Capturing Additional Variability Introduced by Imputation within the Agricultural Resource Management Survey

Darcy Miller* Andrew Dau†

Abstract
The Agricultural Resource Management Survey (ARMS) is a multiple phase survey conducted by the United States Department of Agriculture's (USDA'S) National Agricultural Statistics Service (NASS), and cosponsored by the USDA's Economic Research Service (ERS). The third phase of the survey, ARMS III, collects data to provide an annual snapshot of the financial health of the farm sector and farm household finances. Recently, the ARMS III imputation methodology was updated from using the mean of a stratum group to iterative sequential regression (ISR). With the previous imputation methodology, data that were imputed were treated as observed values. Thus underestimated the true variance of an estimate was underestimated. With ISR, NASS can now evaluate the additional variability due to imputation. Multiple imputation was used to capture the additional variability. Rubin's (1987) method was used to combine the ARMS III estimates from multiple datasets. Variance estimates using pre-calibrated weights and calibrated weights were analyzed to assess the increase in variability due to imputation.

Key Words: nonresponse, imputation, variance estimation, missing data

1. Introduction

The National Agricultural Statistics Service (NASS) is a statistical agency located under the United States Department of Agriculture (USDA). NASS's mission is to provide timely, accurate, and useful statistics in service to U.S. agriculture. To successfully accomplish the agency's mission, NASS conducts numerous surveys every year and publishes more than 400 reports covering virtually every aspect of U.S. agriculture. Some examples of areas covered in NASS's reports are production and supplies of food and fiber, prices paid and received by farmers, farm labor and wages, farm income and finances, chemical use, and rural development. A wide variety of topics are covered within these different areas. The subject matter ranges from traditional crops, such as corn and wheat, to specialty commodities, such as mushrooms and flowers; from agricultural prices to land in farms; from once-a-week publication of cheddar cheese prices to detailed census of agriculture reports every five years. The size of the target population varies from fewer than 50 for a survey to nearly 3 million for the census of agriculture.

The Agricultural Resource Management Survey (ARMS) is conducted by NASS and cosponsored by the USDA's Economic Research Service (ERS). The ARMS is a three-phase survey that provides an annual snapshot of the financial health of the farm sector and farm household finances, and it is the only source of information available for objective evaluation of many critical policy issues related to agriculture and the rural economy. Data uses are wide and vary from univariate to multivariate analyses. Its data are essential to USDA and other federal administrative, congressional, and private-sector decision makers when they must weigh alternative policies and programs or business strategies that touch the farm sector or affect farm families.

*National Agricultural Statistics Service, 1400 Independence Ave SW, Washington, DC 20050
†National Agricultural Statistics Service, 1400 Independence Ave SW, Washington, DC 20050

Response to the ARMS III is voluntary. As with many other surveys, the ARMS III is subject to both unit nonresponse (the sampled record does not respond to the entire questionnaire) and item nonresponse (the respondent does not answer an item(s) on the questionnaire). Both unit and item nonresponse create gaps in the data that need to be addressed prior to the estimation process. Possible systematic differences between respondents and nonrespondents can lead to biased estimators, and the loss of information can lead to a reduction of efficiency for a particular item of interest when only analyzing complete cases.

A common solution to mitigate issues due to item nonresponse is imputation. Multiple imputation provides one useful strategy for dealing with data sets with missing values to properly reflect the uncertainty due to imputing values rather than having valid and true responses from units surveyed. In this paper, we explore the use of multiple imputation capture the additional variability due to imputation in the ARMS III survey.

2. ARMS III Survey

2.1 ARMS III Survey Description

The ARMS is administered in three phases. The first phase is a screening phase for in-scope and in-business farms as well as presence of the on to three targeted commodities for that year; the targeted crop and livestock commodities are selected on a rotational basis, which change from year-to-year. The second phase asks for detailed field-level data. The third phase (ARMS III) is a multi-mode, dual frame survey conducted annually in all states except Alaska and Hawaii. The sample consists of approximately 35,000 farms and ranches. It is selected from NASS's list frame, which attempts to cover all agricultural establishments within the U.S., and an area frame, which is used in ARMS to compensate for the incompleteness of the list frame. The survey questionnaire is mailed to the entire sample, but additional modes of data collection include web, face-to-face, and computer-assisted telephone, although telephone interviews are rare for this survey.

Based on data collected from the ARMS III, NASS publishes estimates of farm production expenditures for the U.S. (except Alaska and Hawaii), five regions, and fifteen leading cash receipt states. Farm production expenditures are also estimated for eight economic sales classes and two farm type categories. In addition to farm production expenditures, the ARMS III also collects data on production practices and costs of production for the targeted crop and livestock commodities. The production practices and cost of production data for these designated commodities are collected in the top producing states while the farm production expenditures data are collected in all states (except Alaska and Hawaii).

Because the survey data are utilized for in-depth analyses of critical policy issues related to agriculture and the rural economy, the ARMS III survey questionnaire is long and complex. The burden is 100 minutes. Some versions of the survey are 28 pages long, with data collected on more than 1000 variables. The survey questions encompass the characteristics, management, income, and expenses of both the farm operation and the farm household. Collecting full responses on all of the items is a challenge. Details concerning expenses of a contractor or landlord are also collected from the respondent and these items are often the most problematic, sometimes with over half of the observations missing. NASS has taken extensive steps to increase awareness of the importance of the survey. In an effort to reduce respondent burden, the sampling procedure reduces the probability of an operation being selected two consecutive years.

2.2 ARMS III Survey Processing

The diagram in Figure 1 illustrates the post-data collection processing of ARMS III survey data. Details of the process are provided in the subsequent sections.

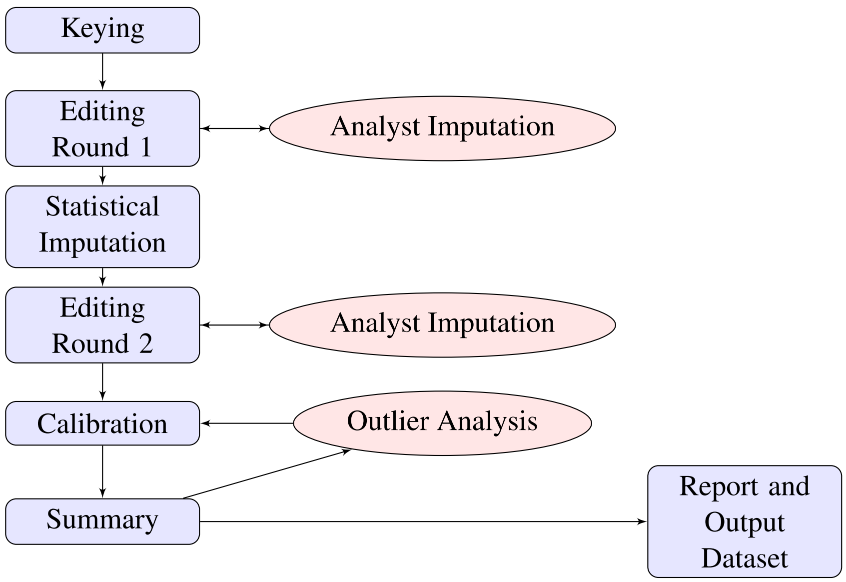

Figure 1: Diagram of post-data collection ARMS III survey processing

2.2.1 Editing Round 1 and Analyst Imputation

After data are keyed into the system, the questionnaire is processed through a computer edit that checks the consistency of the data and verifies that data values fall within a certain range. Then, a statistician reviews all questionnaire items that fail any of the edits. The statistician has the option of manually imputing the data item or marking the item to allow computer generated imputation of the item a later in the process. A manual imputation is typically performed when the statistical analyst has knowledge about the questionnaire item for that operation.

2.2.2 Editing Round 2 and Analyst Imputation

After the imputation routine is complete, the records with imputed data are re-edited to ensure the imputed values are acceptable. Relationships between data items on the current survey are verified, and in certain situations, items are compared to data from earlier surveys to ensure specific relationships are logical. A statistician is required to manually impute any item that fails an edit or could not be imputed. The edit logic also ensures administrative coding, such as the date the questionnaire was administered, follows the methodological rules associated with the survey design. In the case of an administrative code, a statistician is required to manually impute any item that fails an edit.

2.2.3 Calibration

Calibration is a weighting technique used in survey sampling to adjust the survey weights for sampled elements so that the weighted sum of a set of benchmark variables equals a pre-determined set of values for the population. For the ARMS III, the weights generated from the sampling procedures are used as input into the calibration algorithm. Sampling weights are calculated based on numerous factors so that the sample allocation can be representative of the entire population of farms at the state level for the fifteen leading cash receipts states and the five regions for all other states.

Due to survey nonresponse and the possibility of disproportionate responses across different farm types and economic sales classes, weights are adjusted through a calibration algorithm. Calibration adjusts the sampling weights so that the expanded data matches several known commodity, livestock and farm number published totals. The weights are calibrated to approximately 30 targets at specified economic class, geographic, and farm type levels. This ensures that the expense data collected will accurately represent the expense breakdowns for all farm types and farm sizes as well as cover the expenses for the entire target population.

Determination of an operation being in-scope and the economic class information of a unit is updated after incorporating the additional information provided from the imputations. In-scope means that the operation is in the target population. The calibration routine is run after the imputations are made, and in-scope status and economic class of each operation are determined. So, although estimated totals of imputed variables are not directly compared to the published totals in calibration, imputed variables can affect the calibration through the updated economic class assignment and determination of being in-scope.

2.2.4 Outlier Analysis and Summary

Outliers may be caused by aging control data resulting in misstratification, data errors, or the nonresponse and calibration adjustments to the sampling weight. A preliminary calibration and summary are run and any individual record accounting for 0.5 percent of the national estimate for total expenses or 2.5 percent of a regional estimate for total expenses is tagged as an outlier. After verifying the data have not been misrecorded or mishandled, background information on these outliers is compiled and presented to a National Outlier Board. This Board is a team of NASS and ERS analysts that meet to discuss the national outliers and form a consensus on a course of action.

Most outliers trace back to unique situations that do not exist in the target population as often as a large calibrated sample weight indicates. The Board looks at other respondents of the same locality, farm type, and sales class as the reported data on the outlier. The Board examines the weights of the comparable respondents and most often overrides the outlier's weight with the median weight of the comparable respondents. After the most extreme outliers have been addressed, the Board reviews the national totals by expense category following the same methodology and, when necessary, overrides weights of outliers with the median weight of the comparable reports. Finally, staff within NASS examine outliers found at the state level for the published expense categories. A determination is made as to whether a weight adjustment is justified. Adjustments are not made to all outliers, but they are reviewed closely for accuracy. It is important to note that the calibration algorithm is implemented after each stage of the outlier review process.

2.2.5 NASS Estimates: Farm Production Expenditure Report

After the final calibration is performed, more than 400 estimates are generated (typically as sums and ratios), and an estimate may be defined in terms of zero, or one or more imputed component variables. The variance is estimated using a delete-a-group jackknife estimator and coefficients of variation (CV) for the estimates are calculated. NASS publishes a set of 18 key estimates in the annual Farm Production Expenditures report, which is issued annually in August. The final dataset and calibrated weights are passed to ERS for further processing, multivariate analyses, and other reports.

A list of the key NASS Farm Production Expenditure estimates and their corresponding statistical imputation rates are in Table 1.

Farm Production Expenditure
Key Estimates

Estimate	Statistical Imputation Rate
Real Estate Taxes	High
Farm Services	Medium
Total Expenditures	Low
Agricultural Chemicals Expenditures	None
Farm Improvements and Construction	None
Farm Supplies and Repairs	None
Feed Expenditures	None
Fertilizer, Lime and Soil Conditioner Expenditures	None
Fuels Expenditures	None
Interest	None
Labor Expenditures	None
Livestock, Poultry, and Related Expenses	None
Miscellaneous Capital Expenses	None
Other Farm Machinery Expenditures	None
Rent	None
Seeds and Plants	None
Tractor and Self-Propelled Farm Machinery Expenditures	None
Trucks and Autos Expenditures	None

Table 1: 18 key estimates produced by NASS in the Farm Production Expenditure Report

2.3 ARMS III Imputation Methodology

Because ARMS III has many complex multivariate relationships the conditional mean imputation methodology used prior to 2014 generally cannot condition on a sufficiently large set of variables to maintain relationships among the variables imputed and all variables that might be included as related variables in a multivariate analysis. To develop methodology that would incorporate more information when conducting imputation, NASS collaborated with the National Institute of Statistical Sciences (NISS). Iterative sequential regression (ISR) was adapted to ARMS III and implemented for the 2014 survey year.

ISR is founded on the normal distribution. Many of the ARMS III data have a probability mass at zero. Thus, the semi-continuous nature of the ARMS III dataset requires special handling. To handle the probability mass at zero, an indicator variable is constructed for each item to denote whether a value of the item is non-zero or zero. Marginal transformations of the non-zero, continuous portion of each variable are then joined to form a multivariate normal joint density. The multivariate joint density is decomposed into a series of conditional linear models, and a regression-based technique is used. Various criteria utilized by subject-matter experts are used to select the covariates, which allows for flexibility in the selection of the covariates while still providing a valid joint distribution. Parameter estimates for the sequence of linear models and imputations are obtained in an iterative fashion using a Markov-chain-Monte-Carlo (MCMC) sampling method. The ISR method is described as a blend of data augmentation (DA) and fully conditionally specified (FCS) models, having the covariate choice flexibility of the FCS methods but the theoretical background of the DA methods (See Robbins, et al. 2013 for more details).

3. Methods

3.1 Multiple Imputation

Treating imputed values as reported values results in an estimated variance of a point estimate that is smaller than the true variance. Therefore, classical methods to calculate variances are insufficient in the presence of item nonresponse. One method to capture the additional variability due to imputation is multiple imputation. The goal is to make multiple draws from the distribution of the imputation model to obtain estimates that reflect the uncertainty associated with the imputation procedure, itself. The multiple draws create multiple data sets completed through imputation. The variance estimate combines the within data set and between data set variability of the estimate of interest.

3.1.1 Combining Datasets

Let m be the number of datasets created using multiple imputation. Rubin's (1987) method for combining the results from the m datasets is:

1. Calculate estimates from each dataset.

   $\widehat{Q_j}$is an estimate (e.g. an estimated total) obtained from data set $j(j = 1, 2, ..., m)$.

2. Calculate the variance associated with each estimate from each dataset.

   $\widehat{U_j}$is the estimated variance associated with $\widehat{Q_j}$.

3. The overall point estimate is the average of the individual estimates,

$$\overline{Q} = \frac{1}{m}\Sigma_{j=1}^{m}\widehat{Q_j}$$

4. The total variance, T, is

$$T = \overline{U} + \left(1 + \frac{1}{m}\right) B$$

where the within-imputation variance, $\overline{U}$, is

$$\overline{U} = \frac{1}{m}\Sigma_{j=1}^{m}\widehat{U_j}$$

and the between-imputation variance, B, is

$$B = \frac{1}{m-1}\Sigma_{j=1}^{m}\left(\widehat{Q_j} - \overline{Q}\right)^2$$

3.1.2 Measuring Effect on Estimates

Rubin also offers some measures of the effect on variance of imputing values versus having respondent data on estimates, of nonresponse and of the missing information.

The relative increase in variance due to nonresponse, r, is simply

$$r = \frac{\left(1+\frac{1}{m}\right)B}{\overline{U}} = \frac{T-\overline{U}}{\overline{U}}$$

The fraction of missing information, FMI, “depends not only on the missing data for that particular variable, but also on the percentage of missing data for other variables that are correlated with the variable.” (Allison, 2001). Rubin defines this as

$$FMI = \frac{r+\frac{2}{(df_m+3)}}{r+1}$$

where df_m is the degrees of freedom of the t distribution associated with the statistic,

$$\frac{\left(Q-\overline{Q}\right)}{T^{-\frac{1}{2}}}$$

3.1.3 Measuring Number of Imputations Needed

Imputing more than one value increases the time and manpower required for processing; therefore, it is prudent to use only the number of imputations necessary. Rubin also provides a measure, relative efficiency, to inform this decision. The relative efficiency, RE, of using finite m imputations versus infinite imputations based on the fraction of missing information is

$$RE = \left(1 + \frac{FMI}{m}\right)^{-1}$$

3.2 Study Design

ISR was used to multiply impute 2013 ARMS III data using parallel chains. Five hundred iterations were run for each chain before the imputation was drawn. We analyzed the relative efficiency, relative increase due to imputation, and fraction of missing information for m = 5 datasets and m = 10 datasets. Since the calibration routine is affected by the imputed values, we also analyzed these measures using pre-calibrated weights and calibrated weights. In this paper, to quantify the increase in variance due to imputation, we focus on the results from pre-calibrated weights applied to the Farm Production Expenditure Report key estimates. To estimate the additional variability due to the imputed values affecting the calibration routine, we focus on the Farm Production Expenditure Report key estimates that do not include imputation (table 2).

	Pre-Calibrated	Calibrated
m = 5		
Imputed Estimate		
Not Imputed Estimate		
m = 10		
Imputed Estimate		
Not Imputed Estimate		

Table 2: Study focus. Highlighted cells denote areas of focus.

4. Results

We display the results of the three key estimates produced by NASS for the Farm Production Expenditure Report and an estimate produced by NASS for ERS: real estate tax expenses (high imputation rate), farm services expenses (medium imputation rate), total expenses (low imputation rate), and value of assets (high imputation rate). The results for one key estimate produced by NASS, seed expenses, which does not contain any imputed values, is also shown.

4.1 Relative Efficiency

Relative Efficiency m = 5
Select Expenses and Assets

State	$RealEstateTaxExpenses$	$FarmServicesExpenses$	$SeedExpenses$	$TotalExpenses$	$Assets$
AR	0.88	>0.99	>0.99	>0.99	0.93
CA	0.98	0.94	>0.99	>0.99	0.94
FL	0.97	>0.99	>0.99	>0.99	0.90
GA	0.96	0.99	>0.99	>0.99	0.97
IN	0.91	>0.99	>0.99	>0.99	0.91
IA	0.93	>0.99	>0.99	>0.99	0.90
KS	0.94	>0.99	>0.99	>0.99	0.93
MN	0.97	>0.99	>0.99	>0.99	0.96
MO	0.97	0.96	>0.99	>0.99	0.90
NE	>0.99	>0.99	>0.99	>0.99	0.94
NC	0.90	0.94	>0.99	>0.99	0.90
TX	0.89	>0.99	>0.99	>0.99	0.89
WA	0.96	>0.99	>0.99	>0.99	0.97
WI	0.94	>0.99	>0.99	>0.99	0.94
Atlantic	0.87	0.98	>0.99	>0.99	0.96
South	0.98	>0.99	>0.99	>0.99	0.92
Midwest	0.94	>0.99	>0.99	>0.99	0.99
Plains	0.92	>0.99	>0.99	>0.99	0.87
West	0.98	0.95	>0.99	>0.99	0.86
US	0.91	0.97	>0.99	>0.99	0.90

Table 3: Relative efficiency of multiple imputation for m = 5 using calibrated weights.

Results of relative efficiency for the selected variables under five imputations are presented in table 3. The relative efficiency values are close to one and indicate that five imputations are sufficient. Therefore, we proceed to only show results from five imputations for the remaining analysis.

4.2 Relative Increase in Variance

Relative Increase in Variance m = 5
Select Expenses and Assets

State	$RealEstateTaxExpenses$	$FarmServicesExpenses$	$SeedExpenses$	$TotalExpenses$	$Assets$
AR	1.66	<0.01	0	<0.01	0.49
CA	0.06	0.96	0	0.41	0.23
FL	0.18	0	0	<0.01	1.06
GA	0.31	0.02	0	<0.01	0.16
IN	0.53	<0.01	0	<0.01	0.56
IA	0.28	<0.01	0	<0.01	0.57
KS	0.16	<0.01	0	<0.01	0.88
MN	0.13	<0.01	0	<0.01	0.27
MO	0.16	0.07	0	<0.01	1.11
NE	0.14	<0.01	0	<0.01	0.17
NC	0.97	0.34	0	<0.01	0.81
TX	1.47	<0.01	0	<0.01	1.36
WA	0.32	<0.01	0	<0.01	0.17
WI	0.19	<0.01	0	<0.01	0.31
Atlantic	1.67	0.09	0	<0.01	0.25
South	0.23	<0.01	0	<0.01	0.37
Midwest	0.35	<0.01	0	<0.01	0.09
Plains	1.23	<0.01	0	<0.01	2.60
West	0.10	0.95	0	0.32	0.75
US	1.15	0.91	0	0.07	0.48

Table 4: Relative increase in variance for m = 5 using pre-calibrated weights.

We used pre-calibrated weights to test the relative increase in variance so as not to introduce any calibration interaction within the results. The relative increase in variance (table 4) shows nominal increases in variance for most of the estimates and domains. As expected, the estimates that contain more imputed values have a larger increase in variance due to imputation and the estimate with zero imputation has no increase in variance. Florida also has a relative increase of zero for the farm services expenses estimate; no imputations were made for items used in the summary of this estimate.

4.3 Fraction of Missing Information

We used pre-calibrated weights to test the fraction of missing information to prevent introducing noise from calibration interaction within the results. The fraction of missing information (table 5) indicates that real estate tax expenses had the highest fraction of missing information on average. Again, we see that the result for the seed expense estimate for each domain and farm services expense estimate in Florida is zero; these estimates do not contain any imputed values.

4.4 Increase in Variability Using calibrated Weights

As noted earlier, imputed values impact what level a unit is in for calibration (e.g. economic class) as well as in-scope status. We also analyzed the resulting measures in this study using calibrated weights (table 6).

Fraction of Missing Information m = 5
Select Expenses and Assets

State	$RealEstateTaxExpenses$	$FarmServicesExpenses$	$SeedExpenses$	$TotalExpenses$	$Assets$
AR	0.68	<0.01	0	<0.01	0.37
CA	0.06	0.54	0	0.32	0.20
FL	0.16	0	0	<0.01	0.57
GA	0.26	0.02	0	<0.01	0.15
IN	0.38	<0.01	0	<0.01	0.40
IA	0.24	<0.01	0	<0.01	0.40
KS	0.15	<0.01	0	<0.01	0.52
MN	0.12	<0.01	0	<0.01	0.23
MO	0.15	0.07	0	<0.01	0.58
NE	0.13	<0.01	0	<0.01	0.16
NC	0.55	0.28	0	<0.01	0.50
TX	0.65	<0.01	0	<0.01	0.63
WA	0.26	<0.01	0	<0.01	0.15
WI	0.17	<0.01	0	<0.01	0.25
Atlantic	0.68	0.98	0	<0.01	0.21
South	0.20	<0.01	0	<0.01	0.29
Midwest	0.28	<0.01	0	<0.01	0.09
Plains	0.61	<0.01	0	<0.01	0.77
West	0.10	0.54	0	0.26	0.72
US	0.59	0.91	0	0.06	0.48

Table 5: Fraction of missing information for m = 5 using pre-calibrated weights.

Relative Increase in Variance m = 5
Select Expenses and Assets

State	$RealEstateTaxExpenses$	$FarmServicesExpenses$	$SeedExpenses$	$TotalExpenses$	$Assets$
AR	1.77	0.02	<0.01	<0.01	0.30
CA	0.12	0.45	<0.01	0.06	0.43
FL	0.16	<0.01	<0.01	<0.01	1.04
GA	0.28	0.06	<0.01	<0.01	0.18
IN	0.76	<0.01	<0.01	<0.01	0.74
IA	0.53	<0.01	<0.01	<0.01	1.02
KS	0.38	<0.01	<0.01	<0.01	0.48
MN	0.18	<0.01	<0.01	<0.01	0.23
MO	0.20	0.25	<0.01	<0.01	1.03
NE	0.02	<0.01	<0.01	<0.01	0.38
NC	0.92	0.42	<0.01	0.03	0.96
TX	1.52	<0.01	<0.01	<0.01	1.3
WA	0.24	<0.01	<0.01	<0.01	0.17
WI	0.36	<0.01	<0.01	<0.01	0.41
Atlantic	2.46	0.10	<0.01	0.01	0.25
South	0.11	0.03	<0.01	<0.01	0.71
Midwest	0.43	<0.01	<0.01	<0.01	0.08
Plains	0.63	<0.01	<0.01	<0.01	2.69
West	0.13	0.33	<0.01	0.01	2.93
US	0.87	0.20	<0.01	<0.01	0.93

Table 6: Relative increase in variance for m = 5 using Calibrated weights.

We found that estimates containing zero imputed values have a small increase in variance (less than one hundredth). Hence, we have empirical evidence that the imputed values' contribution to determining in-scope status and calibration level adds a negligible amount of variability to estimates with zero imputed values. Estimates with imputed values have an relative increase in variance; however, note that the relative increase in variance using calibrated weights may be larger or smaller than the relative increase due to variance using the pre-calibrated weights. We expect this since the calibrated weight may be smaller or larger than the pre-calibrated weight. The difference in the increase in variance using the calibrated weight versus the pre-calibrated weight is generally small.

5. Conclusion

Multiple imputation is a way to capture the additional variability due to imputation. When applied to the complex ARMS III survey, we found evidence that as few as five imputations may be adequate. For many of the estimates that NASS produces in the Farm Production Expenditure report, the increase in variance is small. However, for the three estimates that contain imputed values, some geographic domains have a large increase in variance. Finally, by analyzing estimates using calibrated weights, we found that estimates containing zero imputation have a negligible increase in variability due to the imputation. The estimates that have some imputation have an increase in variability regardless of which weights are used; however, the increase in variability is not consistently larger using calibrated weights.

REFERENCES

Gelman, A. and Rubin, D.(1992), "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Science* No. 7, pp.457-511.

Lacerda, M., Ardington, C., Leibbrandt, M.(2008), "Sequential Regression Multiple Imputation for Incomplete Multivariate Data using Markov Chain Monte Carlo," *Southern Africa Labour and Development Research Unit - Working paper Series* No. 13.

Miller, D. and O'Connor, T. (2012), "Item Response Rates for the Agricultural Resource Management Survey III in 2006 and 2007," *National Agricultural Statistics Service Research Report*, RDD-11-07.

Miller, D. and Young, Linda J. (2015) "Imputation at the National Agricultural Statistics Service," *United Nations Economic Commission for Europe Work Session on Statistical Editing*, Working Paper 14.

Raghunathan, T., et al (2001), "A multivariate technique for multiply imputing missing values using a sequence of regression models," *Survey Methodology*, 27, pp. 85-96.

Robbins, M., et al (2013), "Imputation in High-Dimensional Economic Data as Applied to the Agricultural Resource Management Survey," *Journal of the American Statistical Association*, 108, pp. 81-95.

Rubin, D.B. (1987), *Multiple-Imputation for Non-Response in Surveys*, John Wiley and Sons, New York.

Schafer, J.L. (1997), *Analysis of Incomplete Multivariate Data*, Chapman Hall/CRC.

Tanner, M.A and Wong, W.H. (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82, pp. 528-550.

Van Buuren , S., Brand , J. P.L., Groothuis-Oudshoorn, C. G.M., and Rubin, D.B. (2006), "Fully conditional specification in multivariate imputation," *Journal of Statistical Computation and Simulation*, 76:12, 1049-1064.

de Waal, T., Pannekoek, J., and Scholtus, S. (2011), *Handbook of Statistical Data Editing and Imputation*, Wiley Handbooks in Survey Methodology, John Wiley Sons, Inc.