EVALUATING A ONE NUMBER APPROACH TO THE AGRICULTURAL CENSUS

Chadd Crouse
National Agricultural Statistics Service

Abstract

The National Agricultural Statistics Service (NASS) conducted the Census of Agriculture for the first time in 1997, and is now the sole USDA agency responsible for agricultural statistics. NASS is considering the adoption of the CALJACK calibration SAS algorithm to insure agreement between census estimates and current annual estimates and thus achieve a "one-number" census. This paper will examine several options in an attempt to determine the best scheme to be used in the calibration process including the choice of initial weight, handling of extreme operators, selection of items on which to calibrate and the desired level of calibration (i.e. state, district or county). There is also a brief examination of some results of the calibration process with a particular focus on those items not involved in the process.

General Background

During 1997 and 1998 the USDA's National Agricultural Statistics Service (NASS) conducted the Census of Agriculture for the first time (it was formerly conducted by the Bureau of Census) with the final results published in February 1999. However, the census estimates did not "agree" with the currently published NASS estimates at the county, state or national level. For example, the 1997 Census of Agriculture estimate of corn for grain produced in the state of Michigan is different from the current NASS estimate for 1997 by over seven percent. Additionally, this current NASS number has been revised based on the 1997 Census of Agriculture estimate (among other things). This presents a problem since both sets of results are now "official" NASS estimates.

The issue of how to handle, and hopefully reduce, differences such as these for the 2002 Census of Agriculture will be discussed at great length within NASS over the next couple of years. These differences result to some extent from definitional differences between the Census and NASS' sample surveys. As much as possible, these will be reviewed and minimized prior to the 2002 Census. After all possible changes are made to increase comparability, however, some differences will remain. If NASS decides to bring its official estimates (based on its ongoing surveys) and the census results closer together, calibration is a possibility for achieving this goal.

Calibration is needed to project State level census totals adjusted for duplication and misclassification to finer levels of aggregation. This is another possible application of the procedures discussed in this paper. The development in this paper describes the most general use of the procedures in using all information available to the Agency (i.e., census results, estimated census adjustments, survey results and official estimates) to come up a best number for each estimated commodity. However, if the decision is made to use the procedures only to distribute census adjustments to lower levels of aggregation, the lessons learned in optimally using the procedures should still be applicable.

This paper will look at the possibility of using Caljack, which is a modified version of the SAS macro CALMAR developed by O. Sautery of INSEE (France), to assign weights to each record. This algorithm calibrates the records to a "known" total for chosen items and establishes a single weight for each record so that the weighted sum achieves these totals. By using this method of assigning weights NASS will be able to solve the problem of disagreements between official estimates. The paper will attempt to determine the best set of parameters to implement the algorithm, and the feasibility of implementation for the 2002 Census of Agriculture by studying various results for the state of Michigan using 1997 census data. Michigan was chosen primarily because of the wide variety of commodities for which state and county level data are available. All calibrations for Michigan data in the rest of this paper will be based upon the NASS estimates prior to any Census revisions.

CALJACK Details

The Caljack algorithm was suggested by Dr. James Gentle, our lead research partner at George Mason University, as a tool to achieve the goal of a one number census. Originally the algorithm ended once the maximum absolute change to

any weight from one iteration to the next reached a given level. To better serve our purposes the program was altered so that it ends once the maximum absolute percentage difference between the official estimates for the items of interest and the weighted totals is less than a given percentage. Prior to making this change the algorithm consistently failed to converge (and thus failed to produce calibration weights) when calibrating more than ten or fifteen items. For the remainder of this paper the items are considered calibrated once the weighted totals are all within five percent of the known totals.

Caljack offers a choice of seven different calibration methods which include: Linear Regression, Raking Ratio, Logit (or Generalized Modified Discrimination Information) and Linear Truncated. For a more detailed description of the methods and computational algorithms used see Singh and Mohl (1996). Also, see Deville, Särndal and Sautory (1993) and Deville and Särndal (1992) for further discussion of calibration methods. The Logit and Linear Truncated methods of calibration have an apparent advantage over the other five methods in that they allow the user to set limits (range restrictions) on the changes to the initial weights. This allows the user to guarantee that all final weights are positive assuming that all initial weights are positive. Without this guarantee some county level weighted item estimates would almost certainly result in negative values which is unacceptable. The difference between these two methods (Logit or Linear Truncated) appears to be minimal. Calibrating the Michigan data at the state level for 47 major commodities resulted in a correlation between the two sets of weights of 0.98. The average absolute difference and average absolute percentage difference between the calibrated values and NASS published estimates at the county level for various items are in the following table. The shaded cells are the cells with the "best" estimates.

Table 1

Item	Mean Absolute Difference		Mean Absolute Percent Difference	
	Logit	Linear Truncated	Logit	Linear Truncated
Corn Grain Acres	4,308.00	4,267.75	0.15901	0.15869
Corn Grain Bushels	543,157.06	535,657.53	0.16744	0.16681
Barley Bushels	5,262.80	5,226.74	0.07715	0.07741
Soybean Acres	4,241.54	4,210.15	0.13256	0.13179
Soybean Bushels	161,501.45	157,926.83	0.14290	0.14154
Dry Bean Acres	352.78	338.35	0.04165	0.04066
Beef Cows	208.64	200.77	0.12502	0.11959
Milk Cows	514.34	528.20	0.12045	0.12211

From this table it seems as though the estimates produced using the Linear Truncated method are best, and this method will be the method of choice for the remainder of this paper. In fact, for the 24 commodities which were compared at the county level the Linear Truncated method was better (based on the absolute difference) in 17 cases. It should be noted that the differences between the two estimates are minimal, and that further testing with data from additional states may be warranted to determine if the Linear Truncated method is in fact the better of the two methods.

Initial Weight

Currently, integer weights of one or two are applied to each census record to adjust for non-response. These weights are essentially the estimated number of non-respondents for a given county and stratum plus the total number of respondents for that county and stratum divided by the total number of respondents for that county and stratum. These weights are then converted to integers using a controlled rounding procedure. One of the first questions that presented itself was whether the calibration should begin with these original non-response weights, or whether every record should be assigned an initial weight of one and calibration applied to that. The benefit of starting with an initial weight of one would be greater control over the final weights. Caljack only allows the user to specify the minimum and maximum final weights in terms of a fraction of the initial weight. By beginning the calibration with a weight of one the user would be able to directly control the range of the final weights. However, there was some concern that the original census weights should be used as the initial weight since those weights led to unbiased estimates.

Calibrated weights were derived using 16 different weighting schemes which could be divided into two distinct groups

of eight. The first group used an initial weight of one, and the second group used an initial weight equal to the original non-response weight. As expected, each case in the second group had a higher correlation to the original non-response weight than every case in the first group. However, the highest correlation in the second group was only 0.24. Therefore, since the relationship between the final weights from the second group and the non-response weight is weak, it seems that it makes little difference which initial weights are used in the calibration process if the only goal is to ensure that the final results are unbiased. The choice of initial weight may make a difference in terms of producing the "best" final estimates, so for the remainder of this paper both choices for the initial weight will be considered.

Extreme Operators

Extreme operators were defined in the 1997 Census of Agriculture as those whose reported item inventory or dollar amount exceeded a pre-determined limit for the item. For some minor commodities this pre-determined limit was merely a presence of the commodity (such as mushrooms in any state except PA or CA). Also, any operation with more than 1,000 acres of harvested cropland, 1,000 or more head of cattle, and depending on the state, either 1,000 or 1,500 or more hogs and pigs were marked as extreme operators. Any operation that was identified as extreme was automatically assigned a non-response weight of one.

The following table details the breakdown of extreme operators based on various major reported items in the state of Michigan for 1997. The third column includes all operations that exceeded the extreme operator cutoff for the items listed in the first column and thus were excluded from potential non-response weighting. The last column includes all operations that exceeded the cutoff for the item listed and at least one other item.

Table 2

Item	Operations reporting this item	Operations Excluded	Excluded, based on this item only	Excluded, based on this item plus
Orchards	2,621	1,472	1,191	281
Total Value of Production	37,248	3,265	787	2,478
Value of Land and Buildings	13,086	349	8	341
Land in Farms	42,084	1,712	201	1,511
Cropland Harvested	34,818	1,227	0	1,227
Milk Cows	3,731	961	152	809
Hogs and Pigs	2,601	247	33	214
Other Criteria	24,694	790	69	721

It is worth noting that the average orchard in Michigan, based on the 1997 Census of Agriculture, was 48.8 acres, and any operation with more than ten acres (56 percent of the population) was automatically excluded from non-response weighting.

To avoid calibration problems caused by the liberal exclusion of extreme operators for some characteristics an alternate extreme operator definition was derived for this analysis. This new definition was designed to identify operations that comprised a significant percentage of any given item or items in any given county. It also was designed to exclude those operations which make up a significant amount of a commodity in a county with virtually no amount of the commodity of interest. In order to achieve this goal two new variables were defined. The first variable (eofactor) was defined as:

$$\text{EoFactor} = \sum [(\text{Quantity Reported on Operation})/(\text{Quantity Reported in County} + 100)]$$

The second variable (maxfactor) was defined as:

$$\text{MaxFactor} = \text{Max}[(\text{Quantity Reported on Operation})/(\text{Quantity Reported in County} + 100)]$$

Both EoFactor and MaxFactor are over 38 items with official NASS estimates available.

A preliminary cutoff was chosen based on the ninety-fifth quantile value of 0.153 for EoFactor and 0.0801 for MaxFactor. After further analyzing the data, final cutoffs were set at 0.2 and 0.08, and only those operations with a total value of products sold greater than $10,000 were identified as extreme operators. Operations with a total value of products sold over $10,000,000, regardless of their EoFactor or MaxFactor values, were also identified as extreme operators. This new definition resulted in only 1,970 extreme operations versus 5,088 using the original census definition. Some additional comparisons to the original census definition are included in the following table. The final two columns in the table are only for those operations that were defined as extreme operators under both the census definition and the new definition. The column labeled "Top" indicates the rank of the last farm that is considered an extreme operator under both definitions when the data are sorted by size of the item. So, for example, the 17 largest Orchards in the state are considered extreme operators under both definitions. The column labeled "Of 100" is simply the number of the 100 largest farms for the given item which are considered extreme operators under both definitions.

Table 3

Item	Average Item Value for Farms Defined as Extreme Operators		Number of Extreme Operators with Item Present		EOs Under Both Definitions	
	Census Definition	"New" Definition	Census Definition	"New" Definition	Top	Of 100
Orchards	88.2	152.2	1,480	352	17	60
VLAB	1,739,680	1,475,844	2,918	1,298	3	56
Cropland Harvested	723.2	600.4	3,904	1,846	3	45
Cattle and Calves	338.6	265.5	1,370	799	14	53
Milk Cows	183.2	168.7	1,022	379	7	40

The most obvious difference between the two definitions is the significant decrease in the number of operations with orchard land and milk cows, and the large increase in average orchard acres. Additionally, 883 operations were identified as extreme operators under both definitions.

Calibrated weights were derived using the 16 different weighting schemes discussed in the Initial Weight section. Eight of these schemes included the extreme operators in the calibration process while the other eight excluded them. Half of the cases used an initial starting value of one, and the other half used the original census weight. However, the actual, reported data of the excluded extreme operators was deducted from the official NASS estimate, and these operations were assigned a final weight of one regardless of their initial census weight. After running Caljack for each of the schemes, the resulting weighted estimates were compared to 24 official NASS published estimates at the county level. In every case, excluding the extreme operators from the calibration process produced "better" results than including them. Additionally, and unexpectedly, when including the extreme operators in the calibration process Caljack occasionally failed to converge to a single answer, and, thus, NO weights were derived. In these cases variables were removed from the calibration process until Caljack converged successfully. The fact that removing the extreme operators resulted in "better" estimates indicates that this is certainly the best procedure to follow.

Items on Which to Calibrate

The selection of items to calibrate poses several problems. If the user chooses to calibrate on too many items it is likely that Caljack will fail to find a solution. Additionally, some NASS estimates are stronger than others, and, thus, the user may choose to calibrate only on the estimates in which the user is most confident.

The current plan is to conduct the 2002 Census of Agriculture and derive weighted totals using historic census procedures and methodology. These weighted totals will then be used as an additional indication, along with any other available data, including current NASS survey data, to determine the official NASS estimates for given items. Then the census data will be calibrated to these new official estimates. Under this scenario the user would likely calibrate for every variable for which an official estimate was produced. It is unlikely that official estimates would be produced for all commodities produced in every state due to the limited production time frame.

In order to determine which items to use for calibration, a comparison study was conducted. Again, weights were derived using 16 different calibration schemes. In eight of these schemes calibration was performed on as many items as possible for which a comparable NASS estimate was available. In the other eight calibration was performed on only the top items in terms of cash receipts. The schemes which used as many items as possible included 47 items. The schemes which used only the top commodities excluded 11 of these items. The top commodities included only the commodities which represented approximately 0.5% or more of the total cash receipts for the state for 1997.

The following table is a comparison of calibrating only top commodities versus all commodities for various items at the state level, excluding extreme operators, and with an initial value equal to the original census non-response weight. Items with no shading are included in both calibration methods. Items with light shading are not included in the "Top Commodities" calibration, and items with dark shading are not included in either calibration.

Table 4

Item	NASS	Census	All Commodities	Top Commodities
Corn for Grain Acres	2,250,000	2,122,283	2,273,516	2,274,488
Corn for Grain Bushels	263,250,000	238,319,129	265,808,862	265,912,476
Soybean Acres	1,890,000	1,694,872	1,904,155	1,903,881
Soybean Bushels	72,765,000	62,242,411	73,224,391	73,216,228
Number of Farms	51,000	46,027	51,151	51,147
Oat Acres	90,000	77,588	90,333	85,465
Oat Bushels	5,490,000	4,624,435	5,505,855	5,206,839
Barley Acres	24,000	18,893	23,996	19,409
Barley Bushels	1,440,000	1,032,383	1,438,424	1,090,909
Rye Acres	16,000	13,469	15,991	13,301
Rye Bushels	416,000	399,027	417,376	401,155
Sheep and Lambs	90,000	72,107	90,024	77,188
Bee Colonies	85,000	72,551	85,027	71,029
Hens and Pullets	5,160,000	4,928,067	5,147,224	4,955,158
Celery Acres	2,100	2,273	2,304	2,306
Onion Acres	6,100	4,725	4,800	4,805
Tomato Acres	6,000	7,779	8,277	8,264

From this table it is obvious that there is a definite difference when commodities are included in the calibration process. However, the newly weighted estimates of items which are omitted from the process are certainly not significantly "worse" than the current Census estimates. Also, the level of variance of the official estimates for the small, relatively insignificant items is greater than for the more important items, so they are not likely better than the Census estimates. The user should calibrate for every official NASS number that is determined under the scenario described in the second paragraph of this section. However the official NASS numbers should perhaps only be determined for the major commodities in each state, and all other official NASS numbers should be derived from the final weighted census numbers. This decision is beyond the scope of this paper, and for the remainder of the paper, calibration for both as many variables as possible and only the top commodities will be considered.

Level of Calibration

The final question to be considered was the level to which the data should be calibrated. For most major agricultural items official estimates are available at the agricultural statistics district and county level, so the option of calibrating down to the county level does exist. However, based on the scenario in which a new official NASS estimate is determined after reviewing census data, it seems unlikely that these new estimates will be determined at the county level. Also, in the past most NASS county data has been significantly revised to levels more "in line" with the census data. The following table for 1997 Michigan corn for grain harvested acres at the district level gives an example of how

NASS estimates tend to "move" towards Census estimates once the Census is released.

Table 5: 1997 Michigan Corn for Grain Acres

District	Original NASS	Census	"New" NASS
Upper Peninsula (10)	13,800	9,704	11,400
Northwest (20)	29,600	28,780	30,100
Northeast(30)	23,600	24,535	25,500
West Central (40)	64,000	58,436	60,000
Central (50)	239,000	218,874	224,000
East Central (60)	450,000	419,871	430,000
Southwest (70)	355,000	348,156	355,000
South Central (80)	715,000	679,711	700,000
Southeast (90)	360,000	334,216	344,000
State Total	2,250,000	2,122,283	2,180,000

Calibration at the county level decreases the likelihood that Caljack will converge to a single solution since there are more variables for the algorithm to solve. In fact, when calibrating to the district level for only corn and soybeans Caljack failed to find a solution when many of the smaller districts were included (i.e. district 10 for corn). Therefore, an attempt to calibrate to the county level for any but the largest counties would be unsuccessful. Despite the fact that calibration at the county level is possible, the fact that NASS estimates eventually "move" towards Census estimates indicates that it is reasonable to calibrate only at the state or possibly district level for a few major items.

Minor Items

An additional concern that must be addressed is the effect that the new weights have on minor commodities or other items for which there are no current NASS published estimates. The following table lists a few of the largest differences for items that were not included in the calibration process, or were not directly related to a calibrated item, and are sorted by the T-value of the differences between the Census and the "New" Census based on an expected value of zero. The "New" Census calibrated totals used the "best" method as described in the next section of this report. The actual Census estimate is included as a point of reference.

Table 6

Item	Census	Sum of Weighted Differences	Standard Error	T-Stat
Horse and Pony Inventory	66,201	8,941	536.1	16.676
Total Tractors less than 40 HP	38,792	1,129	81.5	13.858
Total Motortrucks	76,320	2,819	240.7	11.710
Acres in CRP	287,081	27,369	2,476.1	11.053
Bedding Plants Under Glass Harv.	29,560,166	3,287,367	313,484.0	10.487
Total Government Payments	92,806,090	10,782,782	1,337,386.1	8.063
Grass Silage - Acres Harvested	240,138	-32,822	4,510.6	-7.277

There were approximately 400 variables that were not calibrated or directly affected by calibration. For example, soybean acreage was a calibration variable, but soybean irrigated acreage was not. A change to soybean irrigated acres is a direct result of a change to soybean acres. Of these more than 40 percent of the differences between the Census and the "New" Census were significantly different from zero at the $\alpha = 0.05$ level.

A solution to the potential problem of significant changes to minor items is to include the items with large changes in the extreme operator statistics. However, this process of producing calibrated estimates, checking for items with large, unexpected changes, and then re-calibrating would require a significant amount of extra work. It would also enter user judgement into the process which is undesirable, since results could differ from user to user when the data involved are

identical. There are only two apparent alternatives: 1) include every item in the extreme operator determination step, or 2) ignore the few commodities that have large changes for which we have no official NASS estimate. Neither solution is very appealing since the former would decrease the number of reports available for the calibration process, and the latter is likely to cause problems much like the items listed above.

Final Results and Conclusions

Two methods of measurement error assessment at the county level are displayed in the following tables for the 16 different weighting schemes that were examined. It should be noted here that because only published NASS estimates were compared an error value of zero was assigned in both methods for the cases in which a NASS number was not published and the weighted census number was positive.

The error estimates in the first table were calculated by summing the absolute differences between the published NASS number and the final weighted census number for 24 items for each of the 16 weighting schemes. The mean of these 16 error estimates for each commodity was then calculated and each of the error estimates was divided by this new mean to develop an index estimate for each commodity-scheme combination. These indexes were then averaged across all commodities to produce the displayed weighting scheme index. This method of comparison allows larger counties to have a greater effect on the commodity-scheme index value than smaller counties, and thus attempts to indicate whether or not the calibration scheme produced "good" estimates for the "important" counties for each commodity.

Table 7: Published Estimates - Absolute Error

		District Level Calibration		State Level Calibration	
		Initial = 1	Initial = Census	Initial = 1	Initial = Census
"All" Commodities Calibrated	EOs Excluded	0.9838	0.9694	0.9800	0.9680
	EOs Included	1.0330	1.0735	1.0447	1.0176
"Top" Commodities Calibrated	EOs Excluded	0.9957	0.9756	0.9587	0.9386
	EOs Included	1.0275	1.0378	1.0117	0.9847

The second table uses the same method as the previous table except instead of absolute error the absolute percentage error was computed initially. This method allows the small counties to have a much greater effect in the comparison, and, thus, attempts to indicate whether or not the calibration scheme produced generally "good" estimates for all counties for each commodity.

Table 8: Published Estimates - Percentage Error

		District Level Calibration		State Level Calibration	
		Initial = 1	Initial = Census	Initial = 1	Initial = Census
"All" Commodities Calibrated	EOs Excluded	0.9464	0.9438	0.9596	0.9554
	EOs Included	0.9928	1.0818	1.1000	1.0915
"Top" Commodities Calibrated	EOs Excluded	0.9349	0.9310	0.9368	0.9325
	EOs Included	0.9947	1.0647	1.0661	1.0681

In both tables numbers less than one indicate "better" estimates at the county level when compared to current NASS county estimates. As previously mentioned, in every single case the scheme with extreme operators removed produced better results than the corresponding case with extreme operators included in the calibration process. Also, in both tables it is obvious the state level calibration of the top commodities with the EOs excluded produced the "best" results. Examination of the Percentage Error table seems to indicate that the choice of initial starting value and level of calibration is irrelevant. The following table, which lists the maximum index value across all commodities was examined. It seems reasonable to select the scheme which does not decrease the overall error at the expense of a relatively large error in one or two commodities.

Table 9: Published Estimates - Percentage Error - Maximum Index Value

		District Level Calibration		State Level Calibration	
		Initial = 1	**Initial = Census**	**Initial = 1**	**Initial = Census**
"All" Commodities Calibrated	**EOs Excluded**	1.1066	1.0854	1.1089	1.0809
	EOs Included	1.4538	1.4761	1.4218	1.4261
"Top" Commodities Calibrated	**EOs Excluded**	1.1290	1.1019	1.1018	1.0687
	EOs Included	1.3985	1.4517	1.2348	1.2573

The results in this table, when combined with the previous two tables, lead to the conclusion that the "best" calibration scheme for Michigan is, again, to use an initial weight equal to the original census weight, exclude EOs, and to calibrate only the top commodities at the state level.

Under this calibration scheme Caljack appears to be a very promising tool for use in producing a one number census for the 2002 Census of Agriculture. However, a few minor modifications to the program, such as the handling of convergence failures, would make the program an ideal tool for this application.

References

DEVILLE, J.C. and SÄRNDAL, C.E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87, 376-382.

DEVILLE, J.C., SÄRNDAL, C.E., and SAUTORY, O. (1993). Generalized raking procedures in survey sampling. *Journal of the American Statistical Association*, 88, 1013-1020.

SINGH, A.C. and MOHL, C.A. (1996). Understanding calibration estimators in survey sampling. *Survey Methodology*, 22, 107-115.