

1987

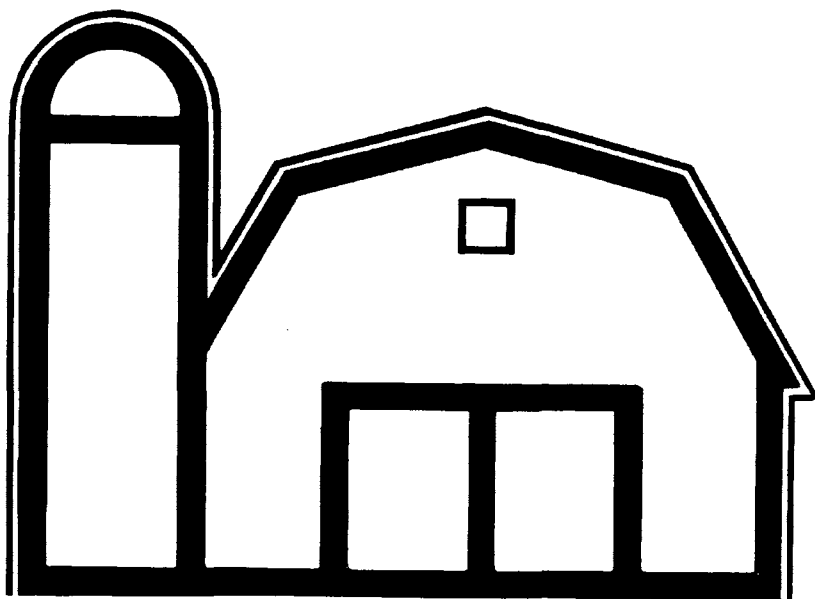
Census of Agriculture

AC87-S-2

Volume 2
SUBJECT SERIES

Part 2

Coverage Evaluation



U.S. Department of Commerce
BUREAU OF THE CENSUS

ACKNOWLEDGMENTS

This report was prepared in the Agriculture Division. Many other divisions contributed to this preparation: Data Preparation performed the clerical processing; Administrative and Publications Services performed planning, design, composition, editorial review, and printing planning and procurement for publications and report forms; Computer Services provided the computer processing facilities; and Economic Programming prepared the computer programs.

Members of the Census Advisory Committee on Agriculture Statistics and representatives of both public and private organizations made significant recommendations which helped establish data content.

Members of various agencies of the U.S. Department of Agriculture provided valuable advice in the planning, publicizing,

and processing phases of the census, and in helping farmers and ranchers complete the report forms.

The U.S. Department of Agriculture, National Agricultural Statistics Service, cooperated with the Bureau of the Census in planning and providing for the use of data from the 1987 June Enumerative Survey as an area sample measurement base for the estimates of farms not on the census mail list.

The press, farm magazines, radio and television stations, and farm organizations were most helpful in publicizing the census and encouraging cooperation of farm and ranch operators.

Special tribute is paid to the millions of farm and ranch operators who furnished the information requested. Only through their cooperation was it possible to collect and publish the data in this report.

If you have any questions concerning the statistics in this report, call:

(301) 763-8555 Division Chief

(301) 763-1113 General Information

(301) 763-8558 Statistical Methodology

1987

Census of

Agriculture

AC87-S-2

Volume 2
SUBJECT SERIES

Part 2

Coverage Evaluation

Issued July 1990



U.S. Department of Commerce
Robert A. Mosbacher, Secretary
Thomas J. Murrin, Deputy Secretary
Michael R. Darby, Under Secretary
for Economic Affairs

BUREAU OF THE CENSUS
Barbara Everitt Bryant, Director



BUREAU OF THE CENSUS
Barbara Everitt Bryant, Director
C. L. Kincannon, Deputy Director

Charles A. Waite, Associate Director for
Economic Programs
Roger H. Bugenhagen, Assistant Director for
Economic and Agriculture Censuses

Thomas L. Mesenbourg, Chief,
Economic Census Staff

AGRICULTURE DIVISION
Charles P. Pautler, Jr., Chief

Library of Congress Cataloging-in-Publication Data

Census of agriculture (1987). Volume 2. Subject series.
1987 census of agriculture. Volume 2. Subject series.
p. cm.
"AC87-S-1 < -6 >"

Contents: pt. 1. Agricultural atlas of the United States—pt.
2. Coverage evaluation—pt. 3. Ranking states and counties—pt.
4. History—pt. 5. Government payments and market value of
agricultural products sold—pt. 6. Code tabulations of selected
items.

Supt. of Docs. no.: C 3.31/12:987/v.2pt.1

1. Agriculture—Economic aspects—United States—Statistics.
2. Agriculture—Economic aspects—United States—States—Statistics.
3. Agriculture—Economic aspects—United States—Territories and
possessions—Statistics. I. United States. Bureau of the Census.
II. Title.

HD1769.C46 1990

338.1'0973'021—dc20

89-600341

CIP

For sale by Superintendent of Documents, U.S. Government Printing
Office, Washington, DC 20402.

CONTENTS

Introduction	Page
General Explanation	V
	VII

FIGURES

1. Map—Regions and Census Divisions of the United States	XIX
2. Estimated Net Census Farm Coverage Error: 1969 to 1987	XX
3. Estimated Census Farm Coverage Error: 1987	XX
4. Estimated Net Census Farm Coverage Error by Value of Sales: 1987	XXI
5. Estimated Undercounted Farms—Farm Count, Land in Farms, and Value of Sales: 1987	XXI
6. Composition of Nation's Estimated Undercounted and Census Farms: 1987	XXII

TABLES

1. Estimated Census Farm Coverage Error for Selected Operation Characteristics: 1987	1
2. Estimated Undercount for Selected Agricultural Items: 1987	3
3. Land in Farms by Sales Group for Estimated Undercounted Farms: 1987	5
4. Market Value of Agricultural Products Sold by Sales Group for Estimated Undercounted Farms: 1987	6
5. Estimated Farms Not on the Mail List: 1987	6
6. Selected Agricultural Items for Esimtated Farms Not on the Mail List: 1987	16

APPENDIXES

A. 1987 June Enumerative Survey Items Used in the Coverage Evaluation Program	A-1
B. Classification Error Survey Report Form	B-1
C. References	C-1

Publication Program	Inside back cover
---------------------------	-------------------

INTRODUCTION

	Page
PURPOSE.....	v
CENSUS AUTHORITY.....	v
FARM DEFINITION	v
ABBREVIATIONS AND SYMBOLS	v

PURPOSE

The Bureau of the Census measures the accuracy and completeness of farm counts and selected data items for each census of agriculture through an independent coverage evaluation program. The program seeks to identify situations that lead to coverage error and to reveal data deficiencies and problems associated with census processes.

CENSUS AUTHORITY

The census of agriculture is required by law under Title 13, United States Code, sections 142(a) and 191. Beginning in 1982, the census has been taken in years ending in 2 and 7.

FARM DEFINITION

Minimum criteria defining a farm for census purposes were first established in 1850. Since 1850, the farm definition has been changed nine times. In 1974, the current farm definition was established as any place from which \$1,000 or more of agricultural products were produced and sold or normally would have been sold during the census year. A place not having sufficient sales to qualify as a farm may qualify on potential sales based on the inventory and production of crops or livestock.

ABBREVIATIONS AND SYMBOLS

The following abbreviations and symbols are used throughout the tables:

- Represents zero.
- (D) Withheld to avoid disclosing data for individual farms.
- (NA) Not available.
- (S) Withheld because estimate did not meet publication standards on the basis of either the response rate (associated relative standard error) or a consistency review.
- (X) Not applicable.
- (Z) Less than half the unit shown.

GENERAL EXPLANATION

	Page
1987 CENSUS OF AGRICULTURE	VII
STUDIES	VIII
THE 1987 PROGRAM	IX
SURVEY DESIGN	IX
ESTIMATION	XI
RESULTS	XII
PRECISION OF THE ESTIMATES	XV
TABLES:	
A. COVERAGE ERROR MODEL	XI
B. PERCENT ESTIMATED CENSUS FARM COVERAGE ERROR FOR SELECTED OPERATION CHARACTERIS- TICS: 1987	XIII
C. ESTIMATED OVERCOUNTED FARMS: 1987	XIV
D. PERCENT ESTIMATED UNDERCOUNTED FARMS FOR SELECTED AGRICULTURAL ITEMS: 1987	XV

1987 CENSUS OF AGRICULTURE

History

The 1987 Census of Agriculture is the 23d taken by the U.S. Department of Commerce, Bureau of the Census. The first agriculture census was taken in 1840 as part of the sixth decennial census of population. From 1840 to 1950, an agriculture census was taken as part of the decennial census. A separate mid-decade census of agriculture was conducted in 1925, 1935, and 1945. From 1954 to 1974, a census of agriculture was taken for the years ending in 4 and 9. In 1976, Congress authorized the census of agriculture to be taken for 1978 and 1982 to adjust the data reference year so that it coincided with the economic censuses covering manufacturing, mining, construction, retail trade, wholesale trade, service industries, and selected transportation activities. This adjustment in timing established the agriculture census on a 5-year cycle collecting data for years ending in 2 and 7.

Mail List

All agriculture censuses beginning with the 1969 census, primarily have used mailout/ mailback data collection. The mail list for the 1987 census was comprised of individuals, businesses, and organizations that could be readily identified as being associated with agriculture. The list was assembled from the records of the 1982 census, administrative records of the Internal Revenue Service (IRS), and the statistical records of the U.S. Department of Agriculture (USDA). In addition, lists of large or specialized operations, such as nurseries and greenhouses, specialty crop farms, poultry farms, fish farms, livestock farms, and cattle feedlot operations, were obtained from State and

Federal agencies, trade associations, and similar organizations. Lists of companies having one or more establishments (or locations) producing agricultural products were obtained from the 1982 census and updated using the information from the Standard Statistical Establishment List maintained by the Census Bureau. Exhaustive record linkage and mathematical modeling yielded a final mail list of 4.1 million names and addresses that had a substantial probability of being a farm operation.

Report Forms

In 1987, three different report forms were used—a two-page, a four-page, and a six-page form to minimize the reporting burden, particularly for small farms and places less likely to be farms. The six-page sample form and the four-page nonsample form are the same, with the exception of sections 23 through 28 which have been added to the sample form to obtain supplemental information from a sample of farms. The information collected in these sections will give the Bureau of the Census a good basis for making estimates of these data for other farms included in the census. The two-page form does not have as many questions or as much detail as the four-page and six-page forms. The four-page form has 11 regional versions and the six-page form has 13 regional versions. Both forms have different crops prelisted. Volume 1, appendix D contains copies of both the two-page and six-page forms.

The six-page form was mailed to 1,104,000 addressees on the mail list, including all those expected to have large (based on expected sales or acreage) or unique operations (farms operated by multiestablishment companies or nonprofit organizations), all those in Alaska and Hawaii, and a sample of other addressees. The two-page form was mailed to 906,000 addressees. These were expected to be small farms or less likely to be farms. The four-page form was mailed to the remaining 2,079,000 addressees. Further discussion of the criteria used to determine which form was mailed to an addressee is provided in the Census Sample Design section of volume 1, appendix C.

Initial Mailing

The report forms were mailed in mid-December 1987 to approximately 4.1 million individuals, businesses, and organizations on the mail list. The mail packages included a report form, a cover letter with a description of the purposes and uses of the census on the reverse side, an information sheet containing instructions for completing the form, and a postage-paid return envelope. Additional special instructions were included with report forms sent to

grazing associations; feedlot operations; institutional organizations; Indian reservations; firms with multiple farm or ranch operations; and producers of poultry under contract, bees and honey, fish, laboratory animals, and nursery and greenhouse crops.

To provide additional help to farmers in completing their reports, copies of an Agriculture Census Guide booklet were sent to vocational agriculture instructors, USDA county offices of the Agriculture Stabilization and Conservation Service, and the Cooperative Extension Service. The Guide contained descriptions and definitions of various items in more detail than the instructions included with each report form. Representatives of the above agencies graciously consented to assist farmers in completing their report forms.

Followup Procedures

A thank you/reminder card was mailed to addressees on the mail list in mid-January 1988. Five followup letters, three of which were accompanied by a report form, were sent to nonrespondents at 4-week intervals starting in mid-February and continuing until early June 1988.

Telephone calls were made to all large farms who had not responded well into the data collection period. In addition, telephone calls were made to a sample of other nonrespondents in counties that had a response rate of less than 75 percent. A nonresponse adjustment procedure was used to represent the final nonrespondent farms in the census results. A description of this procedure is included in the Census Estimation section of volume 1, appendix C.

Data Processing

Selected report forms were reviewed prior to keying the data. These included reports with attached correspondence and reports with remarks or no data on the front page. The data from each report form were subjected to a detailed item-by-item computer edit. The edit performed comprehensive checks for consistency and reasonableness, corrected erroneous or inconsistent data, supplied missing data based on similar farms within the same county, and assigned farm classification codes necessary for tabulating the data. Substantial computer-generated changes to the data were clerically reviewed and verified.

In the computer edit, farms with sales, acreage, or commodities exceeding specified levels were tested for historical comparability. Key items, such as acreage and sales, were compared for substantial changes between 1982 and 1987. Sizeable historical differences were resolved or verified by telephone, if necessary. Respondents who reported sales or acreage above specified levels on non-sample forms were sent correspondence requesting the additional sample data. Prior to publication, tabulated totals were reviewed by statisticians to identify inconsistencies and potential coverage problems. Comparisons were made with previous census data, estimates published by the USDA, and other available data.

STUDIES

Methodology

A number of factors affects census coverage. They include the complexity of farm organizational arrangements, continuing changes in operational status, inadequacies of source lists, and difficulty in communicating census definitions and concepts. Extensive efforts are made each census to obtain a complete and accurate count of all farm operations meeting the census farm definition. Evaluations of questionnaire design, mail list construction, and data collection and processing procedures are conducted to improve coverage and accuracy of census data.

A coverage evaluation program for a census is an important means of measuring census completeness. Data from the coverage evaluation program have provided independent measures of the number and characteristics of farms not on the mail list, farms incorrectly classified as nonfarms, and overcounted farms. The program also helps in identifying problem areas for future improvements in developing the census mail list, in collecting and processing the data, and in developing report form items.

Coverage evaluation programs are designed to measure errors in the census mail list and in farm classification. Mail list error includes a measurement of "farms not on the mail list" and a measurement of "farms duplicated." The first error contributes to census undercount; the second error contributes to census overcount. Classification error includes a measurement of "farms classified as nonfarms" and of "nonfarms classified as farms." Both types of classification errors are referred to as misclassified in tables where undercount or overcount are broken out by component.

The difference between the census count and the true size of the population is defined as net coverage error. The census count may be greater than the true size (an overcount), but usually the error is such that the census count is less than the true size (an undercount). The list error of "farms not on the mail list" dominates other errors for small farms. This component varies considerably by geographic area.

Application of Past Studies to the Census

Since 1945, coverage evaluation studies have been conducted for each census of agriculture. These studies have identified sources of coverage errors. Procedural modifications resulting from these coverage evaluation findings have been introduced into subsequent censuses. Several examples follow.

The 1945 coverage evaluation study showed that operations with sales of less than \$2,500 were a large proportion of the undercounted farms. This was attributed to the instruction interviewers were given to enumerate all places conforming to the farm definition. In 1950, census interviewers were instructed to enumerate all places with specified agricultural activities and the farm definition was applied during the data editing process.

The 1950 evaluation study identified that nonresident farm operators contributed a large proportion of the undercounted farms. Two new techniques were introduced in 1954 to reduce this type of undercount—(1) enumerators in selected counties drew the boundaries of each farm and each nonfarm tract on a township sketch, and (2) a listing book was used to record the location and identification of every residence and every agricultural operation in each enumeration district.

Coverage evaluation studies for the 1969 and 1974 censuses indicated that the source lists acquired for mail data collection did not have complete coverage of small farms with sales of agricultural products of less than \$2,500. In 1969 and 1974, 33 percent (± 2.1)¹ and 27 percent (± 2.0), respectively, of all small farms were undercounted at the U.S. level. The 1978 census mail data collection was supplemented by an area sample for the United States, regional, and State estimates. This dual system estimation reduced the undercount of small farms to about 7 percent (± 0.5) at the U.S. level. Although this methodology was successful in significantly reducing undercount of small farms, budget constraints have prevented subsequent use of area samples as part of the census estimates.

THE 1987 PROGRAM

The coverage evaluation program for the 1987 Census of Agriculture was designed to provide estimates of the components of census coverage at the United States, regional, and State levels. After the 1982 Census of Agriculture and the subsequent coverage evaluation study, data users expressed interest in obtaining State level estimates of farms not on the mail list rather than only United States and regional estimates as produced in 1982. An increase in sample size above the 1982 level was required to provide such State level estimates. Rather than constructing a single purpose area frame, selecting a sample, and conducting a field enumeration survey, the Agriculture Division of the Census Bureau entered into an agreement with the Department of Agriculture, National Agricultural Statistics Service (NASS) to use the 1987 June Enumerative Survey as part of the 1987 Census of Agriculture Coverage Evaluation program. In the agreement, the Census Bureau provided specific requirements for the survey so that the resulting data would be appropriate for use in the census coverage evaluation program. These requirements included specification of additional items for the data collection, an increase in agricultural screening in residential areas, and a 20 percent increase in the sample size in agricultural urban areas.

The objectives of the 1987 program were to:

- provide State estimates of the number of farms not on the mail list,
- provide more reliable national and regional estimates of the number of operations incorrectly classified and the number of duplicate farms than in previous studies,
- provide national, regional, and divisional estimates of selected agricultural characteristics of undercounted farms.

The 1987 June Enumerative Survey (JES) and the 1987 Classification Error Survey (CES) were used to meet the above objectives. The JES was used to provide State level estimates of the number of farms not on the mail list during the 1987 census. For most States, preliminary State level estimates of these farms and their operation and operator characteristics were published in appendix C, table G of Volume 1, Geographic Area Series, 1987 Census of Agriculture and are republished in this report. Divisional level estimates of agricultural production for these farms are published in this report. The CES was used to measure classification error and mail list duplication error. This independent sample was selected from the census mail list, designed to provide regional estimates of classification error with a coefficient of variation of 15 percent. Regional and U.S. level estimates for classification error are published separately, and in addition to the not on the mail list estimates, in this report.

SURVEY DESIGN

Not on the Mail List

The June Enumerative Survey is an annual area sample survey conducted by NASS to measure planted acreage of crops and numbers of livestock. For each State, NASS develops an area frame consisting of land parcels of predetermined size based on easily identifiable boundaries. The size of a sampling unit depends upon a multitude of related factors such as estimation method, data collection costs, data variability among sampling units, population density, concentration of cropland, and the availability of identifiable boundaries for the sampling units.

The basic stratification used by NASS divides all land of a State into about six to eight land use strata such as intensive cultivation, urban areas, and range land. Cultivated land is divided into several strata based on the degree of cultivation throughout the State. Primary sample units (PSU's) are land parcels selected within each land use stratum. Each randomly selected PSU is further divided into several ultimate sample units or segments. The size of the PSU's vary, but on the average, each PSU contains about six to eight segments. Each segment contains on the average three farm operations, one of which includes a resident operator. The JES is a two stage stratified probability area sample of U.S. farm operations. In the first stage, PSU's were selected with probability proportional to

¹Represents 1.65 standard errors above and below estimate. Intervals published in past Coverage Evaluation reports represent 2 standard errors above and below the estimate. See Precision of the Estimates for further discussion of confidence intervals.

the number of segments within each PSU of the substratum. In the second stage, a segment was selected with equal probability from each selected PSU.

Each distinct farm operation within a segment is defined as a tract. Land areas within segments that contain one or more occupied dwellings are designated as residential tracts. These residential tracts are more predominant in the agricultural urban strata (strata that include the perimeter of cities and small towns where residences merge into agricultural areas) but do occur in other strata. Such tracts are of particular interest for measuring farms not on the mail list, as they often include small acreage farms or farm operators not residing at the site of their farm operation. These types of operations are characteristic of those missing from previous censuses because they frequently are not found on census mail list source records.

The Census Bureau requested more intensive screening procedures than NASS generally used in the residential tracts to improve the survey coverage of farms less likely to be on the census mail list. Enumerators were required to compile a listing of all houses in each segment, to sample households according to a given procedure, to ask questions that screen for agricultural activities of the household members, and to inquire whether the resident knew of any neighbors with agricultural activity. The objective of the subsampling procedure was to provide a systematic method for obtaining information about all tract residents. Listing and prescreening were conducted in the fall of 1986 for all residential tracts with 11 or more households in the 1986 JES sample and, in May of 1987, for all smaller residential tracts and other types of tracts in the sample.

To increase the reliability of the estimates of farms not on the mail list, NASS retained in the sample 20 percent of the agricultural urban segments from 1986 that had been scheduled to be rotated out of the sample in 1987. This procedure provided a 20 percent overall increase in the number of agricultural urban segments across all States, as well as an increase in the residential tracts in the sample. In several States, additional minor increases were made in the number of agricultural urban segments to meet census sample size requirements.

The initial files received from the JES data collection consisted of names, addresses, and agricultural data on all sample area segment residents that had any indication of agricultural activity. Appendix A lists the data items on the file received from NASS. An initial match of this data file to the census mail list development file identified all JES area sample records with a special code for census processing. All JES sample records not on the mail list were included in the census data collection.

During the coverage processing, potential match records were clerically reviewed using both census and JES data. In particular, number and type of crops, livestock, total value of products sold, and land usage were used to determine the farm status of area sample cases that did not respond to the census and to resolve cases where the JES and census farm status differed. Coverage classification codes were assigned to each record during census

processing to specify the farm status—farm (in-scope) or nonfarm (out-of-scope)—and relationship to the census mail list—match (on the mail list) or nonmatch (not on the mail list). Both the match and nonmatch farm records constituted the sample for deriving estimates of farms not on the mail list.

The data were keyed, edited, and reviewed for accuracy and consistency prior to tabulation. A final data file was created for estimation. This file was used independently to produce the State and divisional estimates in tables 5 and 6 of this report and with the CES data file to produce the regional and U.S. estimates in all other report tables.

Measurement of Incorrectly Classified Farms

The Classification Error Survey used a specially designed questionnaire entitled "*Evaluation of the 1987 Census of Agriculture.*" The purpose of the report form was to collect agriculture data to measure the accuracy of the census farm counts. A series of screening questions were initially asked to determine if any person in the household had or was associated with any agriculture operations in 1987. The evaluation form contained questions about alternate farm names and addresses used for the operation; farm size, crops, and livestock; and various operator characteristics. See appendix B for a copy of the report form.

The CES used an independent regionally stratified systematic random sample from the final census mail list. Specified operations were excluded from sample selection—all operations in Alaska and Hawaii, all operations with expected sales of \$500,000 or more, and all multiunit or abnormal operations (Indian reservations, research farms, experimental farms, institutional farms). These operations were considered inappropriate for a study of classification error due to their uniqueness or the use of more intensive census processing procedures for the operation. No classification error estimates were made for this portion of the census farm universe. With these exclusions, the universe for the CES consisted of 4,017,213 mail list records.

An initial sample of 18,500 names and addresses was selected from the CES universe. The sampling rate varied by census geographic region: 1 in 71 in the Northeast, 1 in 500 in the Midwest, 1 in 176 in the South, and 1 in 227 in the West. At designated cut-off dates, the census report form check-in status was obtained for all sample addresses using a unique CES evaluation code set in the census data base at the time of sample selection. Two waves of census respondents were selected for mailing the survey questionnaire according to date of response with mailouts in March and July, respectively. A total of 15,331 sample survey cases were mailed in the two waves. The mail data collection procedure for each group required a postcard followup 2 weeks after initial mailout, a form followup 4 weeks after initial mailout, and a telephone followup 6 weeks after initial mailout for nonrespondents only.

A separate technical review of CES sample questionnaires was conducted to classify operations as either farms or nonfarms. A determination of "true" farm status

was made by comparing the CES and census farm status and data. Coverage classification codes were assigned after the survey and census reconciliation to identify farm operations and to define the relationship (match or non-match) between the census farm status and the CES farm status. Approximately 5.8 percent (889) of the sample cases were not classifiable and 0.7 percent (100) cases were survey nonrespondents. These cases were accounted for in the sample weighting. A final data file was created for estimation processing. It was used in conjunction with the final JES data file to produce the regional and U.S. estimates in the text tables and tables 1 through 4 of this report.

ESTIMATION

The "true" universe total (T) of all farms in the United States can be represented as the census published farm count (C) minus the number of overcounted farms (OV) plus the number of undercounted farms (U); i.e.

$$T = C - OV + U \quad (1)$$

The undercount (U) can be split into a component consisting of farms on the census mail list that were incorrectly classified as nonfarms (MCF) and a component consisting of farms not on the census mail list (NML); i.e.

$$T = C - OV + MCF + NML \quad (2)$$

Unbiased weighted sample estimates of OV and MCF, incorporating the nonclassifiable and nonresponding survey cases, are obtained from the CES. These estimates are used to obtain the census farm count adjusted for classification error (C') where,

$$C' = C - OV + MCF \quad (3)$$

Then the "true" universe total of farms (T) can be represented by:

$$T = C' + NML \quad (4)$$

The estimator T is an application of the Petersen Coverage Error Model (Wolter, JASA, 1986). The model uses an independent survey (here, the JES) in conjunction with the census to provide an estimate of the "true" universe total of farms (T). For a variety of reasons, both the census and survey enumerations miss farm operations. The Petersen Coverage Error Model consolidates the estimates from two enumerations to estimate the true total. The model assumes that: (a) both the census, which is observable, and the universe of the JES, which is not observable, attempt to enumerate accurately the complete universe of farms, and farms reported on either list are true farms; (b) the event of being included by the census is independent of being included in the survey; (c) the probability of being missed by either the census or the

survey is the same for all farms within a given size category; and (d) every farm in the complete universe of farms has, independently of every other farm, the same chance of being listed in the census and, independently of listing in the census, the same chance of being listed in JES. Each farm in the universe can be placed into one of the cells in table A.

Table A. Coverage Error Model

Census list farms	In survey universe	Not in survey universe	Total
On mail list	N_{11}	N_{12}	N_{1+}
Not on mail list	N_{21}	N_{22}	
Total	N_{+1}		

where,

N_{11} = the number of farms on the census mail list and in the JES sample resulting from the match of the JES area sample farms to farm records in the census,

N_{12} = the number of farms on the census mail list but not in the JES sample,

N_{21} = the number of farms in the JES sample but not on the census mail list resulting from the match of the JES area sample farms to farm records in the census,

N_{22} = the number of farms not on the census mail list and not in the JES sample,

N_{1+} = the number of farms on the census mail list adjusted for classification error; note that $N_{1+} = C'$,

N_{+1} = the number of farms in the JES sample.

The universe weighted estimator of N_{12} is:

$$\hat{N}_{12} = N_{1+} - \hat{N}_{11} \quad (5)$$

where,

\hat{N}_{11} = the JES sample estimate of number of farms on the census mail list with characteristic x.

The universe weighted estimator of N_{22} is:

$$\hat{N}_{22} = (\hat{N}_{21}) (\hat{N}_{12}) / \hat{N}_{11} \quad (6)$$

where,

\hat{N}_{21} = the JES sample estimate of number of farms not on the census mail list with characteristic x.

The maximum likelihood estimate of total farm count provided by the model is:

$$T = [(\hat{N}_{11} + \hat{N}_{12})(\hat{N}_{11} + \hat{N}_{21})] / (\hat{N}_{11})$$

$$= N_{1+} + (N_{1+}) (\hat{N}_{21} / \hat{N}_{11}) \quad (7)$$

Then, total (T_x) for some characteristic x of farm operations can be estimated by:

$$\hat{T}_x = \hat{C}'_x + (\hat{S}_x) (N_{1+} / \hat{N}_{11}) \quad (8)$$

$$= \hat{C}'_x + (\hat{S}_x / \hat{N}_{21}) (\hat{N}_{21}) (N_{1+} / \hat{N}_{11})$$

$$= \hat{C}'_x + (\hat{S}_x / \hat{N}_{21}) (\hat{N}_{21} + \hat{N}_{22})$$

$$= \hat{C}'_x + \widehat{NML}_x \quad (9)$$

where,

- \hat{C}'_x = the total value for farms in the census with the characteristic x adjusted for classification error; i.e. $\hat{C}'_x = C_x - \widehat{OV}_x + \widehat{MCF}_x$
- N_{1+} / \hat{N}_{11} = ratio of number of census farms to the JES sample estimate of number of farms on the census mail list with characteristic x.
- \hat{S}_x = the unbiased JES sample open segment estimate for farms not on the census mail list with characteristic x.

Thus, the estimate for some characteristic x of farms not on the mail list is:

$$\widehat{NML}_x = \hat{S}_x + (\hat{S}_x / \hat{N}_{21}) (\hat{N}_{22}) \quad (10)$$

Note that for farm count, $\hat{S}_x = \hat{N}_{21}$ and $\widehat{NML}_x = \hat{S}_x + \hat{N}_{22}$.

For a given characteristic x, both the estimated census farm count and the associated census value were adjusted for farm classification error. The components of \hat{C}'_x and \hat{C}'_x were computed at the regional level separately for farms with sales of less than \$2,500 and for farms with sales of \$2,500 or more and summed to produce estimated totals. The ratio of the census farm count to the JES estimate of census farm count, (N_{1+} / \hat{N}_{11}) , was similarly computed separately for these two sales groups at the State, divisional, or regional publication level; each sales break ratio was then multiplied by the JES estimate of farms in JES but not on the census mail list (\hat{S}_x) for that sales break and summed to estimate \widehat{NML}_x at that geographic level. Thus, the sum of the estimates of \hat{C}'_x and \widehat{NML}_x is the estimate of the "true" universe total for the characteristic (T_x). U.S. estimates were obtained by summing the regional estimates.

For a given characteristic x, the JES sample open segment estimate of JES farms not on the census mail list was computed by:

$$\hat{S}_x = \sum_{i=1}^2 \sum_{j=1}^{n_i} e_j \left(\sum_{k=1}^{m_j} s_{ijk} \right) \quad (11)$$

- i = sales group (< \$2,500 and greater than or equal to \$2,500),
- j = segment within group,
- k = farm within segment,
- n_i = number of sample segments in i^{th} group,
- e_j = expansion factor or inverse of the probability of selection for the j^{th} segment,
- m_j = number of farms in j^{th} segment,
- s_{ijk} = value of farm k in j^{th} segment, i^{th} sales group with characteristic x.

As previously mentioned, State level estimates of counts and selected characteristics of farms not on the mail list were published in the State census volumes. As estimates of classification error were not available at the State level, the estimated farm counts used in the volume 1, State reports were not adjusted for classification error. The bias resulting from not adjusting for classification error at the State level was considered to have less impact on mean square error than the increase in variance incurred by deriving a State level estimate including classification error.

RESULTS

The estimated net farm coverage error was 7.2 percent (± 0.5) for the U.S. resulting in an estimated census coverage of 92.8 percent. Approximately 13.2 percent (± 0.5) of estimated total farms were undercounted and approximately 6.0 percent (± 0.5) were overcounted. The net farm coverage error rate was 15.0 percent (± 0.4) in 1969, 10.7 percent (± 0.3) in 1974, 3.4 percent (± 0.4) in 1978, and 9.1 percent (± 2.6) in 1982 as shown in figure 2. The low net farm coverage error in 1978 resulted primarily from the inclusion of the area sample in the census.

Farm counts and values are tabulated by estimated farms, census published farms, undercounted (misclassified and not on the mail list) farms, and overcounted farms. U.S. estimates are presented in tables B, C, and D; U.S. and regional estimates in tables 1 through 4; divisional and State estimates in table 5; and divisional estimates in table 6. Table B presents selected U.S. values on estimated farms, estimated undercounted farms, and estimated overcounted farms as a percent of table 1; U.S. estimated farm totals by sales group, standard industrial classification, size in acres, and operator characteristics. Table C presents count of overcount misclassified and duplicate farms in the four census regions. The base for this percent is census published farms plus estimated undercounted farms minus estimated overcounted farms. Table D presents the U.S. percent estimated undercount for agricultural commodity items provided in table 2. The base for this percent is the total of census published and estimated undercounted farms.

Table B. Percent Estimated Census Farm Coverage Error for Selected Operation Characteristics: 1987

[Detail may not add to total due to rounding]

Characteristics	Estimated farms ¹		Estimated undercounted farms ²		Estimated overcounted farms ⁴		Estimated net farm coverage error ⁵	
	Total (number)	Relative standard error of estimate (percent)	Percent ³	Standard error of percent	Percent ³	Standard error of percent	Percent ³	Standard error of percent
Total	2 243 648	.4	13.2	.3	6.0	.3	7.2	.3
Farms by value of sales:								
Less than \$2,500	662 600	.8	32.3	.8	6.0	.5	26.3	2.0
\$2,500 or more	1 581 048	.4	5.2	.2	6.0	.3	-.8	-
\$2,500 to \$9,999	559 943	.6	9.8	.4	5.6	.5	4.2	.4
Less than \$10,000	1 222 543	.5	22.0	.5	5.8	.3	16.2	1.0
\$10,000 or more	1 021 105	.4	2.8	.2	6.3	.4	-3.6	.3
Farm by standard industrial classification:								
Crops (01)	921 373	.6	9.8	.3	7.9	.5	1.8	.1
Livestock (02)	1 322 274	.5	15.6	.4	4.7	.3	10.9	.7
Farms by size:								
1 to 49 acres	752 362	.8	27.9	.7	6.5	.5	21.4	1.6
50 to 179 acres	672 674	.6	10.0	.4	5.8	.4	4.2	.4
180 to 499 acres	467 925	.6	3.4	.2	5.6	.5	-2.2	.3
500 acres or more	350 689	.7	1.1	.1	6.2	.7	-5.1	.9
Farms by tenure of operator:								
Full owners	1 369 595	.5	17.3	.4	7.5	.4	9.8	.5
Part owners	623 020	.5	5.8	.3	3.4	.4	2.4	.3
Tenants	251 034	1.0	9.6	.7	4.6	.7	5.0	.9
Farms by age of operator:								
Under 35 years	302 485	1.0	14.4	.7	6.2	.8	8.1	1.1
35 to 54 years	959 733	.5	14.9	.4	4.9	.4	10.0	.8
55 years and over	981 430	.5	11.2	.3	7.1	.4	4.2	.3
Farms by principal occupation of operator:								
Farming	1 136 704	.4	5.9	.2	5.7	.4	.1	-
Other	1 106 945	.6	20.8	.5	6.3	.4	14.4	.9

¹Census published farms plus estimated undercounted farms minus estimated overcounted farms.

²Includes misclassified farms (farms classified as nonfarms) and farms not on mail list.

³Base for percent is estimated farms (column 1).

⁴Includes nonfarms classified as farms and multiple enumerations of 1 farm operation.

⁵Difference of estimated undercounted farms and estimated overcounted farms.

For farms with value of agricultural products sold of \$2,500 or more, the net farm coverage error was -0.8 percent,² an overcount. Of the \$2,500 or more sales group, 5.2 percent (± 0.3) were undercounted and 6.0 percent (± 0.5) were overcounted. The net farm coverage error for this group was 6.5 percent (± 0.7) in 1969, 6.8 percent (± 0.5) in 1974, 3.6 percent (± 0.3) in 1978, and 0.5 percent³ in 1982. This group has a low net farm coverage error because larger farms are more likely to be included in census source lists and were subjected to more intensive processing procedures to ensure that they were included.

An estimated 73.7 percent (± 1.2) of farms with value of agricultural products sold of less than \$2,500 were in the census. Of these farms, 32.3 percent (± 1.3) were undercounted and 6.0 percent (± 0.8) were overcounted. The

gross undercounted rate for this group was 33.3 percent (± 2.1) in 1969, 27.3 percent (± 2.0) in 1974, 7.1 percent (± 2.5) in 1978, and 28.6 percent (± 5.5) in 1982.

From tables B and 1, the estimated number of undercounted farms was 296,933 ($\pm 9,309$) or 13.2 percent (± 0.5). About 4 of 5 undercounted farms were not on the mail list. The remainder were farms incorrectly classified as nonfarms (misclassified). While about 27.9 percent of the undercounted farms had value of agricultural products sold of \$2,500 or more, only about 9.5 percent of the undercounted farms were larger farms with sales of \$10,000 or more. Of the undercounted farms, 70.7 percent had 49 acres or less, 22.6 percent had from 50 to 179 acres, 5.4 percent had from 180 to 499 acres, and 1.3 percent had 500 acres or more. Of these farms, 79.8 percent were operated by full owners, 12.1 percent by part owners, and 8.1 percent by tenants. About 2 of 3 undercounted farms were livestock farms and 1 of 3 were crop farms.

²Standard error for estimate is essentially zero.

³Relative standard error was over 100 percent and not published.

Table C. Estimated Overcounted Farms: 1987

[Detail may not add to total due to rounding]

Geographic area	Estimated total		Misclassified		Duplicate	
	Total (number)	Relative standard error of estimate (percent)	Total (number)	Relative standard error of estimate (percent)	Total (number)	Relative standard error of estimate (percent)
United States	135 600	4.2	72 310	5.8	63 290	6.3
Northeast	6 030	9.4	2 430	14.9	3 600	12.3
Midwest	57 776	8.2	29 516	11.5	28 260	11.8
South	50 344	5.1	31 784	6.4	18 560	8.4
West	21 450	9.0	8 580	14.4	12 870	11.7

The estimated number of overcounted farms was 135,600 ($\pm 9,397$) or 6.0 percent (± 0.5). About 1 out of 2 overcounted cases were nonfarms incorrectly classified as farms (misclassified). The remainder were farms with more than one census report (reports duplicated for a single farm or multiple reports for parts of a single farm). Table C presents the estimated overcounted farms included in the census by misclassified and duplicate for the United States and its regions. The distribution of the two components varies considerably by region. Of the overcounted farms, 36.1 percent had 49 acres or less, 28.7 percent had from 50 to 179 acres, 19.2 percent had from 180 to 499 acres, and 16.0 percent had 500 acres or more. Of these farms, 76.0 percent were operated by full owners, 15.5 percent by part owners, and 8.5 percent by tenants. About 2 of 5 overcounted farms were livestock farms and 3 of 5 were crop farms.

Census coverage differed by the census regions as provided in figure 1. The undercount coverage components were lower in the Midwest Region than in the other three census regions as illustrated in figure 3. Overall coverage was substantially more complete for the Midwest Region for all sizes of farms as illustrated in figure 4 using value of sales as the measures of size. This is due primarily to the higher proportion of larger farms in the Midwest Region which are more likely to be included on the source lists for the census.

Table D and tables 2 through 4 present estimated totals and relative standard errors of selected agricultural commodity items (tables D and 2), of acreage (table 3), and market value of agricultural products (table 4) for undercounted farms. Figure 5 provides estimated percent undercounted farms, land in farms, and value of sales for the four census regions. The estimated total (or base for the percents) is the sum of the census published and estimated undercounted farms. Figure 6 provides the composition of undercounted farms by sales, commodity, acreage, ownership of the operation and by age range, principal occupation, and sex of the operator.

The percent estimates for most agricultural commodity data items in table D are much less than the 12.5 percent (± 0.4) estimate of undercounted farms. This can be seen dramatically in figure 5. Thus, it is advisable to consider coverage component estimates relative to the census

agricultural data as well as to the farm count. Estimates of the total number of undercounted farms or the proportion of undercounted farms alone are not a complete indication of the quality of the census data. The undercount estimates for agricultural data items do not represent total error for these characteristics, probably understating it. Comparable estimates for overcounted farms could not be derived from the coverage evaluation sample nor were estimates for incorrect item reporting on these farms measured. The estimates for the data items in these tables from undercounted farms are likely to contribute a larger proportion of total coverage error than from overreporting of these items for farms included in the census.

Table D presents the U.S. estimates of percent undercount for the selected crop and livestock items that are presented in more detail at the regional level in table 2. From table 2, the estimated number of farms not on the mail list was 242,853 ($\pm 7,613$), the estimated number of farms incorrectly classified as nonfarms (misclassified) was 54,080 ($\pm 5,354$) for an estimated total of 296,933 ($\pm 9,309$) undercounted farms. These 296,933 farms have 1.9 percent (± 0.2) of the total census and undercounted farm acreage. The estimated undercounted acres were approximately 18.4 million ($\pm 1,938,000$) with the average size undercounted farm having about 62 acres (± 6.8) compared to 462 acres (± 0.2) for the average census farm.

Table 3 presents estimates of the land in farms by sales group for estimated undercounted farms. It was estimated that 98.1 percent² of the land in farms in the United States was in the census. Undercounted farms accounted for 1.1 percent (± 0.2) of the estimated total acres of farms with sales of \$2,500 or more, and 12.5 percent (± 1.1) of farms with sales of less than \$2,500. Table 4 presents the estimates for the value of agricultural products sold by sales group. It was estimated that 98.8 percent² of the value of agricultural products sold in the United States was in the census. Undercounted farms accounted for 1.1 percent (± 0.2) of estimated farms whose market value of sales were \$2,500 or more, and 24.2 percent (± 3.5) of estimated farms whose sales were less than \$2,500.

²See footnote 2 on p. XIII.

Table D. Percent Estimated Undercounted Farms for Selected Agricultural Items: 1987

[Detail may not add to total due to rounding]

Item	Census published and undercounted farms ¹ (number)	Census published farms (number)	Estimated undercounted farms ² (number)	Percent estimated undercount ³
Farms.....number..	2 379 248	2 082 315	296 933	12.5
Land in farms.....acres..	980 072 515	961 722 372	18 350 143	1.9
Average size of farm.....acres..	411.9	462.0	61.8	(X)
Market value of agricultural products sold.....\$1,000..	137 077 713	135 420 801	1 656 912	1.2
Corn for grain.....farms..	652 550	627 602	24 948	3.8
.....acres..	59 264 383	58 701 505	562 878	.9
Sorghum for grain.....farms..	91 772	89 642	2 130	2.3
.....acres..	9 992 629	9 760 574	232 055	2.3
Wheat for grain.....farms..	358 222	352 231	5 991	1.7
.....acres..	53 713 718	53 224 055	489 663	.9
Soybeans for beans.....farms..	451 625	441 899	9 726	2.2
.....number..	55 787 803	55 291 205	496 598	.9
Hay.....farms..	1 069 759	994 253	75 506	7.1
.....number..	59 330 280	57 943 823	1 386 457	2.3
Tobacco.....farms..	145 565	136 682	8 883	6.1
.....number..	6 544 916	6 333 104	211 812	3.2
Cattle and calves inventory.....farms..	1 323 567	1 175 173	148 394	11.2
.....number..	97 472 804	95 626 821	1 845 983	1.9
Hogs and pigs inventory.....farms..	279 372	242 981	36 391	13.0
.....number..	53 382 391	52 222 983	1 159 408	2.2
Hens and pullets of laying age inventory.....farms..	192 106	141 622	50 484	26.3
.....number..	321 972 049	315 507 008	6 465 041	2.0

¹Differs from estimated farms given in tables B and 1 because overcounted farms are not subtracted from total.

²Includes misclassified farms (farms classified as nonfarms) and farms not on mail list.

³Base for percent is total for census published and undercounted farms (column 1).

Tables 5 and 6 present estimates of farms not on the mail list. Table 5 presents the estimates of farms not on the mail list by census divisions (as shown in figure 1) and state for the number of farms by sales groups, market value of agricultural products sold, standard industrial classification, and selected operator characteristics. Table 6 presents division estimates of selected agricultural commodities for farms not on the mail list.

PRECISION OF THE ESTIMATES

As previously mentioned the estimates in this report are derived from the 1988 June Enumerative Survey and the Classification Error Survey samples. Thus, there are two possible types of errors—sampling and nonsampling. Sampling errors occur because observations are made only on a sample, not on the entire population. The JES and CES sample selection and estimation procedures contributed to the sampling errors. Nonsampling errors encompass all other types of errors associated with estimates from survey data collections that are not related to sample design. These are discussed further in the section on nonsampling errors. The “accuracy” of a survey result is determined by the joint effect of sampling and nonsampling errors.

Sampling Error

Variability in the estimates was due to JES and CES sample selection procedures, sample estimation procedures, and census nonresponse estimation procedures.

For JES, sampling errors for number and characteristics of farms not on the mail list were estimated assuming a random sample with replacement. A Taylor series approximation was used for the ratio estimates. For CES, sampling errors for MCF and OV were estimated assuming a random sample without replacement. The variance of the total “true” farm count and “true” value of characteristics was calculated as the sum of the variance for each error component, derived independently from JES and CES. Zero covariances were assumed for deriving the error component variances and the total variances.

The post stratification based on the sales breaks were used for computing NML variance estimates at the State and divisional levels and MCF and OV variance estimates at the regional level. The NML regional variance estimates are the sum of the corresponding divisional variance estimates. The U.S. variance estimates for all error components are the sum of their respective regional variance estimates. The formulae for the variance estimators are given in Wright (1989).

The samples for the JES and CES were, respectively, each one of a large number of possible samples of the same size that could have been selected using the same sample design. Estimates derived from different samples would differ from each other. The difference between a sample estimate and the average of all possible sample estimates is called the sampling deviation. The standard error or sampling error of a survey estimate is a measure of

the variation among the estimates from all possible samples, and thus is a measure of the precision with which an estimate from a particular sample approximates the average result of all possible samples.

Estimates of sampling variability are expressed as relative standard errors or standard errors in the tables. The estimated relative standard error (percent) of a statistic is derived by dividing the estimated standard error for the statistic by the statistic and multiplying by 100. When the statistic is a percent, the estimated standard error of the percent is the published measure of variability. These estimates only measure the sampling errors associated with the survey. If all possible samples were selected, each of the samples surveyed under essentially the same conditions, and an estimate and its standard error calculated from each sample, then:

- a. Approximately 67 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average value of all possible samples.
- b. Approximately 90 percent of the intervals from 1.65 standard errors below the estimate to 1.65 standard errors above the estimate would include the average value of all possible samples.

The computations necessary to define the above confidence statements are illustrated in the following example. The estimated total number of farms in the United States in table 1 is 2,243,648 with an estimated standard error of 8,975 and an estimated relative standard error of approximately 0.4 percent (i.e., 8,975 divided by 2,243,648 times 100). The chances are about 2 out of 3 (67 percent) that complete coverage using the same methods would yield between 2,234,673 and 2,252,623 farms. Similarly, a 90 percent confidence interval is 2,228,839 to 2,258,457 (i.e. plus or minus 14,809). As calculated, the standard error does not measure the effect of nonsampling errors.

Estimates of the standard error of derived statistics may be of interest to the user. Formulas and examples illustrating their application are provided below for the net coverage error (undercount minus overcount) and for any ratios of estimated components of coverage error to estimated census totals for a given characteristic. The estimated totals and corresponding relative standard errors for the components of coverage error used in these formulas are provided in table 1.

The standard error of the estimated net coverage error (E) for any characteristic in table 1 may be estimated by using the formula:

$$\hat{\sigma}(\hat{E}) = \sqrt{\hat{U}^2 \hat{V}_1^2 + \hat{OV}^2 \hat{V}_2^2} \quad (12)$$

where \hat{E} is the estimated net coverage error, \hat{U} is the estimated undercount, \hat{OV} is the estimated overcount, and \hat{V}_1 and \hat{V}_2 are the estimated relative standard errors of \hat{U} and \hat{OV} , respectively.

Example 1—To compute the standard error of the net coverage error for the total number of farms in the Northeast Region, apply the formula as follows:

$$\hat{U} = 28,994$$

$$\hat{OV} = 6,030$$

$$\hat{V}_1 = .049$$

$$\hat{V}_2 = .094$$

$$\hat{E} = \hat{U} - \hat{OV} = 22,964$$

$$\begin{aligned} \hat{\sigma}(\hat{E}) &= \sqrt{(28,994)^2 (.049)^2 + (6,030)^2 (.094)^2} \\ &= 1,528.9 \end{aligned}$$

The estimated relative percent standard error of \hat{E} is equal to:

$$\begin{aligned} \hat{V}(\hat{E}) &= \frac{\hat{\sigma}(\hat{E})}{\hat{E}} (100) \\ &= \frac{1,528.9}{22,964} (100) \\ &= .0666 (100) \\ &= 6.7 \text{ percent} \end{aligned}$$

The standard error for the ratio (R) of any estimated coverage component (X) such as undercount, the overcount, or the net coverage error for farms to the estimated total farms at the regional level can be approximated by the formula:

$$\hat{\sigma}(\hat{R}) = \sqrt{\hat{T}^{-2} \hat{\sigma}^2(\hat{X}) (1 - 2\hat{R} + \hat{R}^2)} \quad (13)$$

where $\hat{\sigma}^2(\hat{X})$ is the variance of the estimated coverage error component (\hat{X}), \hat{T} is the coverage estimated farm total, and R is the ratio of \hat{X} to \hat{T} . The values for \hat{X} and \hat{T} are given in table 1. The standard error of these ratios at the U.S. level is given in table A for most of the data items in table 1. To calculate the standard error of these ratios at the regional level use formula (13).

Example 2—To compute the standard error of the ratio of the estimated net coverage error for farms to the estimated total farms in the Northeast Region, apply the formula as follows:

$$\hat{U} = 28,994$$

$$\hat{OV} = 6,030$$

$$\hat{E} = \hat{U} - \hat{OV} = 22,964$$

$$\begin{aligned} \hat{T} &= 146,446 \\ \hat{R} &= \hat{E}/\hat{T} = .157 \\ \hat{\sigma}(\hat{E}) &= 1528.9 \text{ (from Example 1)} \\ \hat{\sigma}(\hat{R}) &= (146,446)^{-2} (1528.9)^2 [1-2(.157) + (.157)^2] \\ &= .010 \end{aligned}$$

Nonsampling Error

Nonsampling errors for the coverage estimate can be attributed to such sources as incorrectness of the model used for the estimation, flaws in the questionnaire design, incompleteness of the enumerations, inaccuracies in the census and JES survey list matching procedures, and incorrect or inaccurate data reporting and processing procedures for JES, CES, and the census. Extensive efforts were made to minimize nonsampling errors in the surveys for this study through the use of quality control and other verification measures.

Failure to classify 2.5 percent (619) of the total JES sample cases represents a potential source of nonsampling error in the coverage evaluation study. These unclassified cases resulted from households that did not supply all or some of the required data either in the JES enumeration or on the census report form and could not be contacted after extensive followup procedures. Approximately 8.9 percent (488) of the nonmatched cases and 0.7 percent (131) of the matched cases were not classified. These unclassified cases were removed from the sample estimation procedures. No separate adjustment was made to the coverage estimates for these cases, thus resulting in a potential bias in the estimates of farms not on the mail list. Further examination of these cases (as well as the nonclassifiable and nonresponding survey cases in the CES which were accounted for in the sample estimation) is planned for 1992 to develop alternative processing or estimation procedures.