Comparison of Methods for Updating Census Based Estimates of Number of Farms to Non-Census Years

Michael E. Bellow[1], Phillip S. Kott[2]

[1]USDA-NASS
3251 Old Lee Highway, Room 305, Fairfax, VA 22031

[2] RTI International
6110 Executive Blvd., Suite 902, Rockville, MD 20852

Abstract

The USDA's National Agricultural Statistics Service (NASS) conducts the census of agriculture every five years. On an annual basis, sample surveys including the June Area Survey (JAS) are carried out to obtain estimates of many of the same agricultural commodities as the census. Due to the large number of farming operations surveyed and the complete coverage provided by the census, its numbers are considered more accurate than those derived from the much smaller scale sample surveys. An interesting question is whether census figures for specific survey items can be used in conjunction with survey data to improve estimation accuracy for non-census years. Because of its relative stability over time, the item considered most likely to benefit from such an approach is number of farms.

We evaluate two proposed methods for projecting census counts of number of farms to subsequent non-census years, both at the state level and within categories defined based on value of sales. The first method updates census figures to the current year using JAS data only, while the second makes additional use of official NASS state level estimates of number of farms for the previous year. The two estimators are compared with area frame based and hybrid operational estimators for the 2003-06 period in a study covering most of the lower 48 states. Variances are estimated using an extended delete-a-group jackknife method.

Key Words: census, area sampling frame, delete-a-group jackknife, smoothed alternative.

1. Introduction

The National Agricultural Statistics Service (NASS) publishes estimates of number of farms at the state level as well as within designated sales classes in each state. The agency conducts the census of agriculture (a complete count of US farms and ranches) twice each decade in years ending in '2' and '7'. This paper introduces two related methods of updating census based farm number estimates for years between censuses, using data from NASS's June Area Survey (a comprehensive annual area frame based survey of agricultural activity in the US) and (for one of the methods) official published estimates for the year preceding the current one. While the focus is on estimating the number of farms, the approach may be applicable to other characteristics estimated from the surveys as well. The two methods are referred to as K1 and K2.

Since 1975, NASS has defined a farm as an agricultural operation with at least $1,000 in sales or potential sales for a given year (with certain government agricultural programs included in the category of 'sales'). A key issue is estimation of the number of farms within designated sales classes. In addition to state level estimates of number of farms, NASS is also required to categorize farm numbers based on the following total farm value of sales groups within a state:

1) $1,000 to $9,999
2) $10,000 to $99,999
3) $100,000 to $249,999
4) $250,000 to $499,999
5) $500,000 or greater

There are two things to keep in mind about farm numbers from the June Area Survey (JAS): 1) sales are defined a bit differently for point farms (those having actual sales below $1,000 in the previous year but with sufficient crops and livestock that they could have sold at least $1,000 worth of agricultural products), and 2) the number of farms in a given year is divided into groups based on their sales for the previous year. Both values are available from the JAS.

The two proposed estimators were evaluated in a study carried out for most of the lower 48 states, both at the state level and within sales classes. Estimates of number of farms computed using the two methods for the years 2003-06 based on data from the 2002 Census were compared with corresponding area frame and hybrid operational estimates, with variances of K1 and K2 estimated using an extended delete-a-group jackknife method.

2. Description of Estimators

In the first year after a census of agriculture, the new estimators K1 and K2 of total number of farms in a state and within sales classes are identical:

$$\hat{Y}_t^{(K1)} = \hat{Y}_t^{(K2)} = Y_{t-1}^{(C)} \hat{b}_t , \tag{1}$$

$$\hat{Y}_{tg}^{(K1)} = \hat{Y}_{tg}^{(K2)} = Y_{(t-1)g}^{(C)} \hat{b}_t \tag{2}$$

where:

t = first post-census year,

g = sales class (g=1,…,5),

$Y_{t-1}^{(C)} =$ fully-adjusted (for nonresponse and coverage issues) census number of farms at the state level for year *t-1*,

$Y_{(t-1)g}^{(C)} =$ fully-adjusted census number of farms in sales class g for year *t-1*,

$$\hat{b}_t = \sum_{k \in S_t} w_{kt} y_{kt} \Big/ \sum_{k \in S_t} w_{k(t-1)} y_{k(t-1)} ,$$

$w_{ks} =$ year s expansion factor for segment k (recalculated using only segments in the samples for both year s and year *s-1),*

$y_{ks} =$ number of farms for segment k in year s (which need not be an integer; when a fraction of a farm's entire land area is in segment k, only that fraction – the farm's "tract-to-farm ratio" in k – is a part of this count), and

$S_t =$ set of segments in the samples for both year t and year *t-1.*

The term $\hat{b}_t$ is known as the *change ratio* of overall number of farms between years *t-1* and *t*. For the second, third and fourth post-census years, the K1 estimators of number of farms at the state and sales class levels are:

$$\hat{Y}_t^{(K1)} = \hat{Y}_{t-1}^{(K1)} \hat{b}_t , \tag{3}$$

$$\hat{Y}_{tg}^{(K1)} = \hat{Y}_{(t-1)g}^{(K1)} \hat{b}_t + \sum_{k \in St} w_{kt}(y_{ktg} - \hat{b}_t y_{k(t-1)g}) \tag{4}$$

where:

y_{ksg} = number of farms in sales class g for segment k in year s (where the groups are defined based on sales for year $s-1$).

The corresponding K2 estimators in the second through fourth post-census years are:

$$\hat{Y}_t^{(K2)} = Y_{t-1}^{(B)} \hat{b}_t , \tag{5}$$

$$\hat{Y}_{tg}^{(K2)} = Y_{(t-1)g}^{(B)} \hat{b}_t + \sum_{k \in St} w_{kt}(y_{ktg} - \hat{b}_t y_{k(t-1)g}) \tag{6}$$

where:

$Y_{t-1}^{(B)}$ = official NASS state level estimate of number of farms for year *t-1*, and

$Y_{(t-1)g}^{(B)}$ = official NASS estimate of number of farms in sales class g for year *t-1*.

Note that for the second through fourth years after a census, K2 uses more recent published figures (the previous year's official NASS estimates) as the baseline from which to compute the current year's estimates than K1 does. Both the census numbers and official NASS figures are treated as fixed (zero variance) quantities. The final term in equations (4) and (6) reflects changes in the distribution of farms across sales classes as measured by June Area Survey indications in the current and previous years. These changes net to zero when the group totals are added together. Current ratio methods compute the sales class level (but not the overall level) distributions directly from the JAS. That method is likely to have a much larger standard error than either K1 or K2.

The K1 and K2 estimators were compared with two traditional NASS indications of number of farms, namely the area frame weighted expansion estimator (AF) and the hybrid operational estimator (HYB). NASS's area sampling frame divides the area within a given state into land use strata, then subdivides each stratum into blocks (called primary sample units) with identifiable boundaries (Bush and House, 2003). The primary sampling units are further subdivided into segments of uniform size (generally one square mile for agricultural strata), with sampled segments enumerated during surveys. The area frame estimator uses all agricultural tracts (portions of a segment under a single land operating arrangement) with reported or edited sales of $1,000 or higher. As noted parenthetically above, only the fraction of a farm operation within a sampled segment is part of the farm count for that segment. The acreage values used in computing this fraction, or *tract-to-farm ratio*, include crop land, farmstead acreage, wasteland, woodland, pasture, summer fallow and idle crop land but not public, industrial or grazing association (PIGA) or nonagricultural land. Since it is mathematically equivalent to scale the expansion factor (weight) for each farm by its tract-to-farm ratio (in place of prorating the farm's contribution to the segment farm count), this is called the *weighted expansion estimator*.

The hybrid operational estimators at the state and sales class levels are defined as follows:

$$\hat{Y}_t^{(Hyb)} = Y_{t-1}^{(B)} \hat{b}_t , \tag{7}$$

$$\hat{Y}_{tg}^{(Hyb)} = Y_{t-1}^{(B)} \hat{b}_t (\hat{Y}_{tg}^{(AF)} / \hat{Y}_t^{(AF)}) \tag{8}$$

where:

$\hat{Y}_t^{(AF)}$ = area frame weighted expansion estimator of state level number of farms for year *t*, and

$\hat{Y}_{tg}^{(AF)}$ = area frame weighted expansion estimator of number of farms in sales class *g* for year *t.*

Note that the hybrid estimator at the sales class level is computed by prorating the state level hybrid estimate using group-to-state ratios between corresponding area frame estimates. At the state level (but not at the sales class level), HYB is identical to K2 except for the first post-census year (because of equations (1), (2), (5), (6), (7), and (8)).

A relatively simple way to estimate the variance (more precisely, the mean squared error) of K1 and K2 is via an extended delete-a-group jackknife (Kott, 1998). To that end, each area sample segment is placed into one of R replicate groups. When a segment leaves the sample between one year and the next, a segment from the same substratum (portion of a stratum formed by subdividing it into agriculturally similar areas) takes its place in the replicate group. Although a segment's weight can change from one year to the next based on the number of overlap segments in its substratum, its replicate group designation remains the same. The methodology is basically that described by Kott (2001), but with a slight modification to handle substrata containing one segment.

The empirical study to be discussed in Section 3 uses the state level and within-group (sales class) estimated variances of the area frame estimator, as derived from corresponding operational CVs. Based on equations (1) and (7), if *t* is the first post-census year then the state level variance of the hybrid estimator can be estimated by:

$$v(\hat{Y}_t^{(HYB)}) = [Y_{t-1}^{(B)} / Y_{t-1}^{(C)}]^2 \, v^{(DAG)}(\hat{Y}_t^{(K2)})$$

where:

$v^{(DAG)}(\hat{Y}_t^{(K2)})$ = delete-a-group jackknife variance estimate for K2 (and K1) in year *t.*

The corresponding state level variance of HYB in the second through fourth post-census years is of course identical to that of K2, and can thus be estimated directly by K2's delete-a-group jackknife variance estimate. The within-group variance of HYB is estimated via an approximation using a combination of the delete-a-group jackknife CV estimator for $\hat{b}_t$ and the operational CV estimators for $\hat{Y}_t^{(AF)}$ and $\hat{Y}_{tg}^{(AF)}$:

$$v(\hat{Y}_{tg}^{(HYB)}) = Y_{t-1}^{(B)} [R_g^{(AF)}]^2 \left\{ [CV^{(DAG)}(\hat{b}_t)]^2 + [CV^{(OP)}(\hat{Y}_{tg}^{(AF)})]^2 (1 - 2R_g^{(AF)}) + [CV^{(OP)}(\hat{Y}_t^{(AF)})]^2 \right\}$$

where:

$R_g^{(AF)} = \hat{Y}_{tg}^{(AF)} / \hat{Y}_t^{(AF)}$,

$CV^{(OP)}(\hat{Y}_t^{(AF)})$ = operational CV of state level area frame estimator for year *t,*

$CV^{(OP)}(\hat{Y}_{tg}^{(AF)})$ = operational CV of area frame estimator for sales class *g* in year *t*, and

$CV^{(DAG)}(\hat{b}_t)$ = delete-a-group jackknife CV of $\hat{b}_t$.

The key to this approximation is the simplifying assumption:

$$Cov(\hat{Y}_{tg}^{(AF)}, \hat{Y}_t^{(AF)}) \approx Var(\hat{Y}_{tg}^{(AF)}).$$

A slightly more general form for the estimated number of farms within sales classes (for both K1 and K2) is known as the *smoothed alternative*. Let λ be a variable that can take on any value between 0 and 1. If *t* is the first post-census year, then the estimators are unchanged:

$$\hat{Y}_{tg}^{(K1)}(\lambda) = \hat{Y}_{tg}^{(K2)}(\lambda) = Y_{(t-1)g}^{(C)}\hat{b}_t \, .$$

However, if *t* is the second, third or fourth post-census year then:

$$\hat{Y}_{tg}^{(K1)}(\lambda) = \hat{Y}_{(t-1)g}^{(K1)}(\lambda)\hat{b}_t + \lambda \sum_{k \in S_t} w_{kt}(y_{ktg} - \hat{b}_t y_{k(t-1)g}) \, ,$$

$$\hat{Y}_{tg}^{(K2)}(\lambda) = Y_{(t-1)g}^{(B)}\hat{b}_t + \lambda \sum_{k \in S_t} w_{kt}(y_{ktg} - \hat{b}_t y_{k(t-1)g}) \, .$$

The term λ can be regarded as a smoothing factor. The value λ = 0 forces the group level K1 and K2 estimates to be updated similarly to the state level estimates, while λ = 1 corresponds to equations (4) and (6) above. In order to deflate the impact of outliers, values of λ other than 1 can be tried (with variances again estimated using the modified delete-a-group jackknife). Section 3 includes an empirical evaluation of the smoothed alternative.

3. Results of Study

The K1 and K2 estimators were compared with the area frame weighted expansion and hybrid operational estimators at the state and sales class levels in a study covering most of the lower 48 states. Published figures on number of farms from the 2002 Census of Agriculture were projected to the years 2003-2006 using the methodology described in the previous section. While both the 2002 Census and the full 2003 JAS measure sales in 2002, the former was used to distribute farms into sales groups for 2003 since it is believed to be more accurate.

The following states were excluded from the study entirely: Arizona due to complications associated with estimated-variance computation related to the state's specialized Indian reservation strata, and Delaware, Connecticut, Massachusetts, New Hampshire, Rhode Island and Vermont for an insufficient number of sample segments. Maine and Nevada had enough segments for farm numbers computation but lacked a sufficient number for accurate jackknife variance computations. Thus, those two states were evaluated on estimation error but not on variance. South Carolina was evaluated only for 2003-04 due to a negative K2 estimate in one of the size groups for 2005 (future work will address how to deal with rare negative estimates). Six states in the study (Maine, Nevada, Maryland, New Jersey, West Virginia and Wyoming) were only required to submit estimates for two sales classes ("$1,000-9,999" and "$10,000 or more") at the time of the 2002 Census. Those states are only included in the estimator comparisons for sales class 1 (which coincides with sales class 1 in the states submitting estimates for all five classes) and all sales classes combined (the state level). States that received a new area frame during the 2004-06 period were evaluated only for the years when the old frame was still in operation. Table 1 shows the number of states used for each year of the study, in total and by sales class.

Table 1. Number of States in Study by Year

Sales Class	2003	2004	2005	2006
1	41	38	32	25
2	35	32	28	23
3	35	32	28	23
4	35	32	28	23
5	35	32	28	23
All	41	38	32	25

The variances of K1 and K2 were estimated using the delete-a-group jackknife method referred to in Section 2, with R=15 replicate groups. For each year and sales class, Tables 2 and 3 show the percentage of states where K1 and K2 had a lower estimated variance than the area frame and hybrid estimators, respectively. Table 4 gives the percentage of states where K2 had a lower estimated variance than K1 for 2004-06 (the years for which the two estimators are not identical).

Table 2. Percentage of States where K1 and K2 had a Lower Estimated Variance than AF

Sales	**2003**	**2004**		**2005**		**2006**	
Class	K1, K2	K1	K2	K1	K2	K1	K2
1	97.4	63.9	63.9	51.6	93.3	28.0	64.0
2	100.0	46.9	50.0	28.6	63.0	13.0	56.5
3	97.1	21.9	28.1	7.1	37.0	4.3	34.8
4	97.1	15.6	21.9	7.1	18.5	8.7	17.4
5	94.3	34.4	43.8	17.9	29.6	8.7	52.2
All	89.7	83.3	91.7	71.0	100.0	64.0	92.0

Table 3. Percentage of States for which K1 and K2 Had a Lower Estimated Variance than HYB (K2 Identical with HYB at State Level from 2004-06)

Sales	**2003**	**2004**		**2005**		**2006**	
Class	K1, K2	K1	K2	K1	K2	K1	K2
1	100	38.9	50.0	35.5	70.0	16.0	44.0
2	100	68.8	68.8	46.4	63.0	39.1	52.2
3	100	43.8	53.1	28.6	55.6	26.1	56.5
4	100	43.8	43.8	14.3	48.1	17.4	39.1
5	97.1	56.3	56.3	42.9	59.3	30.4	78.3
All	56.4	30.6	-	22.6	-	12.0	-

Table 4. Percentage of States for which K2 had a Lower Estimated Variance than K1

Sales Class	**2004**	**2005**	**2006**
1	58.3	83.3	80.0
2	71.9	74.1	87.0
3	84.4	85.2	87.0
4	59.4	88.9	91.3
5	62.5	70.4	87.0
All	69.4	77.4	88.0

From Table 2, the percentage of states where K1 had a lower state level estimated variance than AF decreased over years as expected, but was still fairly high (64 percent) in 2006. While K1 compared favorably at the state level with AF in all four years, the corresponding percentages for K2 in the 2004-06 period were considerably higher. The percentages for K1 decreased more sharply over years within sales classes than at the state level. In 2004 and 2005, K1 had a lower estimated variance than AF in at least 50 percent of the states tested only for sales class 1. By 2005, K1 had a lower estimated variance than AF in less than 30 percent of the states for four of the five groups. K2 had a lower estimated variance than AF in 50 percent or more of the states for only two sales classes in 2004 and 2005, but that number increased to three in 2006.

Only K1 can be compared with the hybrid estimator at the state level from 2004-06 since K2 is identical to HYB in those years. Table 3 shows that the percentage of states where K1 had a lower state level estimated variance than HYB also decreased over years, falling to 31 percent as early as 2004. K1 had a lower estimated variance than the hybrid estimator in at least 50 percent of states for only two sales classes in 2004 and none in 2005 or 2006. K2 had a lower estimated variance than HYB in at least half the states for four sales classes in 2004 and 2005, and for three classes in 2006.

Table 4 shows that K2 had a lower estimated variance than K1 in most states tested for each year/sales class combination and at the state level. The superiority of K2 over K1 in terms of estimated variance within sales classes was especially apparent for 2005 and 2006. This finding is not surprising inasmuch as the official NASS estimates used in the computation of K2 were treated as fixed quantities.

Tables 5 and 6 give the percentage of states for which K1 and K2 had a lower absolute error than AF and HYB (respectively) for each year and sales class, while Table 7 shows the percentage of states for which K2 had a lower state level absolute error than K1. Absolute error was computed as the difference between an estimate and the corresponding official NASS figure (either at the state or sales class level), where the official numbers were regarded as 'truth'.

Table 5. Percentage of States Where K1 and K2 Estimators Had a Lower Absolute Error than the Area Frame Estimator

Sales Class	2003	2004		2005		2006	
	K1, K2	K1	K2	K1	K2	K1	K2
1	78.0	73.7	84.2	56.3	71.0	48.0	64.0
2	51.4	46.9	71.9	46.4	74.1	39.1	47.8
3	74.3	68.8	65.6	71.4	74.1	69.6	73.9
4	71.4	62.5	59.4	53.6	51.9	30.4	60.9
5	85.7	75.0	65.6	53.6	74.1	43.5	73.9
All	78.0	68.4	81.6	59.4	68.8	52.0	72.0

Table 6. Percentage of States Where K1 and K2 Estimators Had a Lower Absolute Error than the Hybrid Estimator (K2 Identical with HYB at State Level from 2004-06)

Sales Class	2003	2004		2005		2006	
	K1, K2	K1	K2	K1	K2	K1	K2
1	58.5	55.3	60.5	50.0	74.2	40.0	64.0
2	54.3	37.5	68.8	42.9	59.3	43.5	52.2
3	82.9	78.1	75.0	67.9	77.8	56.5	73.9
4	80.0	65.6	68.8	50.0	48.1	21.7	39.1
5	85.7	71.9	68.8	60.7	74.1	52.2	69.6
All	51.2	36.8	-	34.4	-	32.0	-

Table 7. Percentage of States where K2 had a Lower Absolute Error than K1 (2004-06)

Sales Class	2004	2005	2006
1	68.4	61.3	80.0
2	71.9	70.4	65.2
3	53.1	55.6	60.9
4	59.4	44.4	69.6
5	56.3	66.7	65.2
All	63.2	65.6	68.0

Table 5 shows that (as with estimated variance) the percentage of states for which K1 had a lower state level absolute error than the area frame estimator was above 50 percent in all four years (and the same was true for K2). At the sales class level, K1 had a lower absolute error than AF in at least 50 percent of states for four of the classes in 2004 and 2005, but only one in 2006. The K2 estimator was equal or superior to AF (as measured by the percentage of states with lower absolute error) across every year/sales class combination except for class 2 in 2006.

From Table 6, K1 had a lower absolute error than the hybrid estimator in at least 50 percent of the states for all five sales classes in 2003, four in 2004 and 2005 and two in 2006. By contrast, K2 compared favorably with HYB in terms of absolute error for all year/sales class combinations except for class 4 in 2005 and 2006.

Table 7 shows K2 having a lower absolute error than K1 in more than 50 percent of the states for all year/sales class combinations except for class 4 in 2005. At the state level, K2 had a lower absolute error than K1 in at least 63 percent of states for all three years (2004-06) that the estimators are not identical.

The remainder of this section discusses an evaluation of the smoothed alternative described in Section 2. The goal was to find the optimal value of λ to use in a given situation. To that end, the K1 and K2 estimators and their estimated variances were computed for each state/sales class combination for values of λ between 0 and 1 at increments of 0.05. The results were used to compare estimated variance and absolute error at different values of λ with those of the original ($\lambda = 1$) estimators for 2004-06. The K1 and K2 estimators corresponding to specific values of λ will henceforth be referred to as K1(λ) and K2(λ), with the terms K1 and K2 referring to the two estimation methods in general.

Figure 1 is a set of five plots showing the percentage of states where K1(λ) had a lower estimated variance than K1(1) vs. λ in each sales class (with a separate curve for each of the three years) for the tested values of λ ranging from 0 to 0.95. Similarly, Figure 2 shows the percentage of states where K2(λ) had a lower estimated variance than K2(1) vs. λ for each sales class.

Examination of Figures 1 and 2 shows that the percentage of states where K1(λ) had a lower estimated variance than K1(1) and the percentage where K2(λ) had a lower estimated variance than K2(1) are both non-decreasing with λ for each of the three years. All curves in Figure 1 fall entirely below the 42 percent mark on the vertical axis, meaning there were no cases (with $\lambda < 1$) where K1(λ) had lower estimated variance than K1(1) in at least 42 percent of states tested. From Figure 2, there were no cases where K2(λ) had a lower estimated variance than K2(1) in at least 40 percent of the states.

Figure 3 displays the percentage of states where K1(λ) had a lower absolute error than K1(1) vs. λ for each year/sales class combination, while Figure 4 compares K2(λ) with K2(1) in similar fashion. As was the case for estimated variance, all of the curves are non-decreasing with λ. However, in general the rate of increase was more gradual and there were a number of values of λ for which K1(λ) had a lower absolute error than K1(1) (or K2(λ) had a lower absolute error than K2(1)) in the majority of states tested. For example, Figure 3 shows the K2 curve corresponding to sales class 3 in 2004 increasing from 69 percent (for $\lambda = 0$ through 0.25) to 77 percent (for $\lambda = 0.45$ through 0.95).

To gain further insight, the number of states for which specific values of λ led to the lowest estimated variance (and similarly the lowest absolute error) among all values tested was computed for each year/sales class/estimator combination. The results are shown in Tables 8 (estimated variance) and 9 (absolute error). For each year, sales class and estimator (K1(λ) or K2(λ)) within a given state, λ^* and λ^{**} denote the values of λ that minimize the estimated variance and absolute error, respectively. The third column of Tables 8 and 9 shows the mode of λ^* and λ^{**} (respectively), i.e., the value of λ that minimized the estimated variance or absolute error of K1(λ) in the most states. Percent optimal (fourth column in each table) is the percent of states for which λ^* (or λ^{**}) was the minimizing value, while the mean value of λ^* (or λ^{**}) over states is provided in the fifth column. The final three columns of the tables give the corresponding figures for K2(λ).

Table 8 shows λ=1 to be the most common minimizing value for estimated variance in 14 of the 15 cases (year/sales class combinations) for K1 and in all 15 for K2. The lone exception was sales class 1 in 2006 (for K1) where $\lambda = 0.95$ led to lowest estimated variance more often than any other value (although it was optimal in only 43.5 percent of the states tested). The mean (over states) of λ^* ranged from 0.9 to 0.99 for K1 and from 0.92 to 0.997 for K2.

The situation was very different for absolute error, as shown in Table 9. For K1, the minimizing value of λ was uniquely 0 in 11 cases and uniquely 1 in two, while there were two cases where the values 0 and 1 led to the lowest absolute error in an equal number of states. The minimizing λ for K2 was 0 in 13 cases, and evenly split between 0.6 and 1 or between 0.45 and 1 in the other two. However, there were no cases where λ=0 was optimal in at least 52 percent of the states for K1 or at least 66 percent of the states for K2. The mean of λ^{**} ranged from 0.35 to 0.57 for K1 and from 0.19 to 0.48 for K2.

The above observations reveal $\lambda = 1$ to be (somewhat surprisingly) the clear choice in terms of minimizing estimated variance. Although $\lambda = 0$ led to the lowest absolute error more often than any other tested value, Table 9 and Figures 3 and 4 suggest that the effect of λ on estimator precision is rather marginal. The overall conclusion to be drawn is that the original K1(1) and K2(1) estimators are preferable to those corresponding to lower values of λ.

4. Summary

Two new methods (called K1 and K2) for updating census estimates of number of farms using JAS data were compared with area frame and hybrid operational estimation for the years 2003-06. Using NASS official figures as 'truth' and estimated variance and absolute error as quality measures, comparisons were done both at the state level and within categories defined based on value of sales.

At the state and sales class levels, both K1 and K2 outperformed the area frame and hybrid estimators in terms of estimated variance and absolute error. A direct comparison between K1 and K2 showed the latter to be superior in the same categories. The smoothed alternatives to K1 and K2 were evaluated using values of λ between 0 and 1 at increments of 0.05. The evaluation showed that $\lambda = 1$ (corresponding to the original estimators) was the best choice.

In principle, the formula for $\hat{b}_t$ in Section 2 can be easily modified to estimate the change ratio of land in farms across adjacent years. Similarly, equations (4) and (6) can be adapted to the estimation of land in farms in June of years between censuses. These formulas treat the total in a census year as if it were for June rather than December.

For other land-related variables such as crop land and area planted to a crop, it is not difficult to modify equations (4) and (6) appropriately. However, it is unclear whether the improvement in efficiency due to incorporating census and NASS official totals into the estimation process (for K1 or K2) would be as appreciable for land-related variables as it was for farm counts since the area frame is stratified with such variables in mind. As a consequence, the potential gains from using the regression/difference type estimators defined by equations (1) through (6) may be muted.

References

Bush, J. and House, C. (2003), "The Area Frame: A Sampling Base for Agricultural Surveys", Research Report No. 93-11, US Department of Agriculture, National Agricultural Statistics Service.

Kott, P. (1998), "Using the Delete-a-Group Jackknife Variance Estimator in NASS Surveys", Research Report No. RD-98-01, US Department of Agriculture, National Agricultural Statistics Service.

Kott, P. (2001), "The Delete-a-Group Jackknife", J*ournal Of Official Statistics*, Vol. 17, No. 4, pp. 521-526.

Figure 1: Percentage of States where K1(λ) had a Lower Estimated Variance than K1(1) vs. λ

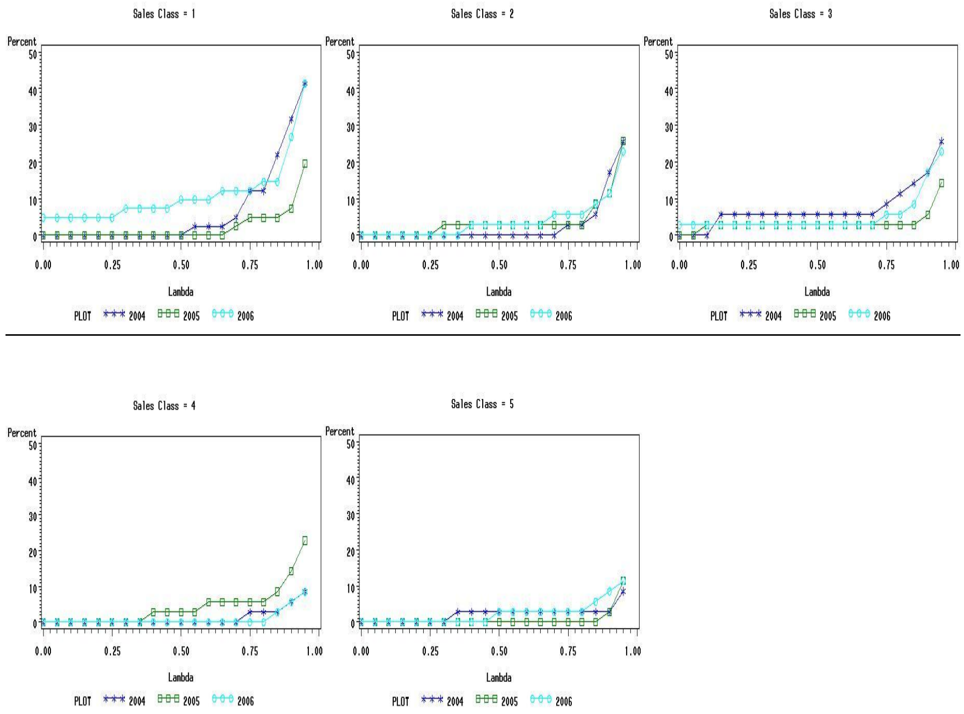

Figure 2: Percentage of States where K2(λ) had a Lower Estimated Variance than K2(1) vs. λ

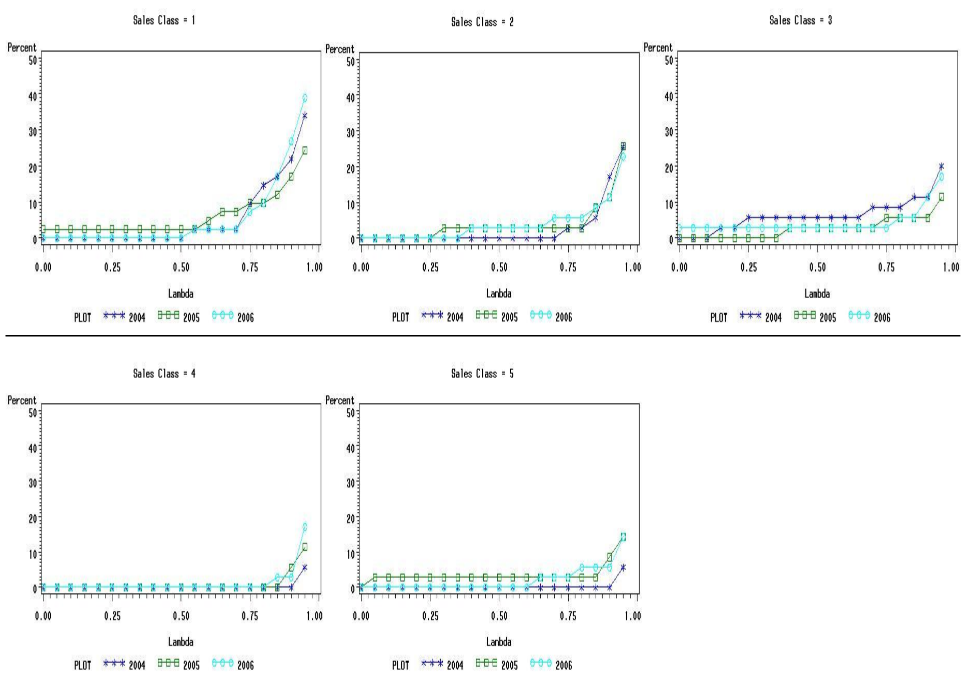

Figure 3: Percentage of States where K1(λ) had a Lower Absolute Error than K1(1) vs. λ

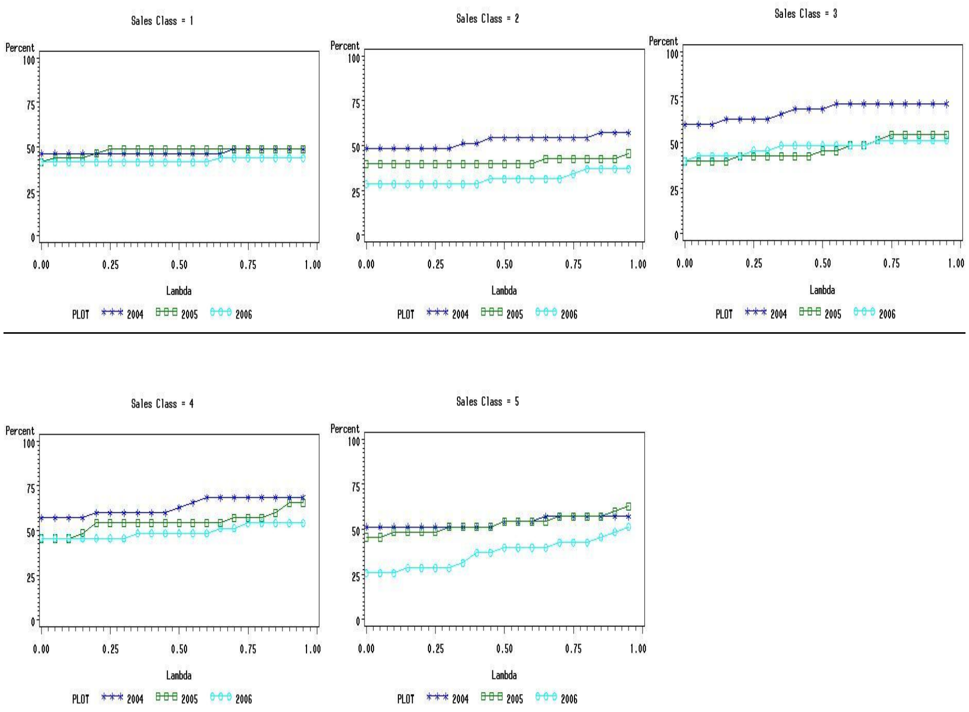

Figure 4: Percentage of States where K2(λ) had a Lower Absolute Error than K2(1) vs. λ

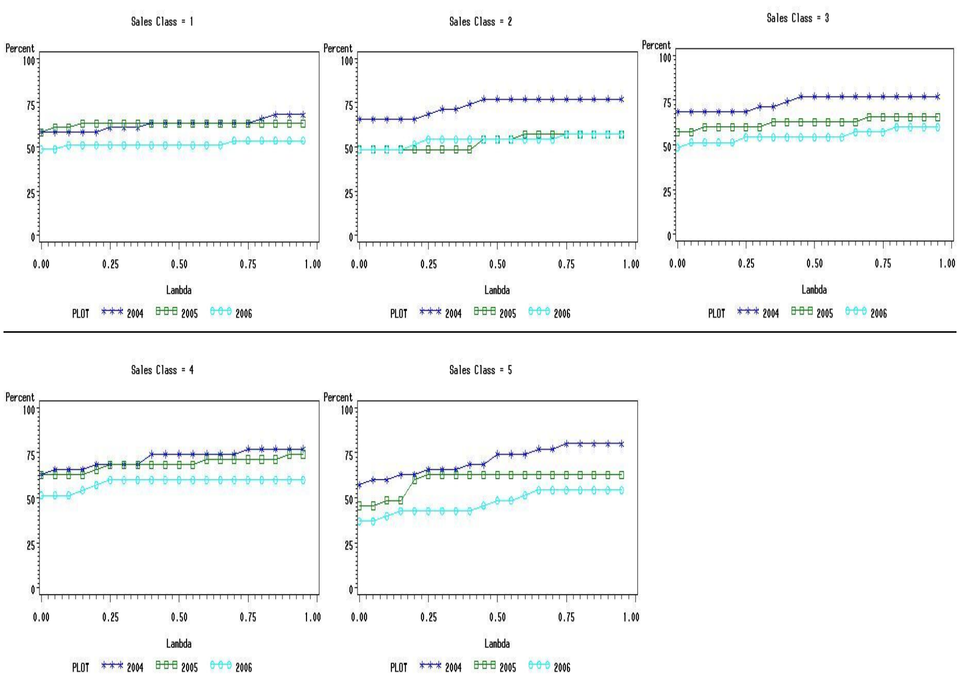

Table 8. Optimality Statistics for Estimated Variance of K1(λ) and K2(λ)

Estimator		K1(λ)			K2(λ)		
Year	Sales Class	Mode (λ^*)		Mean(λ^*)	Mode (λ^*)		Mean(λ^*)
		Value	Percent Optimal		Value	Percent Optimal	
2004	1	1	50.0	0.95	1	58.8	0.96
	2	1	71.9	0.98	1	81.3	0.99
	3	1	71.9	0.95	1	78.1	0.96
	4	1	90.6	0.99	1	93.8	0.997
	5	1	90.6	0.99	1	93.8	0.997
2005	1	1	75.9	0.98	1	69.0	0.92
	2	1	67.9	0.97	1	71.4	0.95
	3	1	82.1	0.98	1	82.1	0.96
	4	1	71.4	0.97	1	82.1	0.97
	5	1	85.7	0.99	1	78.6	0.96
2006	1	0.95	43.5	0.9	1	39.1	0.94
	2	1	65.2	0.97	1	65.2	0.98
	3	1	65.2	0.94	1	73.9	0.94
	4	1	87.0	0.99	1	73.9	0.98
	5	1	82.6	0.98	1	78.3	0.98

Table 9. Optimality Statistics for Absolute Error of K1(λ) and K2(λ)

Estimator		K1(λ)			K2(λ)		
Year	Sales Class	Mode (λ^{**})		Mean(λ^{**})	Mode (λ^{**})		Mean(λ^{**})
		Value	Percent Optimal		Value	Percent Optimal	
2004	1	0, 1	47.1	0.51	0	52.9	0.36
	2	0	46.9	0.46	0	40.6	0.32
	3	0	43.8	0.35	0	46.9	0.28
	4	0	34.4	0.43	0	37.5	0.34
	5	1	37.5	0.51	0	25.0	0.39
2005	1	0	51.7	0.41	0	65.5	0.19
	2	0, 1	42.9	0.5	0	35.7	0.4
	3	0	39.3	0.48	0	32.1	0.34
	4	0	32.1	0.42	0	46.4	0.23
	5	0	28.6	0.48	0.6, 1	17.9	0.41
2006	1	0	47.8	0.4	0	52.2	0.2
	2	1	43.5	0.55	0	47.8	0.29
	3	0	39.1	0.4	0	26.1	0.31
	4	0	43.5	0.36	0	39.1	0.23
	5	0	26.1	0.57	0.45, 1	17.4	0.48