in the value of that estimated data item based on all possible outcomes of the census collection, including variants as to who was on the census list, who returned a census form and who was selected to fill out the sample form.

Data items were classified as either complete count items or sample count items. Sample count items were collected only on the sample version of the census report form. Complete count items were collected from all respondents. Variability in the estimates of complete count items was due only to the nonresponse and coverage estimation adjustment procedures. Variability in the estimates of sample count items was due to both the adjustment procedures and the census sample selection and estimation procedure. Therefore, variability in the sample count item estimates tends to be larger than the variability in the complete count item estimates.

Table B presents the fully adjusted total with the root mean squared error for selected items. The relative root mean squared error is obtained by dividing the root mean squared error by the value of the estimate and then multiplying by 100. The table also includes the percent contribution to the mean squared error (the square of the root mean squared error) from nonresponse adjustment and sampling and from coverage adjustment. Mean squared errors for Hawaii are entirely due to nonresponse adjustment.

Nonsampling error due to mail list incompleteness and duplication as well as misclassification of records on the mail list is called coverage error. The section titled "Classification Error Study" addresses attempts to assess, at least qualitatively, the impact of classification error on the census results.

NONMEASURABLE CENSUS ERROR

The accuracy of the census counts is affected jointly by the measurable errors described in the previous section and by nonmeasurable errors (nonmeasurable in the sense of not being included in root mean squared error estimates). Extensive efforts were made to compile a complete and accurate mail list for the census, to design an understandable report form with instructions, and to minimize processing errors through the use of quality control measures. Despite these efforts, nonmeasurable errors are inevitable and arise from many sources, including respondent or enumerator error, incorrect data capture, editing, and imputing for missing data. These errors are discussed in this section.

Respondent and Enumerator Error

Incorrect or incomplete responses to the census report form or to the questions posed by an enumerator can introduce error into the census data. To reduce reporting error, detailed instructions for completing the report form were provided to each respondent. Questions were phrased as clearly as possible based on previous tests of the report form. Computer-assisted telephone interviewing software included immediate integrity checks of recorded responses so suspect data could be verified or corrected. In addition, each respondent's answers were checked for completeness and consistency by the complex edit and imputation system.

Item Nonresponse

As information flowed from data collection to tabulation, various types of item nonresponses were identified on the census report forms. Nonresponse to particular questions on the form that logically should have been present created a type of nonsampling error

in both complete count and sample count data. In this case, information from a similar farm was used to impute for these missing data items. The resulting data may have been biased if the characteristics of the nonreporting farms were different from those of reporting farms for those items. The section titled "Editing Data and Imputing for Item Nonresponse" provides a detailed explanation of item imputation procedures.

Processing Error

All phases of processing for each census report form were potential sources of nonsampling error. An automated check-in procedure recorded that the report had been returned and excluded it from further followup mailings. Approximately one-third of the mail returns were reviewed to resolve questions dealing with multiple reports, respondent remarks, or no reported data. The remaining mail returns (about two-thirds), along with some of the reviewed cases containing farm data, were batched and sent directly to imaging and data capture. Data were transmitted, formatted, and run through the complex edit and imputation system to ensure within record consistency. About one-fifth of all forms edited were clerically reviewed for inconsistencies, omissions, or questionable values. While reviewing these forms, staff determined if the action taken by the computer edit and imputation system was correct. Additional analysis tools were used to examine data across records for distributional irregularities and extreme values. Edited records were tabulated to the county level. Each county was reviewed and, when necessary, individual records were corrected prior to publication.

Developing accurate processing methods is complicated by the complex structure of agriculture. Among the complexities are the many places to be included, the variety of arrangements under which farms are operated, the continuing changes in the relationship of operators to the farm operated, the expiration of leases and the initiation or renewal of leases, the problem of obtaining a complete list of agriculture operations, the difficulty of contacting and identifying some types of contractor/contractee relationships, the operator's absence from the farm during the data collection period, and the operator's opinion that part or all of the operation does not qualify and should not be included in the census. During data collection and processing of the census, all operations underwent a number of quality control checks to ensure results were as accurate as possible.

COVERAGE ADJUSTMENT

Although much effort was expended making the CML as complete as possible, the coverage of farms was not complete. NASS's goal was to produce agricultural census totals for publication that were fully adjusted for list undercoverage at the county level. Estimates of the undercoverage for a specified set of farm characteristics, called calibration variables, were computed using an area-frame sample. Initial weights were assigned to census respondents to account for nonresponse, and these weights were further adjusted to compensate for estimated state-level undercoverage for each of the calibration variables based on the area frame sample. Each farm with census data was assigned a fully-adjusted weight by this process and county-level totals were generated for every census variable, not just the calibration variables. The section titled "Calibration Algorithm" provides a list of the area frame based calibration variables.

To further improve coverage adjustment, a second set of targets and ranges were added to the calibration effort. These were well established commodity totals for which excellent check data were available for validation. The introduction of these commodity targets strengthened the overall coverage adjustment process by limiting the possible adjustments produced by the area frame based targets to ensure major commodity totals remained within reasonable bounds of established benchmarks.

Most targets were determined at the state level. The one exception was the New England states - Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont - which were combined into one "calibration region." In what follows, "state" refers to the calibration region for New England. Coverage adjustments were not made in Alaska and Hawaii.

Measuring Mail List Undercoverage

Census mail list undercoverage was measured using an independent survey of land segments selected from the NASS area frame. The NASS area frame covers all

land in the U.S. and includes all farms. Each June, NASS conducts a survey that enumerates area frame segments for agricultural activity. The sampled segments are allocated to provide accurate measures of acres planted to widely grown crops and inventories of hogs and cattle.

The 2002 June Agricultural Survey (JAS) was supplemented with the 2002 Agricultural Coverage Evaluation Survey (ACES) to better estimate CML incompleteness. The ACES used a sample of segments allocated in a way that, when pooled with the JAS, ensured accurate measures of number of farms and land utilization could be obtained. Enumerators visited all segments, identified all farms operating land in each segment, and obtained basic data about those farms.

The names and addresses collected in the 2002 JAS and 2002 ACES were matched to the census mail list. Farms that did not match were re-contacted after the census mailout to confirm that they did not receive a census form. Farms that had not received a census form represented the farms not on the mail list (NML). Those who received a census form had been erroneously classified as NML and were removed.

The percentage of farms missed in the census varied considerably by state. In general, farms not on the mail list tended to be small in acreage, production, and sales of agricultural products. Farm operations were missed for various reasons, including the possibility that the operation started after the mail list was developed, the operation was so small as not to appear in any agriculture-related source lists, or the operation was falsely classified as a nonfarm prior to mailout.

Determining Targets to Correct for Undercoverage

The 2002 June Agricultural Survey consisted of 11,075 land segments and the 2002 Agricultural Coverage Evaluation Survey (ACES) added 2,400 segments. Data values a field office associated with NML tracts were used to estimate the state-level undercoverage of the CML for the first set of calibration variables. The state-level totals for these variables were then summed to yield national totals.

The national NML estimate for the number of farms was used directly in determining calibration targets (CML + NML). State-level farm-count estimates based on the NML sometimes had unacceptably high standard errors, as well as apparent systematic biases. These estimates were smoothed across states based on separate NASS surveys and previous analysis.

Other calibration targets were derived from the NML-estimated fractions of farms of certain types (i.e. in a particular sales class or with a principal operator of a particular race). Most of these had unacceptably high state-level standard errors. As a result, more reliable national-level NML estimates were used to smooth state estimates. The smoothed state NML-estimated fraction was computed by taking a weighted average of the actual state estimate and a prediction for the state based on national- and state-level numbers (i.e., the number of NML farms in the state, the fraction of farms with black owners on the state's CML, and the national relative difference between the fraction of black owners on the NML and CML). The weighting factor was chosen to approximately minimize mean squared error under a random effects model. The smoothed NML-estimated fractions were multiplied by the corresponding smoothed NML farm-count estimates described above and added to corresponding CML estimates to obtain coverage-adjusted state-level totals, that served as calibration targets.

Tolerance Ranges

Although full calibration would assure that the weighted total among census respondents equaled its target for each calibration variable in either set, it was not always possible to calibrate to such a large number of target values while keeping all farm weights within a reasonable range (for example, the weight for any farm cannot be less than one). Because of this and because calibration targets are estimates themselves subject to uncertainty, NASS allowed some tolerance in the determination of coverage-adjusted weights. Rather than forcing the total for each calibration variable computed using the coverage-adjusted weights to equal a specific amount, NASS allowed the estimated total to fall within a tolerance range. This tolerance strategy sometimes made it possible for the calibration algorithm to produce a set of satisfactory coverage-adjusted weights that it would not have otherwise.

Ranges for the first set of calibration variables used to adjust for undercoverage were determined differently from the second set used to adjust for measurement error. The number of farms had no tolerance range. The tolerance range for every other variable in the first set was the estimated state total for the variable (CML + NML) plus or minus one-half of one estimated standard error. This choice limited the cumulative deviation from the estimated total for a variable when state-level totals were combined to create a U.S.-level total. These ranges did not have to be symmetric around the target value.

Calibration Algorithm

Coverage adjusted weights were obtained by an algorithm based on the restricted regression algorithm referred to by Singh and Mohl (1996) as the Linear Truncated Method. Coverage adjustments began with the nonresponse weights before integerization. The final coverage-adjusted weights were restricted to the interval [1,6].

The calibration variables were based on the following reported items:

1. Total market value of agricultural products sold and government payments.

$0	$5,000 - $24,999
$1 - $999	$25,000 - $99,999
$1,000 - $2,499	$100,000 - $499,999
$2,500 - $4,999	$500,000 and above

2. Age of principal operator.

   Less than 25 years old
   25 - 34
   35 - 44
   45 - 54
   55 and older

3. Sex of principal operator.

   Female

4. Race of principal operator (selected categories).

   Black, American Indian or Alaska Native, Asian, and Native Hawaiian or Other Pacific Islander

5. Principal operators of Spanish, Hispanic, or Latino origin.

6. Number of farms and land in farms.

7. Number of extreme operations (very large or unusual farms).

8. Selected types of farms by commodity produced.

   All cattle farms
   Dairy farms
   Hog/pig farms
   Horse/Equine farms
   Poultry farms
   Sheep/goat farms

   Fruit/nut/berry farms
   Nursery/horticulture farms
   Tobacco farms
   Vegetable and melon farms

9. Various commodity acreage and production statistics (varies by state).

   Beef cow inventory
   Broiler production
   Cattle on feed inventory
   Layer inventory
   Milk cow inventory
   Total cattle and calves inventory
   Total hog and pig inventory

   Alfalfa acres harvested (South Dakota only)
   Apples acres harvested
   Corn acres harvested
   Cotton bales produced
   Grape acres harvested
   Hay acres harvested
   Lettuce acres harvested
   Potato acres harvested
   Rice acres harvested
   Soybean acres harvested
   Sugarbeet acres harvested
   Sugarcane acres harvested
   Tobacco acres harvested
   Tomatoes acres harvested
   Total orange acres
   Wheat acres harvested
   Durum wheat acres harvested (North Dakota only)

Other spring wheat acres harvested (North Dakota only)

Integerization and Sample Weights

Coverage-adjusted weights were integerized to eliminate the need for rounding estimated counts computed with coverage-adjusted weights. The integerization process minimized county-level impact on the nonresponse and coverage adjustment of number of farms and total land in farms.

Sample weights were computed by multiplying coverage-adjusted weights before integerization with the appropriate sampling factors and adjusting the results to add up to matched census counts as described previously. Sample weights were then integerized.

Measuring the Amount of Coverage Adjustment

Tables A and C display the proportions of selected census data items that are due to nonresponse and coverage adjustments. The section of this appendix on whole farm nonresponse adjustment explained how the nonresponse adjustment values were determined. The coverage adjustment values account for the rest of the differences between the weighted and unweighted totals for these data items. Some estimated coverage adjustments are negative. The use of commodity targets in calibration indirectly exposed some duplication on the census list resulting in negative coverage adjustments.

CLASSIFICATION ERROR STUDY

The 2002 Classification Error Study (CES) was conducted for the conterminous U.S. to study the potential impact of classification error on the census results. The study used data from the 2002 June Agricultural Survey (JAS) and the 2002 Agricultural Coverage Evaluation Survey to examine farms incorrectly classified as nonfarms (undercount), nonfarms incorrectly classified as farms (overcount), and duplication of farms (overcount) in the 2002 Census of Agriculture. The CES was not intended to adjust census farm counts, but rather, to evaluate procedures and to identify potential improvements in list building, data collection, and other activities in preparation for future censuses.

For the evaluation, additional name, address, and telephone information were collected on both the JAS and ACES by adding the following three questions:

1. During the past two years, has the operator received mail for this operation at any address other than the one shown on the face page?

2. Excluding partners and landlords, were any other names associated with this operation in the past two years? (For example, other business names, spouses names, etc).

3. Is any of the land inside the blue tract boundary rented from others? (Include land for which you paid cash rent, land used rent free, or land rented on shares).

The CES consisted of a two phase review process. The initial phase, Review of Possible Matches, used Probabilistic Record Linkage (PRL) to match the additional information collected on the area surveys to the name and addresses on the 2002 Census Mail List (CML) including late adds. PRL is a technique used to identify records that are believed to correspond to a CML record. Records were brought together into link groups, with each link group consisting of all records that possibly represented the same operation. Each link group was classified into one of three distinct types: matches, possible matches and nonmatches. The nonmatches were represented in estimation as part of the undercoverage measure. The CES was primarily concerned with the matches and possible matches. Each field office reviewed the possible matches and determined match or nonmatch status.

Upon completion of the PRL review, the field offices conducted a Farm Classification Resolution review of two additional sets of records. The first of these was comprised of area records matching two or more census records. Reviewing these records helped identify duplication on the CML. The second set consisted of groups of records (area and census) within which the reported acreage differed by more than 25 percent. Analysts reviewed the cases in the second phase. Upon completion of both phases, data were compiled to estimate undercount, overcount and duplication.

The analysis of these data will provide insight into census processes used to accurately determine farm status and identify duplication. Any weaknesses identified in the findings will be addressed for future censuses.

Table A. **Summary of State Nonresponse and Coverage Adjustments: 2002**

[For meaning of abbreviations and symbols, see introductory text]

Item	Total	Percent from nonresponse adjustment, corrected	Percent from coverage adjustment, corrected
Farms ..... number	7,196	16.5	16.1
Land in farms ..... acres	1,369,768	12.9	6.5
Farms by size:			
1 to 9 ..... farms	918	17.6	23.0
acres	4,017	15.9	29.4
10 to 49 ..... farms	1,861	16.4	25.8
acres	47,680	16.3	25.4
50 to 179 ..... farms	2,506	16.3	16.6
acres	253,737	16.7	14.6
180 to 499 ..... farms	1,334	17.5	3.9
acres	383,516	17.6	2.9
500 to 999 ..... farms	393	17.0	-2.8
acres	266,787	16.4	-2.2
1,000 to 1,999 ..... farms	135	6.7	1.5
acres	177,615	6.2	1.6
2,000 or more ..... farms	49	2.0	16.3
acres	236,416	1.5	12.8
Market value of agricultural products sold $1,000	463,603	8.4	-4.2
Farms by value of sales:			
Less than $1,000 ..... farms	2,659	12.3	43.3
$1,000	422	16.6	30.6
$1,000 to $2,499 ..... farms	975	15.4	24.1
$1,000	1,589	15.5	23.3
$2,500 to $4,999 ..... farms	777	19.3	4.6
$1,000	2,736	19.1	4.6
$5,000 to $9,999 ..... farms	682	21.3	-7.2
$1,000	4,783	21.5	-6.8
$10,000 to $19,999 ..... farms	578	18.2	-6.7
$1,000	8,051	18.3	-6.4
$20,000 to $24,999 ..... farms	149	21.5	-7.4
1,000	3,292	21.0	-6.9
$25,000 to $39,999 ..... farms	269	21.9	-11.2
$1,000	8,353	21.9	-11.5
$40,000 to $49,999 ..... farms	118	28.8	-18.6
$1,000	5,231	29.2	-19.0
$50,000 to $99,999 ..... farms	310	21.0	-12.3
$1,000	22,145	21.0	-13.4
$100,000 to $249,999 ..... farms	350	24.9	-15.1
$1,000	56,012	23.7	-13.1
$250,000 to $499,999 ..... farms	163	17.2	-10.4
$1,000	55,647	16.0	-9.8
$500,000 to $999,999 ..... farms	104	4.8	-2.9
$1,000	71,890	4.2	-1.6
$1,000,000 or more ..... farms	62	1.6	0.0
$1,000	223,453	0.7	(Z)
Farms by type of organization:			
Family or individual ..... farms	6,377	16.7	17.5
acres	954,562	15.7	7.2
Partnership ..... farms	321	14.3	4.4
acres	131,218	6.9	0.8
Corporation:			
Family held ..... farms	371	16.7	1.9
acres	257,403	6.1	7.0
Other than family held ..... farms	51	11.8	21.6
acres	3,962	18.7	-3.2
Other - cooperative, estate or trust, institutional, etc ..... farms	76	7.9	14.5
acres	22,623	3.0	3.9
Tenure:			
Full owners ..... farms	5,242	16.0	20.1
acres	715,764	13.7	10.9
Tenure - Con.			
Part owners ..... farms	1,604	17.5	5.3
acres	596,358	11.8	-0.1
Tenants ..... farms	350	19.7	5.7
acres	57,646	13.9	19.8
Principal operator characteristics by-			
Sex of operator:			
Male ..... farms	5,637	16.7	13.9
acres	1,243,921	12.5	6.0
Female ..... farms	1,559	15.8	24.3
acres	125,847	16.5	11.2
Primary occupation:			
Farming ..... farms	3,409	17.5	8.8
Other ..... farms	3,787	15.6	22.7
Spanish, Hispanic, or Latino origin (see text) ..... farms	143	4.2	77.6
acres	19,114	3.5	78.4
Race:			
White ..... farms	7,150	16.5	16.2
acres	1,355,591	12.9	6.6
Black or African American ..... farms	-	-	-
acres	-	-	-
American Indian or Alaska Native ..... farms	17	29.4	-23.5
acres	10,732	7.2	-6.3
Native Hawaiian or Other Pacific Islander ..... farms	1	100.0	-100.0
acres	(D)	(D)	-100.0
Asian ..... farms	8	0.0	37.5
acres	(D)	0.0	(D)
More than one race reported ..... farms	20	15.0	15.0
acres	2,790	16.2	-3.5
Reporting primary occupation as farming by age group:			
Under 25 years ..... farms	20	20.0	45.0
25 to 34 years ..... farms	132	27.3	3.0
35 to 44 years ..... farms	615	14.5	21.3
45 to 54 years ..... farms	1,018	18.1	11.1
55 to 64 years ..... farms	744	18.8	1.5
65 years and over ..... farms	880	16.5	3.5
Reporting primary occupation as other than farming by age group:			
Under 25 years ..... farms	28	0.0	60.7
25 to 34 years ..... farms	143	16.8	30.1
35 to 44 years ..... farms	858	14.0	41.4
45 to 54 years ..... farms	1,282	16.9	22.2
55 to 64 years ..... farms	858	16.4	13.2
65 years and over ..... farms	618	14.1	7.9
All operators by age group [1]:			
Under 25 years ..... farms	253	19.4	17.0
25 to 34 years ..... farms	581	18.4	17.0
35 to 44 years ..... farms	2,510	14.6	28.9
45 to 54 years ..... farms	3,535	17.1	17.4
55 to 64 years ..... farms	2,380	17.1	9.5
65 to 74 years ..... farms	1,270	16.2	4.4
75 years and over ..... farms	614	15.8	4.2

[1] Data were collected for a maximum of three operators per farm.

Table B. **Reliability Estimates of State Totals: 2002**

[For meaning of abbreviations and symbols, see introductory text]

Item	Total	Root mean squared error (RMSE)	Relative RMSE (percent)	Nonresponse and sampling contribution to MSE (percent)	Coverage adjustment contribution to MSE (percent)
Farms ..... number	7,196	347	4.8	0.7	99.3
Land in farms ..... acres	1,369,768	81,079	5.9	1.1	98.9
Farms by size:					
1 to 9 ..... farms	918	72	7.8	4.6	95.4
acres	4,017	331	8.2	6.6	93.4
10 to 49 ..... farms	1,861	114	6.1	3.5	96.5
acres	47,680	2,949	6.2	4.3	95.7
50 to 179 ..... farms	2,506	138	5.5	2.9	97.1
acres	253,737	14,329	5.6	3.1	96.9
180 to 499 ..... farms	1,334	94	7.1	2.9	97.1
acres	383,516	27,967	7.3	2.9	97.1
500 to 999 ..... farms	393	35	8.9	5.0	95.0
acres	266,787	23,502	8.8	5.1	94.9
1,000 to 1,999 ..... farms	135	13	9.4	3.7	96.3
acres	177,615	16,187	9.1	3.3	96.7
2,000 or more ..... farms	49	7	13.6	3.2	96.8
acres	236,416	23,098	9.8	5.4	94.6
Market value of agricultural products sold $1,000	463,603	23,074	5.0	0.7	99.3
Farms by value of sales:					
Less than $1,000 ..... farms	2,659	240	9.0	1.1	98.9
$1,000	422	62	14.7	1.7	98.3
$1,000 to $2,499 ..... farms	975	115	11.8	1.7	98.3
$1,000	1,589	189	11.9	1.7	98.3
$2,500 to $4,999 ..... farms	777	84	10.8	2.2	97.8
$1,000	2,736	294	10.7	2.3	97.7
$5,000 to $9,999 ..... farms	682	67	9.8	2.6	97.4
$1,000	4,783	468	9.8	2.9	97.1
$10,000 to $19,999 ..... farms	578	57	9.8	3.4	96.6
$1,000	8,051	787	9.8	3.6	96.4
$20,000 to $24,999 ..... farms	149	18	12.0	8.7	91.3
1,000	3,292	397	12.1	8.8	91.2
$25,000 to $39,999 ..... farms	269	28	10.2	7.2	92.8
$1,000	8,353	852	10.2	7.1	92.9
$40,000 to $49,999 ..... farms	118	15	12.5	8.4	91.6
$1,000	5,231	655	12.5	8.5	91.5
$50,000 to $99,999 ..... farms	310	30	9.8	6.6	93.4
$1,000	22,145	2,182	9.9	6.5	93.5
$100,000 to $249,999 ..... farms	350	33	9.4	5.8	94.2
$1,000	56,012	5,292	9.4	6.2	93.8
$250,000 to $499,999 ..... farms	163	18	10.8	5.1	94.9
$1,000	55,647	5,956	10.7	4.9	95.1
$500,000 to $999,999 ..... farms	104	10	9.8	2.7	97.3
$1,000	71,890	7,206	10.0	2.2	97.8
$1,000,000 or more ..... farms	62	4	5.8	3.2	96.8
$1,000	223,453	5,593	2.5	2.8	97.2
Farms by type of organization:					
Family or individual ..... farms	6,377	313	4.9	0.9	99.1
acres	954,562	60,722	6.4	1.7	98.3
Partnership ..... farms	321	27	8.5	10.4	89.6
acres	131,218	11,296	8.6	6.2	93.8
Corporation:					
Family held ..... farms	371	28	7.5	9.1	90.9
acres	257,403	21,353	8.3	2.8	97.2
Other than family held ..... farms	51	10	20.0	13.4	86.6
acres	3,962	646	16.3	12.5	87.5
Other - cooperative, estate or trust, institutional, etc ..... farms	76	9	11.6	19.1	80.9
acres	22,623	2,120	9.4	13.9	86.1
Tenure:					
Full owners ..... farms	5,242	266	5.1	1.2	98.8
acres	715,764	41,611	5.8	2.2	97.8
Part owners ..... farms	1,604	104	6.5	2.7	97.3
acres	596,358	44,437	7.5	2.4	97.6
Tenants ..... farms	350	34	9.8	6.4	93.6
acres	57,646	15,485	26.9	1.2	98.8
Principal operator characteristics by-					
Sex of operator:					
Male ..... farms	5,637	300	5.3	0.8	99.2
acres	1,243,921	75,660	6.1	1.2	98.8
Female ..... farms	1,559	152	9.7	1.4	98.6
acres	125,847	16,718	13.3	4.3	95.7
Primary occupation:					
Farming ..... farms	3,409	182	5.4	1.7	98.3
Other ..... farms	3,787	198	5.2	1.9	98.1
Spanish, Hispanic, or Latino origin (see text) ..... farms	143	54	37.8	3.3	96.7
acres	19,114	8,013	41.9	6.4	93.6
Race:					
White ..... farms	7,150	346	4.8	0.7	99.3
acres	1,355,591	80,539	5.9	1.2	98.8
Black or African American ..... farms	-	-	-	-	-
acres	-	-	-	-	-
American Indian or Alaska Native ..... farms	17	8	44.7	2.8	97.2
acres	10,732	2,139	19.9	1.1	98.9
Native Hawaiian or Other Pacific Islander ..... farms	1	1	61.0	0.4	99.6
acres	(D)	(D)	(D)	(D)	(D)

See footnote(s) at end of table.

--continued

Table B. **Reliability Estimates of State Totals: 2002** - Con.

[For meaning of abbreviations and symbols, see introductory text]

Item		Total	Root mean squared error (RMSE)	Relative RMSE (percent)	Nonresponse and sampling contribution to MSE (percent)	Coverage adjustment contribution to MSE (percent)
Principal operator characteristics by- Con.						
Race - Con.						
Asian	farms	8	5	60.9	12.0	88.0
	acres	(D)	(D)	(D)	(D)	(D)
More than one race reported	farms	20	10	49.3	4.9	95.1
	acres	2,790	2,066	74.1	6.6	93.4
Reporting primary occupation as farming by age group:						
Under 25 years	farms	20	10	49.8	3.2	96.8
25 to 34 years	farms	132	33	25.0	3.1	96.9
35 to 44 years	farms	615	73	11.9	2.8	97.2
45 to 54 years	farms	1,018	86	8.5	3.1	96.9
55 to 64 years	farms	744	56	7.5	5.5	94.5
65 years and over	farms	880	67	7.7	3.8	96.2
Reporting primary occupation as other than farming by age group:						
Under 25 years	farms	28	13	47.5	4.3	95.7
25 to 34 years	farms	143	28	19.6	3.1	96.9
35 to 44 years	farms	858	82	9.6	3.7	96.3
45 to 54 years	farms	1,282	103	8.1	2.7	97.3
55 to 64 years	farms	858	62	7.2	5.7	94.3
65 years and over	farms	618	50	8.1	5.2	94.8
All operators by age group [1]:						
Under 25 years	farms	253	33	12.9	6.0	94.0
25 to 34 years	farms	581	77	13.3	2.4	97.6
35 to 44 years	farms	2,510	195	7.8	2.3	97.7
45 to 54 years	farms	3,535	240	6.8	1.8	98.2
55 to 64 years	farms	2,380	147	6.2	3.5	96.5
65 to 74 years	farms	1,270	91	7.2	3.6	96.4
75 years and over	farms	614	51	8.3	5.7	94.3
Net cash farm income of operations (see text) [2]:						
Farms with gains of [3] -						
Less than $1,000	farms	355	72	20.4	48.5	51.5
	$1,000	147	31	20.9	47.7	52.3
$1,000 to $4,999	farms	750	127	16.9	61.3	38.7
	$1,000	1,910	321	16.8	57.7	42.3
$5,000 to $9,999	farms	270	58	21.5	52.0	48.0
	$1,000	2,013	456	22.6	52.6	47.4
$10,000 to $24,999	farms	467	77	16.5	49.7	50.3
	$1,000	7,787	1,315	16.9	50.5	49.5
$25,000 to $49,999	farms	373	74	19.7	41.7	58.3
	$1,000	12,859	2,511	19.5	41.2	58.8
$50,000 or more	farms	448	57	12.7	42.4	57.6
	$1,000	128,935	8,822	6.8	45.2	54.8
Farms with losses of -						
Less than $1,000	farms	487	94	19.3	55.0	45.0
	$1,000	242	59	24.2	56.3	43.7
$1,000 to $4,999	farms	1,883	202	10.7	48.2	51.8
	$1,000	5,279	609	11.5	52.9	47.1
$5,000 to $9,999	farms	987	139	14.1	54.4	45.6
	$1,000	6,756	918	13.6	54.0	46.0
$10,000 to $24,999	farms	841	142	16.9	62.3	37.7
	$1,000	12,911	2,219	17.2	64.4	35.6
$25,000 to $49,999	farms	158	57	36.1	69.3	30.7
	$1,000	5,094	1,727	33.9	68.7	31.3
$50,000 or more	farms	102	27	26.5	55.7	44.3
	$1,000	15,797	2,680	17.0	52.2	47.8

[1] Data were collected for a maximum of three operators per farm.
[2] Data are based on a sample of farms.
[3] Farms with zero net cash income are included as farms with gains of less than $1,000.

Table C. **Summary of Nonresponse and Coverage Adjustments by County: 2002**

[For meaning of abbreviations and symbols, see introductory text]

Geographic area	All farms			Land in farms			Sales		
	Total (number)	Nonresponse adjustment, attributed (percent)	Coverage adjustment (percent)	Total (acres)	Nonresponse adjustment, attributed (percent)	Coverage adjustment (percent)	Total ($1,000)	Nonresponse adjustment, attributed (percent)	Coverage adjustment (percent)
STATE TOTAL									
Maine	7,196	11.1	21.5	1,369,768	10.3	9.1	463,603	7.0	-2.8
COUNTIES									
Androscoggin	334	9.9	20.4	55,782	8.7	5.1	96,649	0.9	-0.2
Aroostook	1,084	10.3	16.0	391,675	9.0	8.6	121,158	9.5	-4.4
Cumberland	596	11.6	24.3	54,455	12.8	11.6	17,594	10.6	-6.0
Franklin	317	9.5	27.1	49,805	12.2	15.0	5,983	15.7	1.1
Hancock	317	12.6	13.2	49,587	12.4	5.8	28,678	2.8	-0.7
Kennebec	575	11.0	24.7	86,168	11.2	9.9	30,229	8.0	-2.1
Knox	275	9.1	21.8	28,581	9.9	8.4	5,046	8.8	-5.8
Lincoln	292	13.4	22.6	30,618	15.8	9.6	7,542	7.7	(Z)
Oxford	469	9.0	26.2	67,406	9.4	17.2	14,651	5.4	-0.9
Penobscot	575	12.2	22.4	107,082	12.0	6.1	28,955	10.6	-5.8
Piscataquis	201	12.9	28.4	39,399	14.6	10.9	3,834	23.8	-19.4
Sagadahoc	158	9.5	25.9	20,171	9.5	22.2	4,271	7.0	7.5
Somerset	504	11.9	20.6	110,124	11.8	9.5	24,293	9.6	-3.3
Waldo	415	11.6	22.4	69,215	12.3	7.6	14,657	10.3	-5.3
Washington	399	11.3	11.0	152,481	4.6	5.3	41,314	2.1	1.2
York	685	11.7	25.8	57,219	15.0	12.0	18,750	16.8	-10.1