Estimating the Number of U.S. Farms with Adjustment for Misclassification

Denise A. Abreu[1], Andrea C. Lamas[1], Shu Wang[2], Linda J. Young[2]

[1]National Agricultural Statistics Service, USDA, 3251 Old Lee Hwy, Fairfax VA 22030
[2] Department of Statistics, University of Florida, Gainesville, FL 32611

Abstract
Five years ago, the National Agricultural Statistics Service (NASS) began a research effort to address an undercount in the estimate of the U.S. number of farms derived from its annual June Area Survey (JAS). Misclassification of farm status was found to be a major cause of the undercount. NASS has evaluated a host of measures and methods to assess, quantify, and account for this misclassification. The approach derived from this process-employs record linkage techniques, logistic regression, and NASS's annual list sampling frame. The methods developed and the subsequent results are presented here.

Key Words: Misclassification Errors, Area Frame, List Frame, Record Linkage, Logistic Regression

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, and the computation includes 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 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 consists of two NASS researchers, two university faculty members, a post doctoral fellow, and a graduate student. The team has 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 (Sang et al. 2011). However, due to resource constraints, the Agency elected not to pursue ALUS at this time.

As a result, a less resource-intensive method was pursued to leverage information contained in the NASS list frame to evaluate JAS misclassification. The challenge with this approach is that the list frame does not have a farm / non-farm status classification. Abreu et al. (2011) have explored the characteristics of the list frame farm status inaccuracies through matching records from the 2009 June Area Survey, the 2009 list frame, and the 2009 Farm Numbers Research Project. They confirmed the presence of farm status inaccuracies on the list frame. Further, the potential for using the list frame for misclassification adjustment of the number of farms indication was found to depend on whether the list frame farm status inaccuracies can be reliably identified and excluded from the adjustment. This paper documents the initial results of using logistic regression methods along with previous Census of Agriculture data to address farm status inaccuracies in the 2011 list frame and providing an adjusted estimate of the number of farms for the 2011 June Area Survey.

2. Estimating the Number of Farms from the June Area Survey

The June Area Survey (JAS) is based on an area-frame 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 in the United States. 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 (the red outlined area in Figure 2).

Through field enumeration, a segment is divided into tracts of land, each representing a unique land operating arrangement (the blue outlined areas in Figure 2). An area screening form that provides an inventory of all tracts within the segment and contains screening questions that determine whether or not each tract has agricultural activity is completed for all sample segments. Using this form, 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 subsequently interviewed using the area version questionnaire, which collects detailed agricultural information about the operator's land, again both inside and outside the segment. Each tract is screened and classified as agricultural or non-agricultural. Non-agricultural tracts belong to one of three categories: (1) non-agricultural with potential, (2) non-agricultural with unknown potential, or (3) non-agricultural with no potential. A tract is considered agricultural if it has qualifying agricultural activity either inside or outside the segment. Otherwise, it is defined as non-agricultural. An agricultural tract will subsequently be classified as a farm if its entire operation (land operated both inside and outside the segment) qualifies with at least $1,000 in agricultural sales or potential sales. All non-agricultural tracts and agricultural tracts with less than $1,000 in sales are classified as non-farms.

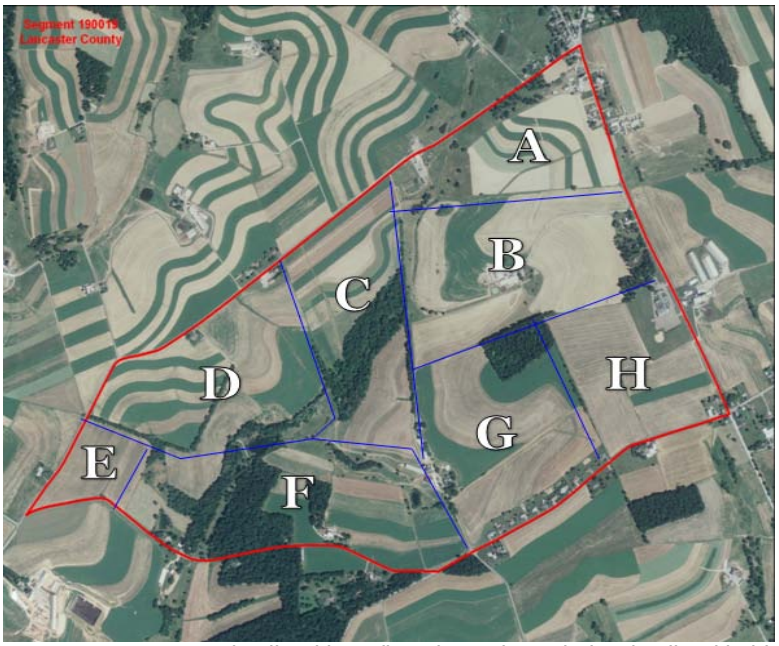

Figure 2: JAS segment (outlined in red) and tract boundaries (outlined in blue)

Because the JAS is a probability-based survey, each tract *i* has an inclusion probability π_i and an expansion factor $e_i = 1/\pi_i$. Within each farm tract, a proportion of a farm is observed (in some cases with smaller farms, the entire operation may reside entirely within the tract). This proportion, the tract-to-farm ratio for tract *i*, is t_i = tract acres / farm acres. Both of these are used in calculating the current JAS estimate for the number of farms (denoted as T), which is defined as follows,

$$T = \sum_{i=1}^{l} \sum_{j=1}^{s_i} \sum_{k=1}^{n_{ij}} e_{ijk} a_{ijk}$$

where

i indexes stratum,
j indexes substratum,
k indexes segment,
l = Number of land-use strata,
s_i = Number of substrata in stratum i,
n_{ij} = Number of segments in substratum j within stratum i,
e_{ijk} = Expansion factor or the inverse of the probability of the selection for each segment in substratum j in land-use stratum i,

$$a_{ijk} = \sum_{m=1}^{x_{ijk}} t_{ijkm}$$

m indexes tract,
x_{ijk} = Number of farm tracts in the given segment,

$$t_{ijkm} = \text{Tract - to - farm ratio of the tract} = \frac{\text{tract acres for the } m^{th} \text{ tract}}{\text{farm acres for the } m^{th} \text{ tract}}$$

The sampling weights are appropriate for the sample design. Therefore, this design-based estimate is unbiased unless misclassification is present.

3. The NASS List Frame

Each year, NASS conducts hundreds of list-based surveys. The agency maintains 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 on an ongoing basis by 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, potential farm (criteria), or inactive. Active list records are assumed to have a high probability of representing active farming operations. Potential farm or criteria records are records whose involvement in agriculture is unknown. 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. Potential farm or criteria records are periodically screened to determine whether or not they are involved in agriculture. Pure active status inaccuracies also exist on the list frame; that is, some records identified as "active" are out-of-business or no longer operate any agricultural land or facilities.

4. Matching 2011 JAS to the 2011 List Frame

Probabilistic record linkage was used to match all 2011 JAS agricultural and non-agricultural tracts to the 2011 list frame records in the 48 conterminous states. 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 a single operation. Subsequently, link groups were classified into one of three distinct types: definite match, possible match or non-match (Broadbent et. al. 1999). Possible matches were sent to Field Office (FO) staff for review and were further classified as matches or non-matches. All non-matches were excluded from further analysis.

The 2011 JAS was comprised of 93,327 tracts with potential names and addresses. These were prepared and standardized for matching to the list frame. For this linkage, all agricultural and non-agricultural tracts were considered. Partner records and records with additional information were also included for each JAS tract to maximize matching results. In addition to the name and address information, existing area-to-area and area-to-list links were used to bring records together. After each June survey, FOs conduct a yearly overlap/non-overlap process in which JAS agricultural tracts are overlapped to the list frame, providing a measure of list incompleteness. JAS identification numbers (IDs) are stored for each list frame record overlapped to the JAS (area-to-list links). In addition, an unduplication of the area frame records is conducted. The ID of any area record matching another area record (area-to-area links) is stored. These identification numbers were used during matching to bring records together that would not have come together solely based on name and address information. Matching the non-agricultural tracts was a departure from standard NASS procedures, primarily because once a non-agricultural tract has been identified as a non-farm it is not accounted for in the number of farms estimate published each year. However, non-agricultural tracts are the primary focus in the identification of misclassification. The procedures for overlapping non-agricultural tracts were slightly different from those for agricultural tracts. Unlike agricultural tracts, many non-agricultural tracts do not have complete name and address information and agricultural data are not collected from them. The procedures for the results of overlap results used for the non-agricultural tracts were fully tested by the Iowa FO staff, prior to dissemination and implementation by all other states.

From the 2011 list frame, 4,678,365 names and addresses were prepared and standardized for matching to the 2011 JAS. This list included active, potential farm (criteria), and inactive records. Every year, certain records are purged from the list frame, usually because they have been inactive for more than five years. The only records excluded were those flagged to be purged from the list frame due to extended inactivity.

For non-agricultural tracts, the overall match rate for all states was 17 percent. After all FO reviews were completed, the results of the linkage yielded 4,264 valid matches to JAS non-farms. These included 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. Each JAS non-farm tract that matches a list frame record can have the modeled list frame probability of being a farm applied to the non-farm tract's appropriate expansion factor. Tracts not matched to a list frame record are assumed to be non-farms and are not used in the adjustment. Table 1 shows the breakdown of the matched tracts by type of agricultural tract as identified in the JAS. Recall that during JAS screening procedures, 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 without potential comprised 70.6 percent of all the matches, while agricultural tracts comprised 18.9 percent.

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

Type of Agricultural Tract	Number Tracts Matched	Percent
Agricultural tracts identified as non-farms	807	18.9
Non-agricultural tracts w/ potential	316	7.4
Non-agricultural tracts w/ unknown potential	132	3.1
Non-agricultural tracts w/out potential	3,009	70.6
Totals	4,264	100.0

Table 2 shows the breakdown of the matched tracts by the type of list frame record. Results show that two-thirds of the matches were to active list frame records, 22.6 percent were matches to inactive records, and only 10.7 percent were matches to criteria records.

Table 2. Matched Non-farm JAS Tracts and List Frame Records by Type of List Frame Record

Type of List Frame Record	Number Tracts Matched	Percent
Active	2,845	66.7
Criteria	456	10.7
Inactive	963	22.6
Totals	4,264	100.0

5. Accounting for Farm Status Inaccuracies on the 2011 List Frame

Modeling the Probability that a List Frame Record is Active

Using the record linkage results, the probability that a list frame record is active given that it matched a JAS non-farm can be modeled using logistic regression. In this section, the development of the logistic model is discussed.

For the logistic regression, the 4,264 JAS non-farm tracts that matched a list frame record are considered. JAS data, list frame data and federal tax information were used to identify a set of explanatory variables to be used in the model. The variables used were the following:

- Type of IRS form (1040, 1121, 943)
- Most recent year of tax form filed
- A flag indicating whether or not the record was a farm in the 2007 Census (non-purge CML flag)
- Marked flag (whether the record was marked for any major NASS survey)
- gross receipts reported on the IRS form
- Year record last went through record linkage process
- Flag indicating records newly added to the list frame since the 2007 Census
- Flag indicating records coming from IRS that were not on the list frame)
- Type of list record came from
- Race
- Gender
- Spanish origin
- US region
- Agricultural/Non-agricultural Status
- Farm size

For some categorical explanatory variables, grouping of categories was used to keep the data from becoming too sparse, which would result in unstable estimated probabilities (*see* Table 3 for groups formed).

Table 3. Grouping of Categories Used for Variables in Matched Data Set

Variable	Groupings
Agricultural tract Type Indicator	Agricultural tracts not grouped Non-agricultural tracts (regardless of potential) grouped
Record Source ID (code descriptors are provided in Attachment A)	643,77,616,102,677,89,42,23,610,21,30,76,9,645,950,82,40, 67,32,44,650, 83,18,37,100,103,655,944 == Group 1 46,631,91,621,928 == Group 2 43 == Group 3 33 == Group 4 1,641 == Group 5 97 == Group 6 607,90,78,453,448,628,34,322,96,65,2,10,615,638,315,49,624,50, 35,929,447,959,31,958,320 == Group 7
Farm Size	Acres <= 0.5 == not grouped 0.5 < Acres <= 0.9 == not grouped 0.9 < Acres <=24.9 == not grouped 24.9 < Acres <= 99.9 == not grouped Acres >99.9 == not grouped
US Region	Region 1 == State FIPS in (CT, IL, IN, IA, KS, ME, MA, MI, NE, NH, NJ, NY, OH, PA, RI, VT, WI) Region 2 == State FIPS in (AL, DE, GA, KY, MD, NC, SC, TN, VA, WV) Region 3 == State FIPS in (AR, FL, LA, MS, MO, NM, OK, TX) Region 4 == State FIPS in (CO, MN, MT, NV, ND, SD, UT, WY) Region 5 == State FIPS in (AK, AZ, CA, HI, ID, OR, WA)
Gross Receipts	Less than $1 $1 to $999 == not grouped $1,000 to 2,499 == not grouped $2,500 to $4,999 == not grouped $5,000 to $9,999 == not grouped $10,000 to $24,999 == not grouped $25,000 to $49,999 == not grouped $50,000 to $99,999 == not grouped $100,000 to $249,999 == not grouped $250,000 to $499,999 with over $1,000,000 $500,000 to $999,999 with missing gross receipts
Type of IRS Form	Form 1040 == not grouped Form 1121== not grouped Form 943 == not grouped No form filled == not grouped

List frame records were divided into three categories (1) active records, (2) inactive records, and (3) potential farms (criteria) records. Using JAS weights, stepwise regression was conducted with the explanatory variables in the final model displayed in Table 4. The analysis was only conducted using only active and inactive records (criteria records were excluded from the initial process) and the modeled probability was assigned to each record. Standard diagnostics were used to evaluate the fit of the model. The Hosmer-Lemeshow goodness-of-fit test statistic was

13.1669, with 8 degrees of freedom. The estimated probability from leave-one-out cross-validation was determined for each record (when it was the one left out) and plotted against the estimated probability based on the model fitted using all records (*see* Figure 3). Nine estimates changed by more than 0.05. The records associated with the estimates changing the most were not from a discernible group. Using the final model, the probability that each criteria record is actually an active one given that it matched a JAS non-farm is estimated.

Table 4. Results for Final Logistic Model of the Probability of a List Frame Record Being Active Given that It Matched a JAS Non-farm

Effect	Degrees of Freedom	Wald Chi-square
Type of agricultural tract	1	11,467.9634
Record Source Id	6	7,307.8663
Gross Receipts	10	2,938,5879
US Region	4	1,269.5771
Farm Size	3	13,514.8491
Type of IRS Form	4	270.5089

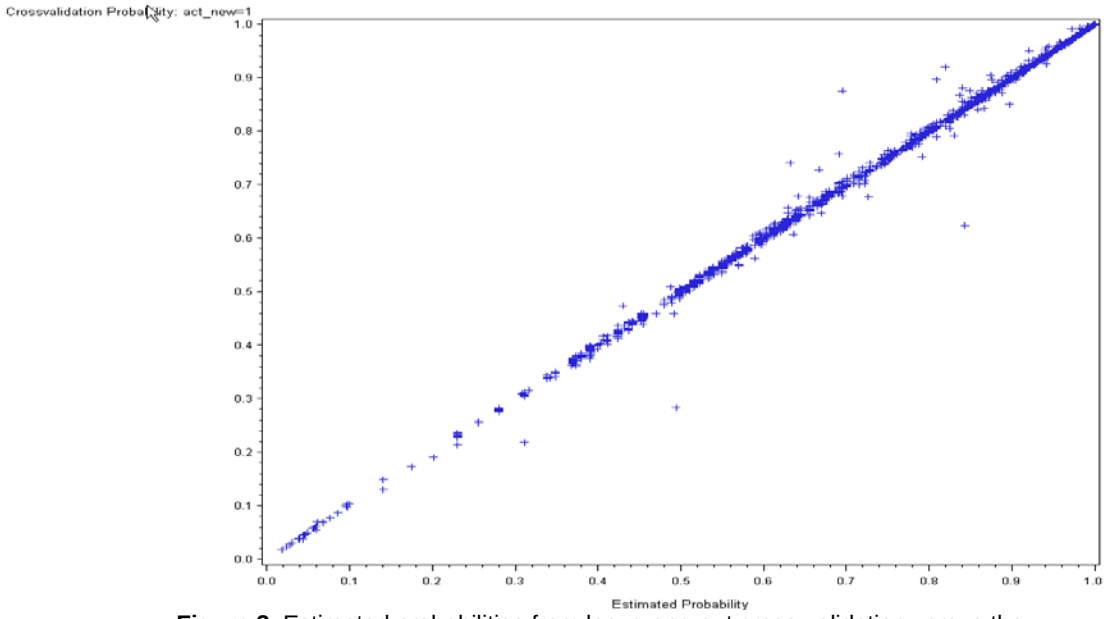

Figure 3. Estimated probabilities from leave-one-out cross-validation versus the estimated probabilities from the full model

It is important to note that the probabilities resulting from the logistic regression modeled the probability that a list frame record is active given that it matched a JAS non-farm tract. This is not equivalent to a list frame record being a farm. Recall that for this endeavor, the goal is to use the list frame in the adjustment of the number of farms estimate derived from the JAS. After the 2007 Census of Agriculture, the farm/non-farm status of the 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 contains active records that are not associated with farming operations (farm status inaccuracies).

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

These farm status inaccuracies exist because farms go in and out of business with some regularity. Also, the majority of the records are only contacted at the time of the census. The presence of these farm status inaccuracies was also confirmed on the 2009 list frame by Abreu et al. (2011). In the 2007 Census of Agriculture, the proportion of active status list frame records that were farms was recorded for each state. Using this information, the modeled probabilities derived from the logistic regression modeling were adjusted to produce an estimate of the probability a record is a farm given that it matched to a JAS non-farm. All inactive records were assigned a farm rate of zero because they are not included in building the CML. Therefore, only active and criteria records were adjusted and included in the adjustment of the 2011 JAS number of farm estimates.

6. Results

From Section 2, the current JAS estimate for the number of farms (denoted as T), is defined as follows,

$$T = \sum_{i=1}^{l} \sum_{j=1}^{s_i} \sum_{k=1}^{n_{ij}} e_{ijk} a_{ijk}$$

This estimate is unbiased unless misclassification is present. However, misclassification has been found to be present on the JAS. Using the modeled probabilities presented in the Section 5, an estimator for the number of farms from the JAS with an adjustment for misclassification (highlighted portion) can be constructed as follows:

$$T_1 = T + \sum_{i=1}^{l} \sum_{j=1}^{s_i} \sum_{k=1}^{n'_{ij}} e_{ij} \sum_{m=1}^{z_{ijk}} t_{ijkm} p_{ijkm}$$

where

n'_{ij} = Number of segments where a list frame record was matched to a non-farm
z_{ijk} = Number of matched non-farm tracts in the given selected segment
p_{ijkm} = Probability that the list frame record is active given that it matched a JAS non-farm tract and adjusted by the 2007 Census farm rates
T = JAS estimator without considering the misclassified non-farms

Under this framework, using the estimated probabilities, the 4,264 JAS non-farm tracts that matched to a list frame record had the probability applied to its appropriate expansion factor leading to a misclassification adjustment of 9.4 percent at the U.S. level. Due to the confidentiality nature of how NASS derives the JAS annual estimate of the number of farms no further results can be presented here.

9. Standard Errors by Bootstrapping

Bootstrapping was used to obtain the standard error of the estimate. In general bootstrap samples are drawn, with replacement, from the empirical distribution of the data. These bootstrapped samples are used to quantify uncertainty. Here two datasets must be considered, and the methods

are similar to that of Kovacevic, *et al.* (2008). Dataset 1 is comprised of the active records and inactive records in the list frame records dataset. This dataset was used to find the best logistic model, which provided the parameter estimates. Dataset 2 consists of the potential farm (criteria) records. The model developed for dataset 1 was applied to each record in dataset 2 to obtain the estimated probability of a criteria record being active status, which was then used in the adjustment of the estimated number of farms.

Suppose there are *B* sets of bootstrap samples from dataset 1 and *D* sets of bootstrap samples from dataset 2. Note that, because the JAS data is a sample with associated weights, the bootstrap weights need to be adjusted when obtaining the parameter estimates based on each bootstrap sample. For each bootstrap sample, $(n_i - 1)$ individuals in each stratum with replacement, where n_i is the size of stratum *i*. Let w_{ij} denote the initial sampling weight of the *j*th individual in the *i*th stratum. For a given bootstrap sample, the bootstrap weights are given by

$$w_{ij}^b = \frac{n_i}{n_i - 1} w_{ij} m_{ij}$$

where m_{ij} represents the number of times the jth individual in the *i*th stratum is selected in the bootstrap sample, for $b = 1,2,\ldots,B$.

From dataset 1, $\hat{\beta}$ is obtained from logistic regression and, from the *B* bootstrap samples, $\hat{\beta}^{(b)}, b = 1,2,\ldots,B$ are computed. Let $\hat{N}(\hat{\beta})$ be the estimated number of farms from dataset 2, and let $\hat{N}^{(d)}(\hat{\beta})$ be the estimate of the number of farms from the *d*th bootstrap sample from dataset 2, $d = 1,2,\ldots,D$. Then, the variance of $\hat{N}(\hat{\beta})$ may be expressed as

$$\mathrm{Var}\left(\hat{N}(\hat{\beta})\right) = \mathrm{E}\left[\mathrm{Var}\left(\hat{N}(\hat{\beta}) \mid \hat{\beta}\right)\right] + \mathrm{Var}\left[\mathrm{E}\left(\hat{N}(\hat{\beta}) \mid \hat{\beta}\right)\right] \quad (1)$$

Consider the first term. First, $\mathrm{Var}\left(\hat{N}(\hat{\beta}) \mid \hat{\beta}\right)$ is estimated and then its expectation determined. $\mathrm{Var}\left(\hat{N}(\hat{\beta}) \mid \hat{\beta}\right)$ can be estimated using

$$\hat{\mathrm{Var}}\left(\hat{N}(\hat{\beta}) \mid \hat{\beta}\right) = \frac{1}{D} \sum_{d=1}^{D} \left[\hat{N}^{(d)}(\hat{\beta}) - \hat{N}(\hat{\beta})\right]^2$$

Then, taking the expectation with respect to $\hat{\beta}$,

$$\hat{\mathrm{E}}\left[\hat{\mathrm{Var}}(\hat{N}(\hat{\beta}) \mid \hat{\beta})\right] = \frac{1}{B} \sum_{b=1}^{B} \frac{1}{D} \sum_{d=1}^{D} \left[\hat{N}^{(d)}(\hat{\beta}^{(b)}) - \hat{N}(\hat{\beta}^{(b)})\right]^2 \quad (2)$$

Now consider the second term in equation (1). Denoting

$$\hat{\mathrm{E}}(\hat{N}(\hat{\beta}) \mid \hat{\beta}) = \frac{1}{D} \sum_{d=1}^{D} \hat{N}^{(d)}(\hat{\beta}) = \hat{\bar{N}}(\hat{\beta})$$

and, similarly, $\hat{\bar{N}}(\hat{\beta}^{(b)}) = \frac{1}{D}\sum_{d=1}^{D}\hat{N}^{(d)}(\hat{\beta}^{(b)})$ so that

$$\hat{\mathrm{Var}}\left(\hat{\mathrm{E}}(\hat{N}(\hat{\beta})\,|\,\hat{\beta})\right) = \hat{\mathrm{Var}}\left(\hat{\bar{N}}(\hat{\beta})\right) = \frac{1}{B}\sum_{b=1}^{B}\left[\hat{\bar{N}}(\hat{\beta}^{(b)}) - \hat{\bar{N}}(\hat{\beta})\right]^2$$

Therefore, the bootstrap estimate of the variance of $\hat{N}(\hat{\beta})$ may be expressed as

$$\hat{\mathrm{Var}}\left(\hat{N}(\hat{\beta})\right) = \frac{1}{B}\sum_{b=1}^{B}\frac{1}{D}\sum_{d=1}^{D}\left[\hat{N}^{(d)}(\hat{\beta}^{(b)}) - \hat{N}(\hat{\beta}^{(b)})\right]^2 + \frac{1}{B}\sum_{b=1}^{B}\left[\hat{\bar{N}}(\hat{\beta}^{(b)}) - \hat{\bar{N}}(\hat{\beta})\right]^2 \quad (3)$$

where $\hat{N}(\hat{\beta}^{(b)})$ is the estimate of the number of farms from dataset 2 using $\hat{\beta}^{(b)}$ and $\hat{N}^{(d)}(\hat{\beta}^{(b)})$ is the estimated number of adjusted farms based on the dth bootstrap sample using $\hat{\beta}^{(b)}$.

Based on 1,000 bootstrap samples, the standard error of the adjustment, which equals the square root of the estimated variance of the adjustment, is 2,725 farms.

10. Conclusions and Future Work

The overall results of this research indicate potential exists for using the list frame to evaluate misclassification on the JAS. In addition, the list frame is a more viable option than the ALUS due to its cost efficiencies. The use of logistic regression modeling provides a solid, reproducible technique to modeling the probability that a list frame record is active given that it matches a JAS non-farm. Adjusting the probability of active status to obtain the probability of farm was based on the proportion of active status records that were found to be farms during the 2007 Census of Agriculture could be improved with further research.

11. References

Abreu, Denise A., Andrea C. Lamas, Hailin Sang, Kenneth K. Lopiano, Pam Arroway, 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.

Abreu, Denise A., Andrea C. Lamas, Hailin Sang, Pam Arroway, Kenneth K. Lopiano, Linda J. Young. (2011). Is It Feasible to Use a Sampling List Frame to Evaluate Misclassification Errors on an Area-Frame-Survey? Proceedings of the 2011 Joint Statistical Meetings.

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.

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.

Garber, Samuel Chad (2009). Census Mail List Trimming using SAS Data Mining. Research and Development Division. RDD Report Number RDD-09-02.

Kovacevic, Milorad S., Lenka mach, and Georgia Roberts. 2008. Bootstrap variance estimation for predicted individual and population-average risks. *Proceedings of the Section on Survey Research Methods, JSM 2008*. Pp. 2289-2296.

Sang, Hailin, Pam Arroway, Kenneth K. Lopiano, Denise A. Abreu, Andrea C. Lamas, and Linda J. Young (2011). Annual Land Utilization Survey (ALUS): Design and Methodology. Research and Development Division. RDD Report Number RDD-11-02.

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, Hailin Sang, Pam Arroway, Kenneth K. Lopiano, and Linda J. Young (2011). Adjusting an Area Frame Estimate for Misclassification Using a List Frame. Proceedings of the 2011 Joint Statistical Meetings.

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., Andrea C. Lamas, Denise A. Abreu, Pam Arroway, and Linda J. Young (2011). Adjusting the June Area Survey Estimate for the Number of U.S. Farms for Misclassification and Non-response. Research and Development Division. RDD Report Number RDD-11-04.

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.

Appendix A -- Record Source Id – List of Source Codes

Recsourceid	Name
1	AGRICULTURAL STATISTICS SERVICE
2	AGRICULTURE COMMODITY COMMISSION
3	ECONOMIC RESEARCH SERVICE (ERS)
4	IMPORT/EXPORT MARKET LIST
7	STATE TAX LIST
9	APHIS
10	INFOUSA LIST
11	AELOS LANDLORDS
18	BLM LEASES
19	FOREST SERVICE PERMIT HOLDERS
20	STATE WATER USE LIST
21	STATE POLLUTION CONTROL LIST
22	STATE LIQUOR CONTROL BOARD
23	STATE DEPT OF REVENUE
24	STATE CROP INSURANCE
27	FOOD & DRUG ADMINISTRATION
28	RURAL DEVELOPMENT
29	EPA
30	COOPERATIVE EXTENSION SERVICE
31	NATURAL RESOURCES CONSERVATION SERVICE (NRCS)
32	LOCAL MUNICIPALITY TAX LIST
33	COUNTY TAX RECORDS LIST
34	COUNTY PLANNING AND ZONING
35	STATE DEPT OF ENVIRONMENTAL QUALITY
36	SOIL EROSION AND WATER QUALITY LIST
37	DEPARTMENT OF AGRICULTURE
38	DEPARTMENT OF HEALTH
39	DEPARTMENT OF NATURAL RESOURCES
40	DEPARTMENT OF LANDS
41	DEPARTMENT OF LABOR STATISTICS
42	FARM BUREAU LIST
43	FARM SERVICE AGENCY (FSA)
44	FARMERS MARKET LIST
45	DIRECT MARKETING ASSOCIATION
46	DIVISION OF MOTOR VEHICLES
47	RISK MANAGEMENT AGENCY (RMA)
48	DEPARTMENT OF WILDLIFE & FISHERIES
49	AMISH LIST
50	ADDITIONAL NAME FROM A SURVEY
51	COMMUNITY SUPPORTED AGRICULTURE
52	FARM ORGANIZATION

Recsourceid	Name
53	PRACTICAL FARMER LIST
54	MARKET NEWS BULLETIN
56	PRAIRIE BOUNTY LIST
65	STATE FARM CENSUS
67	PESTICIDE APPLICATOR'S LIST
68	EMPLOYMENT DEVELOPMENT DIVISION
69	CHEMIST LIST
75	SMALL FARM INITIATIVE
76	STATE BUSINESS LICENSE LIST
77	COGGINS LIST
78	STATE VETERINARIAN LIST
79	SUSTAINABLE AGRICULTURE
80	SMALL LIVESTOCK
82	TRADE MAGAZINES, NEWSLETTERS,WEBSITE & PERIODICALS
83	TRADE SHOW
84	PURCHASED LIST
85	FIRST AMERICAN SPATIAL
89	IRRIGATION LIST
90	BIA IRRIGATORS, GRAZING, LEASES, ETC
91	RESERVATION OPERATORS
94	EXPORT/IMPORT LIVESTOCK LIST
96	CAMA LIST
97	CENSUS MAIL LIST
99	FEDERATION OF SOUTHERN COOPERATIVES
100	MINORITY LIST BUILDING - GENERAL
101	AFRICAN AMERICAN CHURCH PROJECT (RESEARCH DIV.)
102	MAKE SURE I'M COUNTED WEBSITE
103	MINORITY FARM REGISTER
104	NATIONAL BLACK FARMER ASSOCIATION
105	CENSUS DAY
106	FSA MINORITY
130	COMPANY GROWER LIST
303	ALFALFA SEED GROWER'S LIST
305	SUGARBEET GROWER LIST
310	CORN GROWERS ASSOCIATION
312	HOP COMMISSION
313	BARLEY COMMISSION
315	WHEAT GROWERS ASSOCIATION
320	BOLL WEEVIL LIST
322	COTTON PRODUCER'S ASSOCIATION
326	CERTIFIED SEED PRODUCER'S LIST

Recsourceid	Name
334	DRY BEAN DEALER ASSOCIATION
347	GRAIN PRODUCER'S ASSOCIATION
349	HAY GROWER'S ASSOCIATION
356	OILSEED GROWER LIST
369	POTATO GROWER'S ASSOCIATION
370	STATE POTATO LIST
380	SORGHUM PRODUCER LIST
410	NUT GROWER'S ASSOCIATIONS
411	APPLE COMMISSION
412	FRUIT TREE LIST
419	BLUEBERRY GROWERS ASSOCIATION
427	CHERRY ASSOCIATION
434	CRANBERRY COMMISSION
446	FRUIT & VEGETABLE GROWER'S ASSOCIATION
447	STATE VEGETABLE LIST
448	GRAPE GROWER'S ASSOCIATION
452	MACADAMIA NUT GROWER'S ASSOCIATION
453	MAPLE PRODUCER'S LIST
457	MINT GROWER'S ASSOCIATION
458	COFFEE GROWER'S LIST
462	PEAR ASSOCIATION
466	PECAN GROWER'S ASSOCIATION
467	HAZELNUT PRODUCER LIST
470	STRAWBERRY LIST
471	RASPBERRY COMMISSION
480	CHERRY SWEET AND TART GROWER'S LIST
483	TROPICAL FRUIT GROWER'S ASSOCIATION
485	WATERMELON GROWER'S ASSOCIATION
486	WINE PRODUCER'S ASSOCIATION
487	DEPARTMENT OF TREASURY (WINE - FED ALCOHOL TAX)
558	VEGETABLE AND BERRY GROWER'S ASSOCIATION
560	ONION GROWER'S ASSOCIATION
563	PEA & LENTIL ASSOCIATION
570	INTERNET FARM INDEX
571	LAB ANIMALS
572	POLLINATOR LIST
573	SEED RESOURCES
574	SUPPLIERS OF BENEFICIAL INSECTS
575	WOMEN IN SUSTAINABLE AG DIRECTORY
576	WOMEN FARM OPERATORS
600	ASSOC NATURAL BIOCONTROL PROD

Appendix A -- Record Source Id – List of Source Codes

Recsourceid	Name
601	BUTTERFLY BREEDERS
602	HERITAGE BREEDERS CONSERVANCY
603	HONEY SUPPLIERS
604	LIVESTOCK AUCTION/COMMISSION
605	LIVESTOCK BOARD
607	AMS LIST
608	ANGUS ASSOCIATION
609	AMERICAN BREEDER CLUB
610	APIARY LICENSE
611	STATE DAIRY LIST
612	AQUACULTURE GROWER ASSOCIATION
613	STATE AQUACULTURE LIST
615	BEEF PRODUCERS ASSOCIATION
616	BEEKEEPERS ASSOCIATION
617	BISON PRODUCER'S LIST
618	DEER ASSOCIATION LIST
619	GAMEBIRD LICENSE LIST
620	ELK ASSOCIATION LIST
621	BRAND LIST
622	BRUCELLOSIS LIST
623	CATFISH GROWERS ASSOCIATION
624	CATTLEMEN'S ASSOCIATION
628	EMERGENCY FEED ASSISTANCE
630	BRAHMA LIST
631	CATTLE LIST
635	DAIRY ASSOCIATION
636	DAIRY GOAT ASSOCIATION
637	GOAT ASSOCIATION
638	BOER GOAT ASSOCIATION
640	EMU ASSOCIATION
641	EQUINE LIST
642	MULE GROWER'S LIST
643	UNIVERSITY VET EQUINE LIST
645	NATIONAL GOAT LIST
650	HORSEMEN'S ASSOCIATION
651	LIVESTOCK BREEDER'S ASSOCIATION
655	MILK MARKETING ORDER
656	MINK GROWER'S COOPERATIVE
657	OSTRICH GROWER'S LIST
658	OLD ENGLISH GAME BANTAM - ROOSTERS
659	HAMPSHIRE BREEDERS DIRECTORY

Recsourceid	Name
643	UNIVERSITY VET EQUINE LIST
645	NATIONAL GOAT LIST
650	HORSEMEN'S ASSOCIATION
651	LIVESTOCK BREEDER'S ASSOCIATION
655	MILK MARKETING ORDER
656	MINK GROWER'S COOPERATIVE
657	OSTRICH GROWER'S LIST
658	OLD ENGLISH GAME BANTAM - ROOSTERS
659	HAMPSHIRE BREEDERS DIRECTORY
667	HOG CONTRACTEE LIST
668	PORK PRODUCER'S ASSOCIATION
669	HOG WASTE FEEDER LIST
670	POULTRY PRODUCER'S ASSOCIATION
671	POULTRY CONTRACTEE LIST
672	RATITE ASSOCIATION
673	RABBIT ASSOCIATION
674	RHEA PRODUCER'S ASSOCIATION
676	SHEEP PRODUCER'S ASSOCIATION
677	LLAMA AND ALPACA ASSOCIATION
687	WOOL GROWER'S ASSOCIATION
710	DAIRY PRODUCTS SYSTEM
920	STATE LAND PRESERVATION
928	CHRISTMAS TREE ASSOCIATION
929	CHRISTMAS TREE - STATE LIST
943	HORTICULTURE LIST
944	FLORICULTURE LICENSE
945	FORESTRY ASSOCIATION
946	NURSERY AND LANDSCAPE ASSOCIATION
950	TB LIST
955	PAPAYA GROWER LIST
956	HERB GROWER ASSN
957	LAVENDER (HERBS) GROWERS ASSN
958	NURSERY GROWER LIST
959	NURSERY LICENSE LIST
960	FLORAL CENSUS
961	ORCHID GROWER'S ASSOCIATION
962	ORGANIC GROWER'S ASSOCIATION
963	CERTIFIED ORGANIC
973	NATIONAL DEER, ELK, AND REINDEER LISTS
984	TURF GRASS SURVEY/ASSOCIATION
985	FORAGE, PASTURE, OR GRASSLAND ASSOCIATION