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

1. Introduction

The USDA's National Agricultural Statistics Service (NASS) conducts the Census of Agriculture (a complete count of US farms and ranches) 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 quantities as the Census. Due to the large number of operators 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 in a state. Two related methods were developed for the purpose of updating Census estimates of number of farms. The first method (called K1) extrapolates the Census figures using JAS data only, while the second (K2) makes additional use of official NASS state level estimates of number of farms for the previous year (if it wasn't a Census year). The two estimators are identically defined for the first post-Census year.

The proposed methods were tested in a study carried out for most of the lower 48 states, both overall and within categories defined based on farm value of sales. The research involved first computing number of farms estimates using the two methods for the years 2003-06 based on data from the 2002 Census, then comparing them with area frame and hybrid operational estimates as well as official NASS figures. Variances were estimated using an extended delete-a-group jackknife method.

2. Description of Estimators

A key issue facing NASS is estimation of the number of farms within designated sales classes. In addition to state level estimates of number of farms, NASS is required to set estimates in the following classes based on total farm value of sales:

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

Since 1975, NASS has defined a farm as an agricultural operation with at least $1,000 in value of sales for a given year (with certain exceptions made for establishments not having a normal production year). There are two things to keep in mind about farm numbers from the 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 (which is all the information available from the JAS).

Traditional estimators that NASS has used for number of farms include the area frame weighted expansion (AF) and the hybrid operational estimator (HYB). NASS's area sampling frame (Bush and House, 2003) divides the area within a given state into land use strata, then subdivides each stratum into blocks (called primary sample units) with identifiable boundaries. The primary sampling units are further subdivided into segments of uniform size (generally one square mile in agricultural strata), with sampled segments being the actual areas enumerated during surveys. The AF estimator of number of farms uses positive data for all agricultural tracts (subsets of a segment corresponding to land under a single operation or management).with reported or edited sales of $1,000 or higher. The specific weight used in the expansion is the ratio between tract and farm level acres. The acreage values used to compute the weights include crop land, farmstead acreage, wasteland, woodland, pasture, summer fallow and idle crop land but not PIGA (public, industrial or grazing association) land or nonagricultural land.

The hybrid operational estimators of number of farms at the state and sales class levels in year t are defined as follows:

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

$$\hat{Y}_{tg}^{(Hyb)} = Y_{t-1}^{(B)} \hat{b}_t (\hat{Y}_{tg}^{(AF)} / \hat{Y}_t^{(AF)}) \qquad \text{(g=1,…,5)},$$

where:

g = sales class

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

$\hat{Y}_t^{(AF)}$ = 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 category *g* for year *t.*

In the first post-Census year, the new estimators K1 and K2 of total number of farms in a state and within sales classes are identical:

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

$$\hat{Y}_{tg}^{(K1)} = \hat{Y}_{tg}^{(K2)} = \hat{Y}_{(t-1)g}^{(C)} \hat{b}_t \ , \tag{3}$$

where:

$\hat{Y}_{t-1}^{(C)}$ = coverage-adjusted Census estimate of state level number of farms for year *t-1* (the Census year),

$\hat{Y}_{(t-1)g}^{(C)}$ = coverage-adjusted Census estimate of number of farms in category *g* for year *t-1*,

$$\hat{b}_t = \sum_{S_t} w_{kt} y_{kt} / \sum_{S_t} w_{k(t-1)} y_{k(t-1)} \qquad \text{(change ratio)},$$

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

y_{ks} = number of farms in segment k for year s (tract-to-farm weighted count), and

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

For the second through fourth post-Census years, the K1 estimators of number of farms at the state and sales class levels are as follows:

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

$$\hat{Y}_{tg}^{(K1)} = \hat{Y}_{(t-1)g}^{(K1)} \hat{b}_t + \sum_{S_t} w_k\left(y_{ktg} - \hat{b}_t y_{k(t-1)g}\right), \tag{4}$$

where:

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

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

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

$$\hat{Y}_{tg}^{(K2)} = Y_{(t-1)g}^{(B)} + \sum_{S_t} w_k\left(y_{ktg} - \hat{b}_t y_{k(t-1)g}\right), \tag{6}$$

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.

By substituting the previous year's official NASS estimate for the corresponding updated Census figure, the K2 method provides a timelier baseline from which to compute the current year's estimate. The final term in equations (4) and (6) reflects changes in the distribution of farms across sales class as measured by June Agricultural 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 size-group distributions (but not the overall level) directly from the JAS. That method is likely to have a much larger standard error than K1 or K2. Note that at the state level (but not the sales class level), K2 is identical to the hybrid operational estimator except for the first post-Census year (because of equations (1), (2) and (5)).

A relatively simple way to estimate the variance (more precisely, the mean squared error) of the proposed estimators of number of farms is via an extended delete-a-group jackknife (Kott, 1998). To that end, each sample segment is placed into one of R replicate groups. When a segment leaves the area sample between one year and the next, a segment from the same substratum (subdivision of a stratum used in the stratification process) 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 value R=15 was used for the estimated-variance computations in this study. K1 and K2 were compared with AF and HYB.

The state-level and within-group estimated variances of the area frame estimator were derived from the corresponding operational CVs. The state-level estimated variance of the hybrid estimator is of course identical to that of K2, while its within-group variance 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 $\hat{Y}_{tg}^{(AF)}$:

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

where:

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

$CV^{(DAG)}(\hat{b}_t)$ = delete-a-group jackknife CV of change ratio for year *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. If *t* is the first post-Census year, then the estimators are unchanged:

$$\hat{Y}_{tg}^{(K1)}(\lambda) = \hat{Y}_{tg}^{(K2)}(\lambda) = \hat{Y}_{tg}^{(K1)} \quad \text{(from equation (3))} \qquad \text{where } 0 \le \lambda \le 1.$$

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

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

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

The term λ can be regarded as a smoothing factor. The value $\lambda = 0$ forces K1's sales class level estimates to be updated similarly to the state level estimates and K2's to match the previous year's official figures, while $\lambda = 1$ 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 4 describes a detailed investigation of the smoothed alternative.

3. Results of Estimator Comparison

Table 1 shows the number of states used for each year in the study (2003-06), both overall and by sales class. 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. Arizona was not included in the study due to complications associated with an estimated variance computation for specialized Indian reservation strata. Since the area frames for Delaware, Connecticut, Massachusetts, New Hampshire, Rhode Island and Vermont contained fewer than ten sample units for each year in the study, those states were excluded as well. Maine and Nevada had a sufficient number of sample units to compute the number of farms estimates but not enough for accurate jackknife variance computations, so the estimators in those two states were only evaluated for estimation accuracy (not variance).

There are six states in the study (Maine, Nevada, Maryland, New Jersey, West Virginia and Wyoming) that were only submitting 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 class 1 in the states submitting estimates for all five classes) and all sales classes combined.

Table 1. Number of States Used For Each 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

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 AF and hybrid estimators, respectively. Table 4 shows the percentage of states where K2 had a lower estimated variance than K1 for 2004-06 (the three years for which the two estimators are not identically defined).

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 with AF in all four years, the corresponding percentages for K2 were considerably higher (especially for the last two years).

The percentages for K1 decreased more sharply over years within sales classes than at the state level. In 2004 and 2005, only sales class 1 showed K1 with lower estimated variance than AF in at least 50 percent of the states tested. By 2005, K1 had lower estimated variance than AF in less than 30 percent of states for four of the five classes. The corresponding percentages for K2 were 50 or higher for just two of the classes in 2004 and 2005, but increased to three classes 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 lower state-level estimated variance than HYB also decreased over years and was down to 31 as early as 2004. At the sales class level, 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, whereas K2 compared favorably with HYB for four classes in 2004 and 2005 and three 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 in 2005 and 2006. This finding is not surprising inasmuch as the official NASS estimates used in the computation of K2 are treated as fixed (zero variance) quantities.

Table 2. Percentage of States Where K1 and K2 had Lower Estimated Variance than AF Estimator

Sales Class	**2003**	**2004**		**2005**		**2006**	
	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

* - K1 and K2 identical for first post-Census year

Table 3. Percentage of States Where K1 and K2 Had Lower Estimated Variance than Hybrid Estimator

Sales Class	**2003***	**2004**		**2005**		**2006**	
	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	_

**- K2 identical with hybrid estimator for 'All' after first post-Census year

Table 4. Percentage of States Where K2 had Lower Estimated Variance than K1 (2004-06)

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

Tables 5 through 7 regard the official NASS figures as truth and thus as a means of measuring the absolute error of an estimator. Tables 5 and 6 display the percentage of states where K1 and K2 had a lower absolute error than AF and HYB (respectively) for each year and sales class, while Table 7 gives the percentage of states where K2 had a lower absolute error than K1. Absolute error is computed as the difference between an estimate generated by a particular method and the corresponding NASS official figure (either at the state or class level).

Table 5 shows that (as with estimated variance), the percentage of states where K1 had a lower state level absolute error than the area frame estimator was above 50 for all four years (and the same was true for K2). At the sales class level, K1 was better than AF in terms of absolute error for four of the classes in 2004 and 2005 but only one in 2006. There was only one year/sales class combination for which K2 had lower absolute error than AF in fewer than half of the states (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 18 of the 20 year/sales class combinations (with the exceptions being class 4 in 2005 and 2006).

Table 7 shows that K2 had a lower absolute error than K1 in more than 50 percent of the states for 14 of the 15 year/sales class combinations, and lower state level absolute error for all three years (2004-06) where the two estimators are not identically defined.

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

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	_

*- K2 identical with hybrid estimator for 'All' after first post-Census year

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

4. Evaluation of the Smoothed Alternative

This section discusses results of an empirical study of the smoothed alternative described in Section 1. The goal is 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 in the study for values of λ between 0 and 0.95 at increments of 0.05. The results were used to generate plots comparing 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 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 percent of states where K1(λ) had lower estimated variance than K1(1) vs. λ in each sales class (with a separate curve for each of the three years 2004-06) for tested values of λ ranging from 0 to 0.95. Similarly, Figure 2 shows percent of states where K2 (λ) had lower estimated variance than K2 (1) vs. λ for each sales class.

Examination of Figures 1 and 2 shows that the percent of states where K1 (λ) had lower estimated variance than K1(1) and the percent where K2(λ) had lower estimated variance than K2(1) are both non-decreasing with λ for each of the three years (2004-06) tested. All five 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. Figure 2 shows that there were no cases where K2(λ) had lower estimated variance than K2(1) in at least 40 percent of states.

Figure 3 is a set of five graphs showing percent of states where K1(λ) had 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 manner. As was the case for estimated variance, all of the curves are non-decreasing with λ. However, the rate of increase was more gradual in general and there were a number of instances where K1(λ) had lower absolute error than K1(1) or K2(λ) had lower absolute error than K2(1) in the majority of states tested. For example, Figure 4 shows the K2 curve corresponding to sales class 3 in 2004 increasing from 69 percent (for $\lambda = 0$ through 0.25) to 71 percent (for $\lambda = 0.3$ and 0.35), 74 percent (for $\lambda = 0.4$) and 77 percent (for $\lambda = 0.45$ and higher).

To gain additional insight, the number of states for which specific values of λ led to lowest estimated variance and lowest absolute error (respectively) among all values tested was computed for each year/sales class/estimator combination. The results are shown in Tables 8 (for estimated variance) and 9 (for absolute error). For each year, sales class and estimator (K1(λ) or K2(λ)) within a given state, the value of λ minimizing estimated variance is referred to as λ^* and the one minimizing absolute error as λ^{**}. The third column of Table 9 shows the mode of λ^*, i.e., the value of λ that minimized the estimated variance of K1(λ) in the most states. Similarly, the third column of Table 10 gives the mode of λ^{**} (value or values of λ that minimized estimated absolute error of K1(λ) in the most states). The fourth column in each table (percent optimal) shows the percent of states for which λ^* (or λ^{**}) was the minimizing value. The fifth column indicates the mean value of λ^* (or λ^{**}) over states. Columns six through eight provide the corresponding figures for K2(λ).

Table 8 shows that $\lambda = 1$ was the most common minimizing value for estimated variance in 14 of the 15 year/sales class combinations for K1 and 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 (but was only optimal in 43.5 percent of states tested). The percent of states for which $\lambda = 1$ was optimal ranged from 50 to 90.6 for K1 and 58.8 to 93.8 for K2, while the mean (over states) of λ^* ranged from 0.9 to 0.99 for K1 and 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 0 for eleven year/sales class combinations and 1 for only two, while there were two cases where the values 0 and 1 led to 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 48 percent of states for K1 or at least 66 percent of 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 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 and K2 estimators corresponding to $\lambda = 1$ are preferable to those corresponding to lower values of λ.

5. Summary

Two proposed 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 2003-06 period with regard to estimated variance and absolute error in most states. The 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 were found to outperform the area frame and hybrid estimator in terms of having lower 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 both K1 and K2 were evaluated using values of λ between 0 and 1 at increments of 0.05, with $\lambda = 1$ (corresponding to the original estimators) found to be the best choice.

For land-related variables such as crop land and area planted in a crop, it is a simple matter to modify equations (4) and (6) appropriately. However, whether the improvement in efficiency due to incorporating Census and official figures into the estimation process (for K1 or K2) would be as appreciable for land-related variables as farm counts is unclear since area frame sample stratification is carried out with such variables in mind. As a consequence, the potential gains from using the regression/difference type estimators in (4) and (6) may be muted. Future research could address this question.

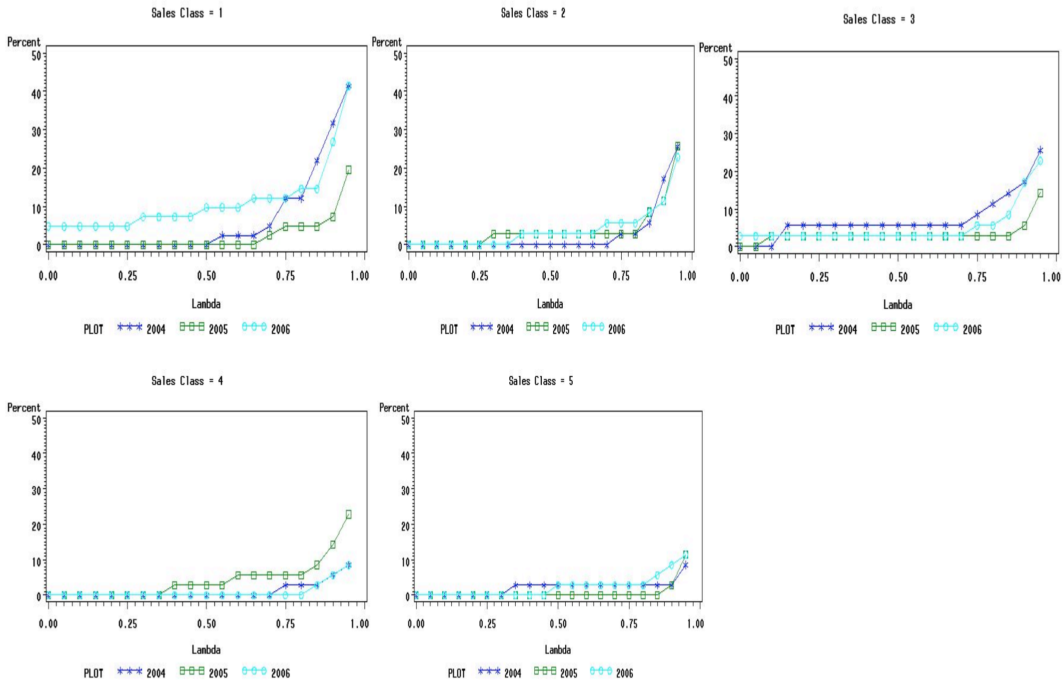
References

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

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.

<u>Figure 1</u>: Percent of States where K1(λ) had Lower Variance than K1(1) vs. λ



<u>Figure 2</u>: Percent of States where K2(λ) had Lower Variance than K2(1) vs. λ

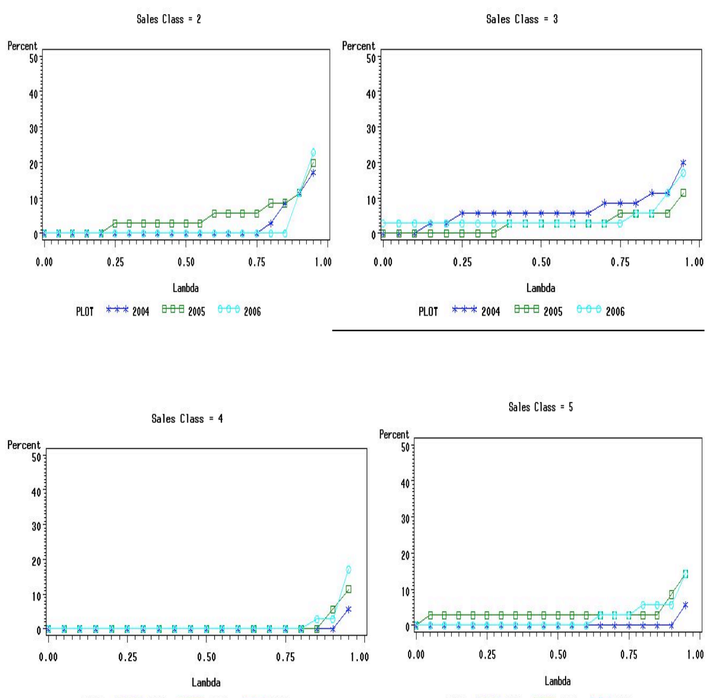


Figure 3: Percent of States where K1(λ) had Lower Absolute Error than K1(1) vs. λ

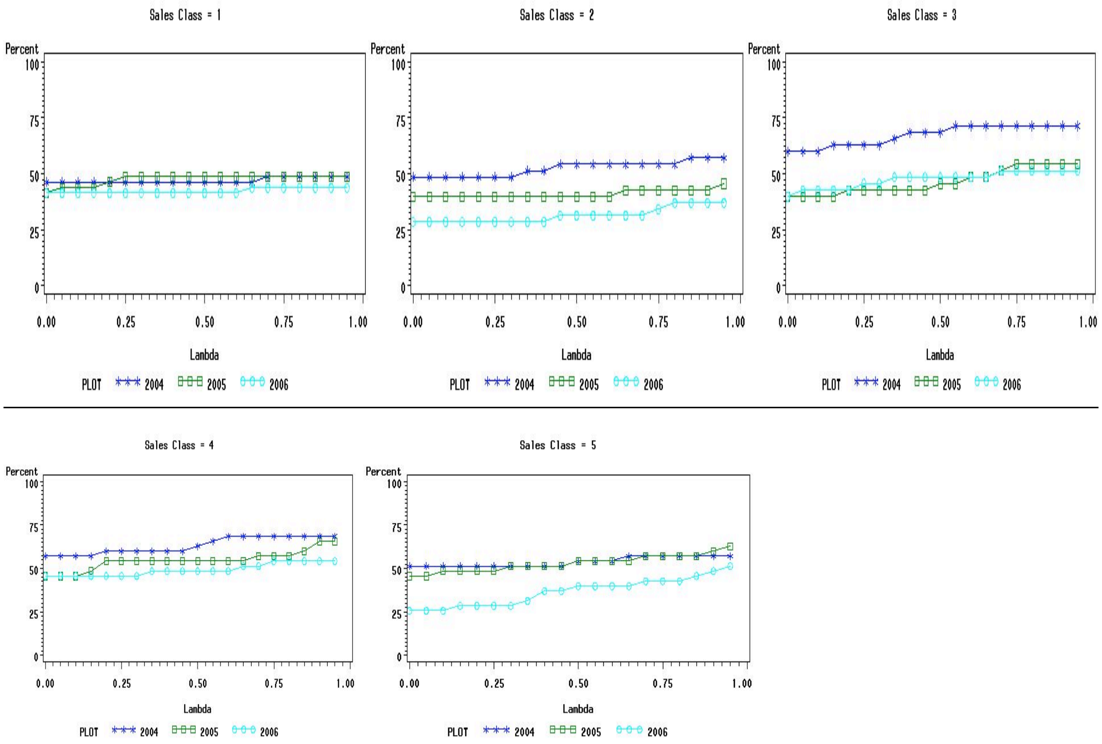


Figure 4: Percent of States where K2(λ) had Lower Absolute Error than K2(1) vs. λ

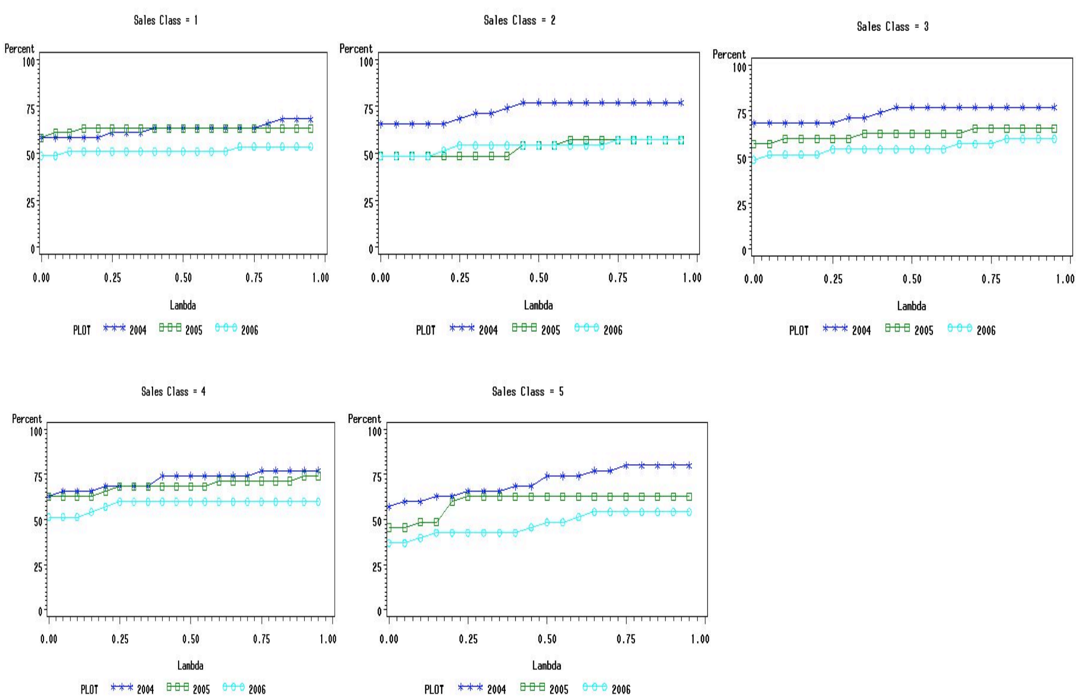


Table 8. Optimality Statistics for Estimated Variance of K1(λ) and K2(λ)

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