

Using the Census of Agriculture List Frame to Assess Misclassification in the June Area Survey

Denise A Abreu¹, Pam Arroway², Andrea C. Lamas¹, Kenneth K. Lopiano³,
Linda J. Young³

¹National Agricultural Statistics Service, USDA, 3251 Old Lee Hwy, Fairfax VA 22030

²Department of Statistics, North Carolina State University, Raleigh, NC 27695

³Department of Statistics, University of Florida, Gainesville, FL 32611

Abstract

The National Agricultural Statistics Service (NASS) conducts an annual area frame based survey, the June Area Survey (JAS). Also, the quinquennial Census of Agriculture is conducted in years ending in 2 and 7. The census has a dual frame: an independent list frame and the area frame from the JAS. Both surveys produce an indication of the number of farms, and it is expected that the farm/non-farm classification for operations should generally agree between the two sources. In 2007, an evaluation of differences in classification for specific farms across the two surveys revealed that most classification errors occurred within the JAS, not the census. The study was later expanded to evaluate all operations which were misclassified on either frame. The characteristics of the farms that were misclassified by either the JAS or the census are discussed.

Key Words: Classification Errors, Area Frame, List Frame, Record Linkage

1. Introduction

Each year the National Agricultural Statistics Service (NASS) publishes an estimate of the number of farms in the United States (U.S.) based on the June Area Survey (JAS). A farm is defined as a place from which \$1,000 or more of agricultural products were produced and sold, or normally would have been sold, during the year, including government payments. An independent estimate of the number of farms is also published from the quinquennial Census of Agriculture, which is conducted in years ending in 2 and 7. At the end of each five year period, the JAS number of farm estimates are adjusted based on intercensal trends. The annual estimate of the number of farms from the JAS has been declining steadily between censuses (especially between the 2002 and 2007 Censuses) as depicted in Figure 1. Furthermore, the 2007 estimate from the JAS was significantly below that from the census; and the required intercensal trend adjustment to the JAS was unexpectedly large as shown by the circled area in Figure 1. The discrepancy between the two surveys' estimates was large enough that could not be simply attributed to the sampling error.

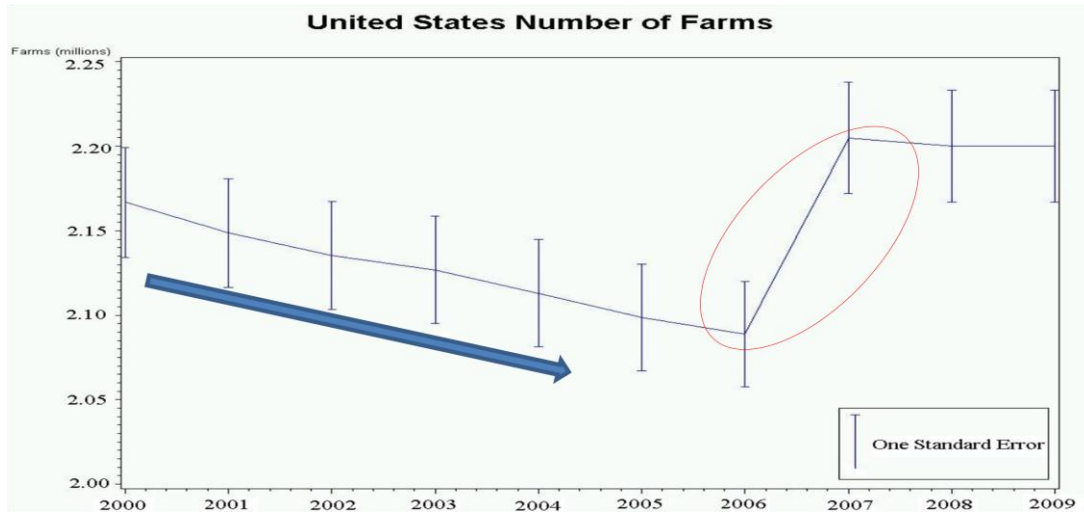


Figure 1: Published estimates of the number of U.S. farms from 2000 to 2009 and bars with length one standard error on either side of the estimate.

Historically, after each census, an evaluation has been conducted to measure misclassification of farms on the census mail list. This evaluation involves either recontacting a sample of census respondents or overlap matching the census mail list to the JAS area frame. In the 2007 Census of Agriculture, classification errors were measured by comparing an operation's status on the census to its status on the JAS. The 2007 evaluation of census classification errors showed that most of the discrepancies were actual errors which occurred in the JAS, not the census (Abreu et. al, 2007). These results suggested that with misclassification present in the JAS, the number of farm estimates resulting from the JAS were biased downward. This provided the first indication of an underlying cause that could help explain the discrepancy in published estimates presented in Figure 1. However, the results of the study were based on a small sample of 67 respondents. In the framework of this report, the independence between the 2007 Census list frame and the 2007 JAS area frame is used to assess misclassification in the JAS for all tracts for the first time. This research presents a deviation for NASS. Prior to 2007, NASS assumed that there was little or no misclassification error in the JAS, primarily because it's a survey conducted with face-to-face interviews. This report outlines the characteristics of farms misclassified on the JAS and proposes potential ways to correct the number of farms estimates for this type of survey error.

2. Background

NASS conducts an annual area-frame-based survey which collects information about U.S. crops, livestock, grain storage capacity, and type and size of farms. The distribution of crops and livestock can vary considerably across a state in the United States. Therefore, the precision of the survey indications or statistics can be substantially improved by dividing the land in a state into homogeneous groups or strata and optimally allocating the total sample to the strata. The basic stratification employed by NASS involves: (1) dividing the land into land-use strata such as intensively cultivated land, urban areas and range land, and (2) further dividing each land-use stratum into substrata by grouping areas that are agriculturally similar. The June Area Survey (JAS) uses a sample comprised of designated land areas (segments) selected from this stratification which field enumerators visit to collect data on all agricultural activity occurring therein.

A typical segment is about one square mile, which is equivalent to 640 acres. Each segment is outlined on an aerial photo which is provided to the appropriate field enumerator (See red outlined area in Figure 2).

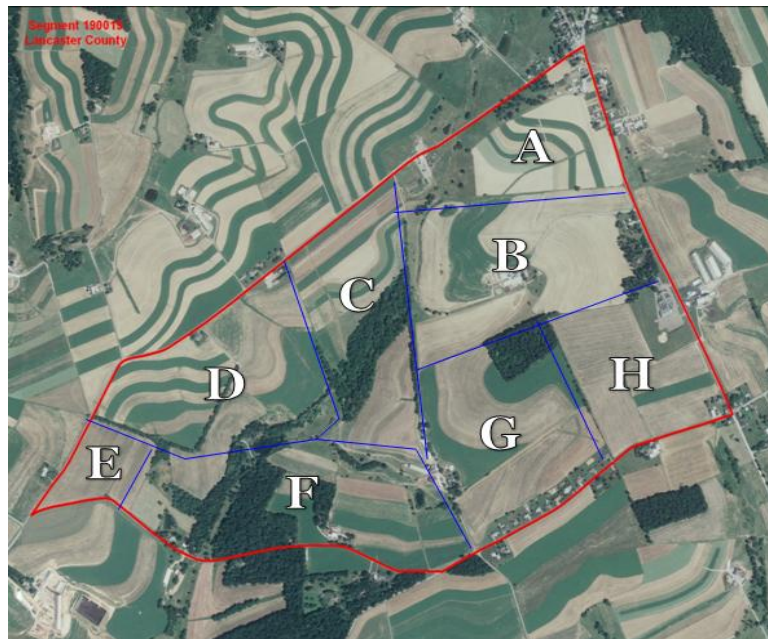


Figure 2: JAS Segment and Tract Boundaries.

Through field enumeration, a segment is divided into tracts of land, each representing a unique land operating arrangement (Refer to blue outlined areas in Figure 2). An area screening form is completed for all sample segments. It inventories all tracts within the segment and contains screening questions that determine whether or not each tract has agricultural activity. In this way, all land inside the segment is screened for agricultural activity and the screening applies to all land in the identified operating arrangement (both inside and outside the segment). Those operations (tracts) that qualify as agricultural are interviewed using the area version questionnaire, which collects detailed agricultural information specifically about the operator's land, again both inside and outside the segment.

The area frame is a theoretically complete sampling frame with every acre of land having a known chance of selection. As such, it can be used to estimate the number of farms and land in farms independently of the list frame, as well as measure incompleteness in the list. The area frame uses a replicated sample design. A sample rotation scheme is used to reduce respondent burden caused by repeated interviewing and to avoid the expense of selecting a completely new area sample each year. Sample rotation is accomplished each year by replacing segments from specified replicates in each land-use stratum with newly selected segments. Approximately 20 percent of the replicates in each land-use stratum are replaced annually.

In addition to the JAS and the yearly list-frame based surveys, NASS conducts a Census of Agriculture every five years (for years ending in 2 and 7). The Census of Agriculture is a complete count of U.S. farms and ranches and the people who operate them. The census collects data on land use and ownership, operator characteristics, production practices, income and expenditures, and many other characteristics. The outcome, when

compared to earlier censuses, helps to measure trends and new developments in the agricultural sector of our nation's economy. The information is used only for statistical purposes and data are published only in tabulated totals. Census forms are sent to all known and potential agricultural operations in the U.S. The census provides the most uniform, comprehensive agricultural data in the nation down to the county level. It employs a dual frame, with the area frame from the JAS serving as a measure of incompleteness for an independent census list frame.

In the past, the JAS has been used to estimate the number of farms not on the census mail list or coverage error, as well as the number of farms misclassified. The evaluation of misclassification has primarily involved either recontacting a sample of census respondents or overlap matching the census mail list to the area frame. For the 1997 and 2002 Censuses of Agriculture, classification errors were measured by comparing an operation's status on the census to its status on the JAS. In cases where there were discrepancies between the two, the JAS was assumed to be correct, and the operations were counted as misclassified on the census (Abreu 2007; Johnson 2000). For 2007, the primary focus of the misclassification evaluation was to identify reasons for discrepancies between the JAS and the census reports. The 2007 Classification Error Survey (CES) was a qualitative examination of why classification and reporting errors occur. The 2007 CES results showed that most of the discrepancies were actual errors which occurred in the JAS, not the census (Abreu et. al, 2007). These results suggested that with misclassification present in the JAS, the number of farm indications resulting from the JAS were biased. However, the results of the 2007 CES were based on a small sample of 67 respondents and thus further research was needed to fully understand the extent of misclassification on the JAS.

3. Methodology

The census list frame is created independently from the JAS area frame and as a result can be used to assess misclassification in the JAS. The approach taken here is to link or match each 2007 JAS tract to its 2007 Census of Agriculture report. For this endeavor, the focus is on records for which the JAS and census farm status disagreed. Each JAS agricultural tract was identified as a farm or non-farm in June based on whether it had \$1,000 in sales of agricultural products or 1,000 points based on the potential for agricultural products produced (if sales were less than \$1000). All non-agricultural tracts were considered non-farms. Disagreement in farm status between the JAS and census occurs when (1) tracts identified as non-farms in the JAS are subsequently identified as farms in the census or (2) tracts identified as farms in the JAS are subsequently identified as non-farms in the census. In this framework, it is assumed that a tract that is identified as a farm in either the JAS or the census is a farm.

2007 JAS and 2007 Census reports were matched in one of the following three ways: (1) through their response on the Not-on-Mail-List (NML) process, (2) using matched IDs which are obtained via NASS's yearly overlap/non-overlap process, or (3) matching via record linkage. The farm/non-farm status of the matched records was compared and farm status disagreement identified. The logic used to identify misclassification may best be followed through the diagram in Attachment A. Although misclassification of both types (1) and (2) are represented in Attachment A's diagram, only the tracts identified as non-farms in the JAS are discussed here.

All JAS tracts were first split into those that were identified as farms and those that were identified as non-farms in June. The left side of the diagram shows all tracts identified as non-farms by the JAS. These non-farms consist of both agricultural and non-agricultural tracts.

A portion of the JAS non-farms were part of the 2007 Census of Agriculture Not-on-Mail List (NML) domain. These tracts (mostly agricultural) were identified as missing from the census mail list during list building efforts for the 2007 Census. At that time, the NML domain tracts were mailed a census of agriculture questionnaire. The results of the questionnaire were used to determine their final farm/non-farm status on the census and subsequently provided a coverage adjustment estimate for the 2007 Census. Since the NML indicator and census farm/non-farm status were available for a portion of the JAS non-farms, it was possible to use this information to identify farm status disagreement between the JAS and the census reports.

Another portion of the JAS non-farm tracts that were not in the NML domain were linked to their census reports via the yearly overlap/non-overlap process. JAS agricultural tracts are overlapped to the list frame yearly to measure list incompleteness. JAS identification numbers are stored for each list frame record overlapped. It is important to note that not all JAS agricultural tracts can be matched to census records because there are some that are simply not on the census list, which is why the area frame is used to evaluate list undercoverage. For JAS non-farms tracts that had matching identification numbers available, however, farm status disagreement was determined between them and their census report.

Recall that all non-agricultural tracts are considered non-farms in the JAS. Furthermore, non-agricultural tracts are classified into the following three types: potential for agriculture unknown; having potential for agriculture; and not having potential for agriculture. Non-agricultural tracts with unknown potential and those having potential for agriculture are often overlapped to the list frame. However, non-agricultural tracts without potential for agriculture are not considered for this yearly overlap process.

The remaining non-farm tracts (also the majority) that did not have their farm status resolved via the NML domain or the yearly overlap process were non-agricultural tracts. Probabilistic record linkage was used to match all three types of non-agricultural tracts to the 2007 Census Mail List (CML) records. Non-agricultural tracts were matched for the 48 contiguous states to identify those believed to correspond to a CML record. One major issue with non-agricultural tracts was that 30 percent of them contained poor quality name and address information. As such, they were removed from further matching. Records were brought together into link groups, each of which possibly represented the same operation. Routinely, link groups are classified into one of three distinct types: definite match, possible match or non-match (Broadbent et. al., 1999). Possible matches are identified for field office (FO) staff to review. However, for this study, no FO review was conducted in the interest of saving time and resources. Instead, only two distinct types of matches were identified: match and non-match. Eliminating the FO review from the linkage process led to a more conservative approach in identification of matches and non-matches. That is, to maximize the quality of the final results, all possible matches were treated as non-matches. Consequently, possible matches representing farms were missed resulting in potentially fewer farms being identified.

4. Results

The 2007 JAS was comprised of 107,990 tracts. Of these, 54.8 percent were identified as non-farm tracts that were either non-agricultural tracts or agricultural tracts that did not have \$1,000 in sales or potential sales of agricultural products produced. Table 1 presents a breakdown of the non-farm tracts by type of agricultural tract. The results discussed in this report will pertain to the JAS non-farm tracts since these are not properly accounted for in the number of farms estimates published annually by the JAS (refer to the left side of the diagram presented in Attachment A). In other words, these non-farm tracts are misclassified on the JAS and not included in the number of farms estimates published annually.

Table 1: JAS Non-farm Tracts by Type of Agricultural Tract

Tract Type	Total	Percent
Agricultural tract	5,370	9.1
Non-agricultural tract with potential	1,813	3.1
Non-agricultural tract with unknown potential	1,079	1.8
Non-agricultural tract without potential	50,961	86.0
Total	59,223	100.00

JAS non-farm tracts were linked to their census reports in one of three possible ways: (1) through their response on the NML process, (2) using matched IDs identified during the yearly overlap/non-overlap process, or (3) matching via record linkage. Table 2, shows the results of the matching or linking based on the source used to identify the JAS non-farm tract's census report. The evaluation of all JAS non-farm tracts resulted in 11,295 matches to the census farm and non-farm reports. The results showed that 72.8 percent of the JAS non-farm tracts were correctly classified as non-farms on the JAS. The remaining 27.2 percent of the JAS non-farm tracts were actually farm tracts and thus were misclassified as non-farms during the JAS.

Table 2: JAS Non-farm Tracts by Source Used to Match to the Census

Matching Source	Matches Census Farm	Matches Census Non-Farm	Total
NML Procedure 1	748	5,477	6,225
Overlap Process	342	1,318	1,660
Record Linkage	1,978	1,432	3,410
Total	3,068	8,227	11,295

The issue of misclassifying a tract as non-farm when in fact it represents a farm is a problem of substantial importance in the JAS because it indicates that farms were missed

¹ There were 13,066 JAS tracts in the NML domain for the 2007 Census of Agriculture coverage evaluation. Table 1 just presents those tracts that were identified as non-farms in JAS.

in the JAS and not properly accounted for in the number of farms estimates. Table 3 shows the results of the JAS misclassified non-farm tracts by the source used to match to the census. The results show that a substantial number of farms were misclassified. The majority of the farms misclassified were identified through use of record linkage techniques. The NML procedure also identified a reasonable number of farms that were misclassified. Even though the overlap process did not identify as many farms as the other two sources, it still identified a generous number of farms.

Table 3: JAS Non-farm Tracts by Matching Source

Matching Source	JAS Non-Farms that Matched Census Farms	Expanded Number of Farms
NML Procedure	748	62,810
Overlap Process	342	31,225
Record Linkage	1,978	84,611
Total	3,068	178,646

The research has identified that a sizeable number of farms are misclassified on the JAS. Thus, it is important to determine how the types of tracts are distributed among the expanded farms missed. Table 4 presents the results of the misclassified JAS non-farm tracts by type of agricultural tract and source used for matching. The results revealed that the majority of the misclassified non-farm tracts were pre-screened as non-agricultural. This suggests that these tracts were not screened correctly during the screening procedures of the JAS, raising major concerns. Even more importantly is the fact that the vast majority (over 90%) of the non-agricultural tracts misclassified were screened as having *no potential* for agriculture. From Table 4, it is evident that non-agricultural tracts without potential are not considered for any yearly processing or even the NML process. However, they contributed substantially to the number of farms misclassified. These results, and the fact that over 30 percent of all non-agricultural tracts contained poor quality name and address information and were precluded from the matching process indicate that potentially more farms are missed by the JAS than is being reported here.

Table 4: JAS Non-farm Tracts by Type of Tract and Matching Source

Tract Type	NML procedure	Overlap Process	Record Linkage	Total Non-farm Tracts Misclassified	Expanded Number of Farms Misclassified
Agricultural tract	505	342	n/a	847	76,240
Non-agricultural tract with potential	167	n/a	71	238	16,539
Non-agricultural tract with unknown potential	76	n/a	40	116	7,749
Non-agricultural tract without potential	n/a	n/a	1,867	1,867	78,118
Total	748	342	1,978	3,068	178,646

Characteristics of Non-agricultural Tracts without Potential that are Misclassified on the JAS

The results of this research have indicated that non-agricultural tracts without potential need to be studied carefully and in more detail. These non-agricultural tracts are not being screened correctly during the JAS. An evaluation of their characteristics revealed that most of these operations were primarily family-owned or individual operations (83.4 percent), and, nearly all of them (95.8 percent) were in operation prior to 2002 which means that they are not “new” farms. In all previous studies of misclassification on the census, part-time operators were those most often missed or misclassified. However, based on this study, that was not the case for the operations misclassified on the JAS. Almost 65 percent of the operations misclassified were full-time operations. The characteristics of these operations did coincide with those of the census in terms types of farm and total value of production. For the most part, these were small operations with \$10,000 or less in total value of products sold or produced, and they were mostly cattle & calves and other crops & hay types of operations. Tables displaying all these results are shown in Attachment B.

5. Discussion and Future Research

This work has provided additional quantifiable evidence of the magnitude of misclassification present in the JAS. Unlike the 2007 CES results that were based on a small sample of 67 respondents in five states, this research was conducted nationwide providing stronger validation to the issue at hand. The issue being that misclassifying a tract as non-farm when in fact it represents a farm is a problem of substantial importance because it directly impacts the number of farms published annually by the JAS.

In census years, this research has provided the basis for a methodology by which the number of farms that were misclassified in the JAS can be directly estimated. However, during non-census years, the effect of misclassification needs to be quantified, and the JAS estimate of the number of farms adjusted. Consequently, a framework for modeling the effect of misclassification in the JAS merits further research.

6. Acknowledgements

The authors of this paper are all members of a team brought together under a cooperative agreement between the National Institute of Statistical Sciences and USDA’s National Agricultural Statistics Service.

7. References

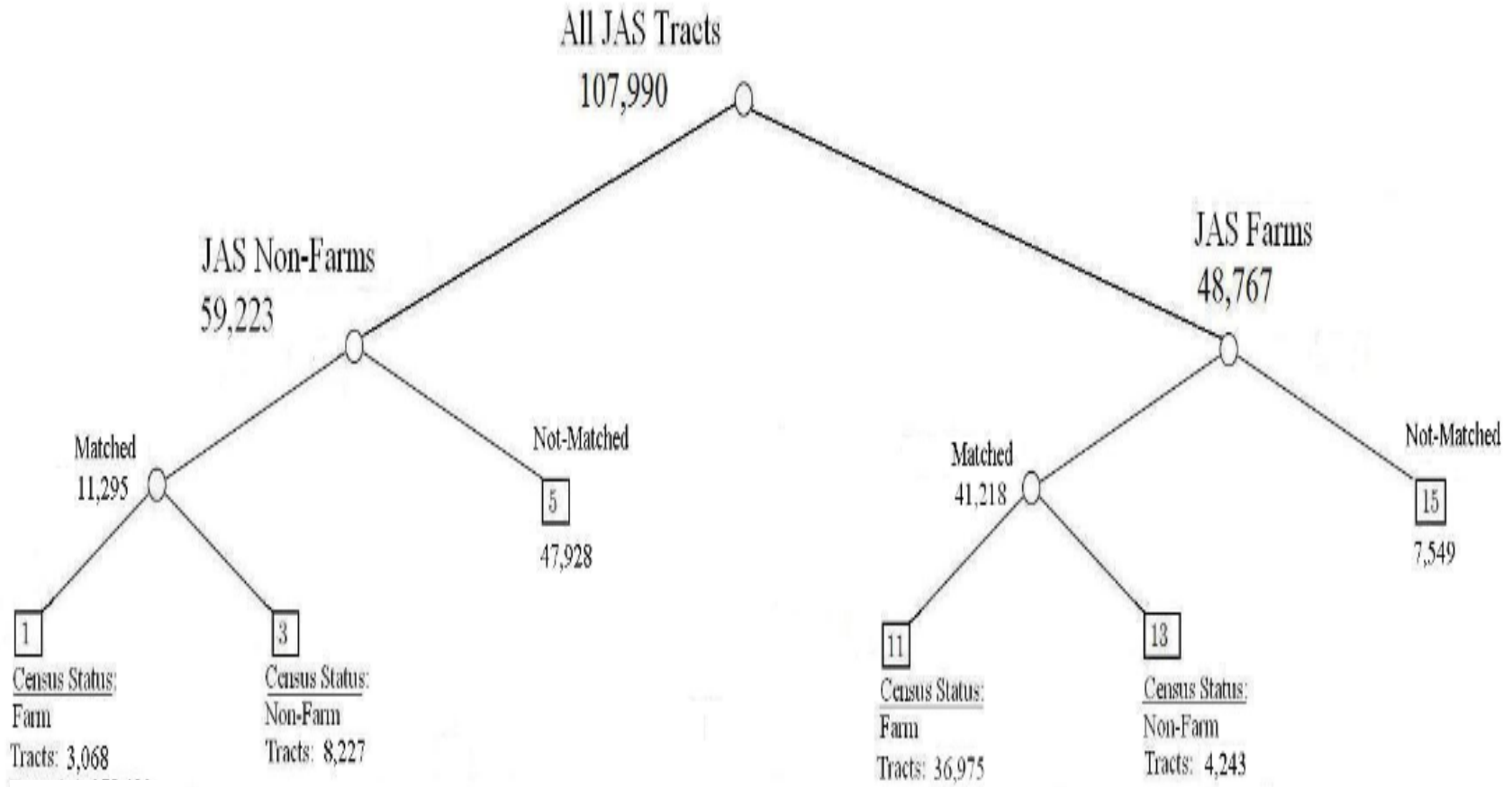
Abreu, D. A., N. Dickey and J. McCarthy (2009). 2007 Classification Error Survey for the United States Census of Agriculture. RDD Research Report # RDD-09-03. Washington, DC:USDA, National Agricultural Statistics Service.

Abreu, D. A. (2007). Results from the 2002 Classification Error Study. Research and Development Division. RDD Research Report #RDD-07-03. Washington, DC:USDA, National Agricultural Statistics Service.

Broadbent, K and Iwig, W. (1999), “Record Linkage at NASS Using Automatch”. *FCSM Research Conference*, <http://www.fcsm.gov/99papers/broadbent.pdf>

Johnson, J.V. (2000). Agricultural Census Classification Error Estimation Using an Area Frame Approach. Data Quality Research Section Unpublished Manuscript. Washington, DC:National Agricultural Statistics Service, USDA.

ATTACHMENT A



ATTACHMENT B

Table B1: JAS Non-agricultural Tracts without Potential Misclassified by Total Value of Agricultural Products

Sales of Agricultural Products	Number of Non-agricultural tracts without Potential	Expanded Number Farms	Percent of Expanded Farms
Less than \$10,000	978	54,543	69.8
\$10,000-\$24,999	282	11,304	14.5
\$25,000-\$49,999	148	5,841	7.5
\$50,000-\$99,999	108	2,521	3.2
\$100,000-\$499,999	205	2,453	3.1
\$500,000 +	146	1,456	1.9
Totals	1,867	78,118	100.0

Table B2: JAS Non-agricultural Tracts without Potential Misclassified by Type of Farm

Type of Farm	Number of Non-agricultural tracts without Potential	Expanded Number Farms	Percent of Expanded Farms
Aquaculture	8	482	0.6
Cattle & Calves	502	18,610	23.8
Christmas Trees, etc.	28	1,587	2.0
Cotton & Cottonseed	13	43	0.1
Fruit, Tree Nuts, etc.	63	3,213	4.1
Grains, Oilseeds, etc.	262	4,856	6.2
Hogs & Pigs	19	849	1.1
Horses, Ponies, etc.	63	4,205	5.4
Milk & Other Prods.	62	835	1.1
Nursery & Greenhouse	27	2,247	2.9
Other Animals & Prods.	62	3,661	4.7
Other Crops & Hay	613	27,711	35.5
Poultry & Eggs	51	2,579	3.3
Sheep, Goats & Prods.	33	1,920	2.5
Tobacco	7	178	0.2
Vegetables, Melons, etc.	29	1,250	1.6
Unknown	25	3,892	4.9
Totals	1,867	78,118	100.0

ATTACHMENT B

Table B3: JAS Non-agricultural Tracts without Potential Misclassified by Year Began Operation

Year Began Operation	Number of Non-agricultural tracts without Potential	Expanded Number Farms	Percent of Expanded Farms
Prior to 2002	1,667	74,803	95.8
After 2002	200	3,315	4.2
Totals	1,867	78,118	100.0

Table B4: JAS Non-agricultural Tracts without Potential Misclassified by Type of Organization

Type of Organization	Number of Non-agricultural tracts without Potential	Expanded Number Farms	Percent of Expanded Farms
Family Owned/Individual	1,456	65,184	83.4
Partnerships	204	4,389	5.6
Incorporated	121	2,806	3.6
Other	66	5,739	7.3
Totals	1,867	78,118	100.0

Table B5: JAS Non-agricultural Tracts without Potential Misclassified by Principal Occupation of Operator

Principal Occupation	Number of Non-agricultural tracts without Potential	Expanded Number Farms	Percent of Farms
Farm or Ranch	814	26,612	34.1
Other (Farming part-time)	1,053	51,506	65.9
Totals	1,867	78,118	100.0