United States
Department of
Agriculture

National
Agricultural
Statistics
Service

Research and
Development Division
Washington DC 20250

RDD Research Report
RDD-10-02

November 2010

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

Michael E. Bellow
Phillip S. Kott

EXECUTIVE SUMMARY

The 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. Since the number of operations from which data are obtained in the census is so much larger than in the Agency's sample surveys, and since the census provides complete coverage of US farms, its numbers for comparable characteristics are often considered more accurate than those derived from the 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 subsequent non-census years.

Due to its relative stability over time, the survey 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) updates the census figures using JAS data only, while the second (K2) additionally makes use of previous, non-census year Agricultural Statistics Board (ASB) estimates.

The two methods were evaluated in a study carried out for most of the lower 48 states, based on estimation accuracy both at the state level and by farm value of sales categories within a state. The research involved computing number of farms estimates for the years 2003-06 based on data from the 2002 Census and comparing them with the area frame and hybrid operational estimators as well as published ASB estimates. Variances were estimated using an extended delete-a-group jackknife method. The criteria for evaluating the methods were estimation error (regarding the ASB figures as 'truth') and variance.

Both K1 and K2 were found to be superior to the area frame and hybrid estimators with regard to the two efficiency criteria, with K2 appreciably better than K1 at the state level as well as within sales classes.

RECOMMENDATIONS

Based on results of the study, the following recommendations are made:

1. Implement the K2 estimator (which makes use of previous year official estimates) in NASS's procedure for estimating number of farms, at both the state level and by sales class within a state.

2. Explore the potential use of the K2 (and possibly K1) estimation methodologies to improve the indications of other agricultural characteristics estimated from surveys.

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

By Michael E. Bellow and Phillip S. Kott[1]

Abstract

The National Agricultural Statistics Service (NASS) conducts the census of agriculture (a complete count of US farms and ranches) every five years. Sample surveys, including the June area survey (JAS), are carried out annually 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 survey item considered most likely to benefit from such an approach is number of farms in a state.

Two proposed methods for projecting census counts of number of farms to subsequent non-census years are evaluated. The first method updates the census figure 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 (if it wasn't a census year). The two methods are identical for the first post-census year. The proposed estimators are compared with area frame based and hybrid operational estimators for the years 2003-06 in a study covering most of the lower 48 states, both at the state level and within categories defined by farm value of sales. Variances are estimated using an extended delete-a-group jackknife method.

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

[1] Michael E. Bellow is a mathematical statistician in the Research and Development Division of NASS, Room 305, 3251 Old Lee Highway, Fairfax, VA 22030. Phillip S. Kott is a senior research statistician at RTI International, 6110 Executive Blvd., Suite 902, Rockville, MD 20852

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. This report introduces two methods to update census of agriculture based farm number estimates for years between censuses, using June area survey (JAS) data 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 methodology may be applicable to other characteristics estimated from the surveys as well. The two estimators 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. Certain government agricultural program payments received are included in the category of 'sales'. Specifically, what qualifies as agricultural *activity* has undergone some revision over recent years. For example, operations with five or more horses or ponies and no agricultural sales were added to the number of farms estimates as horse farms in 1995. Two new industries - maple syrup and short-rotation woody-crop places - were added as agricultural activities in 1997 due to the new North American Industry Classification System (NAICS). After a review of census and annual survey procedures in 2002, JAS procedures were modified to include establishments having 100 acres of pasture land but less than $1,000 of annual agricultural sales as farms. These changes were made to establish comparability between census of agriculture numbers and annual published estimates for the purpose of providing a consistent data series.

A key issue facing NASS is estimation of the number of farms within designated sales classes for a given state. In addition to state level estimates, NASS field offices (FOs) are also required to categorize farm numbers based on the following total farm value of sales categories:

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 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 would normally have sold at least $1,000 worth of agricultural products), and 2) the number of farms for a given year is divided into groups based on sales from the previous year (which is all that is available from the JAS).

While apportioning coverage-adjusted census numbers into sales classes using the JAS size definitions for farms is a simple matter, they're not tabulated that way in census publications.

2. METHODS

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:

$$\hat{Y}_t^{(K1)} = \hat{Y}_t^{(K2)} = Y_{t-1}^{(C)}\hat{b}_t , \tag{2.1}$$

$$\hat{Y}_{tg}^{(K1)} = \hat{Y}_{tg}^{(K2)} = Y_{(t-1)g}^{(C)}\hat{b}_t \tag{2.2}$$

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* (the census year),

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

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

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 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 from year *t-1* to year t. 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:

$$\hat{Y}_t^{(K1)} = \hat{Y}_{t-1}^{(K1)} \hat{b}_t \ , \tag{2.3}$$

$$\hat{Y}_{tg}^{(K1)} = \hat{Y}_{(t-1)g}^{(K1)} \hat{b}_t + \sum_{k \in St} w_{kt} (y_{ktg} - \hat{b}_t y_{k(t-1)g}) \tag{2.4}$$

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:

$$\hat{Y}_t^{(K2)} = Y_{(t-1)}^{(B)} \hat{b}_t \ , \tag{2.5}$$

$$\hat{Y}_{tg}^{(K2)} = Y_{(t-1)g}^{(B)} \hat{b}_t + \sum_{k \in St} w_{kt} (y_{ktg} - \hat{b}_t y_{k(t-1)g}) \tag{2.6}$$

where:

$Y_{t-1}^{(B)} =$ Agricultural Statistics Board (ASB) state level estimate of number of farms for year *t-1*, and

$Y_{(t-1)g}^{(B)} =$ ASB 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 ASB figures are treated as fixed (zero variance) quantities. The final term in equations (2.4) and (2.6) reflects changes in the distribution of farms across sales class as measured by JAS 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 June survey. 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 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 in 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 agricultural activity) that have reported or edited sales of $1,000 or higher. 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, or 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:

$$\hat{Y}_t^{(Hyb)} = Y_{t-1}^{(B)} \hat{b}_t , \tag{2.7}$$

$$\hat{Y}_{tg}^{(Hyb)} = Y_{t-1}^{(B)} \hat{b}_t \left\{ \hat{Y}_{tg}^{(AF)} / \hat{Y}_t^{(AF)} \right\} \tag{2.8}$$

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 of the 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 (2.1), (2.2) and (2.5) through (2.8)).

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 (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 year to year based on the number of 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. Appendix A describes the extended delete-a-group jackknife in detail.

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 the corresponding operational CVs. Based on equations (2.1) and (2.7), 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 thus can 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 $\hat{Y}_{tg}^{(AF)}$:

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

where:

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

$CV^{(OP)}(\hat{Y}_t^{(AF)})$ = 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:

$$\hat{Y}_{tg}^{(K1)}(\lambda) = \hat{Y}_{(t-1)g}^{(K1)}(\lambda)\hat{b}_t + \lambda \sum_{k \in St} w_{kt}(y_{ktg} - \hat{b}_t y_{k(t-1)g}) \ , \tag{2.9}$$

$$\hat{Y}_{tg}^{(K2)}(\lambda) = Y_{(t-1)g}^{(B)}\hat{b}_t + \lambda \sum_{k \in St} w_{kt}(y_{ktg} - \hat{b}_t y_{k(t-1)g}) \ , \tag{2.10}$$

The term λ can be regarded as a smoothing factor. The value $\lambda = 0$ forces the group level K1 and K2 estimates to be updated similarly to the state level estimates, while $\lambda = 1$ corresponds to equations (2.4) and (2.6) above. In order to deflate the impact of outliers, values other than 1 can be tried (with variances again estimated using the modified delete-a-group jackknife). Section 3 includes discussion of a detailed evaluation of the smoothed alternative.

3. RESULTS

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 an empirical 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 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. There were six states in the study (Maine, Nevada, Maryland, New Jersey, West Virginia and Wyoming) which 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 use. Table 1B (in Appendix B) gives the specific years for which each state was included in the study as well as the total number of states for each year.

The variances of K1 and K2 were estimated using the delete-a-group jackknife method described in Appendix A, with R=15 replicate groups. For each year and sales class, Tables 1 and 2 show the percentage of states for which K1 had a lower estimated variance than the area frame and hybrid estimators, respectively. Table 3 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).

A negative K2 estimate for 2005 was encountered for sales class 4 in South Carolina, while sales class 3 in Utah had negative K1 and K2 estimates for 2004. For that reason, none of South Carolina's sales class level K2 estimates for 2005 or Utah's sales class level K1 and K2 estimates for 2004 were used in the computation of summary statistics (over all states tested) provided in this section. In addition, Utah's sales class level K1 estimates for 2005-06 were excluded since the current year's K1 estimate for a given class is computed based on that of the previous year (which is not the case for K2). Future work will address how to deal with occasional negative estimates.

Table 1. Percentage of States for which K1 and K2 had a Lower Estimated Variance than AF

Sales Class	**2003**	**2004**		**2005**		**2006**	
	K1, K2	K1	K2	K1	K2	K1	K2
1	97.4	63.9	63.9	51.6	93.5	28.0	61.5
2	100.0	46.9	50.0	28.6	64.3	13.0	54.2
3	97.1	21.9	28.1	7.1	39.3	4.3	37.5
4	97.1	15.6	21.9	7.1	17.9	8.7	20.8
5	94.3	34.4	43.8	17.9	28.6	8.7	50.0
All	89.7	83.8	91.9	71.9	100.0	61.5	88.5

Table 2. Percentage of States for which K1 and K2 had a Lower Estimated Variance than HYB (K2 Identical with HYB at State Level)

Sales Class	**2003**	**2004**		**2005**		**2006**	
	K1, K2	K1	K2	K1	K2	K1	K2
1	100	38.9	50.0	35.5	67.7	16.0	46.2
2	100	68.8	68.8	46.4	64.3	39.1	54.2
3	100	43.8	53.1	28.6	57.1	26.1	58.3
4	100	43.8	43.8	14.3	50.0	17.4	41.7
5	97.1	56.3	56.3	42.9	57.1	30.4	79.2
All	56.4	32.4	-	25.0	-	15.4	-

Table 3. 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	67.6	75.0	84.6

Examination of Table 1 reveals that 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 (61.5 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's estimated variance was lower than that of AF in 50 percent or more of the states for only two sales classes in 2004 and 2005, but that number increased to three classes in 2006.

Only K1 could be compared with the hybrid estimator at the state level from 2004-06 since K2 was identical to it in those years. Table 2 shows that the percentage of states where K1 had a lower state level estimated variance than HYB also decreased over years, falling to 32 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. Alternatively, K2 had a lower estimated variance than HYB in at least half the states for four sales classes in 2004, all five classes in 2005 and three classes in 2006.

From Table 3, K2 had a lower estimated variance than K1 in at least 58 percent of the 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 quantities.

Tables 4 and 5 show 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 6 shows the percentage of states for which K2 had a lower absolute error than K1. Absolute error was computed as the difference between an estimate and the corresponding ASB figure (either at the state or sales class level), where the official numbers were regarded as 'truth'.

Table 4. Percentage of States for which K1 and K2 had a Lower Absolute Error than AF

Sales Class	**2003**	**2004**		**2005**		**2006**	
	K1, K2	K1	K2	K1	K2	K1	K2
1	78.0	73.7	84.2	56.3	68.8	48.0	65.4
2	51.4	46.9	71.9	46.4	75.0	39.1	45.8
3	74.3	68.8	65.6	71.4	75.0	69.6	75.0
4	71.4	62.5	59.4	53.6	50.0	30.4	58.3
5	85.7	75.0	65.6	53.6	71.4	43.5	75.0
All	78.0	69.2	79.5	60.6	69.7	50.0	69.2

Table 5. Percentage of States for which K1 and K2 had a Lower Absolute Error than HYB (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	71.9	40.0	65.4
2	54.3	37.5	68.8	42.9	60.7	43.5	50.0
3	82.9	78.1	75.0	67.9	78.6	56.5	75.0
4	80.0	65.6	68.8	50.0	46.4	21.7	37.5
5	85.7	71.9	68.8	60.7	71.4	52.2	70.8
All	51.2	38.5	-	36.4	-	30.8	-

Table 6. Percentage of States for which K2 had a Lower Absolute Error than K1

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	61.5	63.6	69.2

Table 4 reveals 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 at least 50 percent in all four years, while the corresponding percentage for K2 was always above 69 percent. At the sales class level, K1 had a lower absolute error than AF in most states tested for all five classes in 2003 and for four of the classes in 2004 and 2005, but for only one class 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 5, 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, for four classes in 2004 and 2005 and for two classes in 2006. By contrast, K2 was at least as good as HYB in terms of absolute error for all year/sales class combinations except for class 4 in 2005 and 2006.

Table 6 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 61 percent of the states for all three years (2004-06) where the estimators were not identical.

Tables 2B and 3B (in Appendix B) give the mean absolute relative error (MARE) and mean squared relative error (MSRE) of K1, K2, AF and HYB at the state and sales class levels over all states tested. Relative error is defined as the ratio of a state's estimation error to the corresponding ASB estimate of number of farms. Within each sales class, the four estimators were ranked from 1 (best) to 4 (worst) for both metrics in each year (except for 2003 where K1 and K2 were regarded as the same estimator so the largest rank was 3). At the state level, estimators were ranked from 1 to 3 in all years since K2 is identical with K1 for 2003 and identical with HYB for 2004-06. The sales class level rankings were averaged over the five groups. Table 7 gives the average sales class level ranks as well as the state level ranks.

Table 7 shows that K2 had a smaller (i.e., better) average rank at the sales class level than HYB and AF with respect to both MARE and MSRE in all four years. K1's average rank for both metrics was smaller than that of both HYB and AF in 2003 and 2004, but larger (or at best equal to) that of HYB and AF in 2005 and 2006. At the state level, K2 had the smallest (best) rank with respect to both metrics among the four estimators from 2004-06.

Figures 1B through 4B (in Appendix B) are sets of vertical bar charts showing the estimated relative efficiency (RE) of K2 with respect to the AF estimator for each state in the study, where RE was computed as the ratio of the estimated variance of AF to that of K2. For ease of interpretation, separate bar charts are provided for three regions of the US - eastern, central and western. For each state, years are indicated on the horizontal axis by their fourth digit (3, 4, 5 or 6). The scale of the vertical axis is limited to 0 through 30 on each chart, with values of the RE exceeding 30 displayed numerically at the top right of the corresponding bars. The charts in Figures 1B through 3B are by sales class within states, while those in Figure 4B are at the state level.

The remainder of this section discusses an evaluation of the smoothed alternative described in Section 2. The generalized K1 and K2 estimates 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). Estimated variances and absolute errors at different values of λ were compared with those of the original $(\lambda = 1)$ estimators for the years 2004-06. The estimators corresponding to specific values of λ will henceforth be denoted by K1(λ) and K2(λ), with the terms K1 and K2 referring to the two estimation methods in general.

Table 7. Rank Based Comparison of Estimators Based on Two Accuracy Measures (MARE = mean absolute relative error, MSRE = mean squared relative error)

Year	**Estimator**	**Average Rank (Sales Class Level)**		**Rank (State Level)**	
		MARE	MSRE	MARE	MSRE
2003	K1, K2	1	1	2	2
	HYB	2.6	2.6	1	1
	AF	2.4	2.4	3	3
2004	K1	2.2	2.6	2	2
	K2	1.2	1.6	1	1
	HYB	3.4	3.0		
	AF	3.2	2.8	3	3
2005	K1	3	3.2	2	2
	K2	1.6	1.6	1	1
	HYB	2.4	2.2		
	AF	3	3	3	3
2006	K1	3.2	3.2	2	2
	K2	1.8	1.8	1	1
	HYB	2	2.4		
	AF	3	2.6	3	3

Figure 1 is a set of five plots showing the percentage of states for which 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. Note that all curves in each plot are non-decreasing with λ. There are only two data points exceeding 50 percent on the vertical axis, both corresponding to sales class 1 in 2006 where K1(0.95) had a lower estimated variance than K1(1) in 68 percent of the states tested and K2(0.95) had a lower estimated variance than K2(1) in 62 percent of the states.

Figure 3 is a set of five plots showing 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 75 percent (for $\lambda = 0$ through 0.25) to 84 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 given 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 columns of the two tables show 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 defined as 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 sixth through eighth columns of Tables 8 and 9 show the corresponding figures for K2(λ).

Table 8 shows $\lambda = 1$ 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 cases for K2. The lone exception was sales class 1 in 2006 (for K1) where $\lambda = 0.95$ led to the lowest estimated variance more often than any other value (although it was only optimal in 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 0.99 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 twelve cases and uniquely 1 in one, 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 uniquely 0 in 13 cases and evenly split between 0.6 and 1 or between 0.45 and 1 in the other two cases. However, there were no cases where $\lambda = 0$ 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 value tested, 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.

Figure 1: Percentage of States for which K1(λ) had a Lower Estimated Variance than K1(1) for Various Values of λ

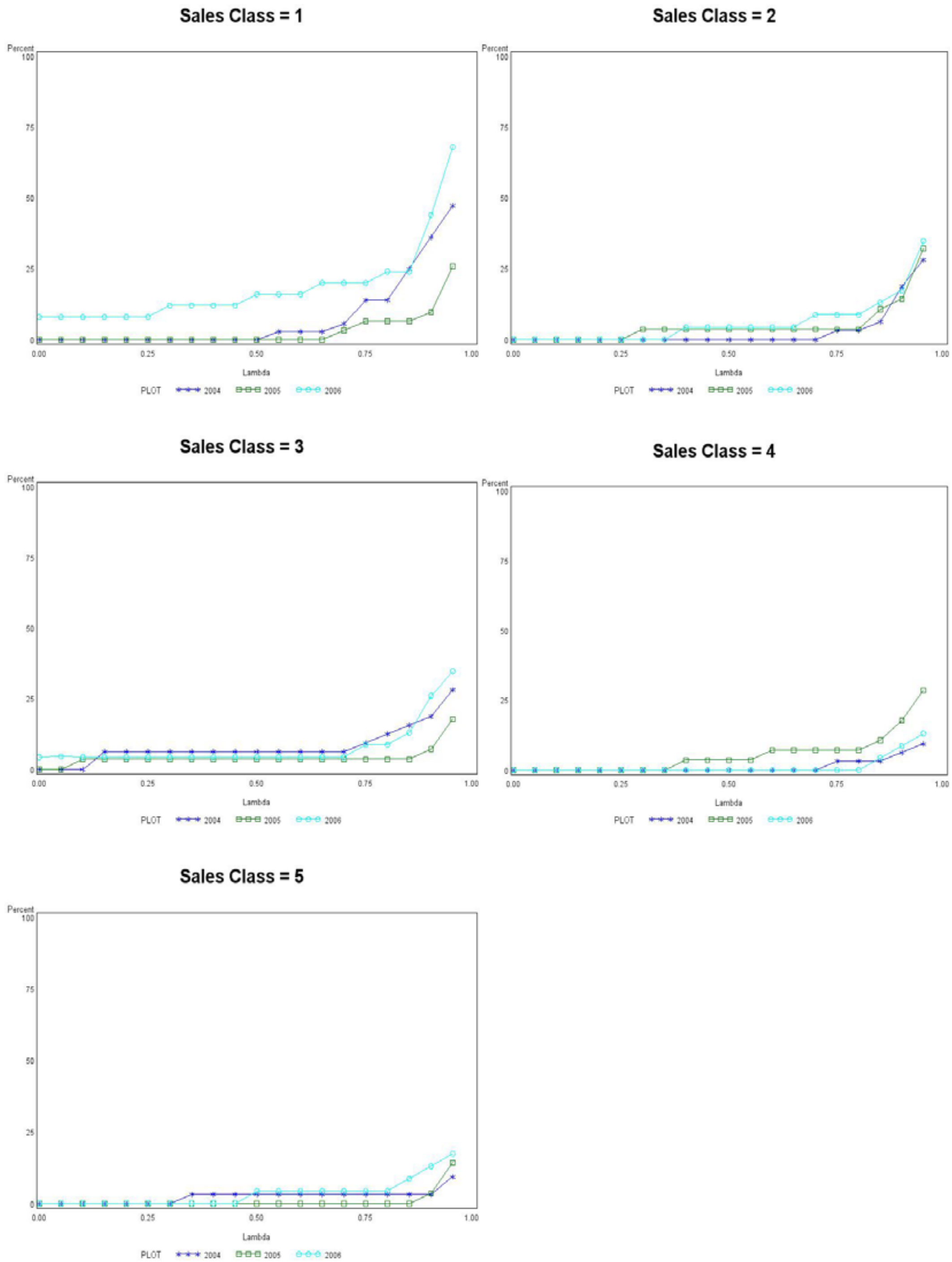


Figure 2: Percentage of States for which $K2(\lambda)$ had a Lower Estimated Variance than $K2(1)$ for Various Values of λ

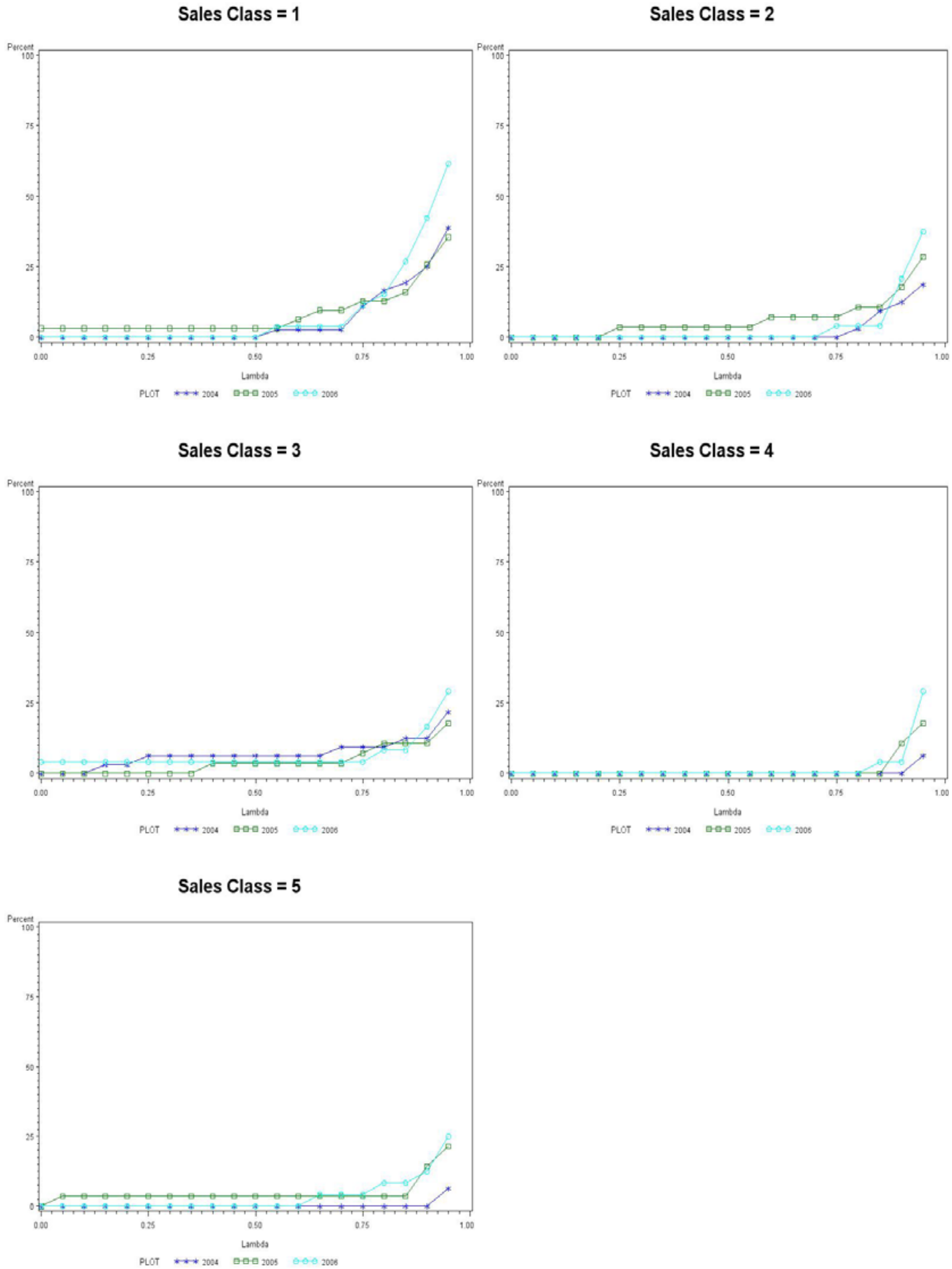


Figure 3: Percentage of States for which K1(λ) had a Lower Absolute Error than K1(1) for Various Values of λ

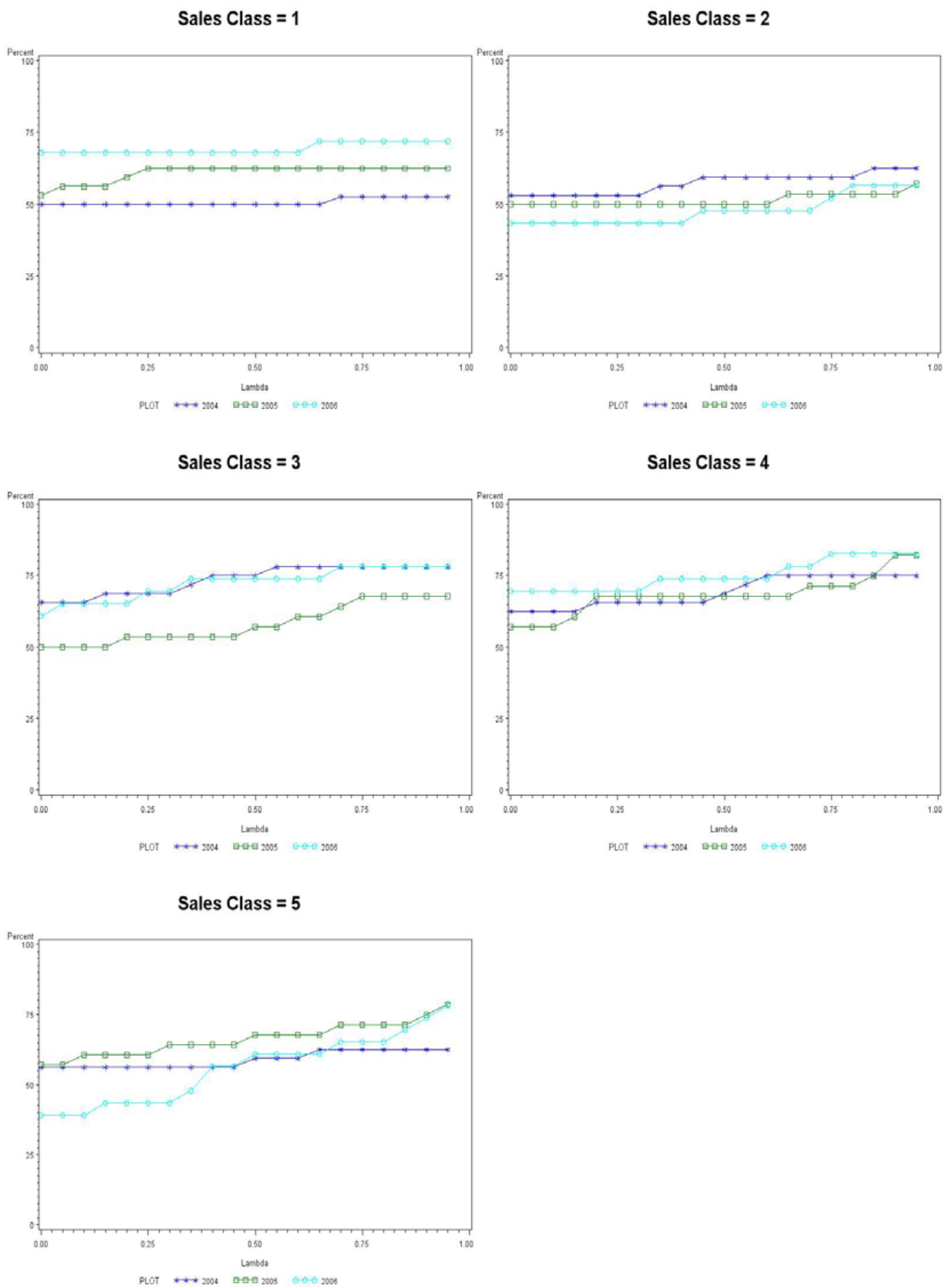


Figure 4: Percentage of States for which K2(λ) had a Lower Absolute Error than K2(1) for Various Values of λ

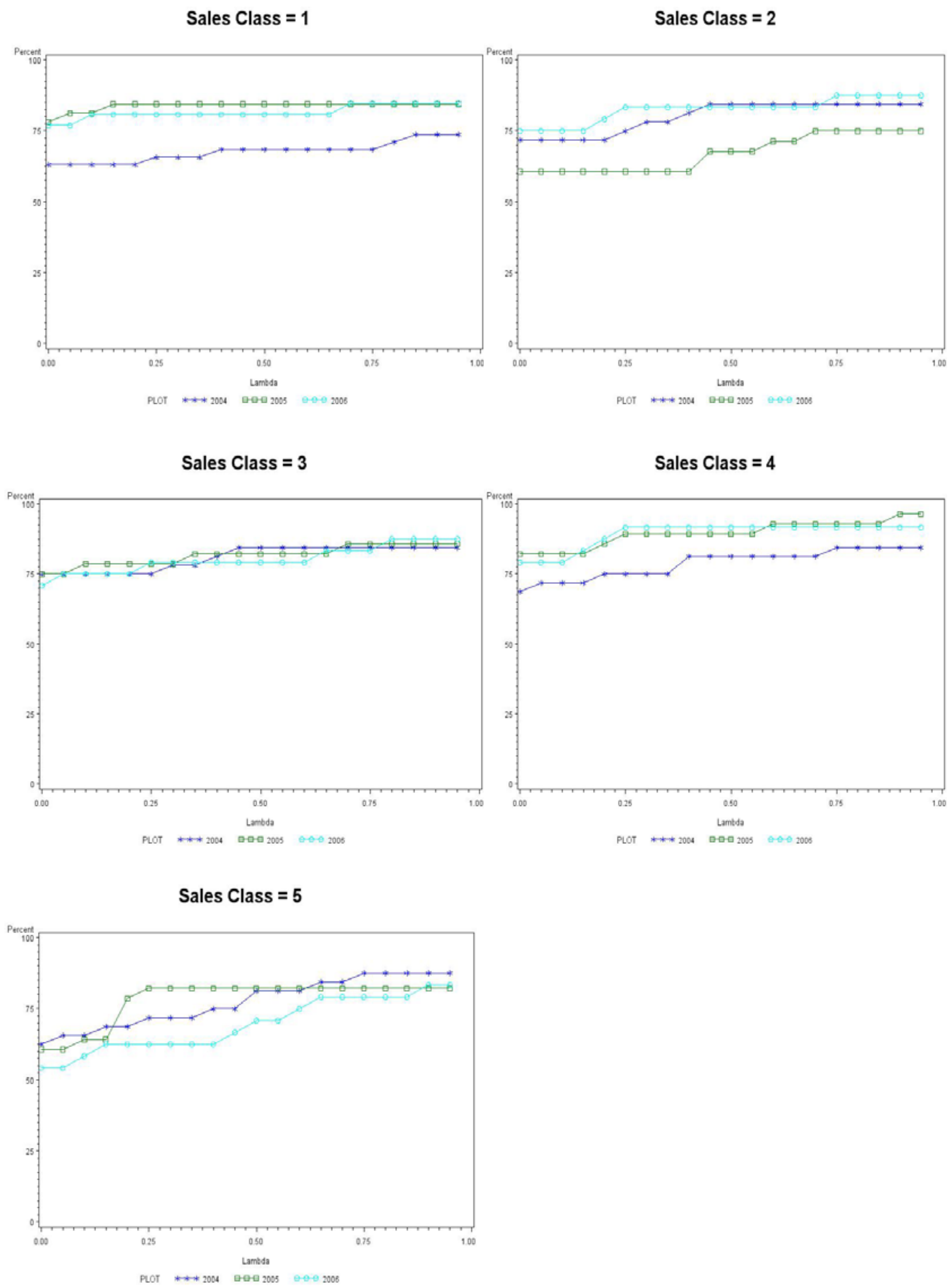


Table 8. Optimality Statistics for the Estimated Variance of K1(λ) and K2(λ)

Estimator		K1(λ)			K2(λ)		
Year	Sales Class	Mode (λ^*)		Mean(λ^*)	Mode (λ^*)		Mean(λ^*)
		Value	Pct. Opt.		Value	Pct. Opt.	
2004	1	1	48.6	0.95	1	57.1	0.96
	2	1	69.7	0.98	1	78.8	0.98
	3	1	69.7	0.95	1	75.8	0.96
	4	1	87.9	0.99	1	90.9	0.99
	5	1	87.9	0.98	1	90.9	0.99
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 the 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	Pct. Opt.		Value	Pct. Opt.	
2004	1	0	48.6	0.49	0	54.3	0.35
	2	0	45.5	0.47	0	42.4	0.31
	3	0	42.4	0.35	0	45.5	0.28
	4	0	36.4	0.42	0	39.4	0.33
	5	0, 1	36.4	0.5	0	27.3	0.38
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	47.8	0.21
	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

4. SUMMARY AND CONCLUSIONS

Two proposed methods (called K1 and K2) for updating census estimates of number of farms using June area survey data were compared with corresponding area frame and hybrid operational estimates for the years 2003-2006. Regarding Agricultural Statistics Board figures as ‘truth’ and using estimated variance and absolute error as quality measures, comparisons were done both at the state level and within categories defined by farm value of sales.

At the state and sales class levels, both K1 and K2 outperformed the area frame and hybrid estimator 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 both 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 (2.4) and (2.6) can be adapted to the estimation of land in farms in June of years between censuses. Note that 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 (2.4) and (2.6) appropriately. However, whether the improvement in efficiency due to incorporating census and ASB totals into the estimation process (for K1 or K2) would be as appreciable for land-related variables as it was for farm counts is unclear 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 (2.1) through (2.6) may be muted.

5. RECOMMENDATIONS

Based on the results of the empirical study, the following recommendations are made:

1. Implement the K2 estimator (which makes use of previous year official estimates) into NASS’s procedure for estimating number of farms at the state level and within sales classes, thus providing an additional improved indication. This recommendation is made based on K2 being found to compare favorably with the area frame and hybrid operational estimators as well as the K1 estimator.

2. Explore the potential extension of the K2 (and possibly K1) estimation method to other agricultural items. This research may require some modification of the formulas depending on the specific item being estimated.

6. REFERENCES

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.

United States Department of Agriculture (2009), *NASS Estimation Manual*.

APPENDIX A: Variance Estimation Using the Delete-A-Group Jackknife

In the study discussed in this report, variances (more precisely, mean squared errors) of the K1 and K2 estimators as well as within-group variances of the hybrid operational estimator were estimated using an extended delete-a-group (*edag)* jackknife. This appendix describes the *edag* jackknife estimation procedure in detail.

The first step is to place every area frame sample segment into one of *R* replicate groups. When a segment leaves the area sample, another segment from the same substratum takes its place in the replicate group. Although a segment's weight can change from year to year (based on the number of sample segments in its substratum), its replicate group designation must remain the same.

The *edag* jackknife variance estimator for an arbitrary estimator *u* has the following form:

$$\hat{v}(u) = [R/(R-1)] \sum_{r=1}^{R} (u - u^{(r)})^2]$$

where $u^{(r)}$ is the r'th replicate analogue of *u.* The r'th replicate analogue of the state level K1 (and K2) estimator of number of farms in the first year after a census is as follows:

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

where:

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

$S_t =$ set of segments in the samples for both year *t* and year *t-1* (as in Section 2),

$w_{kt}^{(r)} =$ jackknife replicate weight for segment *k*, year *t*, replicate group *r*.

Formulas for the replicate weights will be given shortly. The corresponding replicate analogue within sales classes is:

$$\hat{Y}_{tg}^{(K1)(r)} = \hat{Y}_{tg}^{(K2)(r)} = Y_{(t-1)g}^{(C)} \hat{b}_t^{(r)} \quad (g=1,\ldots,5).$$

The r'th replicate analogues of the generalized K1 and K2 estimators for the second, third and fourth post-census years at the sales class level (defined by equations (2.9) and (2.10)) are as follows:

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

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

Let $S_t^{(hr)}$ denote the set of segments in S_t, substratum h, replicate group r. For each segment k in substratum h, the jackknife weight for replicate group r is computed as follows:

$$\begin{aligned}
w_{kt}^{(r)} &= w_{kt} && \text{if } S_t^{(hr)} \text{ is empty,} \\
&= w_{kt}[1 - (n_h - 1)\sqrt{Z}] && \text{if } n_h > 1 \text{ and } k \in S_t^{(hr)}, \\
&= w_{kt}(1 + \sqrt{Z}) && \text{if } n_h > 1,\ S_t^{(hr)} \text{ is not empty and } k \notin S_t^{(hr)}, \\
&= w_{kt}\{1 - \sqrt{0.5R/(R-1)}\} && \text{if } n_h = 1 \text{ and } k \in S_t^{(hr)}
\end{aligned}$$

where:

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

n_h = number of segments from substratum h in S_t ,

$Z = \min\{R/[(R-1)n_h(n_h-1)], 1/(n_h-1)^2\}.$

These formulas are basically the ones given in Kott (2001), with a slight modification to handle substrata containing one segment in the usual NASS manner. Note that if $n_h \geq R$, the equation $\sqrt{Z} = 1/(n_h - 1)$ holds and the standard delete-a-group jackknife obtains. In particular:

$$\begin{aligned}
w_{kt}^{(r)} &= 0 \quad \text{if } k \in S_t^{(hr)}, \\
&= w_{kt}\left[n_h/(n_h-1)\right] \quad \text{otherwise.}
\end{aligned}$$

APPENDIX B: Supplementary Tables and Charts

Table 1B shows the years for which each state was included in the research study. The number of states used for each year is given in parentheses in the top row. As mentioned in Section 3, states receiving a new area frame during the 2004-06 period could only be evaluated for the years when the old frame was still in operation.

Table 1B. States in Research Study by Year

State	**No. Sales Classes**	**2003** (41)	**2004** (39)	**2005** (33)	**2006** (26)
Alabama	5	x	x	x	x
Arkansas	5	x	x	x	
California	5	x	x	x	x
Colorado	5	x	x	x	
Florida	5	x	x	x	x
Georgia	5	x	x	x	x
Idaho	5	x	x	x	x
Illinois	5	x	x	x	
Indiana	5	x	x		
Iowa	5	x	x	x	x
Kansas	5	x	x	x	x
Kentucky	5	x			
Louisiana	5	x	x	x	x
Maine	2	x	x		
Maryland	2	x	x	x	
Michigan	5	x	x	x	x
Minnesota	5	x	x	x	x
Mississippi	5	x	x		
Missouri	5	x			
Montana	5	x	x	x	x
Nebraska	5	x	x	x	x
Nevada	2	x	x	x	
New Jersey	2	x	x	x	x
New Mexico	5	x	x	x	
New York	5	x	x	x	x
North Carolina	5	x	x	x	x
North Dakota	5	x	x	x	x
Ohio	5	x	x	x	x
Oklahoma	5	x	x	x	x
Oregon	5	x	x		
Pennsylvania	5	x	x	x	x
South Carolina	5	x	x	x	x
South Dakota	5	x	x	x	x
Tennessee	5	x	x	x	x
Texas	5	x	x	x	
Utah	5	x	x	x	x
Virginia	5	x	x		
Washington	5	x	x	x	x
West Virginia	2	x	x		
Wisconsin	5	x	x	x	x
Wyoming	2	x	x	x	x

Table 2B shows the mean absolute relative error (MARE) of K1, K2, AF and HYB at the state and sales class levels for 2003-06 (where the ASB numbers are regarded as truth), while Table 3B shows the corresponding mean squared relative error (MSRE) values. Figures 1B through 4B are bar charts displaying sales class and state level estimated relative efficiency of K1 and K2 with respect to AF (see Section 3 for a detailed explanation).

Table 2B. Mean Absolute Relative Error of Estimators (K1 and K2 Identical for First Post-Census Year)

Year	**Sales Class**	**Estimator**			
		K1	**K2**	**AF**	**HYB**
2003	1	0.101		0.245	0.134
	2	0.098		0.121	0.137
	3	0.071		0.221	0.282
	4	0.085		0.194	0.221
	5	0.08		0.302	0.301
	All	0.065		0.161	0.063
2004	1	0.142	0.086	0.249	0.123
	2	0.136	0.091	0.139	0.134
	3	0.134	0.136	0.2	0.289
	4	0.23	0.217	0.197	0.282
	5	0.157	0.144	0.255	0.27
	All	0.082	0.055	0.16	0.055
2005	1	0.157	0.099	0.227	0.141
	2	0.136	0.084	0.133	0.079
	3	0.111	0.092	0.232	0.27
	4	0.204	0.18	0.169	0.171
	5	0.233	0.207	0.277	0.258
	All	0.104	0.064	0.137	0.064
2006	1	0.188	0.12	0.237	0.17
	2	0.146	0.117	0.114	0.103
	3	0.155	0.132	0.261	0.283
	4	0.257	0.2	0.177	0.166
	5	0.207	0.108	0.211	0.205
	All	0.117	0.071	0.133	0.071

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

Table 3B. Mean Squared Relative Error of Estimators (K1 and K2 Identical for First Post-Census Year)

Year	**Sales Class**	**Estimator**			
		K1	**K2**	**AF**	**HYB**
2003	1	0.022		0.109	0.031
	2	0.013		0.022	0.031
	3	0.009		0.106	0.208
	4	0.013		0.056	0.076
	5	0.011		0.17	0.15
	All	0.012		0.044	0.011
2004	1	0.038	0.015	0.145	0.037
	2	0.033	0.02	0.029	0.032
	3	0.037	0.041	0.066	0.133
	4	0.279	0.252	0.092	0.212
	5	0.069	0.052	0.103	0.101
	All	0.014	0.006	0.053	0.006
2005	1	0.042	0.015	0.152	0.033
	2	0.031	0.016	0.025	0.009
	3	0.02	0.013	0.103	0.178
	4	0.075	0.068	0.044	0.053
	5	0.109	0.09	0.113	0.101
	All	0.017	0.008	0.038	0.008
2006	1	0.051	0.024	0.147	0.054
	2	0.034	0.024	0.023	0.019
	3	0.036	0.032	0.152	0.216
	4	0.134	0.076	0.048	0.049
	5	0.086	0.024	0.073	0.068
	All	0.026	0.012	0.036	0.012

Figure 1B. Estimated RE of K2 with respect to AF for Eastern States by Sales Class

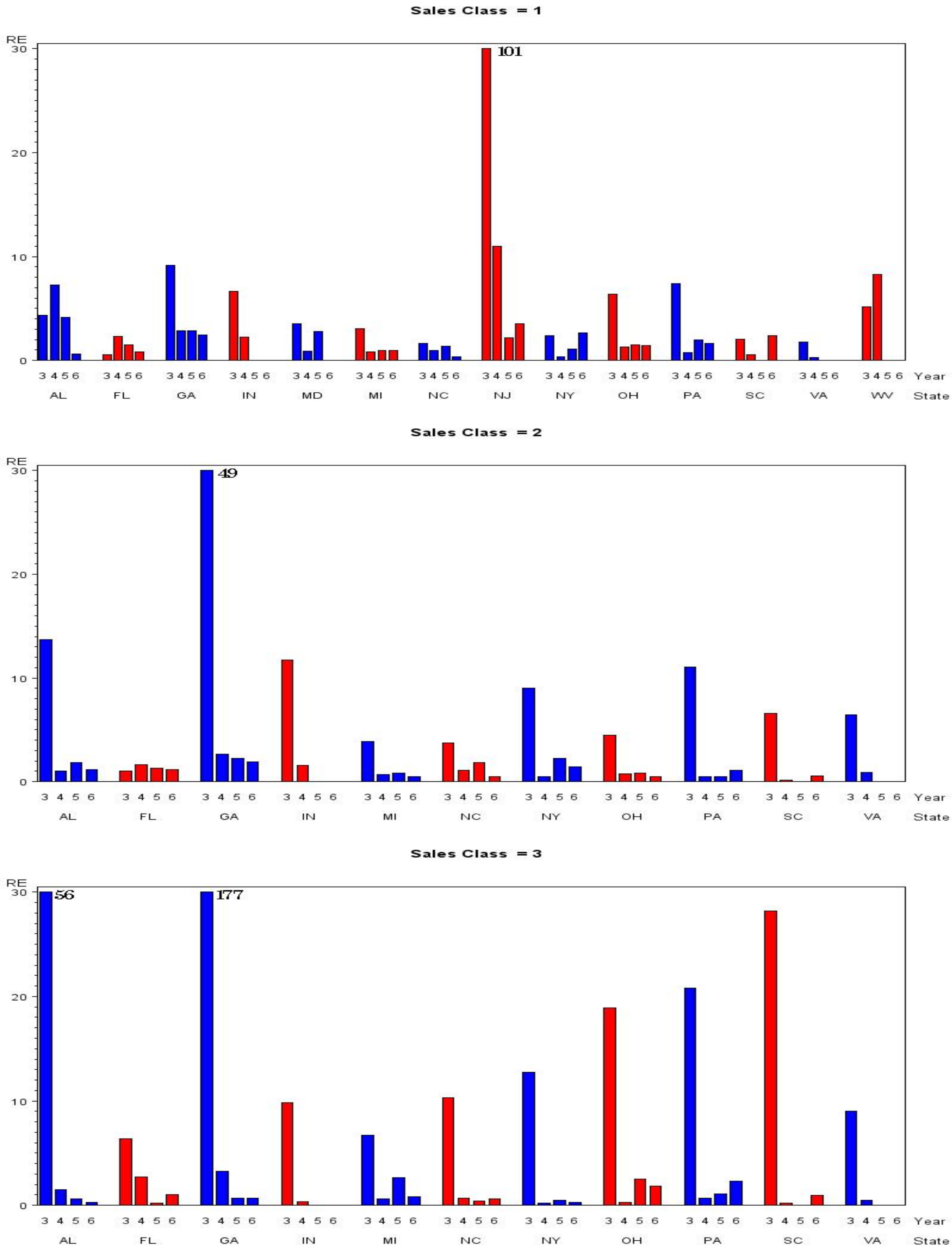
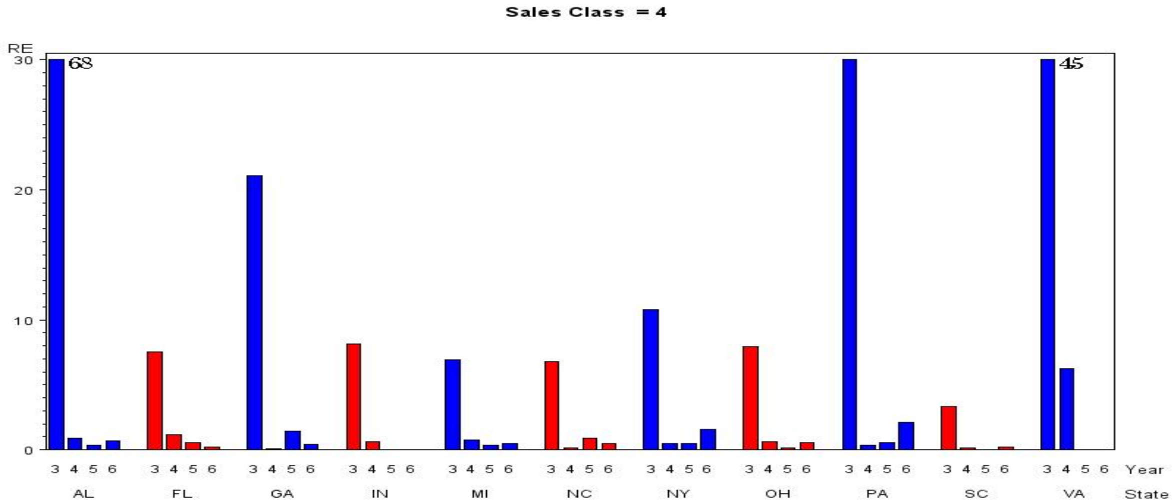


Figure 1B. Estimated RE of K2 (with respect to AF) for Eastern States by Sales Class (cont.)



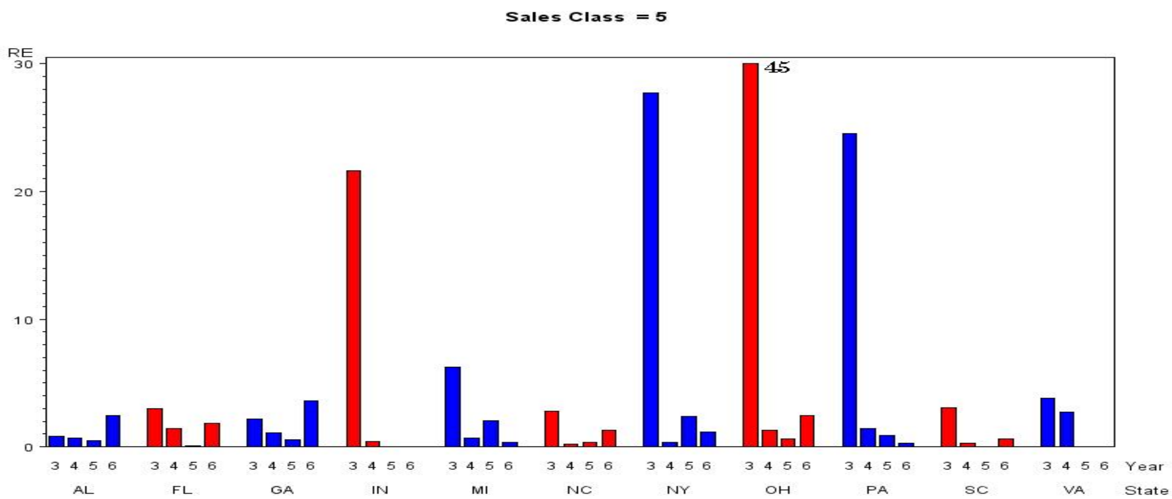
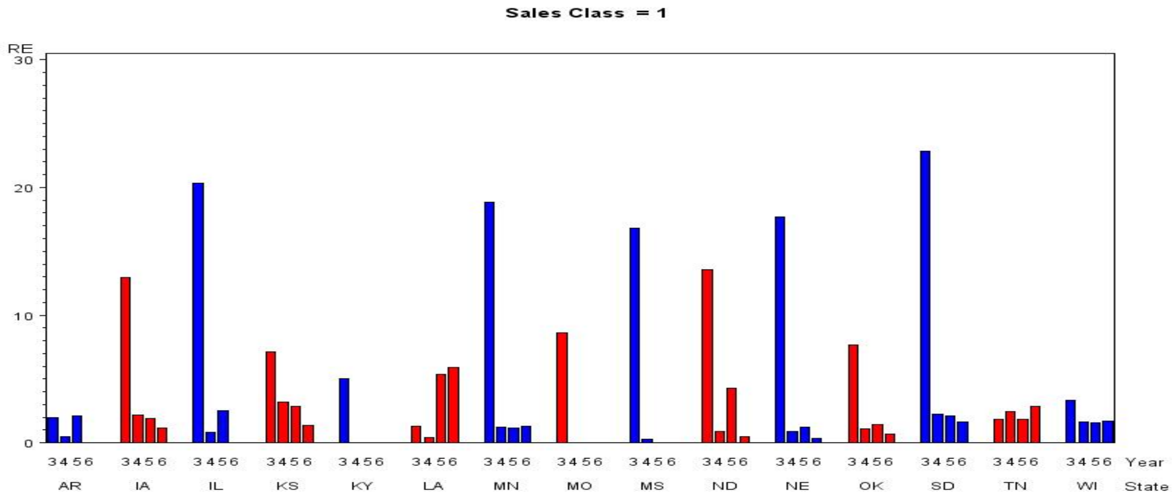


Figure 2B. Estimated RE of K2 (with respect to AF) for Central States by Sales Class



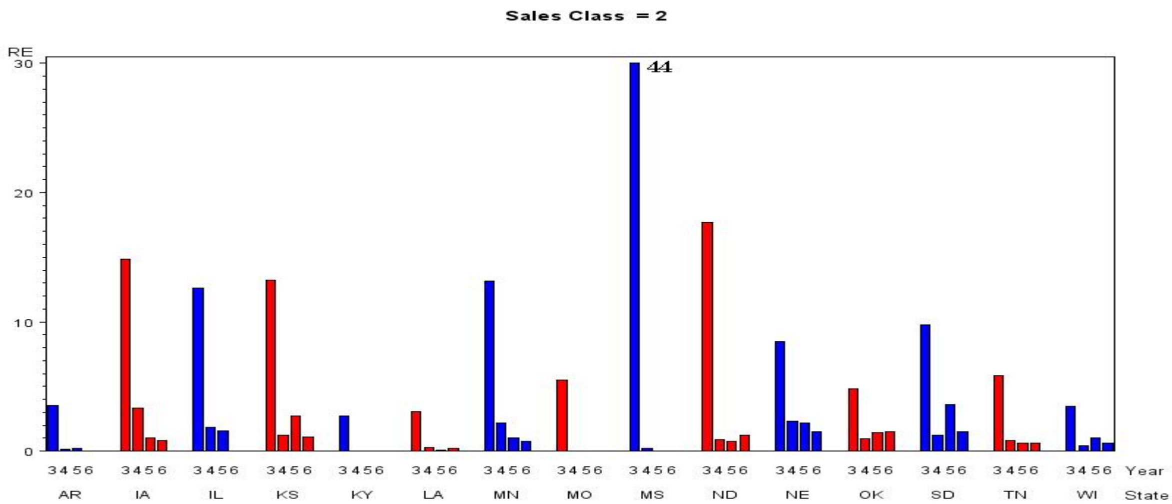
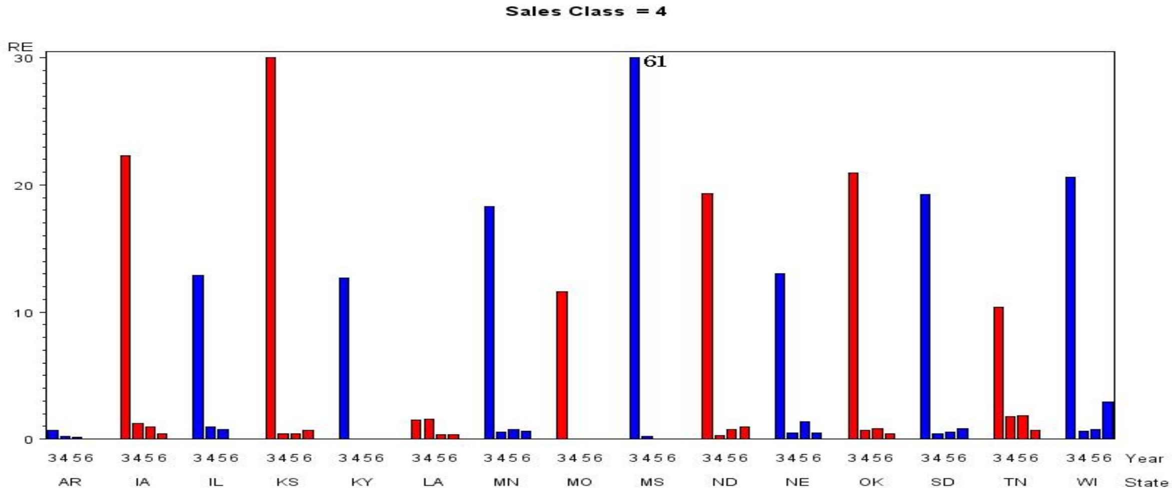


Figure 2B. Estimated RE of K2 (with respect to AF) for Central States by Sales Class (cont.)



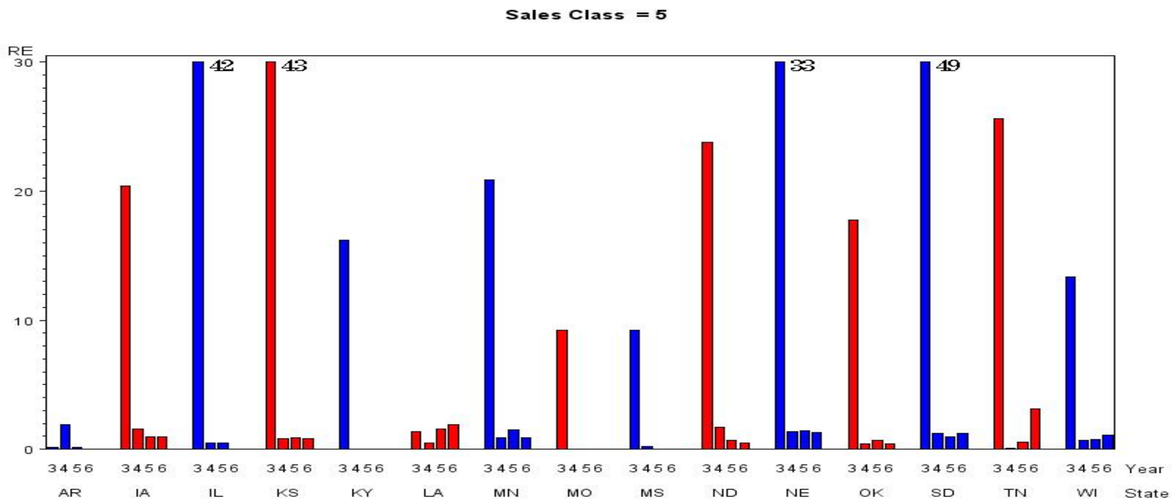


Figure 3B. Estimated RE of K2 (with respect to AF) for Western States by Sales Class

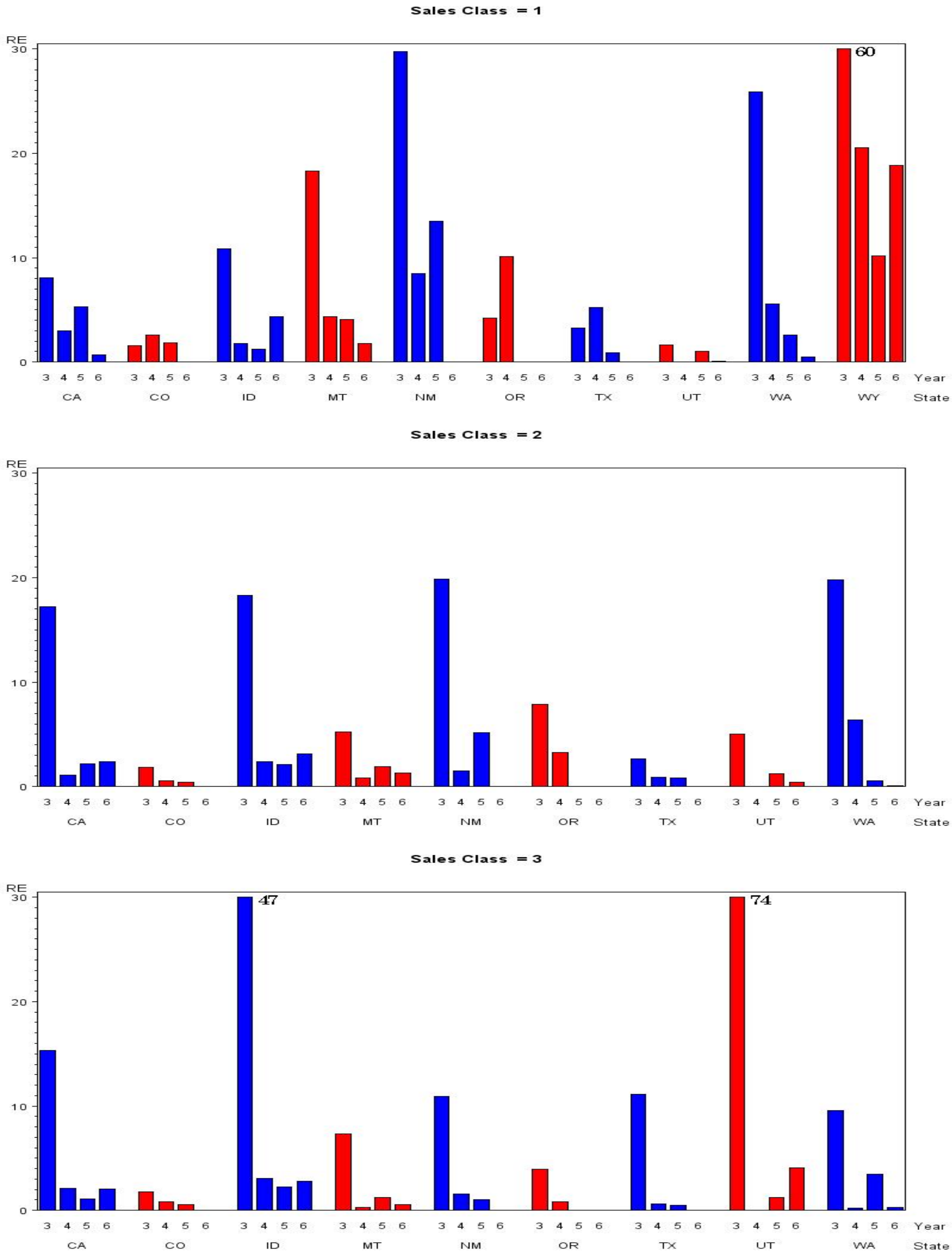
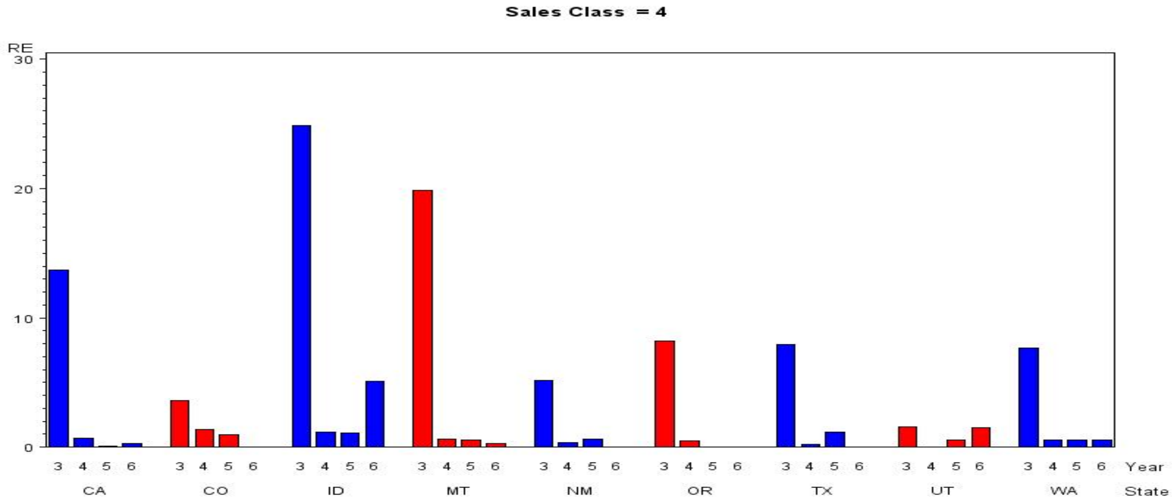


Figure 3B. Estimated RE of K2 (with respect to AF) for Western States by Sales Class (cont.)



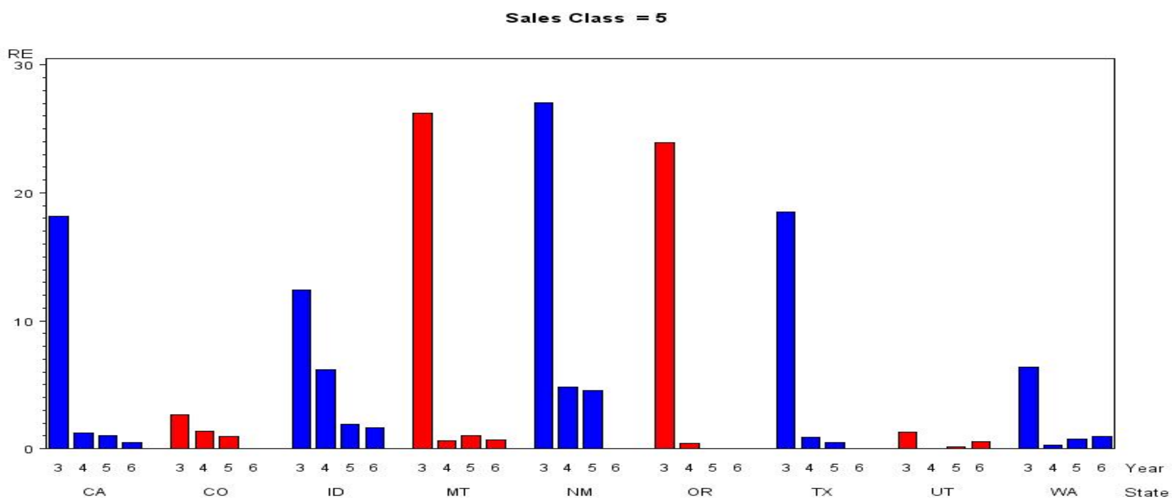


Figure 4B. Estimated RE of K2 (with respect to AF) at State Level