

Is It Feasible to Use a Sampling List Frame to Evaluate Misclassification Errors on an Area-Frame-Survey?

Denise A. Abreu¹, Andrea C. Lamas¹, Hailin Sang², Pam Arroway³,
Kenneth K. Lopiano⁴, Linda J. Young⁴

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

²National Institute of Statistical Sciences, 19 T.W. Alexander Dr., Research Triangle, NC 27709

³Department of Statistics, North Carolina State University, Raleigh, NC 27695

⁴Department of Statistics, University of Florida, Gainesville, FL 32611

Abstract

During the past three years, the National Agricultural Statistics Service (NASS) has made an effort to address, quantify and adjust for an undercount in the number of farms indication on its annual June Area Survey (JAS), which is based on an area frame. This undercount is a direct result of the misclassification of agricultural tracts as non-agricultural. The 2007 Census of Agriculture mailing list (CML) was evaluated as a potential source to assess misclassification on the 2007 JAS. This evaluation revealed that the CML was a rich source from which to quantify the undercount of farms on the JAS. However, the CML is only available every five years and misclassification on the JAS should be assessed each year. Independently of the area frame, NASS maintains a list of agricultural operators, referred to as the list frame. Yearly list-based samples are selected from the list frame. In addition, the list frame serves as the foundation for building the CML. The list frame is updated on an on-going basis and operators are categorized as either active or inactive. This paper discusses the feasibility of using the list frame to assess misclassification on the JAS.

Key Words: Misclassification Errors, Area Frame, List Frame, Record Linkage, Re-screening Survey

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 any government agricultural payments received. An independent estimate of the number of farms is 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 annual estimates based on the JAS number of farms indication 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. In 2007, the 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.

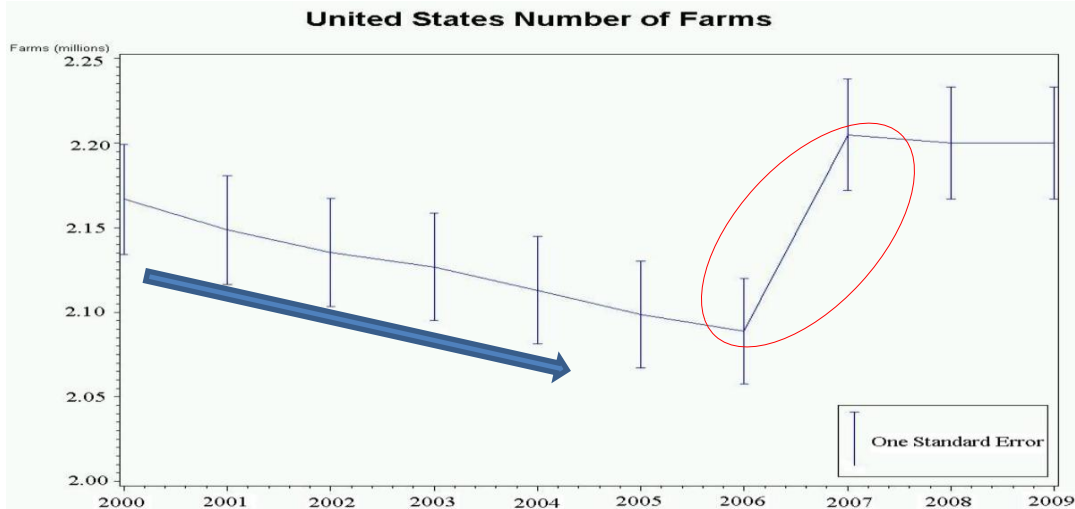


Figure 1: Published estimates of the number of U.S. farms from 2000 to 2009 with one standard error on either side of the estimate.

During previous studies conducted by NASS, misclassification was identified as a source of the underestimation in the JAS (Abreu 2007; Johnson 2000). Misclassification occurs (1) when an operating arrangement with qualifying agricultural activity is incorrectly identified as a non-farm, or (2) when a non-farm arrangement is incorrectly identified as a farm. One study of misclassification (Abreu, Dickey and McCarthy, 2009) revealed that some agricultural operations were incorrectly classified as non-agricultural during JAS pre-screening. These results led to more intensive efforts to understand the source and extent of misclassification in the JAS so that it could be addressed. One effort was the Farm Numbers Research Project (FNRP), based on an intensive post-June survey re-screening in 2009 (Abreu, McCarthy and Colburn, 2010). Concurrently, this undercount issue was also addressed by a team of researchers formed to review the methodology associated with the JAS and to recommend changes, through a collaborative agreement with the National Institute of Statistical Sciences (NISS). This latter team consisted of two NASS researchers, two university faculty members, a post doctoral fellow, and a graduate student. The team considered several measures to address the issue of misclassification on the JAS. Through matching the JAS to the Census of Agriculture list frame, the team evaluated misclassification on the JAS (Abreu, et al. 2010) and then developed appropriate methodology to adjust for misclassification during non-census years (Lamas, et al. 2010). In addition to misclassification, the team identified non-response as another source contributing to the JAS undercount. In Lopiano, et al. (2010), the effect of estimation of agricultural activity for some JAS sampled units is discussed, and methodology for adjusting for both non-response and misclassification is developed. Because the census is only conducted every fifth year, the team further proposed a yearly follow-on survey to the JAS called the Annual Land Utilization Survey (ALUS) (Arroway et al. 2010). ALUS would make the JAS a two-phase sample. In addition to providing information about misclassification of farms and non-farms, it would allow for proper assessment of misclassification and result in an improvement in all JAS indications. However, because ALUS would lead to greater costs associated with the JAS, alternative methods that would not require enumerators to collect further data are attractive. One possibility is to use NASS's annual list frame to assess misclassification in the JAS during non-census years. The team's current effort is focused on evaluating the potential for this approach. The initial results are discussed in this report.

2. Background of the June Area Survey (JAS)

The June Area Survey (JAS) is based on an area frame covering the contiguous U.S., and 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 within each state. Therefore, the precision of the survey indications can be substantially improved by dividing the land within each state into homogeneous groups (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 JAS uses a sample comprised of designated land areas (segments) selected from this stratification. A typical segment is about one square mile (i.e., 640 acres). Each segment is outlined on an aerial photo that is provided to the appropriate field enumerator (See red outlined area in Figure 2).

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 tracts within a segment and contains screening questions that determine whether or not each tract has agricultural activity. The screening applies to all land operated (both inside and outside the segment). Through this screening, any operation (tract) showing any potential for agricultural activities is subsequently interviewed using the area version questionnaire, which collects detailed agricultural information about the operator's land, again both inside and outside the segment.

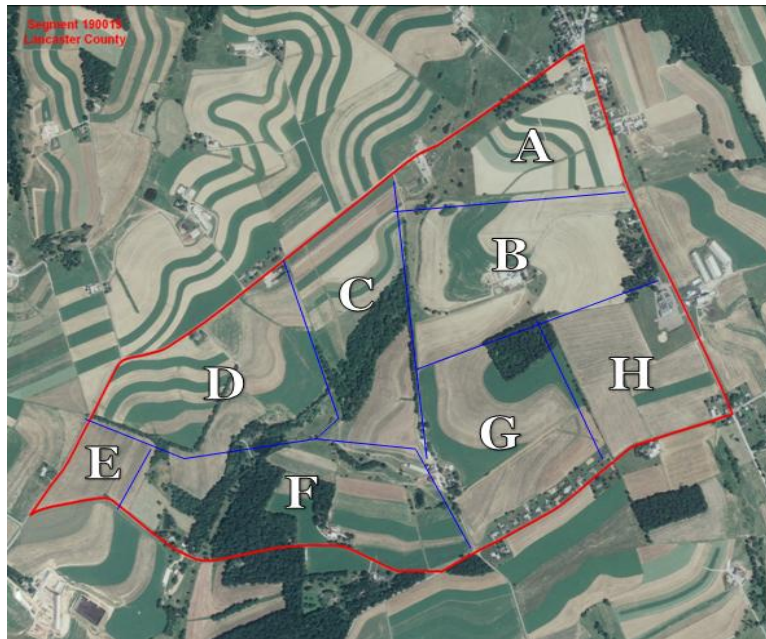


Figure 2: JAS Segment (outlined in red) and Tract Boundaries (outlined in blue)

The area frame is a theoretically complete sampling frame with every acre of land having a known probability of selection. As such, it is used to estimate the number of farms and land in farms independent of the list frame. The area frame also provides a measure of incompleteness in the list. The JAS uses a replicated sample design. The sample rotation scheme reduces respondent burden caused by repeated interviewing and the expense of selecting a completely new area sample each year. Annually, approximately 20 percent of the sampled segments in each land-use stratum are replaced with newly selected segments. Once selected, a segment stays in the sample

for five years. Full descriptions of the JAS design and analysis procedures may be found in Davies (2009).

3. Potential Source to Evaluate Misclassification -- The NASS List Frame

Each year, NASS conducts hundreds of list-based surveys. The agency maintains, on an ongoing basis, a list of farmers and ranchers from which the samples for these list-based surveys are selected. This list frame also serves as the foundation for the development of the Census Mail List (CML). NASS builds and improves the list by regularly obtaining outside source lists. Sources include lists from state and federal government agencies, producer associations, seed growers, pesticide applicators, veterinarians, marketing associations, and a variety of other agricultural sources. NASS also obtains special commodity lists to address specific list deficiencies. These outside source lists are matched to the NASS list using record linkage programs. Most names on newly acquired lists are already on the NASS list. Records not on the NASS list are treated as potential farms until NASS can confirm their existence as a qualifying farm. Each operation on the list frame is categorized as active or inactive. Active list records are assumed to have a high probability of representing active farming operations. Inactive list records may be associated with landlords, deceased operators, farms no longer in business, etc. Many of the active records represent agricultural establishments that operate land but do not have sufficient production to be classified as a farm in a specific year. However, they are maintained on the list frame as active records to help ensure high coverage of farms for the Census of Agriculture every five years. There are also pure active status inaccuracies that exist on the list frame. That is, there are records identified as "active" that are out-of-business or no longer have the potential to operate any agricultural land or facilities.

The question being considered here is whether the NASS list frame can be used to assess misclassification in the JAS in non-census years. After the 2007 Census of Agriculture, the farm/non-farm status of 2007 list frame records was evaluated. Seventy-two percent of the active list frame records matched to farms on the census. The remaining 28 percent were found to be non-farms,⁵ indicating that the census list frame contained active records that were not associated with farming operations (farm status inaccuracies). If these list frame farm status inaccuracies are not considered when adjusting for misclassification on the JAS, the adjustment for misclassification will be larger than it should be. Thus, for the list frame to be useful in assessing misclassification in the JAS, a method of properly accounting for the list frame farm status inaccuracies must be developed.

4. The Farm Numbers Research Project (FNRP) Evaluated Misclassification on the JAS

In 2009, due to the intercensal farm numbers adjustment required following the 2007 Census, NASS conducted the Farm Numbers Research Project (FNRP) to evaluate misclassification on the JAS. FNRP was a one-time follow-on survey to the 2009 JAS, which sampled all newly rotated-in segments (Abreu, McCarthy and Colburn, 2010). Recall the sample design of the JAS includes 20 percent new segments that rotate in each year, with the 20 percent oldest segments rotating out. Thus, the 2009 JAS sample contained segments from years 2009, 2008, 2007, 2006 and 2005, with the FNRP targeting only the newly rotated 2009 JAS segments. For the FNRP, all tracts in the 2009 JAS segments that were classified as non-agricultural or manually estimated by an enumerator were revisited.

⁵ Internal analysis conducted by Thomas Jacob of the Information Management Group.

⁶ A \$5M row is not present for this table because there were not any list frame records with sales exceeding \$5M that matched to a FNRP record. The bars displayed on the table represent the percent contribution of

The results of the FNRP indicated a significant number of JAS non-agricultural tracts were misclassified during the operational screening efforts. It cost nearly \$1M to conduct FNRP and because of this, the Agency is unable to continue conducting the re-screening effort on an annual basis. This is the primary reason for using the list frame to evaluate misclassification on the JAS, as it represents an attractive, inexpensive alternative. In addition, within our present framework, FNRP information could subsequently be used to verify the accuracy of the farm/non-farm status of the 2009 list frame records. That is, FNRP would provide the “gold standard” on farm status for 2009 list frame records associated with the 20 percent of the 2009 JAS segments previously investigated.

5. Matching 2009 JAS to the 2009 List Frame

Probabilistic record linkage was used to match all 2009 JAS agricultural and non-agricultural tracts to the 2009 list frame records in 42 states. The analysis excluded the New England states because those files were not available at the time the match was processed. The JAS is only conducted in Hawaii during census years, and Alaska does not have an area frame. 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 typically reviewed by Field Office (FO) staff. However, in the interest of saving time and resources, the FO review was not conducted. 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 the identification of matches. That is, to maximize the quality of the final results, all possible matches were treated as non-matches. Consequently, some true matches went unmatched.

From the 83,203 original 2009 JAS tracts, 92,152 names and addresses were identified. These were prepared and standardized for matching to the list frame. For this linkage, all agricultural and non-agricultural tracts were considered. From the 2009 list frame, 4,683,345 names and addresses were prepared and standardized for matching to the 2009 JAS. This list included both active and inactive records. More information on this process is provided by Abreu, et al. 2011.

6. Results

The results of the linkage yielded 35,254 matches. Table 1 shows the breakdown of the matched tracts by type of agricultural tract as identified in the JAS. The vast majority of the matches were to agricultural tracts (86.8 percent). This is not surprising because the list frame is targeted for agricultural operations, and agricultural tracts have the most complete name and address information. During JAS screening procedures, non-agricultural tracts are classified into the following three types: having potential for agriculture, potential for agriculture unknown, and not having potential for agriculture. Non-agricultural tracts without potential comprised slightly over 11 percent of all the matches.

Table 1. Matched JAS and List Frame Records by Type of Agriculture as Identified by the JAS

Type of Agricultural Tract	Number Tracts Matched	Percent
Agricultural tracts	30,587	86.8
Non-agricultural tracts w/ potential	476	1.3
Non-agricultural tracts w/ unknown potential	201	0.6
Non-agricultural tracts w/out potential	3,990	11.3
Totals	35,254	100

7. Comparing the List Frame’s Farm/Non-Farm Status to the JAS Farm/Non-farm Status

Each JAS agricultural tract is identified as a farm or non-farm in June based on whether it had at least \$1,000 in sales of agricultural products or at least 1,000 points based on their potential for agricultural products produced. This computation includes any government agricultural payments received. If neither condition is met, the agricultural tracts are considered non-agricultural and thus non-farms.

Identifying farms on the list frame is important because list frame records lack a farm/non-farm status. Subject-matter experts from the List Frame Section recommended assigning farm status based on the type of active business status and total value of sales. This approach was adopted and led to the following algorithm for assigning farm status to list frame records:

- 1) For active records, an operation was identified as a farm if sales met or exceeded \$1000; otherwise it was taken to be a non-farm.
- 2) All inactive records were identified as non-farms based on enumerator information, regardless of any sales value on the records.
- 3) Operations participating in any Conservation Reserve Program were considered farms.
- 4) Operations headquartered out of state were considered farms.
- 5) Operations that have a major name changes were considered non-farms.

Farm status was assigned to each JAS tract that matched to a list record. In Table 2, the number and percent of list frame farms and non-farms by their agricultural/non-agricultural status on the JAS are shown. Using the farm/non-farm algorithm, 12 percent of all matching list frame records were non-farms and 87.9 percent were farms.

Table 2. Farm Status Assignment for 2009 List Frame Records

Type of JAS Tract	List Non-Farm	Percent	List Farm	Percent	Total
Agricultural tracts	1,911	6.2	28,626	93.8	30,587
Non-agricultural tracts w/ potential	195	41.0	281	59.0	476
Non-agricultural tracts w/ unknown potential	73	36.3	128	63.7	201
Non-agricultural tracts w/out potential	2,108	52.8	1,882	47.2	3,990
Totals	4,287	12.1	30,967	87.9	35,254

The farm/non-farm status of matched records for both the list frame and the JAS is shown in Table 3 below. The cells highlighted in green, represent the areas where the list frame and the JAS farm status agree. Based on these results, there is an 87.9% agreement rate, in which both sources agree in their determination of farms and non-farms. The cells highlighted in red represent the area of disagreement between these two sources. Overall, there was a 12.1 percent disagreement rate between the list frame and the JAS. Of particular importance are the 2,694 records identified as farms on list frame and non-farms on the JAS. These will be the JAS misclassified farms and are not accounted for in the annual estimate of the number of farms. However, as noted earlier, about 28 percent of the 2007 list frame records identified as active were not farms in the 2007 Census of Agriculture⁵. Thus, the farm/non-farm status of the 2009 list frame records also had to have some misclassification, which is not clearly identified here. If

the list frame is to be used to assess misclassification in the JAS, then being able to identify the list frame farm status inaccuracies is important

Table 3. Farm/Non-farm Status of Matched List Frame and JAS Records

Farm/Non-farm Status	List Non-Farm	List Farm	Total
JAS Non-Farm	2,729 (7.7%)	2,694 (7.6%)	5,423
JAS Farm	1,558 (4.5%)	28,273 (80.2%)	29,831
Total	4,287	30,967	35,254

8. Using FNRP as a Validation Source of the List Frame Misclassification

The results of the FNRP, discussed earlier, are used here to help evaluate the list frame farm/non-farm status. NASS procedures define a tract as a unique land operating arrangement. However, for densely populated tracts, multiple operations (places of interest) may have been erroneously included for any particular tract during the JAS survey enumeration. For the FNRP, the concept of subtracts was introduced to address tracts that had multiple places of interest. For a selected tract, all places of interest were considered subtracts. For enumeration purposes, if eight or more subtracts were present within a tract, these subtracts were sub-sampled at pre-determined rates. The FNRP sample consisted of 10,204 JAS tracts, which resulted in a total of 17,191 subtracts. Of these, only 1,941 FNRP subtracts (5.5 percent) matched to the list frame, the remaining 94.5 percent were not in FNRP (Table 4). This is primarily because the FNRP only constituted 20 percent of the 2009 JAS.

Table 4. Distribution FNRP for Matched List Frame and JAS Records

	JAS/List Frame Matched Records	Percent
In FNRP	1,941	5.5
Not in FNRP	33,313	94.5
Total	35,254	100

Of the 1,941 matching records, 483 were list non-farms and 1,458 list farms (Table 5). These were further separated into their corresponding JAS and FNRP farm/non-farm status. As mentioned earlier, the FNRP results are the gold standard and assumed to be accurate. The areas highlighted in green represent where all three sources agree in their farm/non-farm status, and represent 69.8 percent of all the matches. However, for purposes of addressing misclassification on the JAS, we are primarily interested in the areas where the FNRP and the list frame disagree in their farm/non-farm classification (areas highlighted in red), 19.6 percent of all the matches.

Table 5. A Comparison of Farm/Non-Farm Status on 2009 JAS, 2009 List Frame, and 2009 FNRP

		List non-farm	List farm	Total
JAS non-farm	FNRP non-farm	356	246	602
	FNRP farm	61	188	249
JAS farm	FNRP non-farm	18	25	43
	FNRP farm	48	999	1,047
Total		483	1,458	1,941

Table 6 (below), focuses on evaluating the list frame farm/non-farm status of JAS non-farms. When not properly accounted for, these will be missed and not included in the number of farms estimate published annually by the JAS. Furthermore, misclassification on the JAS is intrinsically focused around JAS non-farms.

Table 6. Evaluation of JAS Non-farms: Comparison of Farm/Non-Farm Status Between the 2009 List Frame and FNRP

		List non-farm	List farm	Total
JAS non-farm	FNRP non-farm	356 (85.4%)	246 (56.7%)	602
	FNRP farm	61 (14.6%)	188 (43.3%)	249
Total		417 (100.0%)	434 (100.0%)	851

From the table, there were 851 JAS non-farms. Of these, 417 were classified as non-farms by the list frame and 434 were classified as farms. From those identified as non-farms by the list frame (417), FNRP indicates that 85.4 percent were correctly classified as non-farms, while 14.6 percent were misclassified as non-farms by the list frame. Evaluating the list frame farm status, FNRP indicates that the list frame correctly classified 43.3 percent of these as farms, and misclassified 56.7 percent as farms. From these results, it is clear that the list frame tends to over-classify non-farms as farms.

When evaluating the characteristics of the 188 operations that were JAS non-farms, but identified as farms on both FNRP and the list frame, it is clear that these are correctly classified on the list frame. Nearly all of them came from JAS tracts identified as non-agricultural without potential. Because value of sales was available for both FNRP and list frame farms, agreement of farming operations from both sources was evaluated. The highlighted cells in Table 7 below represent the agreement between list sales class and FNRP sales class. The two sources only agree about a third of the time. Most of the time, the list frame value of sales is higher than that reported in FNRP, indicating that the list frame is over-estimating sales and thus categorizing operations as having more sales than was reported in FNRP. It is important to note that the current list frame procedure dictates that the largest value of sales is assigned when different sales values are reported across surveys. Thus, it is not surprising that it overstates sales when compared to FNRP. It does point out that value of sales alone should not be used to determine farm/non-farm status on the list frame.

Table 7. A Comparison of Sales Class Values for Matched FNRP and List Frame Records with Highlighted Cells Indicating Agreement Between the Two Sources.⁶

List Sales Class	FNRP Sales Class												Total
	\$1-\$999	\$1,000-\$2,499	\$2,500-\$4,999	\$5,000-\$9,999	\$10,000-\$24,999	\$25,000-\$49,999	\$50,000-\$99,999	\$100,000-\$249,999	\$250,000-\$499,999	\$500,000-\$999,999	\$1M-\$2.5M	\$5M+	
\$1-\$999	0	1	0	0	0	0	0	0	0	0	0	0	1
\$1,000-\$2,499	0	20	2	2	1	1	0	0	0	0	0	0	26
\$2,500-\$4,999	0	10	3	3	4	0	1	0	0	0	0	0	21
\$5,000-\$9,999	0	5	3	6	4	1	1	0	1	0	0	0	21
\$10,000-\$24,999	0	10	8	6	10	1	1	0	0	1	0	0	37
\$25,000-\$49,999	0	3	7	7	8	5	3	1	0	0	0	0	34
\$50,000-\$99,999	0	2	1	2	3	3	3	1	0	0	0	0	15
\$100,000-\$249,999	0	1	1	1	2	1	2	8	1	0	0	0	17
\$250,000-\$499,999	0	1	0	0	1	0	0	1	2	0	0	0	5
\$500,000-\$999,999	0	0	0	0	0	1	0	1	1	1	2	1	7
\$1M-\$2.5M	0	0	0	0	0	0	1	0	0	1	2	0	4
Total	0	53	25	27	33	13	12	12	5	3	4	1	188

9. Conclusions and Discussion

The overall results of this research indicate that there is potential in using the list frame to evaluate misclassification on the JAS. In addition, the list frame is a more viable option due to its cost efficiencies. Using the FNRP as the “gold standard” to accurately identify list frame farm status inaccuracies was also informative. It is clear from comparisons with FNRP, that the list frame has a tendency to over-classify non-farms as farms and also to estimate value of sales higher than actual sales due to current list frame protocol. As such, value of sales alone should not be a determining factor in the farm/non-farm status of records on the list frame. These farm status inaccuracies are an issue that needs to be addressed if the list frame is to be used to adjust for misclassification on the JAS, as the JAS number of farms indication could be upwardly biased. The potential for using the list frame for misclassification adjustment of the number of farms indication from the JAS merits further research; including whether the list frame farm status inaccuracies can be reliably identified and accounted for in the adjustment.

10. References

Abreu, Denise A., Andrea C. Lamas, Hailin Sang, Kenneth K. Lopiano, Pam Arroway, and Linda J. Young (2011). On the Feasibility of Using NASS’s Sampling List Frame to Evaluate Misclassification Errors of the June Area Survey. Research and Development Division. RDD Research Report #RDD-11-01. Washington, DC: USDA, National Agricultural Statistics Service.

Abreu, Denise A., Pam Arroway, Andrea C. Lamas, Kenneth K. Lopiano, and Linda J. Young (2010). Using the Census of Agriculture List Frame to Assess Misclassification in the June Area Survey. Proceedings of the 2010 Joint Statistical Meetings.

Abreu, D. A., J. S. McCarthy, and L. A. Colburn (2010). Impact of the Screening Procedures of the June Area Survey on the Number of Farms Estimates. Research and Development Division. RDD Research Report #RDD-10-03. Washington, DC: USDA, National Agricultural Statistics Service.

⁶ A \$5M row is not present for this table because there were not any list frame records with sales exceeding \$5M that matched to a FNRP record. The bars displayed on the table represent the percent contribution of the cell to the column total.

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.

Arroway, Pam, Denise A. Abreu, Andrea C. Lamas, Kenneth K. Lopiano, and Linda J. Young (2010). An Alternate Approach to Assessing Misclassification in JAS. Proceedings of the 2010 Joint Statistical Meetings.

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

Davies, Carrie (2009). Area Frame Design for Agricultural Surveys. Research and Development Division. RDD Research Report #RDD-09-06. Washington, DC: USDA, National Agricultural Statistics Service.

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.

Lamas, Andrea C., Denise A. Abreu, Pam Arroway, Andrea C. Lamas, Kenneth K. Lopiano, and Linda J. Young (2010). Modeling Misclassification in the June Area Survey. Proceedings of the 2010 Joint Statistical Meetings.

Lopiano, Kenneth K., Denise A. Abreu, Pam Arroway, Andrea C. Lamas, and Linda J. Young (2010). Modeling Misclassification in the June Area Survey. Proceedings of the 2010 Joint Statistical Meetings.

Young, Linda J., Denise A. Abreu, Pam Arroway, Andrea C. Lamas, and Kenneth K. Lopiano (2010). Precise Estimates of the Number of Farms in the United States. Proceedings of the 2010 Joint Statistical Meetings.