USDA United States Department of Agriculture

National Agricultural Statistics Service

Research Division

RD Research Report Number RD-97-02

June 1997

Evaluation of the Multiyear Method for Estimation of Crop and Livestock Totals

Michael E. Bellow
Charles R. Perry
Lih-Yuan Deng

EVALUATION OF THE MULTIYEAR METHOD FOR ESTIMATION OF CROP AND LIVESTOCK TOTALS, by Michael E. Bellow, Charles R. Perry, Jr. and Lih-Yuan Deng, Sampling and Estimation Research Section, Research Division, National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, D.C. June, 1997. NASS Research Report No. RD-97-02.

ABSTRACT

In the annual June Agricultural Survey conducted by the National Agricultural Statistics Service (NASS), area frame sampling is based on multiyear rotation designs with approximately twenty percent of the sample units replaced each year. Only the current year's sample data are used for estimation of agricultural commodities of interest. A multiyear estimation method was developed based on an analysis of variance model that utilizes the successive sampling of units in the area frame across years. In a previous report, multiyear estimates of hog inventories and soybean acreage in several States were shown to be more precise and robust than corresponding single year estimates.

This report extends the earlier research by comparing the multiyear method with single year estimation for a number of crop area, grain stocks, and hog inventory items at the State and national levels. The multiyear method is evaluated with respect to amount of variance reduction achieved and proximity to official NASS estimates.

KEY WORDS

Analysis of variance model, rotation sample design, relative efficiency

---

ACKNOWLEDGMENTS

The authors thank Bill Iwig for guidance and support of this project, Dale Atkinson for reviewing the draft manuscript, and Jim Burt for computer assistance.

TABLE OF CONTENTS

SUMMARY

With the exception of a current year-previous year ratio estimator for sampling units (segments) that are in the sample both years, NASS uses only the current year's survey data to compute commodity estimates. Area frame sampling for the June Agricultural Survey involves rotation with about 80 percent overlap of segments from one year to the next. Since segments exhibit some degree of consistency over years, an approach that makes effective use of two or more years of survey data would be expected to improve estimation efficiency.

Chhikara and Deng (1992) developed a multiyear estimation method based on ideas of Hartley (1980). The method uses an analysis of variance (ANOVA) model to describe multiyear survey data. The survey reported value for a given sample segment is modeled as the sum of a year effect, segment effect and random error term. Chhikara and Deng conducted an empirical study, comparing their method with single year estimation for two commodities in four States. The results were promising for the multiyear method.

The research described in this report extends the earlier work of Chhikara and Deng. A large scale evaluation of the multiyear estimation method was carried out. The method was tested for a number of crop acreage, grain stocks, and hog inventory items at the State and U.S. levels. The measures used to compare the multiyear method with single year estimation were estimated relative efficiency (RE) and deviations from final NASS estimates. A bootstrap study demonstrated the validity of a model-based estimate of relative efficiency as a performance measure.

At the State level, the multiyear method showed appreciable reductions in variance over single year estimation only in a small number of cases. At the U.S. level, no item had an estimated (model-based) RE above 1.2. The values of estimates generated using the two methods tended to be in close proximity.

Since the gains from multiyear estimation were marginal and the variance computations complex, plans for operational testing of the method on the 1996 June Agricultural Survey were cancelled.

INTRODUCTION

The National Agricultural Statistics Service (NASS) conducts probability sample surveys to estimate many agricultural commodities in the United States. NASS's annual June Agricultural Survey (JAS) uses a multiple frame approach to sampling. The area frame is based on a land use stratification of a State's area. This frame provides full coverage of the 48 conterminous States but is inefficient for rare commodities and those represented by extremely large farms. The list frame, consisting of a list of known farm operators in a State, is much more efficient than the area frame for most commodities. However, it is usually incomplete and difficult to maintain. The multiple frame approach takes advantage of the strengths of both sampling frames.

NASS generally uses only the current year's survey data to estimate commodities, and never uses more than two years of data. The area frame sampling involves multiyear rotation designs with approximately 20 percent replacement of sample units each year. Since roughly 80 percent of units remain in the sample from one year to the next when the same frame is in place, an estimation approach that uses multiple years of survey data can augment the current year's information and effectively increase the sample size. The sampling variance of estimates is thereby reduced. In fact, the 80 percent overlap ratio estimator has been used operationally for many years.

Chhikara and Deng (1992) proposed an approach that applied an analysis of variance (ANOVA) model to commodity estimation using two or more years of area frame survey data. They tested the method using 1987-90 soybean acreages and hog inventories in Indiana, Iowa, Ohio, and Oklahoma (Chhikara et al., 1993). The main conclusion was that the multiyear model led to a more efficient estimate than the single year method. The multiyear method also produced a much more stable yearly variance estimate than the single year method.

One application for which multiyear estimation was thought to have potential is in the revision of previous year estimates. The way the model is devised, commodity estimates for a given year can be calculated using survey data of following years as well as previous years. Thus, for example, commodity estimates for 1993 could conceivably be revised four years later using multiyear data from 1989 through 1997. However, that approach assumes that the relevant conditions (farming practices, weather patterns, etc.) do not change dramatically over the time period involved.

This report documents the extension of the earlier research on multiyear estimation. The method was compared with single year estimation for a number of crop area, grain stocks, and hog inventory items in the 48 conterminous States. The evaluation was done both at the individual State and national levels.

NASS ESTIMATORS

The commodity estimators currently used by NASS are described in detail by Chhikara and Deng (1992), so only a brief account is given here.

The sampling unit of NASS's area frame is known as the *segment*, a piece of land with identifiable boundaries and generally between 0.1 and 4 square miles. The reporting unit is the *tract*, an area within a segment that is under a single operation or management. An estimator of the State total for a commodity can be obtained by summing the survey data over tracts within a segment, multiplying by an expansion factor, summing over segments within strata, and finally aggregating the stratum totals to the State level. This unbiased estimator, known as the *area tract estimator*, is considered reliable for estimating crop acreages since it uses the accurately determined tract data. However, it does not work well for livestock items or commodities associated with a farm operation. In such cases, the *area weighted estimator* is preferred. This estimator is derived from sample inventories of farms totally or partially within the sample segments. The weight is the ratio of the within-segment tract acreage to the corresponding farm acreage.

Multiple frame estimators use data from both frames, but favor the list frame. The overlap domain is the set of farm operators in both the area and list frames. The nonoverlap (NOL) domain is the set of farm operators in the area frame but not the list frame. A multiple frame estimator is the sum of a list frame estimator (imputed or adjusted) in the overlap domain and an area frame estimator (closed or weighted) in the NOL domain. In general, multiple frame estimators are more efficient than estimators that use only the area frame, especially for livestock items and specialty crops. Those items are poorly correlated with land area, so the substitution of data from a properly stratified list frame for data from a land use stratified area frame usually causes a reduction in variance.

MULTIYEAR ESTIMATION MODEL

As mentioned earlier, NASS's area frame sampling has about 80 percent overlap of segments from one year to the next. There is a degree of consistency in area segment characteristics from year to year. The factors that remain fairly constant over years are the prevalent soil types in segments and the capabilities of certain operators to grow crops. Factors that vary across years include weather and economic conditions. Multiyear estimation should achieve the largest gains in efficiency for those commodities that are most consistent over time.

Hartley (1980) proposed an analysis of variance approach to crop area estimation using multiyear data acquired from earth orbiting satellites. Lycthuan-Lee (1981) implemented Hartley's idea, estimating North Dakota wheat acreage with three years of satellite data. Chhikara and Deng (1992) adapted this methodology to estimation of commodities using multiyear survey data collected from the area frame.

The multiyear ANOVA model is given by:

$$y_{tk} = \alpha_t + b_k + e_{tk}$$

$$(k=1, 2, \ldots, n_t;\ t=1,2,\ldots,T)$$

where:

α_t = fixed effect for year t

b_k = random effect for segment k

e_{tk} = random error for year t, segment k

T = number of years

n_t = sample size in year t

In matrix form, the model can be written as:

$$\mathbf{y} = \mathbf{X}\alpha + \mathbf{U}\mathbf{b} + \mathbf{e} \quad (3.1)$$

where:

$\mathbf{y} = (y_{11}, y_{12}, \dots, y_{Tn_T})$

$\alpha = (\alpha_1, \dots, \alpha_T)$

$\mathbf{b} = (b_1, \dots, b_S)$

S = number of distinct segments sampled over T years

$\mathbf{e} = (e_{11}, e_{12}, \dots, e_{Tn_T})$

X is a design matrix consisting of 0's and 1's accounting for the fixed year effect α. **U** is design matrix of 0's and 1's specified according to the rotation sampling scheme and accounting for the random segment effect **b**. The dimensions of X and U are NxT and NxS, respectively, where:

$$N = \sum_{t=1}^{T} n_t$$

The assumptions are that **b** has mean **0** and covariance matrix $\sigma_b^2 \mathbf{I}$, and the random error term **e** has mean **0** and covariance matrix $\sigma_e^2 \mathbf{I}$. The total error $\mathbf{q} = \mathbf{U}\mathbf{b} + \mathbf{e}$ has mean **0** and covariance matrix $\sigma_e^2 \mathbf{W}$, where:

$$\mathbf{W} = \mathbf{I} + \gamma \mathbf{U}\mathbf{U}', \quad \gamma = \sigma_b^2 / \sigma_e^2$$

The parameter γ is usually not known and must be estimated from survey data. The estimator used here is:

$$\hat{\gamma} = (S/N)[(MSB/MSE) - 1]$$

where MSB is the mean square due to segment and MSE is the mean square due to error. Although the model is applied separately within each substratum (subdivision of a stratum), γ is estimated at the stratum level to stabilize estimation across substrata.

The weighted least squares estimator of α is:

$$\hat{\alpha} = (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{-1}\mathbf{y}$$

The covariance matrix of $\hat{\alpha}$ is:

$$\mathscr{C}(\hat{\alpha}) = (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}\sigma_e^2$$

The single year estimator used by NASS is obtained by assuming no segment effect, i.e., setting γ=0:

$$\tilde{\alpha} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

The covariance matrix of $\tilde{\alpha}$ under the multiyear model (i.e., where γ is not assumed to be zero) is given by:

$$\mathscr{C}(\tilde{\alpha}) = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{W}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}\sigma_e^2 \quad (3.2)$$

(Chhikara et al., 1993). The diagonal elements of this matrix are the single year variances for years 1,...,T under the multiyear model. Alternatively, the single year variance for a given year can be estimated by the standard formula using only the current year's survey data. That estimator will be referred to as the survey-based single year variance estimator, and the one computed from equation (3.2) as the model-based single year variance estimator.

The multiyear estimator for the final year, $\hat{\alpha}_T$, is of primary interest. The variance of this estimator is always less than or equal to the model-based variance of the single year estimator.

Assuming that γ is known, $\hat{\alpha}_T$ is the best unbiased estimator (BLUE) of α_T under the model. A simulation study performed by

Chhikara and Deng (1992) showed the multiyear estimation method to be fairly robust to moderate misspecification of γ, i.e., varying γ did not affect estimator performance appreciably.

The optimal number of years of survey data to use depends upon the percentage of sample segments replaced each year. By comparing results for T=2,3,4 and 5 in a simulation study, Chhikara and Deng concluded that under NASS's current sample design, the highest gains would be achieved for T=5. However, the results for T=4 were nearly as good as for T=5.

When multiple frame estimation is used, the multiyear method applies only to the area frame (NOL) component of estimates. The list frame (overlap) component is the same as for single year estimation. Consequently, the gains due to the multiyear method should be lower for multiple frame estimation than for area frame estimation. The consistency of survey observations across years has a strong influence on these gains. If a survey item is consistent, then a reduction in variance is expected. If the item is not consistent, then the single year estimate becomes very unstable, while the multiyear estimate produces a "smoother" result across years for the estimate and variance.

STATE LEVEL EVALUATION

Multiyear estimates of eight crop area items and four hog inventory items were computed for the 48 conterminous States, using JAS data from the four-year period 1992-95. Since an area tract estimator and a multiple frame estimator were computed for each crop area item, there was a total of 20 commodity/estimator combinations. All computing was done in SAS/IML.

Table 1 compares the single year and multiyear area tract estimates of corn planted area in the five States that had the highest final 1995 NASS estimates. The ratios between the multiyear and single year estimates are shown, along with the standard errors. Table 2 gives the results for total hog and pig inventory in the top five States for 1995, per the final NASS estimates. The estimator used was the adjusted list weighted NOL estimator, which is the sum of the area weighted estimator in the NOL domain and a nonresponse adjusted list frame estimator in the overlap domain. Table 2 also shows the NOL component of the single year and multiyear estimates and their standard errors.

Three separate estimates of relative efficiency (RE) are shown in the tables. The survey-based RE is the ratio of the survey-based variance of the single year estimator to the multiyear variance. The model-based RE is the ratio of the model-based variance of the single year estimator to the multiyear variance. The survey-based RE is a questionable measure of the improvement achieved by using multiyear estimation, since it is highly sensitive to outliers and underestimation of the single year standard error. This fact is illustrated by Table 2, where the survey-based RE in the NOL domain is less than one for four of the five States. The multiyear variance estimate is much more stable over years, and hence more reliable. The model-based estimate of single year variance, obtained from equation (3.2), is also much more stable over years than the corresponding survey-based estimate. Hence, assuming that the multiyear model is valid, the model-based RE should be

Table 1: 1995 Corn Planted Area Estimator Comparison for Top Five States (SE values in thousand acre units).

State	Domain	Ratio - M/S (%)	Ratio - S/F (%)	Ratio - M/F (%)	Single Year SE (SB)	Multiyear SE	RE (SB)	RE (MB)	RE (Boot)
Iowa	AF	100.7	101.5	102.2	260.3	232.4	1.25	1.07	1.08
Illinois	AF	100.6	100.0	100.6	247.9	244.8	1.03	1.04	1.04
Nebraska-	AF	102.5	97.3	99.8	314.9	239.5	1.73	1.29	1.30
Minnesota	AF	100.5	74.6	75.0	199.1	196.0	1.03	1.02	1.01
Indiana	AF	99.3	97.1	96.5	183.9	173.6	1.12	1.05	1.03

(AF - area frame, M - multiyear estimate, S - single year estimate, F- final NASS estimate, SB - survey-based, MB - model-based)

Table 2: 1995 Total Hogs and Pigs Estimator Comparison for Top Five States (SE values in thousand head units).

State	Domain	Ratio - M/S (%)	Ratio - S/F (%)	Ratio - M/F (%)	Single Year SE (SB)	Multiyear SE	RE (SB)	RE (MB)	RE (Boot)
Iowa	NOL	88.2			234.4	476.0	0.24	1.02	1.01
	MF	98.9	92.7	91.7	395.8	572.9	0.48	1.01	
North Carolina	NOL	622.2			14.7	145.0	0.01	1.28	1.24
	MF	101.9	98.9	100.7	115.0	184.5	0.39	1.17	
Illinois	NOL	96.2			196.5	165.1	1.42	1.01	1.02
	MF	99.6	97.3	96.9	301.5	282.1	1.14	1.00	
Minnesota	NOL	94.9			74.0	92.7	0.64	1.03	1.03
	MF	99.8	102.4	102.2	160.0	169.5	0.89	1.01	
Indiana	NOL	132.0			56.6	118.6	0.23	1.15	1.18
	MF	101.3	97.3	98.5	144.4	178.1	0.66	1.07	

(NOL - non-overlap, MF - multiple frame, other abbreviations - see Table 1) a more reliable measure of effectiveness than the survey-based RE. However, little is known about robustness of the model-based RE against departure from model assumptions. One way of addressing this issue is to use a resampling method. After consideration of several options, a balanced bootstrap on model residuals was chosen as the most feasible method to apply. Bootstrapping regression residuals is described by Efron and Tibshirani (1993) and Shao and Tu (1995). The balanced bootstrap, where each observed residual is constrained to appear the same number of times in the set of all bootstrap samples, improves the efficiency of the results.

Direct application of the bootstrap method is difficult due to the correlated error structure of the multiyear model. Therefore, a transformation was applied to the data so that an equivalent model with diagonal error covariance matrix could be used (Seber, 1977).

The nonsingular matrix V satisfying W=VV' was first computed using the eigenvalues and eigenvectors of W. Premultiplication of model (3.1) by V^{-1} yields the transformed model:

$$\mathbf{z} = \mathbf{B}\alpha + \mathbf{f}$$

where $\mathbf{z} = \mathbf{V}^{-1}\mathbf{y}$, $\mathbf{B} = \mathbf{V}^{-1}\mathbf{X}$, and $\mathbf{f} = \mathbf{V}^{-1}(\mathbf{Ub} + \mathbf{e})$. The error term $\mathbf{f}$ has mean $\mathbf{0}$ and covariance matrix $\sigma_e^2 I$.

The adjusted residuals of the transformed model are given by:

$$\hat{\mathbf{f}}_a = [N/(N\text{-tr}(\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'))]^{1/2}(\mathbf{z} - \mathbf{B}\hat{\alpha})$$

One can show that $(\hat{\mathbf{f}}_a'\hat{\mathbf{f}}_a)/N$ is an unbiased estimator of the random error variance σ_e^2 (see Appendix).

A balanced set of 500 bootstrap samples was selected from the empirical distribution assigning probability 1/N to each of the N adjusted, transformed residuals. For each replication, the selected bootstrap residuals were substituted into the transformed model to construct bootstrap values of **z**. These values were premultiplied by V to obtain bootstrap values of **y**. The model was then fitted to obtain both multiyear and single year bootstrap estimates of α.

The above procedure was applied separately within each substratum. The bootstrap values of α were summed over substrata and strata to obtain the bootstrap State level totals. The means and variances of those State totals over all replications were then computed. The bootstrap relative efficiency was computed as the ratio between the bootstrap single year and multiyear variances.

Tables 1 and 2 show the bootstrap RE values. For the total hogs and pigs estimates, the bootstrap RE is given only for the NOL domain. Comparison of the model-based and bootstrap RE's shows close agreement, with the largest discrepancy being 0.04 for total hogs and pigs in North Carolina. From these results and others not shown here, the model-based estimated RE can be considered a reliable measure.

For corn planted area, of the five States listed in Table 1, only Nebraska showed a model-based RE (1.29) appreciably greater than one. Nebraska was also the only State where the difference between the single year and multiyear estimates exceeded one percent. For total hogs and pigs in the MF domain (Table 2), North Carolina and Indiana had model-based RE's of 1.17 and 1.07, respectively, while the other three States had RE's not exceeding 1.01. However, the North Carolina result is an anomaly caused by two segments: one having an extremely high reported value in 1992 and then being rotated out of the sample, the other with an extremely high value in 1994 but zero in 1993 and 1995. As a result, the multiyear estimate was more than six times as large as the single year estimate in the NOL domain.

Figures 1 through 8 display State level estimates and estimator performance measures for eight crop area items. The items are alfalfa hay harvested, all hay harvested, corn harvested, corn planted, soybeans planted, durum wheat harvested, winter wheat harvested, and all wheat harvested. Each figure consists of four bar charts. The first bar chart shows the final NASS estimates of the item for each State where NASS published an estimate, in descending order of magnitude. The remaining three charts show the States in the same order as the first, to facilitate comparisons between the charts. The performance measures are displayed for both the area tract (AT) and multiple frame (MF) estimation schemes. The second chart shows the estimated model-based relative efficiencies (RE). The third chart shows the percent deviation of the single year and multiyear indications from the final NASS estimates (the numerical values are suppressed as administratively confidential). The fourth chart shows the percent proximity difference from the final NASS estimates, defined as follows:

$$PPD = (|SYR-FNL| - |MYR-FNL|)/FNL$$

where:

SYR = single year indication
MYR = multiyear indication
FNL = final NASS estimate

The PPD is a scaled measure of the difference between the distances of the two indications from the final NASS estimate. It shows which indication was closer, and percentage-wise how much closer. If the PPD is positive, then the multiyear indication is closer than the single year indication to the final estimate. If the PPD is negative, the multiyear estimate is further from the final estimate.

Four States received new area frames during the 1992-95 period: Oklahoma (1993), California (1994), New York (1995), and South Carolina (1995). Thus the multiyear estimator used three years of survey data for Oklahoma, two years for California, and one year for New York and South Carolina. Those four States are indicated with an asterisk in the second, third, and fourth charts on each page, since they cannot be compared directly with the States where four years of data were used. Of course, when only one year of data is used, the single and multiyear indications are identical.

Examining the charts of estimated RE, it is clear that the area tract estimator has higher RE than the multiple frame estimator in almost all cases. For the AT estimator, no State had an RE exceeding 1.4 for any of the eight crop area items. For the MF estimator, the RE was below 1.2 for every item with the exception of corn harvested in Montana, which had negligible acreage. The highest

State level RE's occurred for the area tract estimators of corn planted and all hay harvested acreage.

The percent deviation charts in Figures 1 through 8 indicate that there was little appreciable difference between the single year and multiyear indications in most cases. Furthermore, both indications tended to fall below the final NASS estimates.

The PPD charts reinforce the notion that the two indications do not differ by much. If most of the PPD values were positive, one could hypothesize that the multiyear estimator is less biased than the single year estimator. However, there are roughly equal numbers of positive and negative PPD values for all items.

Figures 9 through 12 show the same bar charts for four hog inventory items: pig crop (Dec.-Feb.), sows farrowed (Dec.-Feb.), total breeding stock, and total hogs and pigs. The MF estimator used is known as the "adjusted list modified weighted" estimator. States not listed individually are grouped together as "OTH". The RE's are less than 1.4 for all items in all States, and less than 1.2 except for Ohio. As with crop acreages, the percent deviation charts show little difference between the single year and multiyear indications. Both estimators tended to overestimate pig crop and sows farrowed but underestimate total breeding stock and total hogs and pigs. There are roughly equal numbers of positive and negative PPD values for each item, so neither estimator appears to be less biased than the other.

Overall, the multiyear method showed noteworthy improvement over single year estimation in only a small number of cases.

NATIONAL LEVEL EVALUATION

The 1995 State level multiyear and single year estimates of the twelve commodities from the previous Section were aggregated to the U.S. level. In addition, estimates of three grain stocks items were computed at the State level and aggregated to the U.S. level. These items were corn stocks, soybean stocks, and wheat stocks. As discussed earlier, the multiyear estimates for the four States that changed area frames during 1992-95 were computed using only those years when the new frames were in effect, e.g., 1993-95 for Oklahoma.

Tables 3 through 5 compare the single year and multiyear estimation methods at the national level for the fifteen items. Both the area tract and multiple frame estimation schemes were used for each crop area item. The SE's and RE's shown are the model-based values, in light of the bootstrap results discussed earlier. The tables also show all three ratios among the multiyear indications, single year indications, and final NASS estimates.

Table 3 shows that for crop area estimation, the highest estimated RE's occurred for the area tract estimates of durum wheat harvested (1.18), alfalfa hay harvested (1.17), and all hay harvested (1.17). Those three items also showed the largest percent differences between the single year and multiyear estimates. As expected, the estimated RE for each crop was higher with the area tract estimator than the multiple frame estimator.

Table 4 shows that the estimated RE for each grain stocks item was 1.01 or lower. The single year and multiyear estimates were in

Table 3: U.S. Level Estimator Comparison for Crop Area (SE values in thousand acre units). SE and RE values model-based.

Item	Type	Ratio - M/S (%)	Ratio - S/F (%)	Ratio - M/F (%)	Single Year SE	Multiyear SE	RE
Alfalfa Hay Harvested	AF	101.1	84.0	84.9	535.4	499.9	1.15
	MF	99.8	82.5	82.3	411.7	404.5	1.04
All Hay Harvested	AF	101.6	91.5	93.0	859.7	796.4	1.17
	MF	100.1	86.1	86.2	688.4	671.9	1.05
Corn Harvested	AF	100.3	99.9	100.2	716.9	682.2	1.10
	MF	100.0	84.0	84.0	549.7	544.9	1.02
Corn Planted	AF	100.4	96.5	96.9	725.1	688.0	1.11
	MF	100.0	82.5	82.5	564.5	559.7	1.02
Soybeans Planted	AF	99.9	98.6	98.6	662.9	641.5	1.07
	MF	99.6	79.8	79.4	606.8	603.0	1.01
Winter Wheat Harvested	AF	100.6	102.7	103.3	701.6	677.9	1.07
	MF	99.7	76.1	75.9	481.1	476.7	1.02
Durum Wheat Harvested	AF	102.3	96.7	98.9	224.1	205.9	1.18
	MF	100.1	78.4	78.5	153.7	153.5	1.00
All Wheat Harvested	AF	100.5	99.8	100.3	839.2	803.3	1.09
	MF	99.9	76.1	76.0	593.5	587.9	1.02

(M - multiyear estimate, S - single year estimate, F - final NASS est., AF - area frame, MF - multiple frame)

Table 4: 1995 U.S. Level Estimator Comparison for Grain Stocks (SE values in thousand bushel units). SE and RE values model-based. See Table 3 for abbreviations key.

Item	Type	Ratio - M/S (%)	Ratio - S/F (%)	Ratio - M/F (%)	Single Year SE	Multiyear SE	RE
Corn Stocks	MF	99.9	72.7	72.6	25,691.5	25,616.1	1.01
Soybean Stocks	MF	99.9	65.1	65.0	6,390.1	6,368.5	1.01
Wheat Stocks	MF	100.1	64.2	64.3	4,917.5	4,912.1	1.00

Table 5: 1995 U.S. Level Estimator Comparison for Hog Items (SE values in thousands head units). SE and RE values model-based.

Item	Type	Ratio - M/S (%)	Ratio - S/F (%)	Ratio - M/F (%)	Single Year SE	Multiyear SE	RE
Pig Crop (Dec.-Feb.)	MF	100.0	107.1	107.1	462.9	456.6	1.03
Sows Farrowed (Dec.-Feb.)	MF	100.0	106.7	106.7	56.1	55.4	1.03
Total Breeding Stock	MF	100.0	93.5	93.5	149.4	147.0	1.03
Total Hogs and Pigs	MF	100.1	96.6	96.8	818.5	804.1	1.04

(M - multiyear estimate, S - single year estimate, F - final NASS estimate, MF - multiple frame)

close proximity, but much less than the final NASS estimates.

From Table 5, the estimated RE's of the four hog inventory items were all below 1.05, and the single year and multiyear indications were in close proximity.

Figure 13 consists of three bar charts showing the estimated RE, percent deviation from final NASS estimates (numerical values not shown), and percent proximity difference from final NASS estimates, respectively, of the national level indications. The percent deviation chart shows that the single year and multiyear estimates were in close proximity and tended to fall below the final NASS estimates of crop area and grain stocks. The PPD was below three percent for all items, and well below one percent for the grain stocks and hog inventory items.

CONCLUSIONS

The multiyear estimation method was evaluated for a number of crop acreage, grain stocks, and hog inventory items at the State and national levels using four years of June Agricultural Survey data. Estimated relative efficiencies, estimator ratios and deviations from final NASS estimates were used to compare multiyear estimation with single year estimation. The estimated RE values may in fact be slightly optimistic since they depend to some degree on model accuracy, which has never been validated. The bootstrap study showed that the model-based estimate of relative efficiency was a more feasible measure to use than the corresponding survey-based estimate. State level results showed that the multiyear method caused appreciable gains in efficiency only in a small number of cases. At the national level, the estimated model-based RE's applied to the area tract estimator of crop area ranged from 1.07 to 1.18. Gains for the multiple frame estimators of crop area, grain stocks and hog inventories were less than 1.05. In general, the two indications were in close proximity.

RECOMMENDATIONS

The results discussed in this report do not warrant operational use of the multiyear estimation technique by NASS at this time. The gains in efficiency at the State and national levels were not sufficiently large to justify such a major change to NASS's estimation methodology.

REFERENCES

Chhikara, R.S. and Deng, L.Y. (1992). "Estimation Using Multiyear Rotation Design Sampling in Agricultural Surveys". *Journal of the American Statistical Association*, 87, 924-932.

Chhikara, R.S., Deng, L.Y. , Yuan, Y., Perry, C.R. and Iwig, W.C. (1993). "Estimation of Totals Using Multiyear June Agricultural Statistics Data". U.S. Dept. of Agriculture, NASS Report No. SRB-93-09.

Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. New York, N.Y., Chapman & Hall.

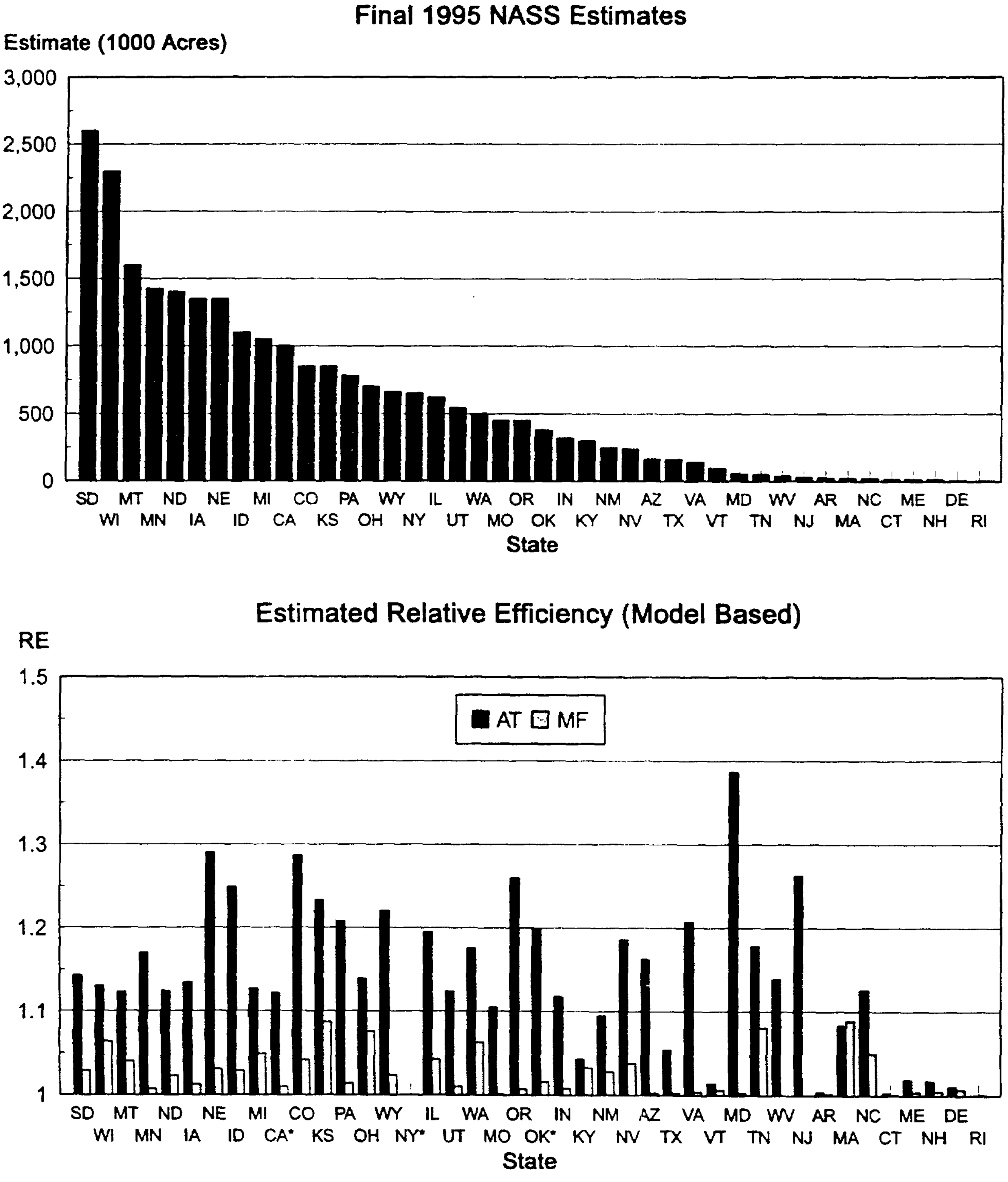
Hartley, H.O. (1980). *A Survey of Multiyear Estimation Procedures*. Technical Report DS1, Duke University, Dept. of Mathematics.

Lycthuan-Lee, T.G. (1981). *Development of Rotation Sample Designs for the Estimation of Crop Acreage*. Technical Report No. 15409, Lockheed Engineering and Management Services Company, Inc.

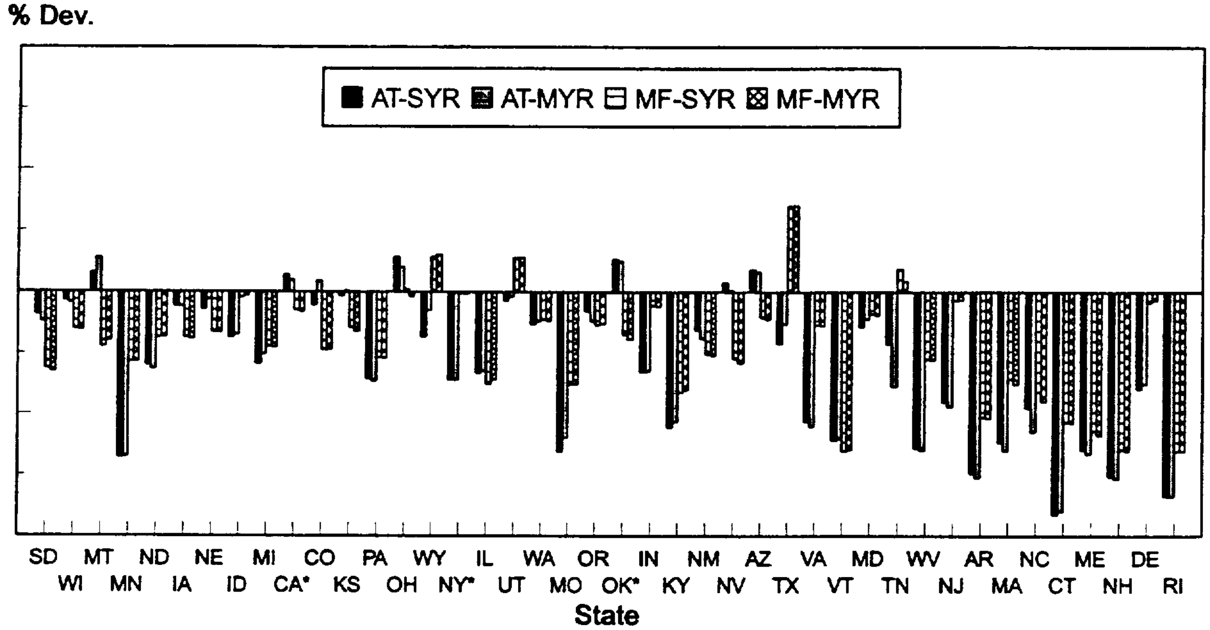
Seber, G.A.F. (1977). *Linear Regression Analysis*. New York, N.Y., John Wiley & Sons.

Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*. New York, N.Y., Springer.

Figure 1. State Level Estimator Comparison for Alfalfa Hay Harvested Area

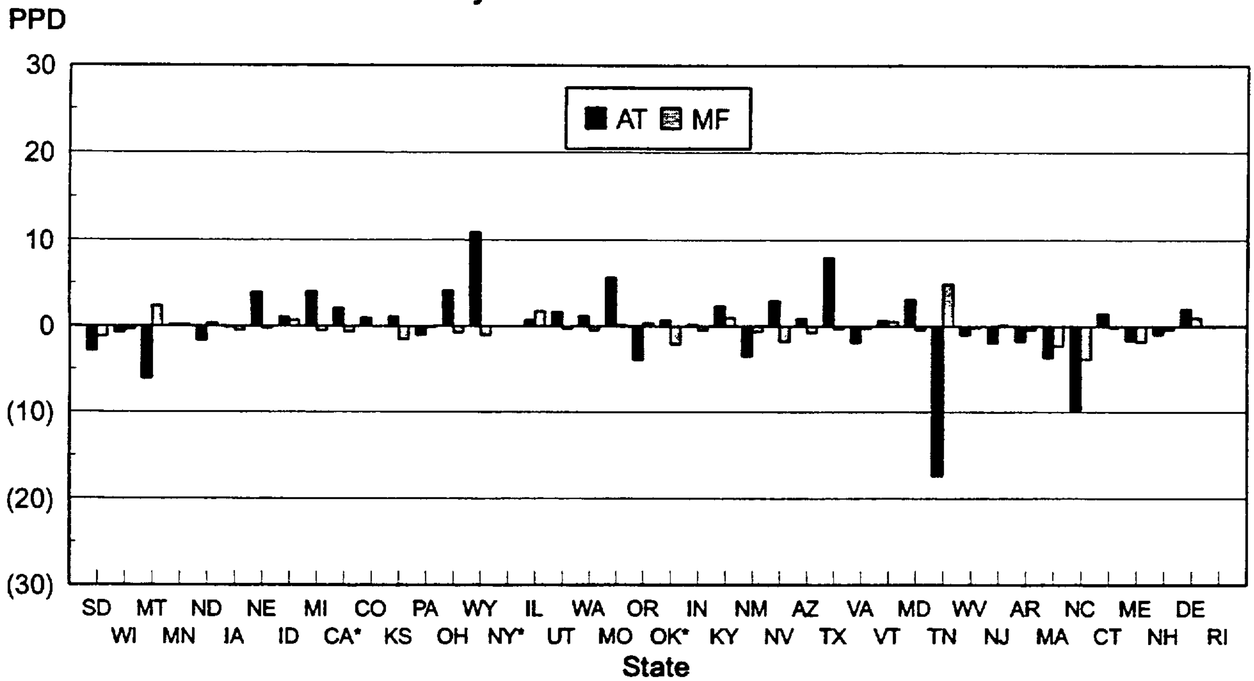


Percent Deviation of Indications from Final NASS Estimates



(scale suppressed as adminstratively confidential)

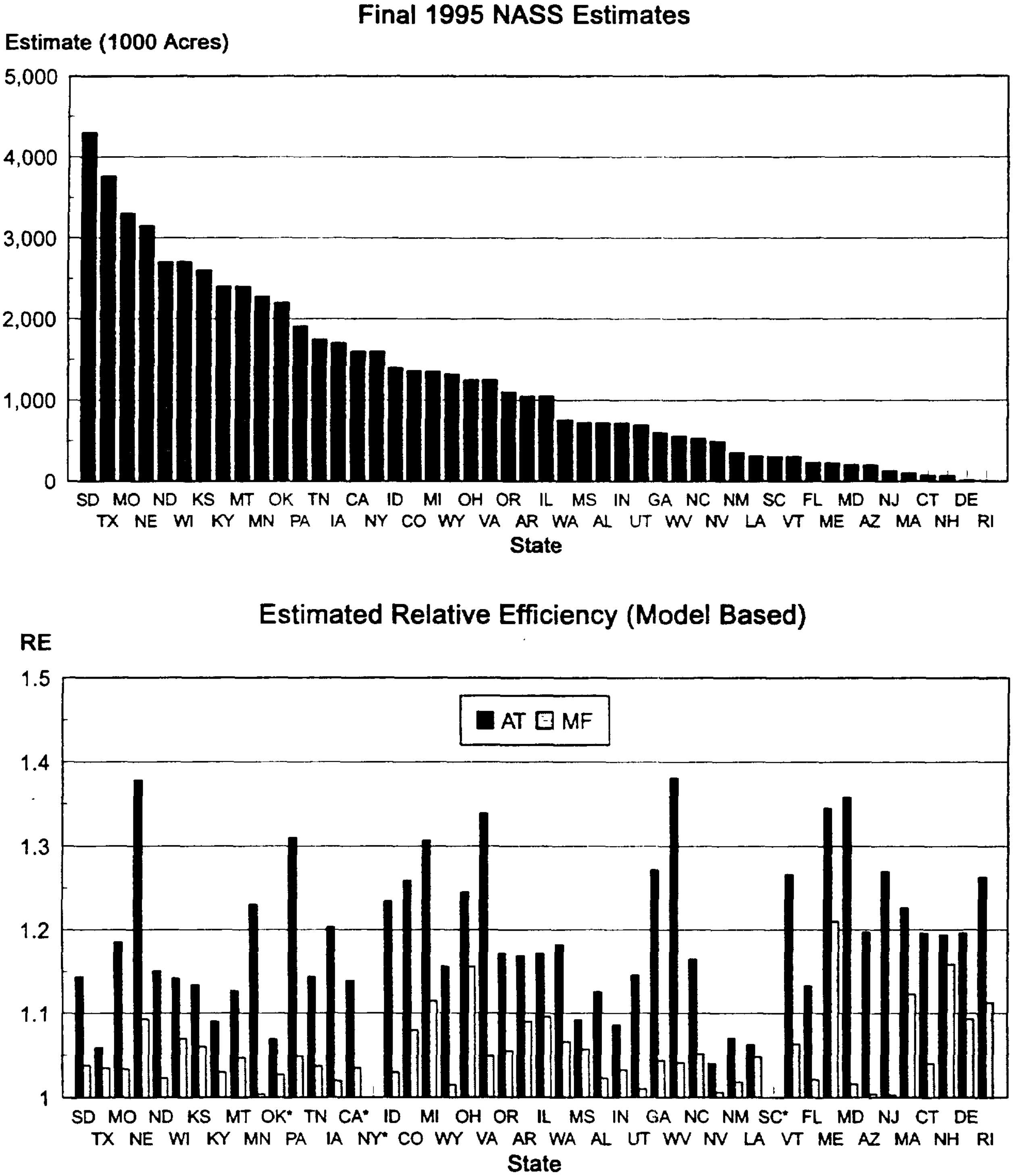
Percent Proximity Difference from Final NASS Estimates



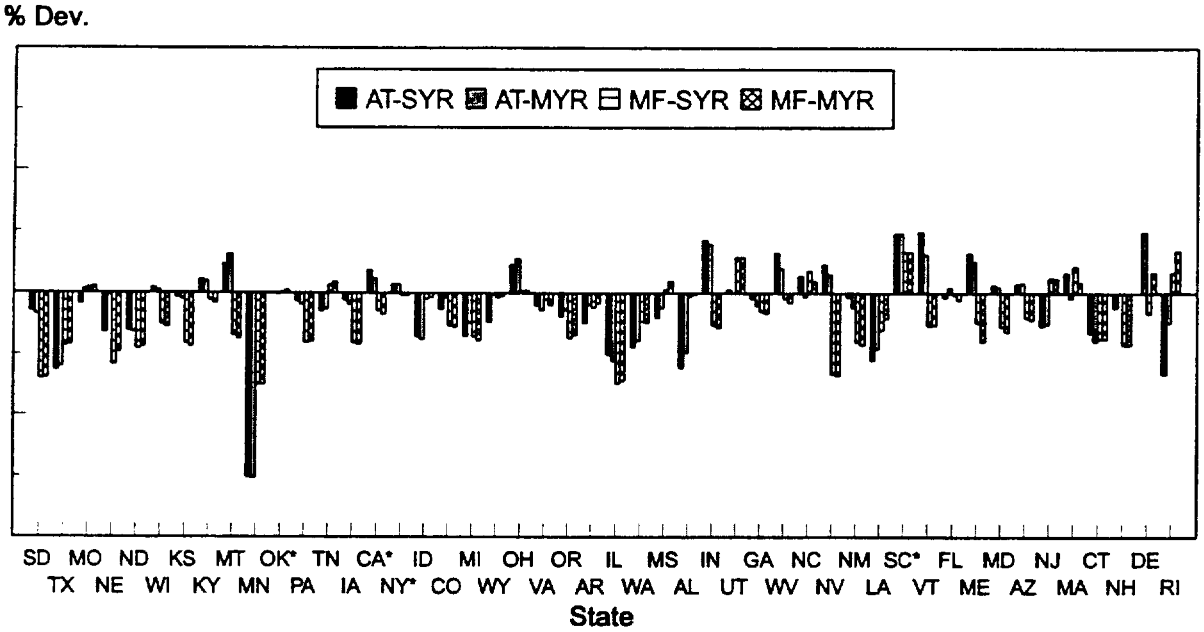
* - state received new area frame between 1992-95

AT - area tract estimator, MF - multiple frame estimator, SYR - single year, MYR - multiyear

Figure 2. State Level Estimator Comparison for All Hay Harvested Area

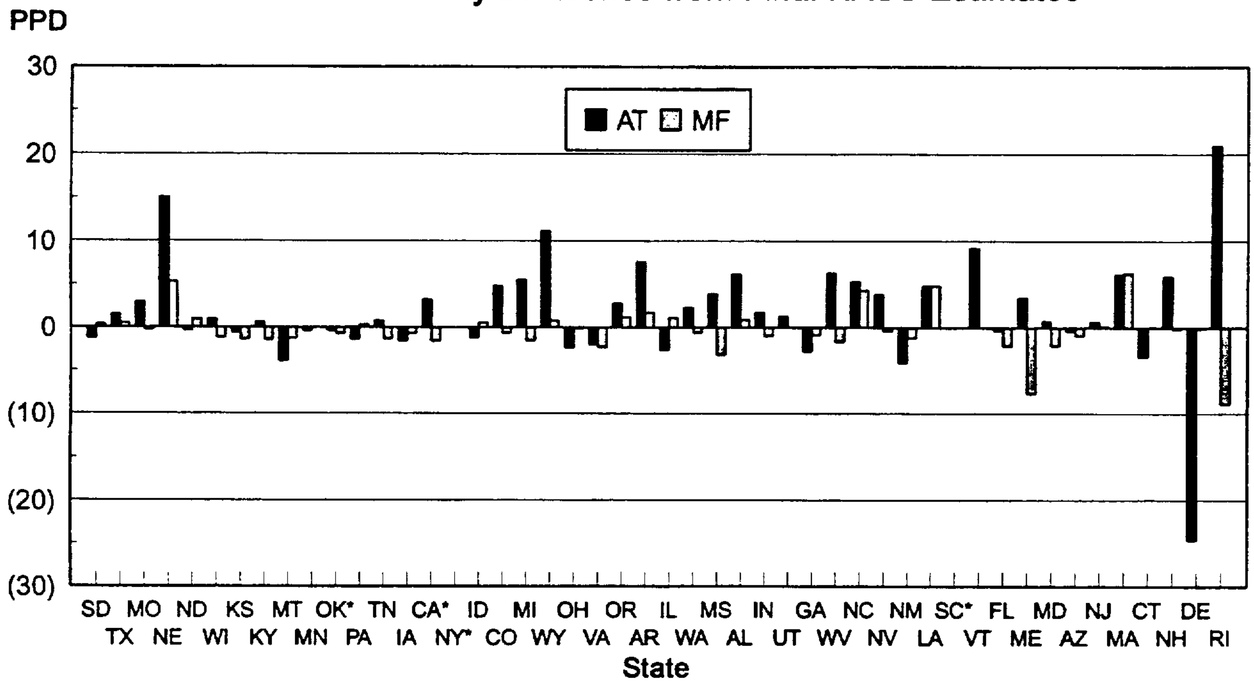


Percent Deviation of Indications from Final NASS Estimates



(scale suppressed as adminstratively confidential)

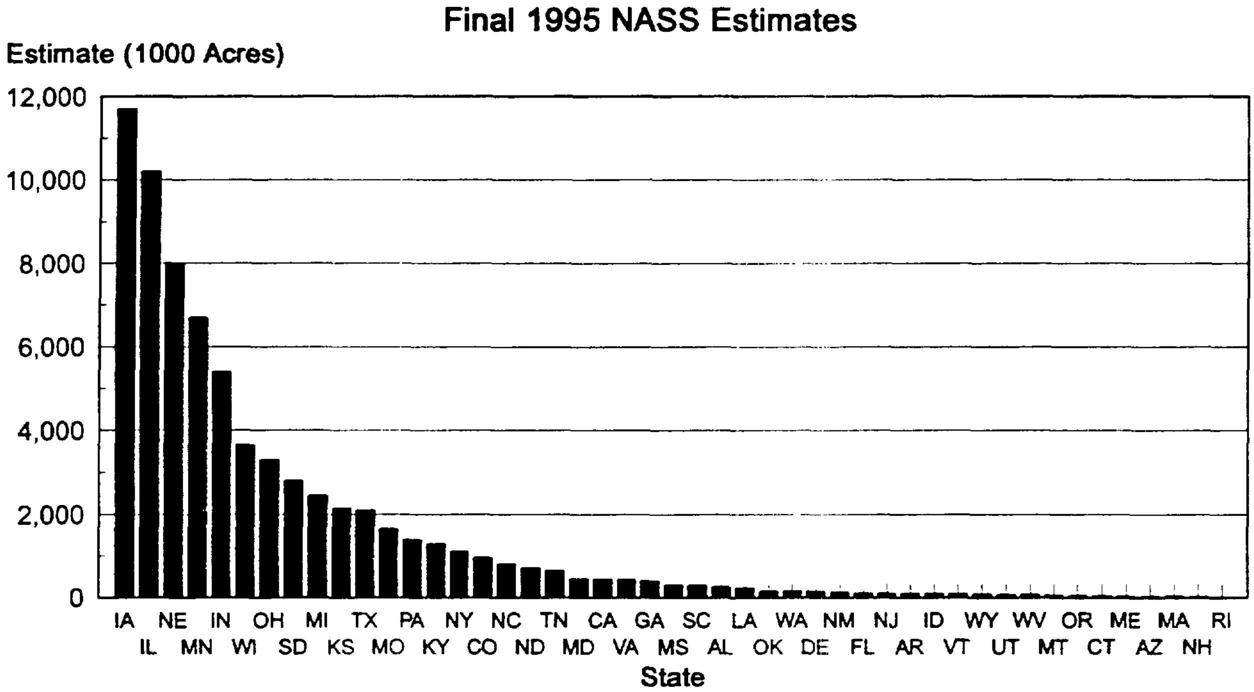
Percent Proximity Difference from Final NASS Estimates

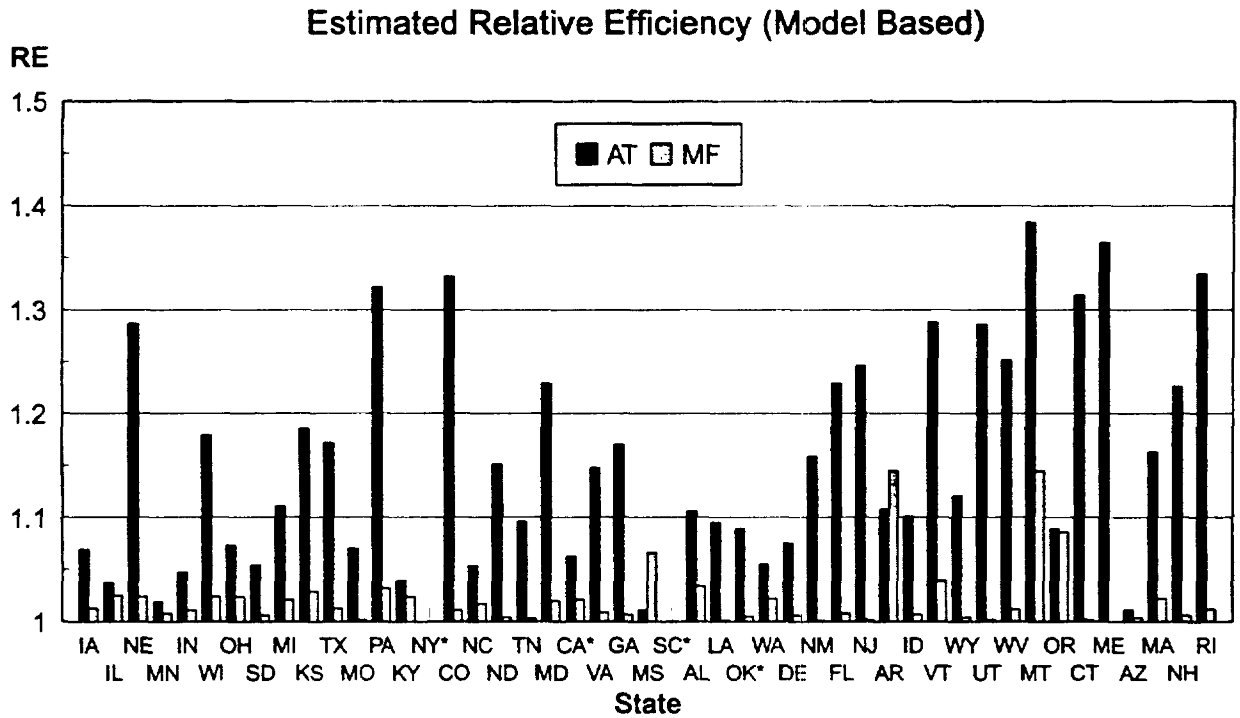


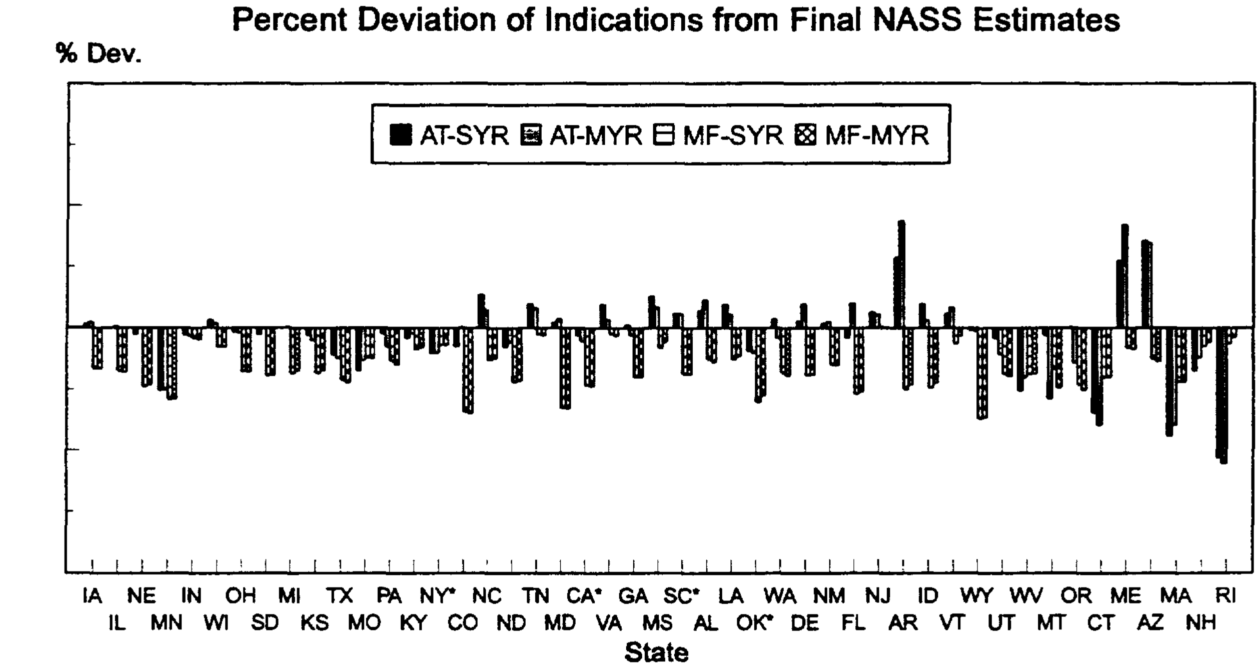
* - state received new area frame between 1992-95

AT - area tract estimator, MF - multiple frame estimator, SYR - single year, MYR - multiyear

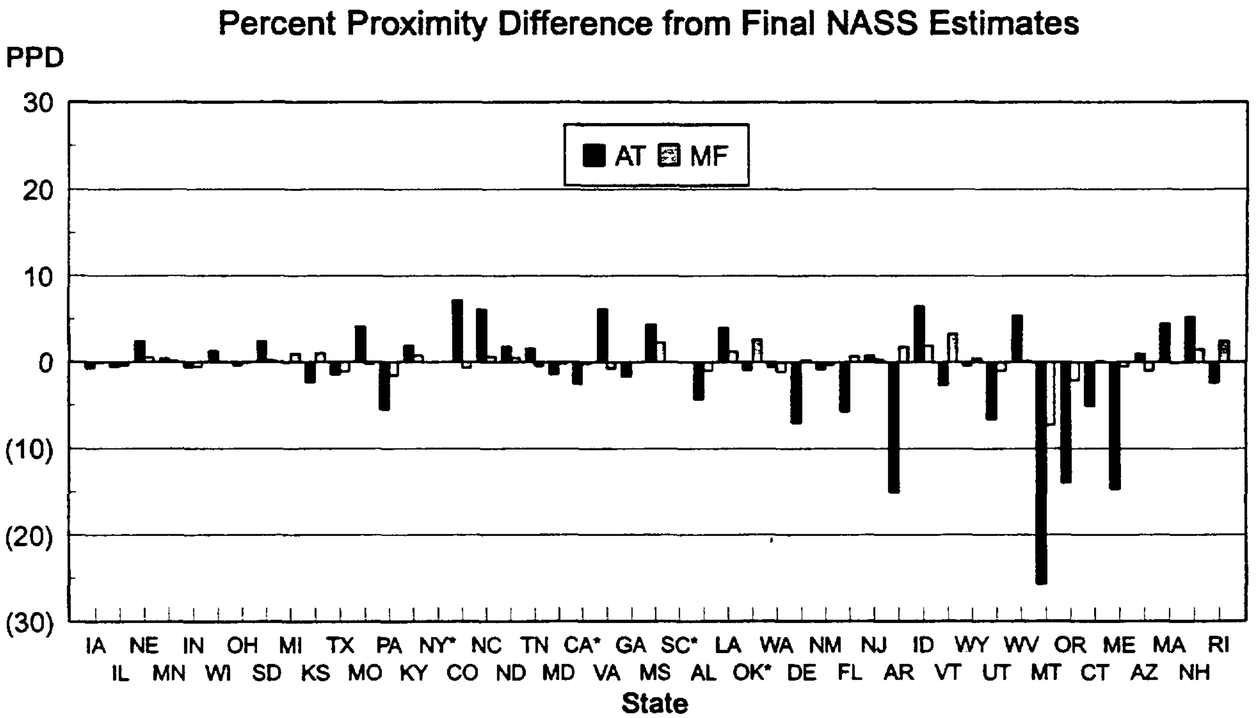
Figure 3. State Level Estimator Comparison for Corn Planted Area







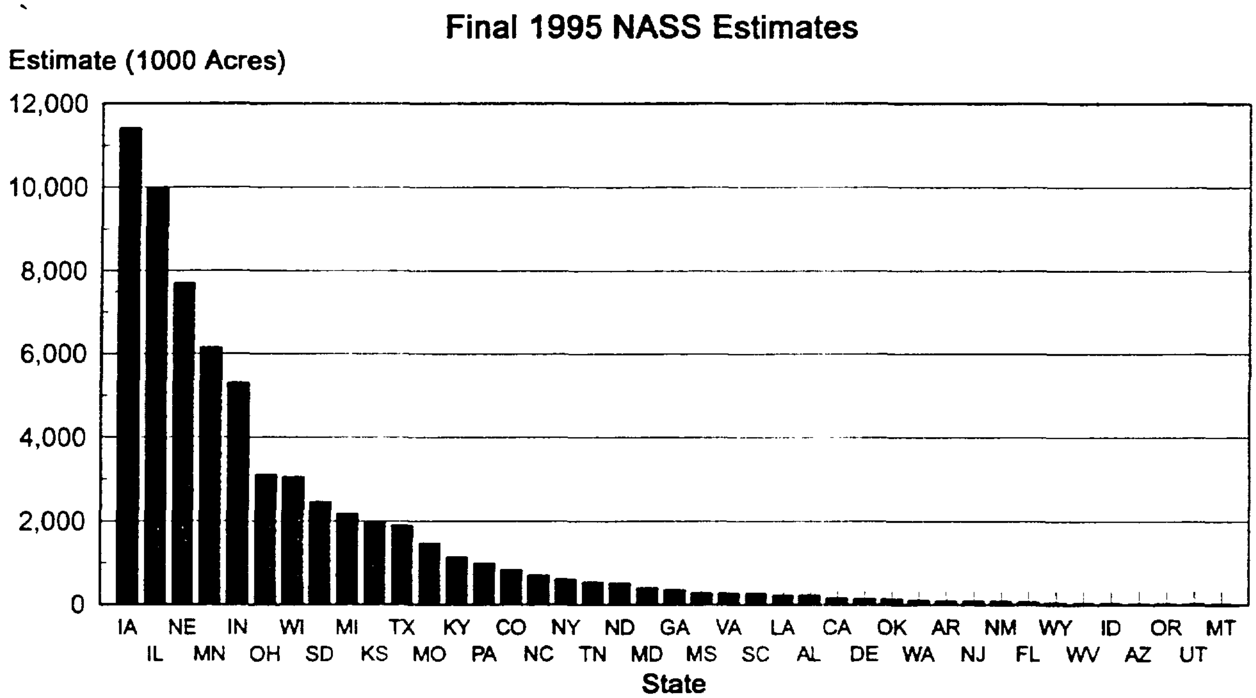
(scale suppressed as adminstratively confidential)

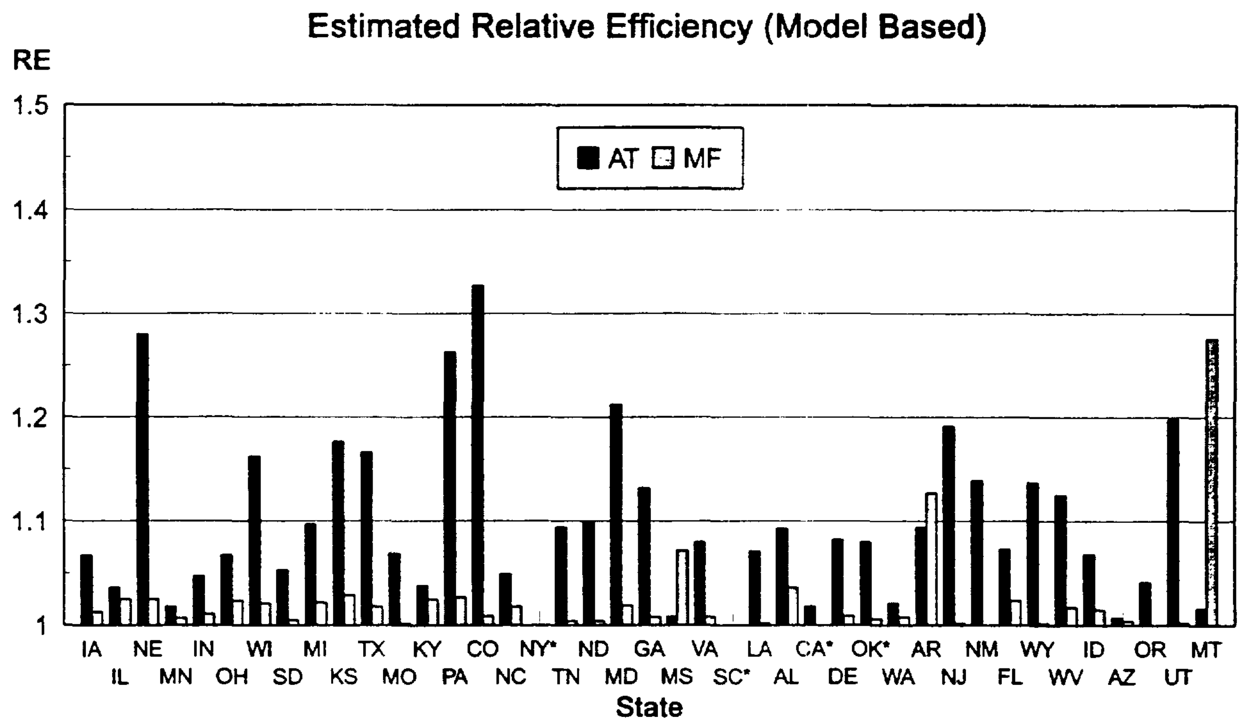


* - state received new area frame between 1992-95

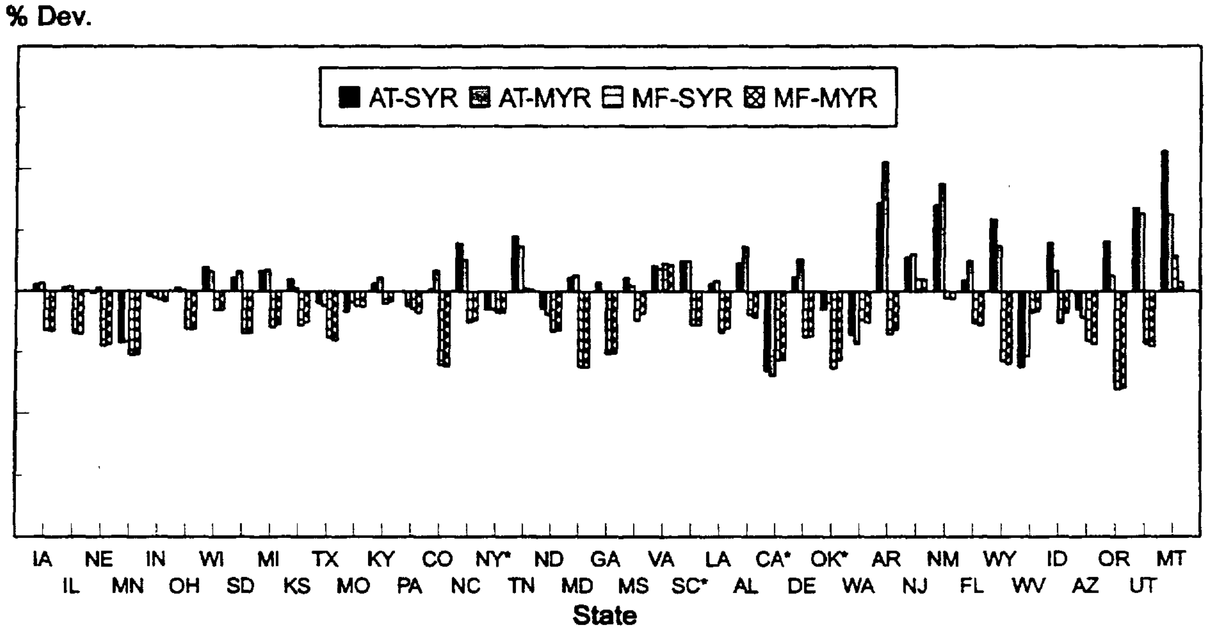
AT - area tract estimator, MF - multiple frame estimator, SYR - single year, MYR - multiyear

Figure 4. State Level Estimator Comparison for Corn Harvested Area



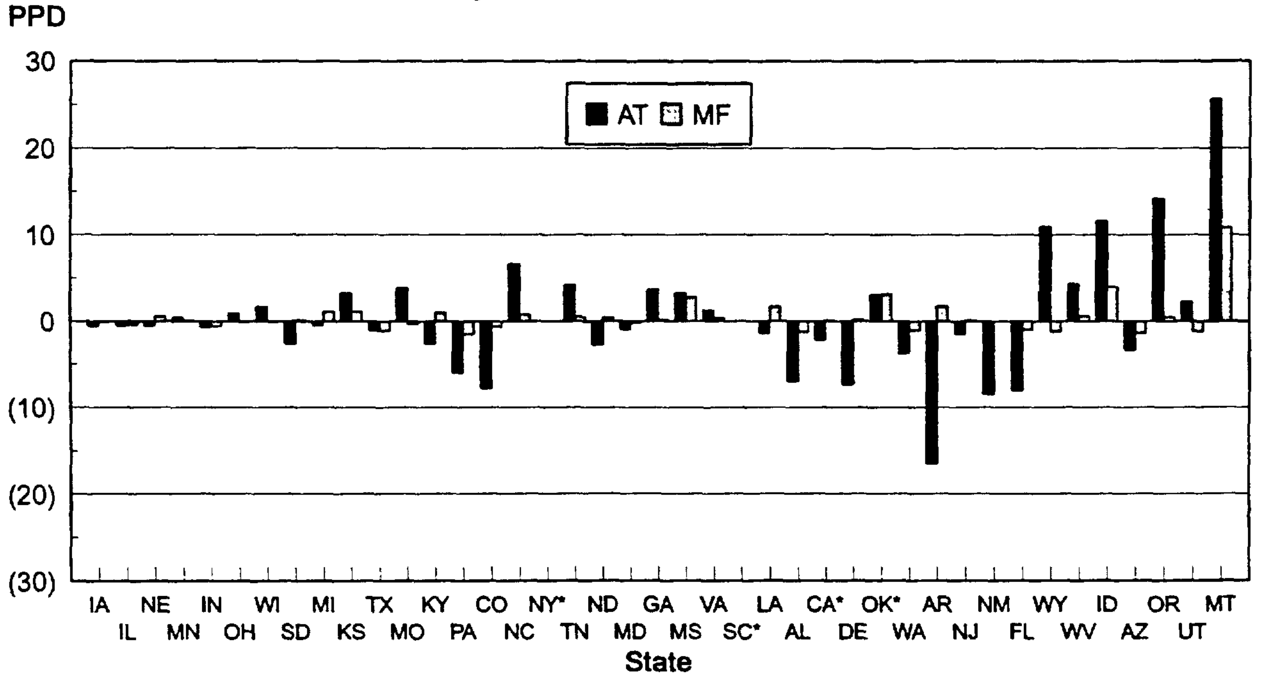


Percent Deviation of Indications from Final NASS Estimates



(scale suppressed as adminstratively confidential)

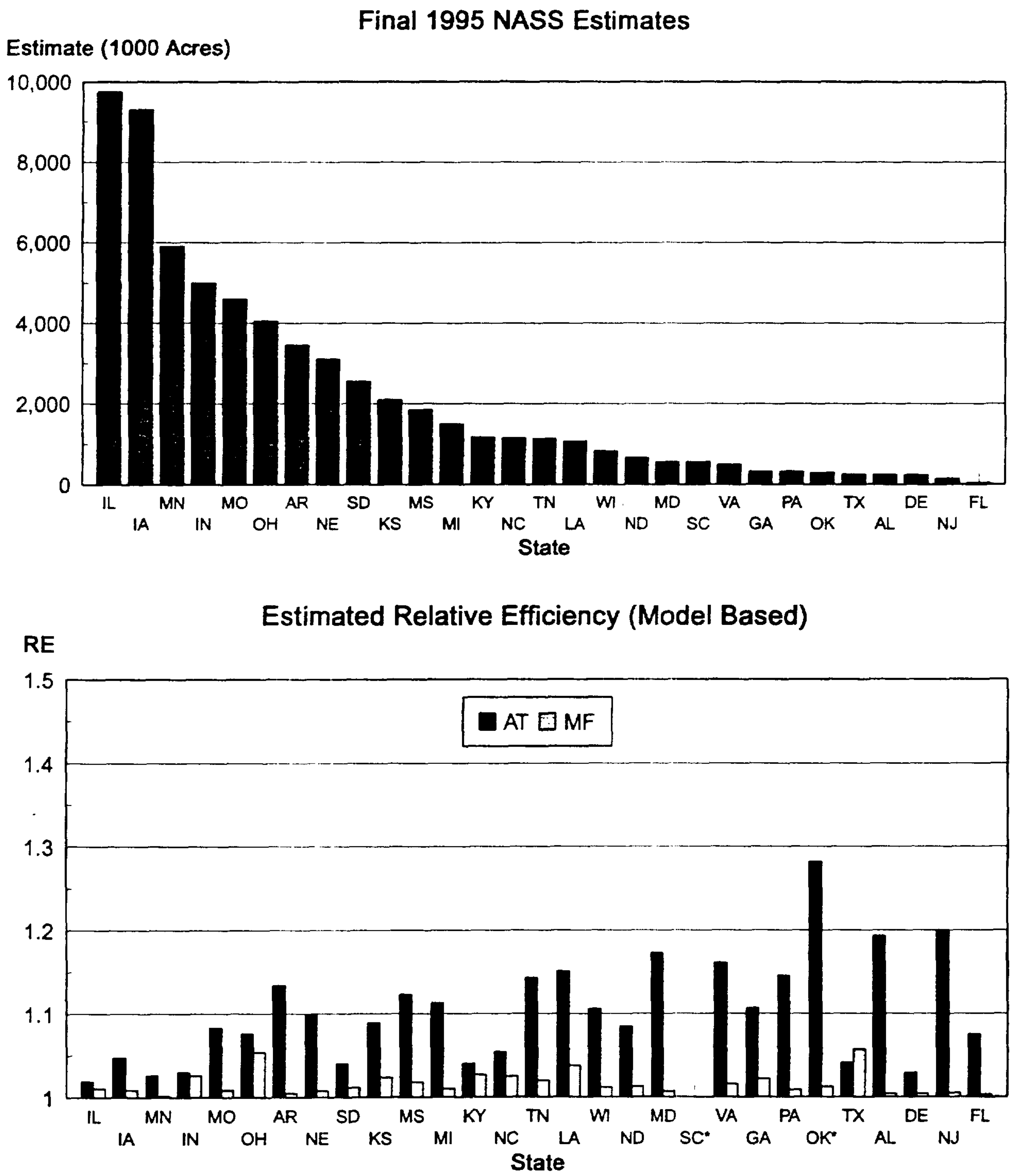
Percent Proximity Difference from Final NASS Estimates



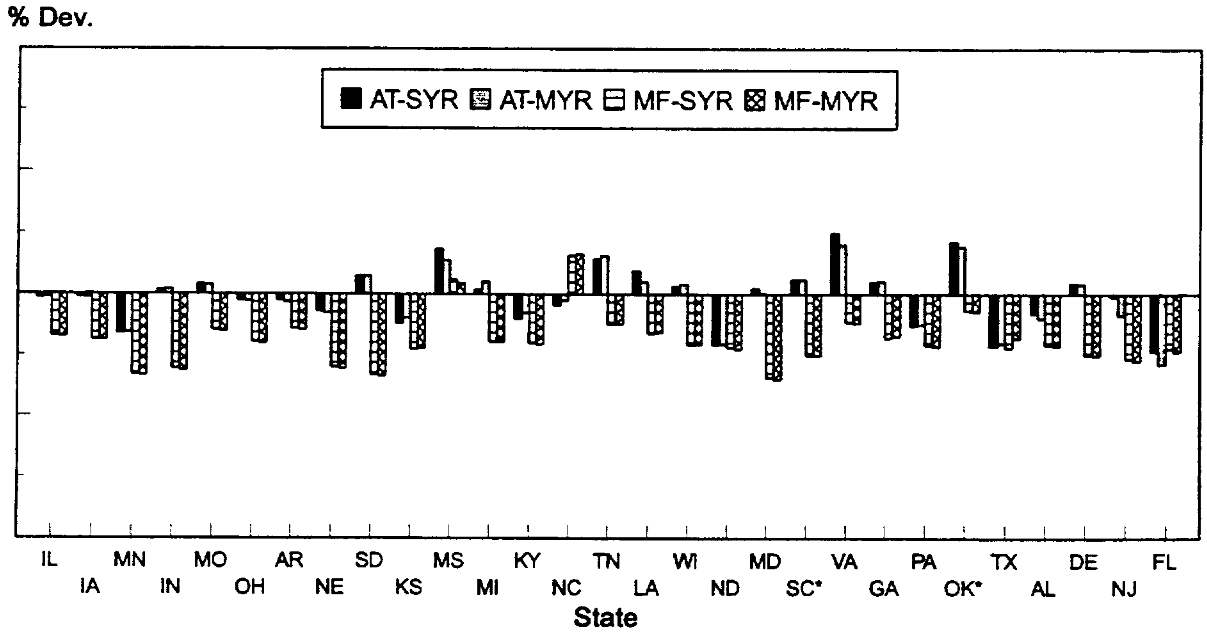
* - state received new area frame between 1992-95

AT - area tract estimator, MF - multiple frame estimator, SYR - single year, MYR - multiyear

Figure 5. State Level Estimator Comparison for Soybean Planted Area

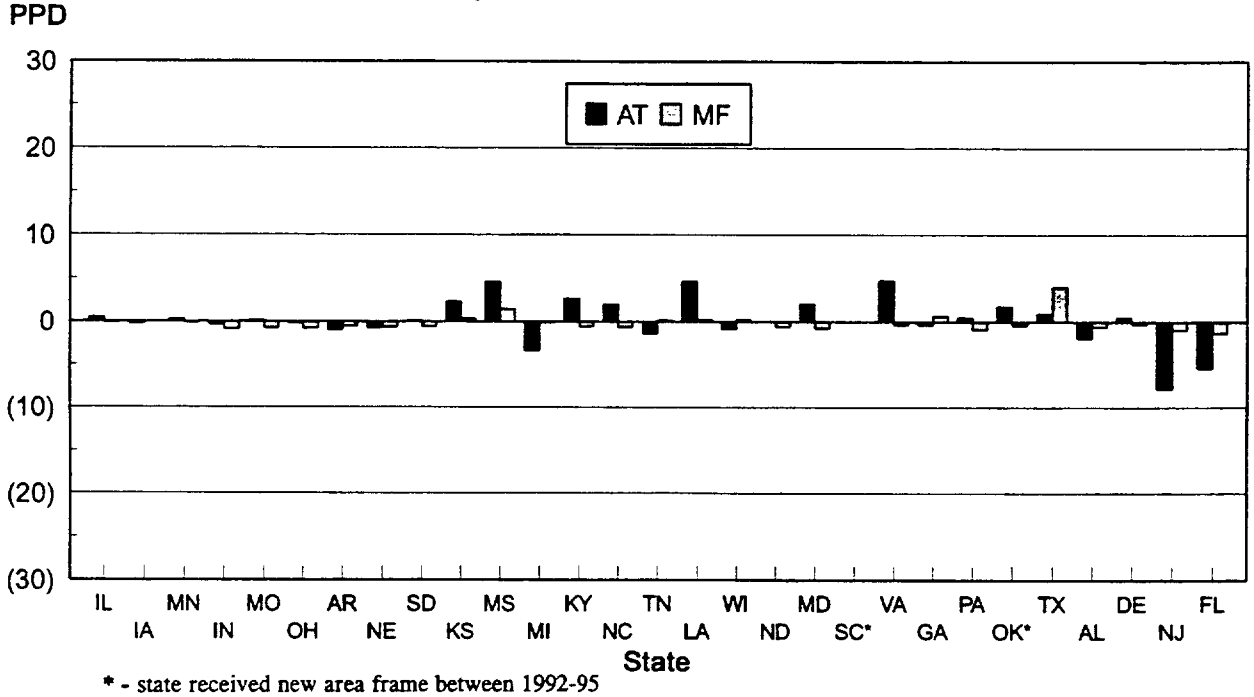


Percent Deviation of Indications from Final NASS Estimates



(scale suppressed as adminstratively confidential)

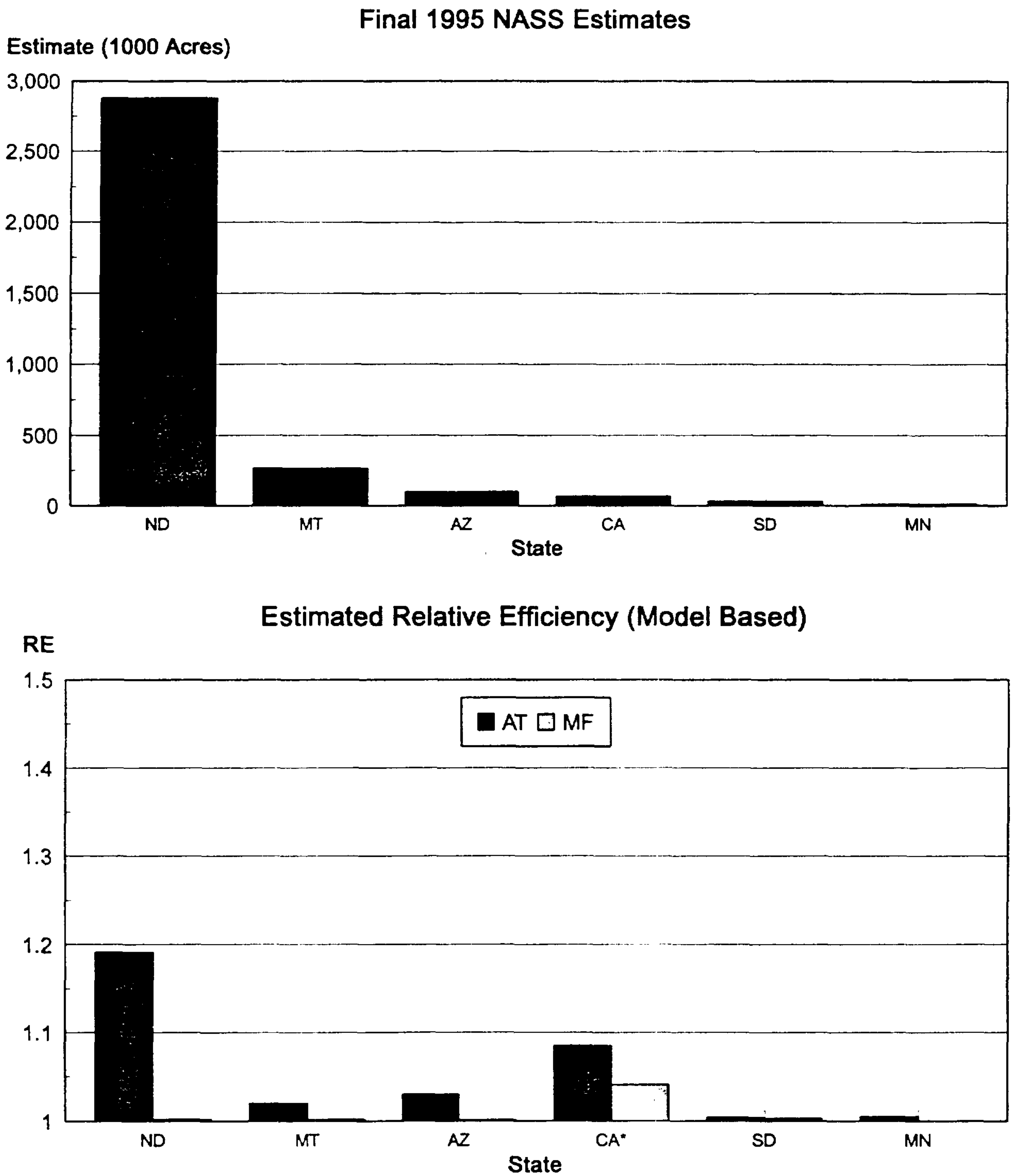
Percent Proximity Difference from Final NASS Estimates



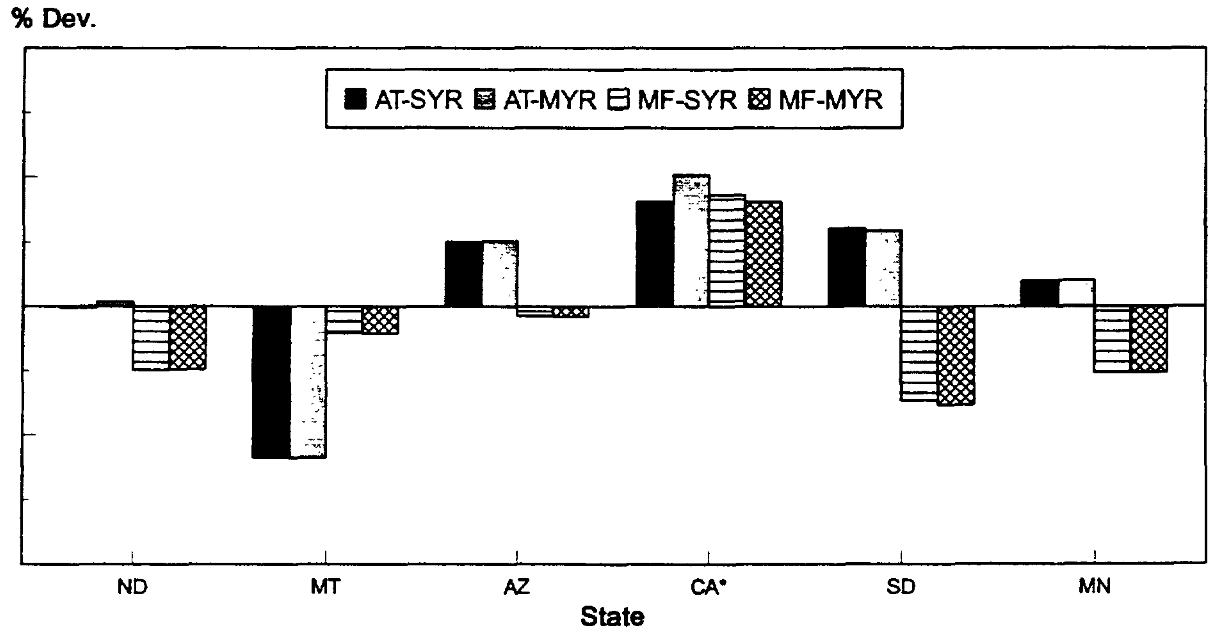
* - state received new area frame between 1992-95

AT - area tract estimator, MF - multiple frame estimator, SYR - single year, MYR - multiyear

Figure 6. State Level Estimator Comparison for Durum Wheat Harvested Area

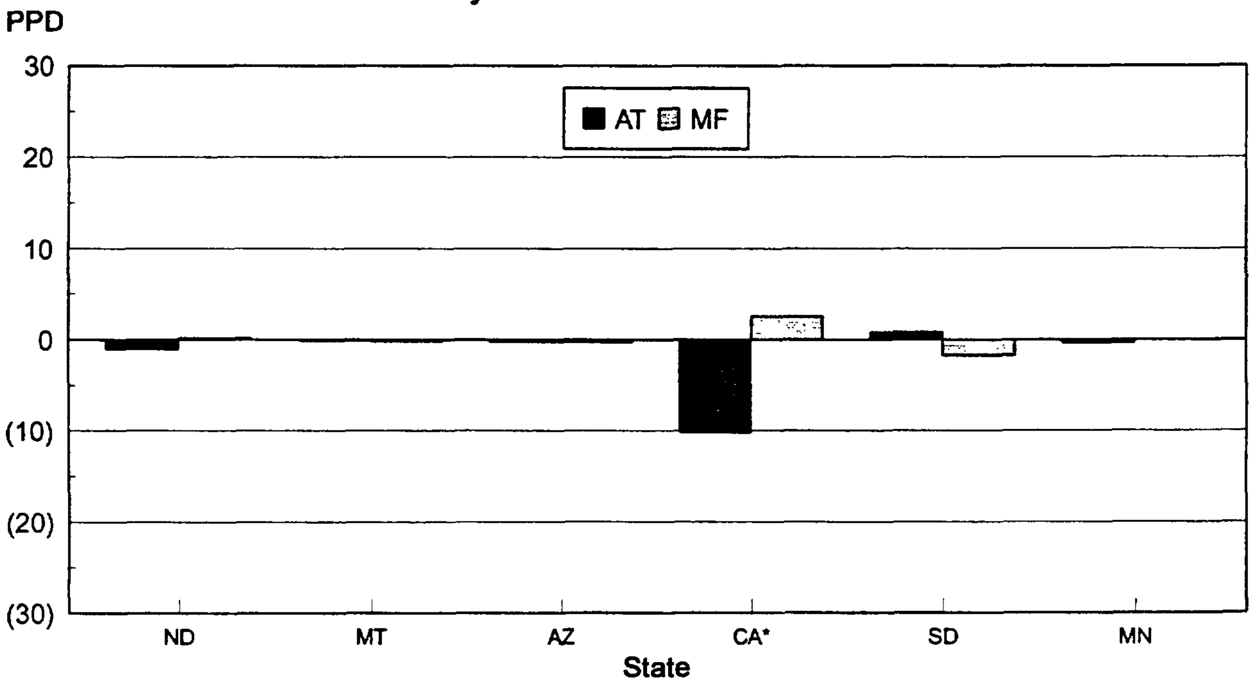


Percent Deviation of Indications from Final NASS Estimates



(scale suppressed as adminstratively confidential)

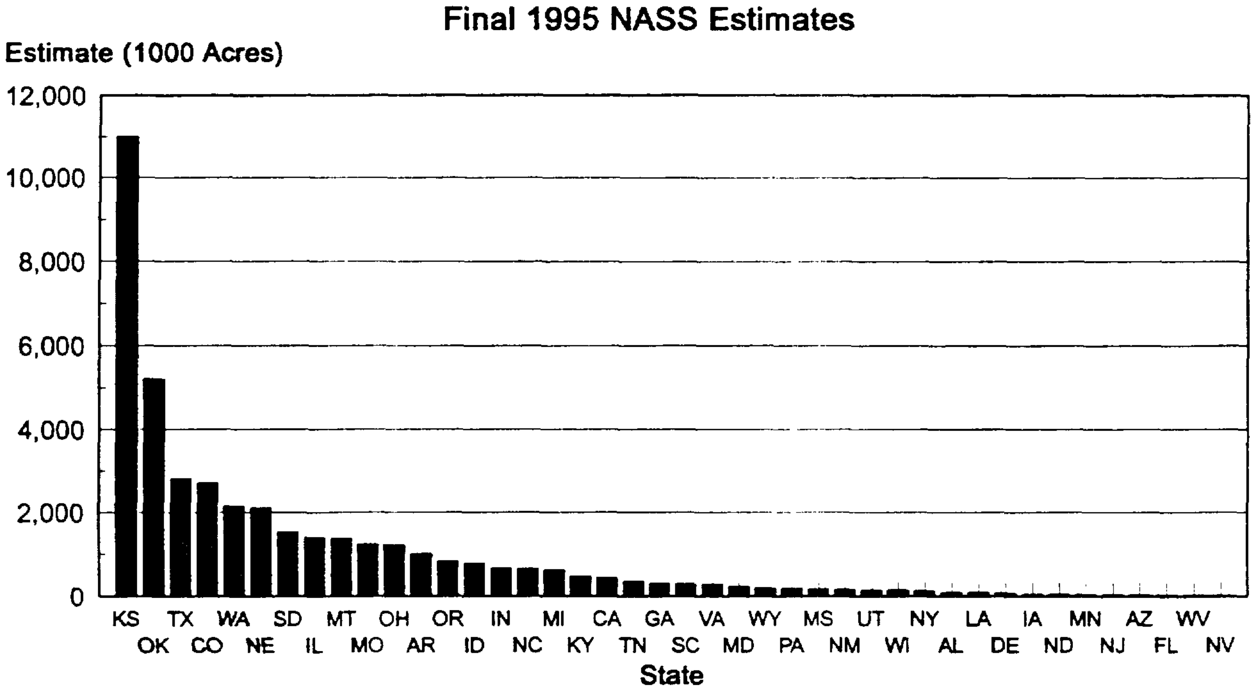
Percent Proximity Difference from Final NASS Estimates

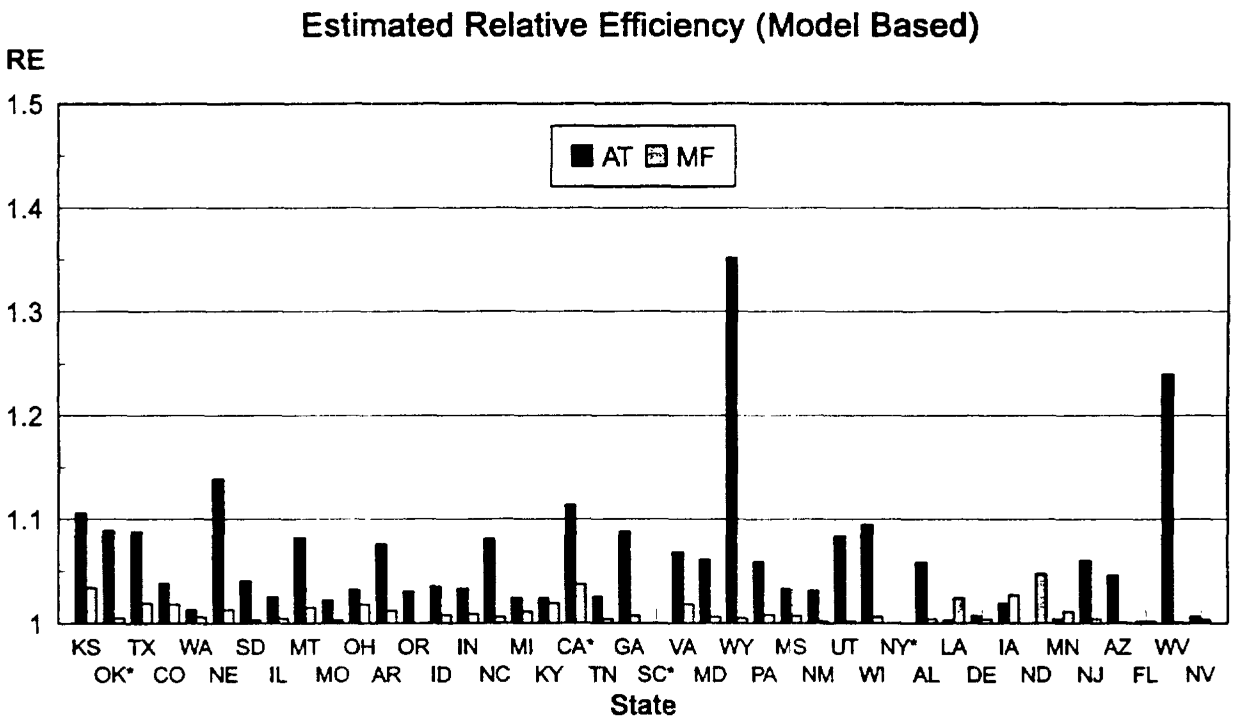


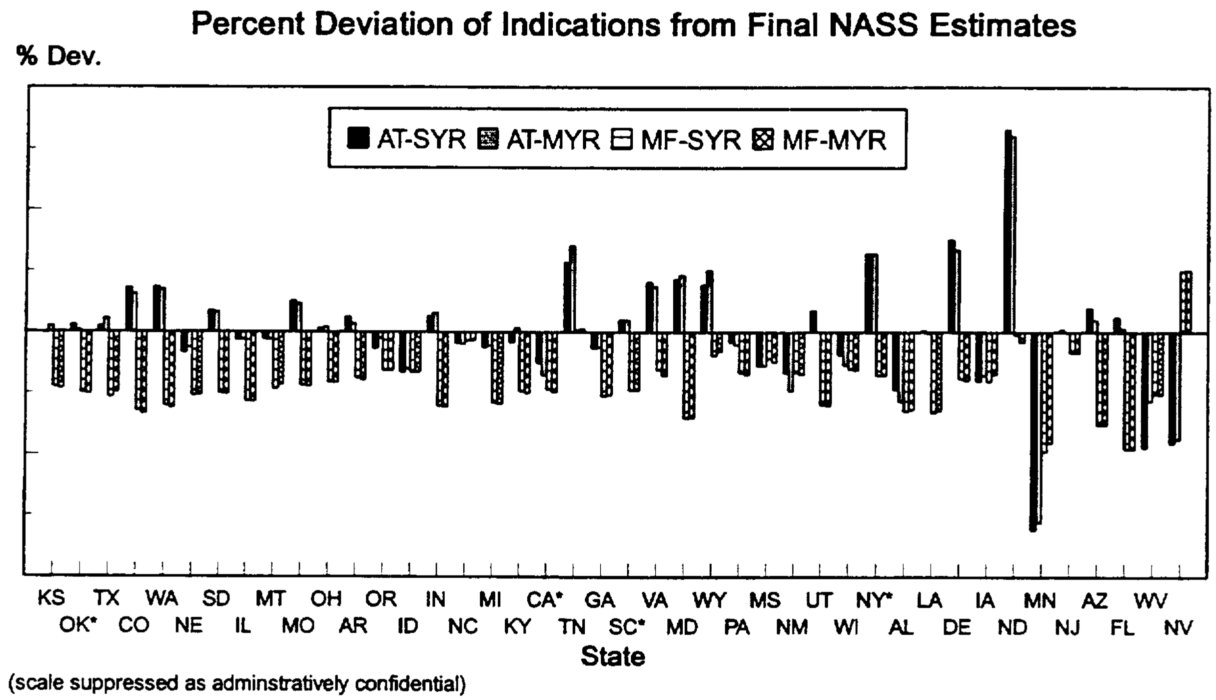
* - state received new area frame between 1992-95

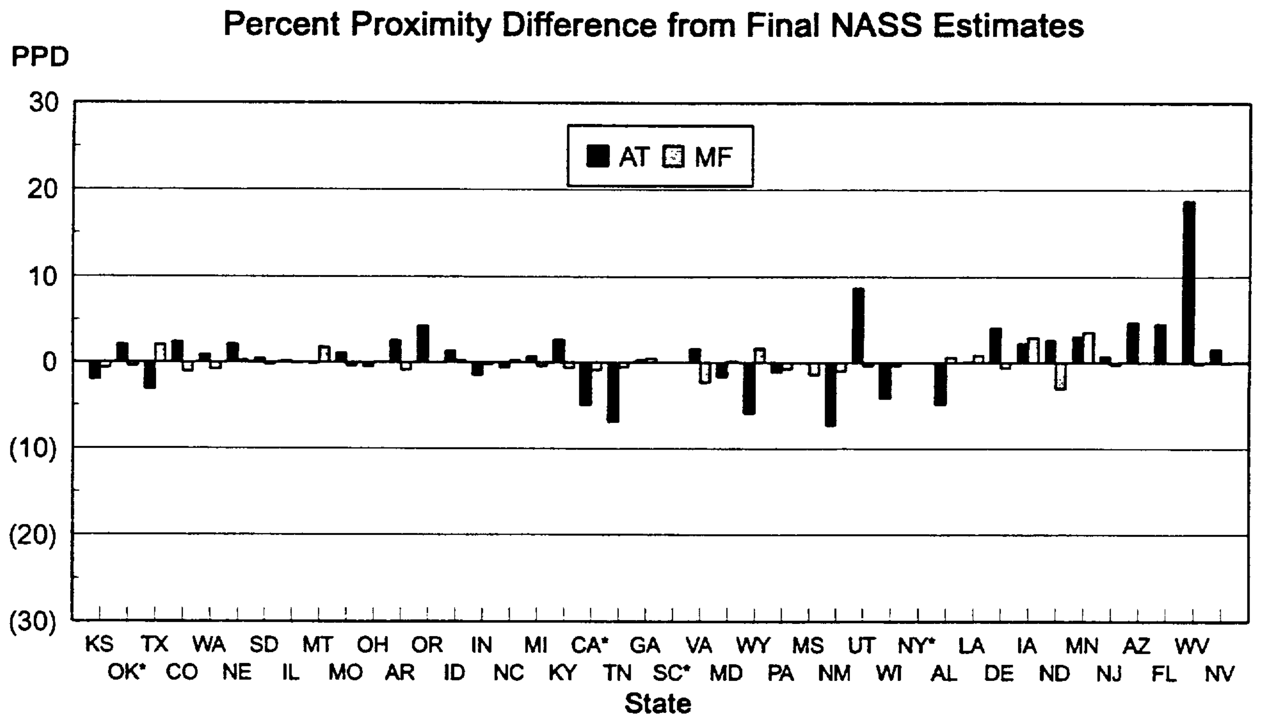
AT - area tract estimator, MF - multiple frame estimator, SYR - single year, MYR - multiyear

Figure 7. State Level Estimator Comparison for Winter Wheat Harvested Area





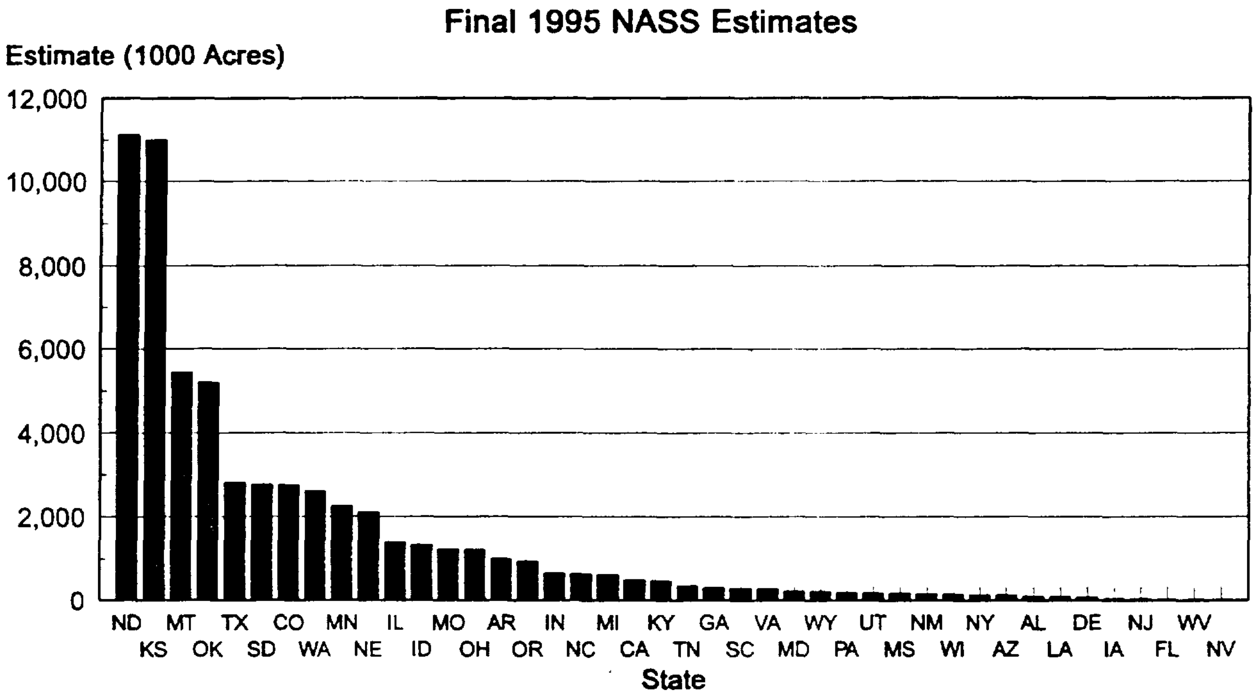


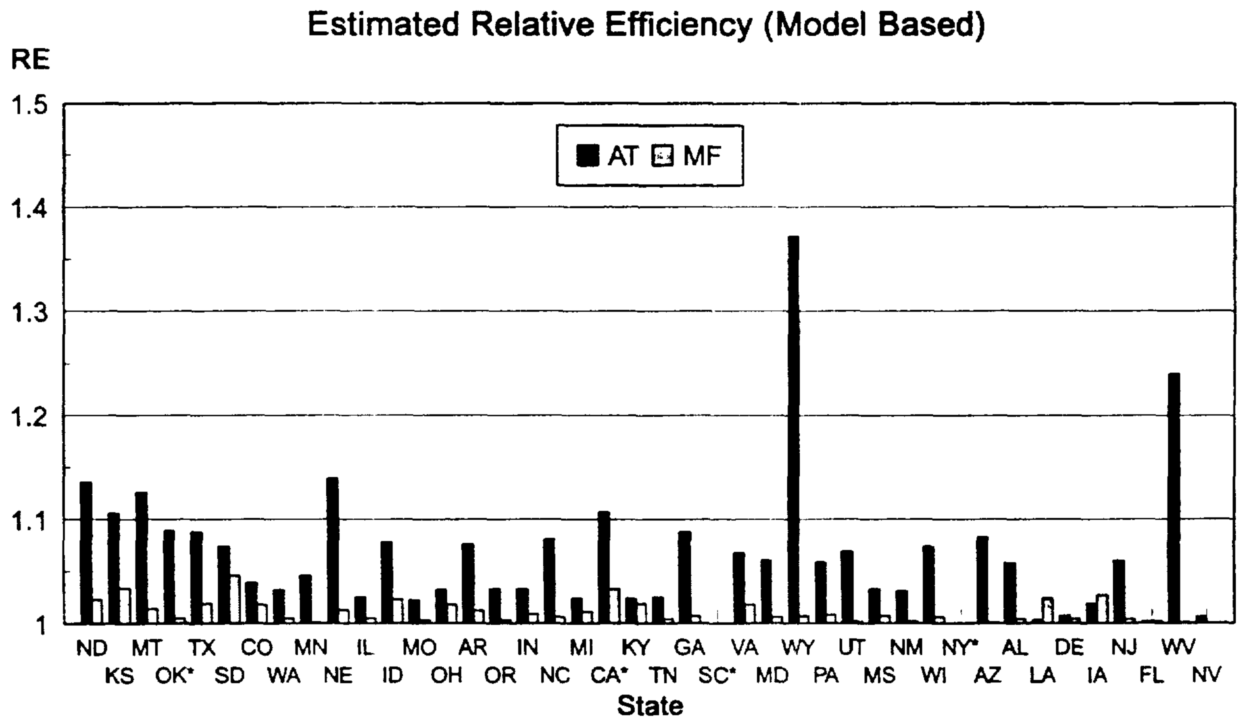


* - state received new area frame between 1992-95

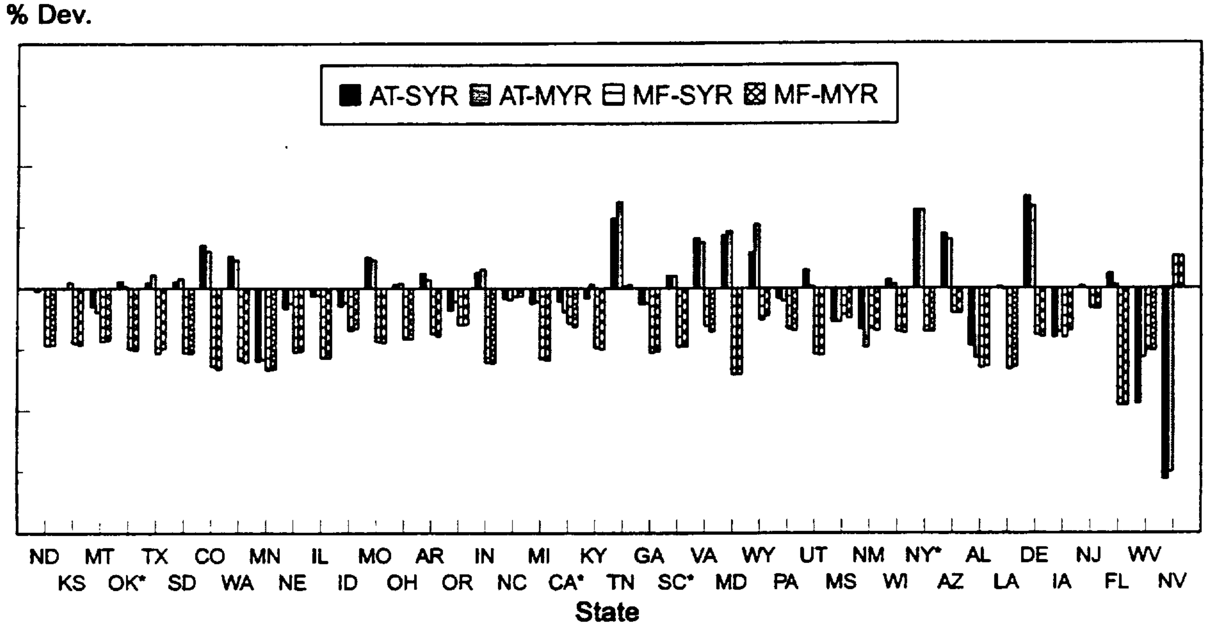
AT - area tract estimator, MF - multiple frame estimator, SYR - single year, MYR - multiyear

Figure 8. State Level Estimator Comparison for All Wheat Harvested Area



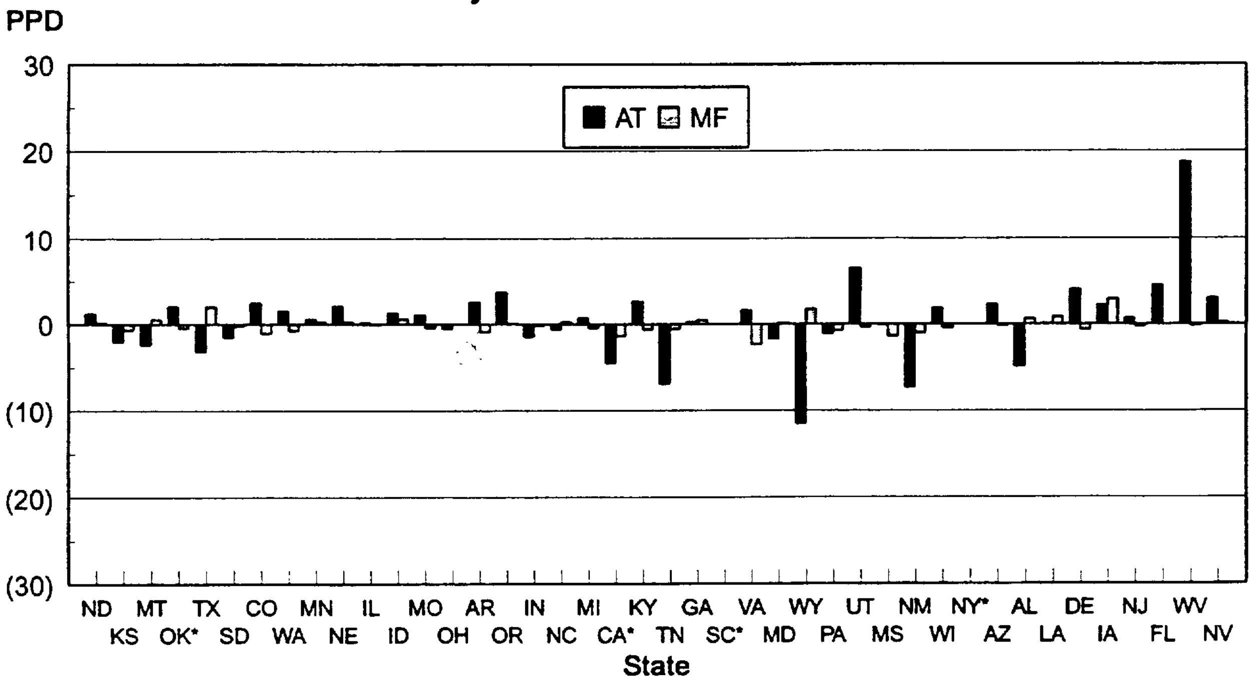


Percent Deviation of Indications from Final NASS Estimates



(scale suppressed as adminstratively confidential)

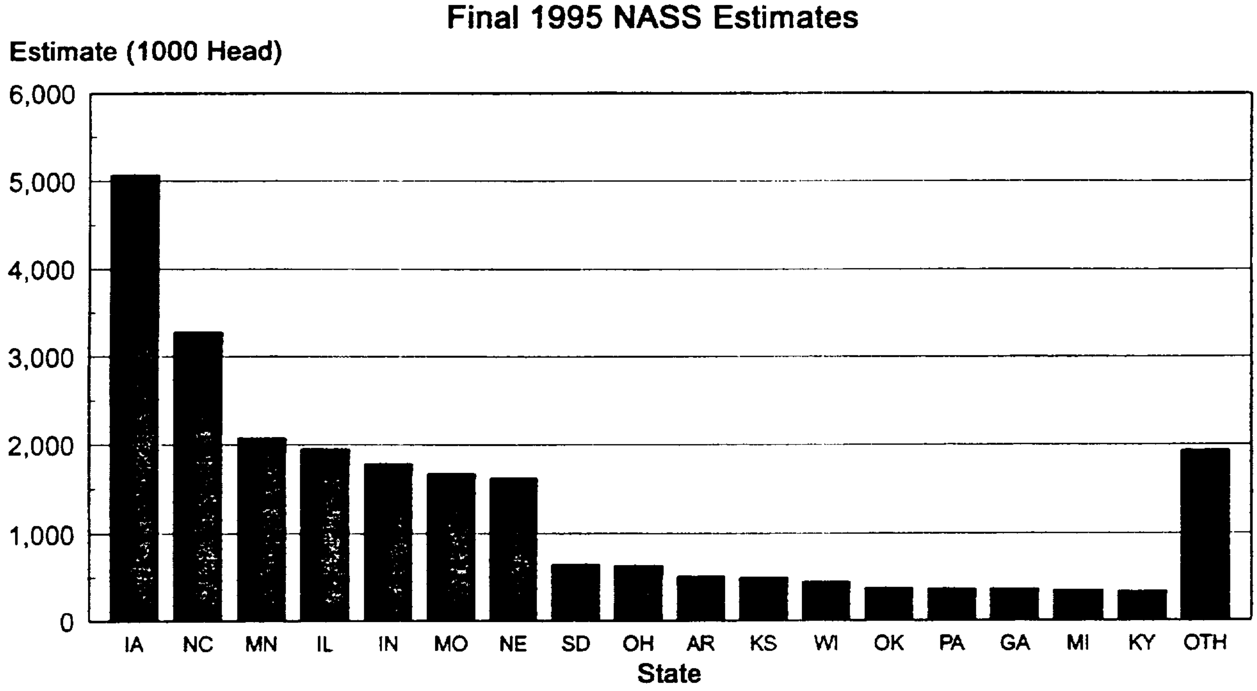
Percent Proximity Difference from Final NASS Estimates

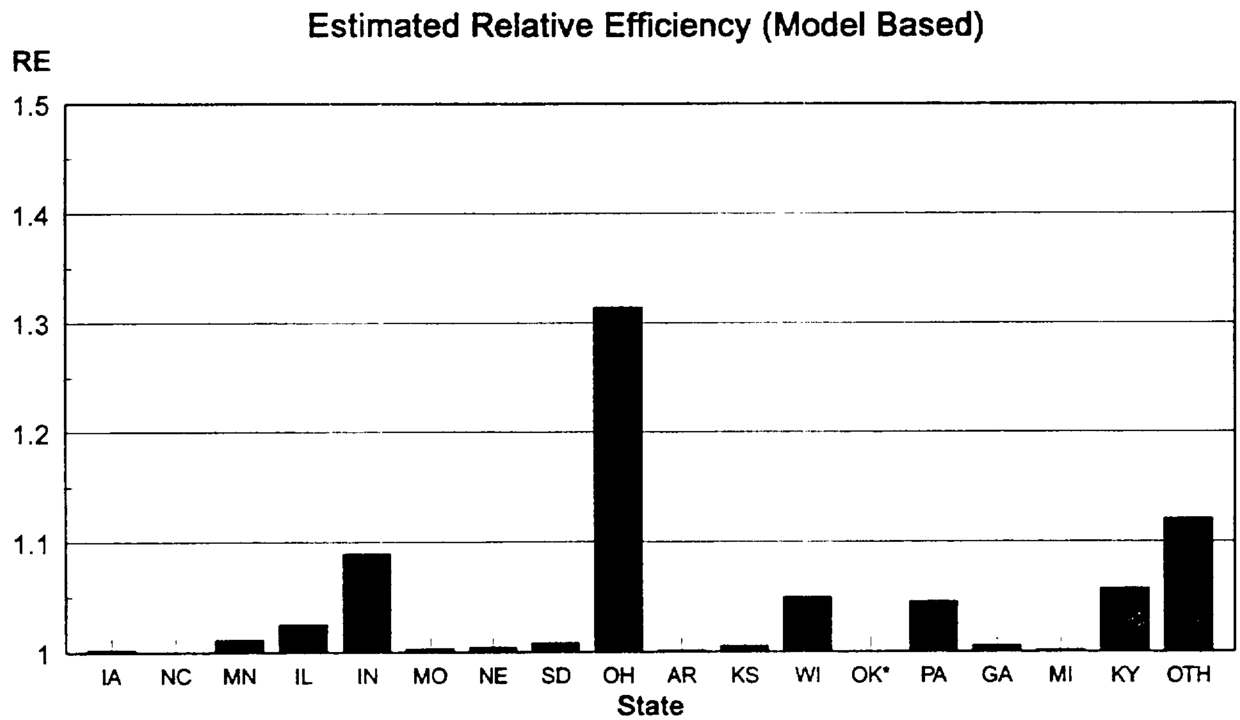


* - state received new area frame between 1992-95

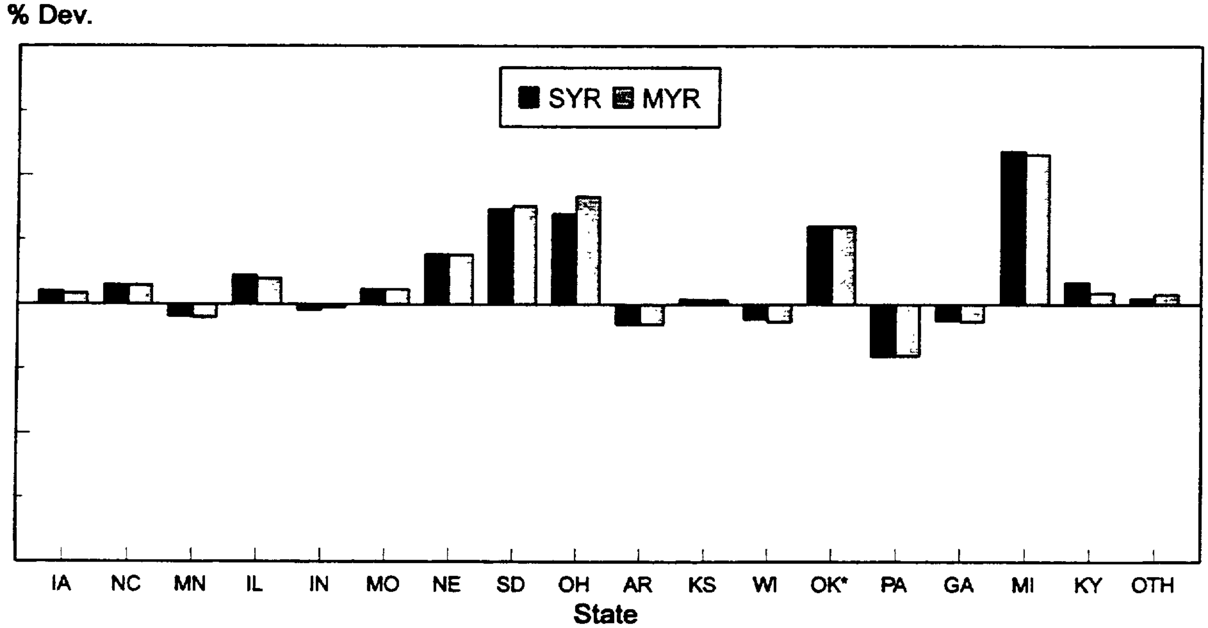
AT - area tract estimator, MF - multiple frame estimator, SYR - single year, MYR - multiyear

Figure 9. State Level Estimator Comparison for Pig Crop (Dec.-Feb.)



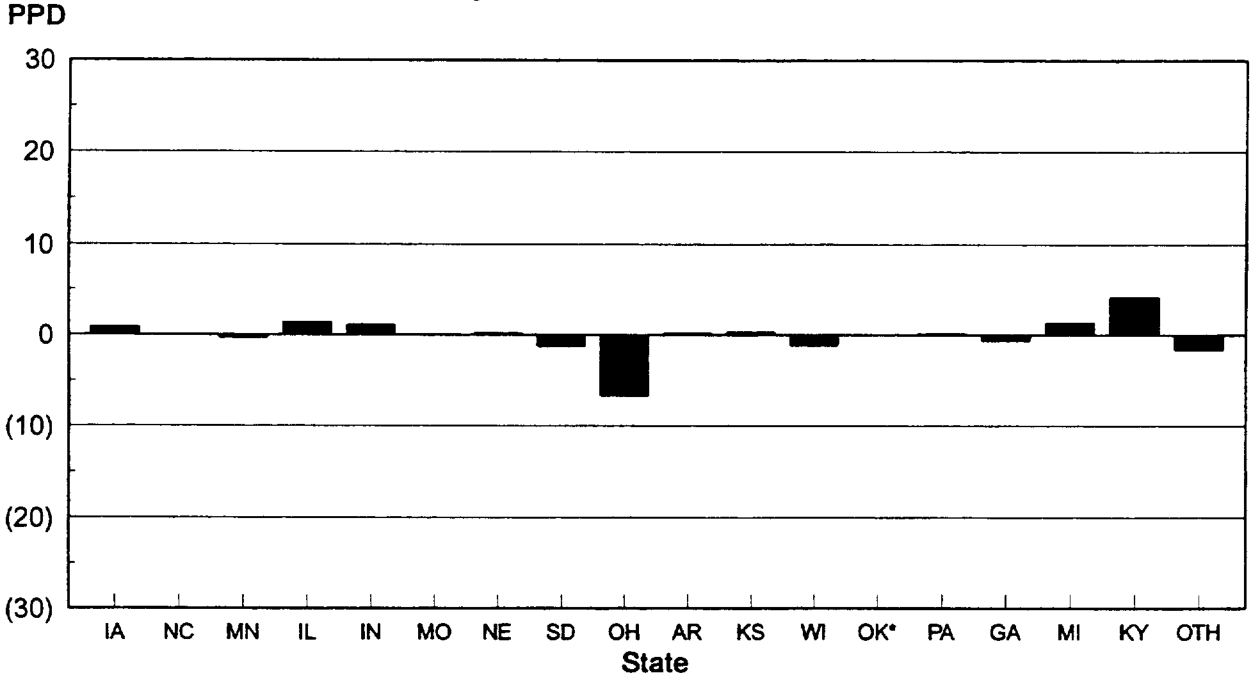


Percent Deviation of Indications from Final NASS Estimates



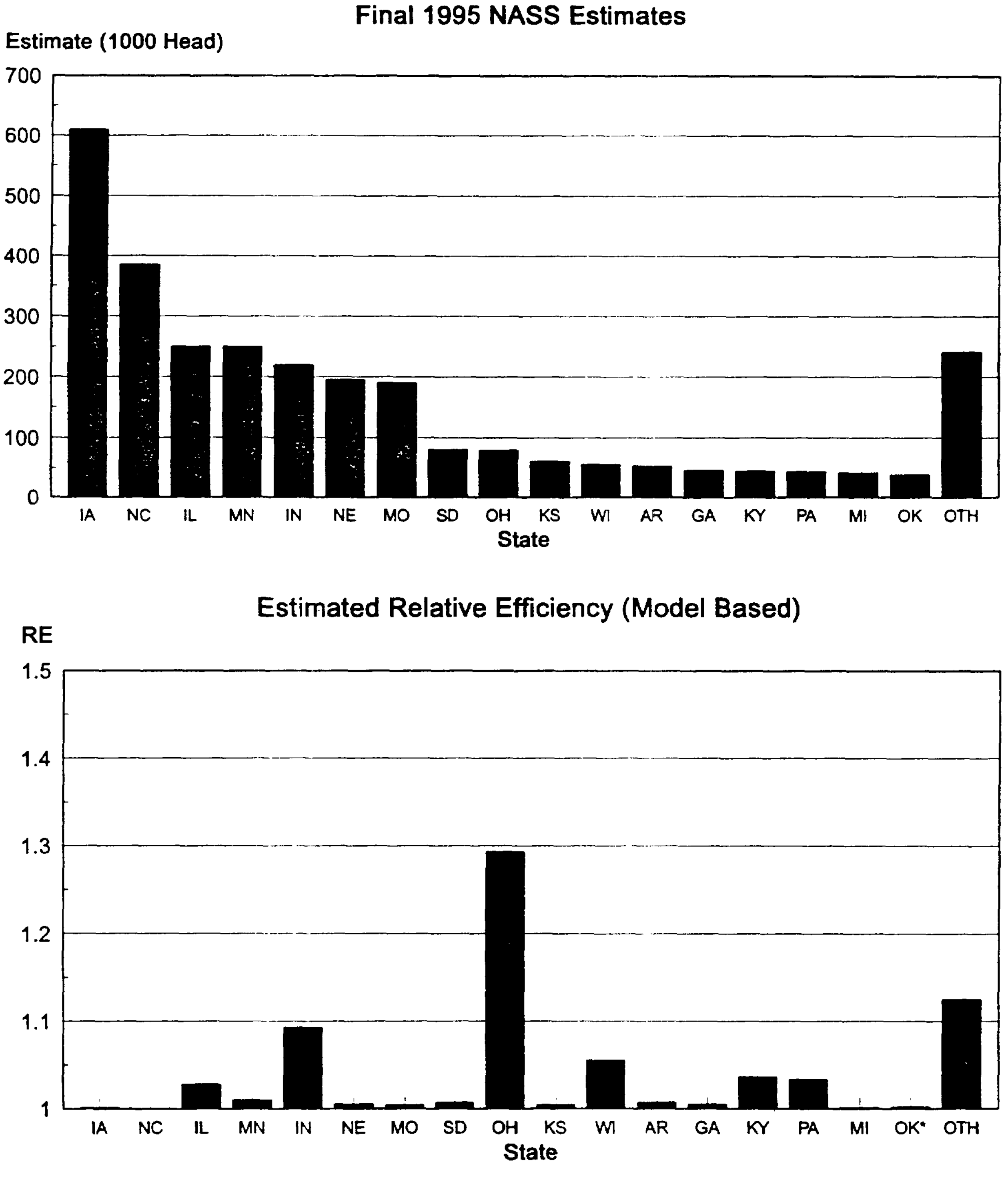
(scale suppressed as adminstratively confidential)

Percent Proximity Difference from Final NASS Estimates

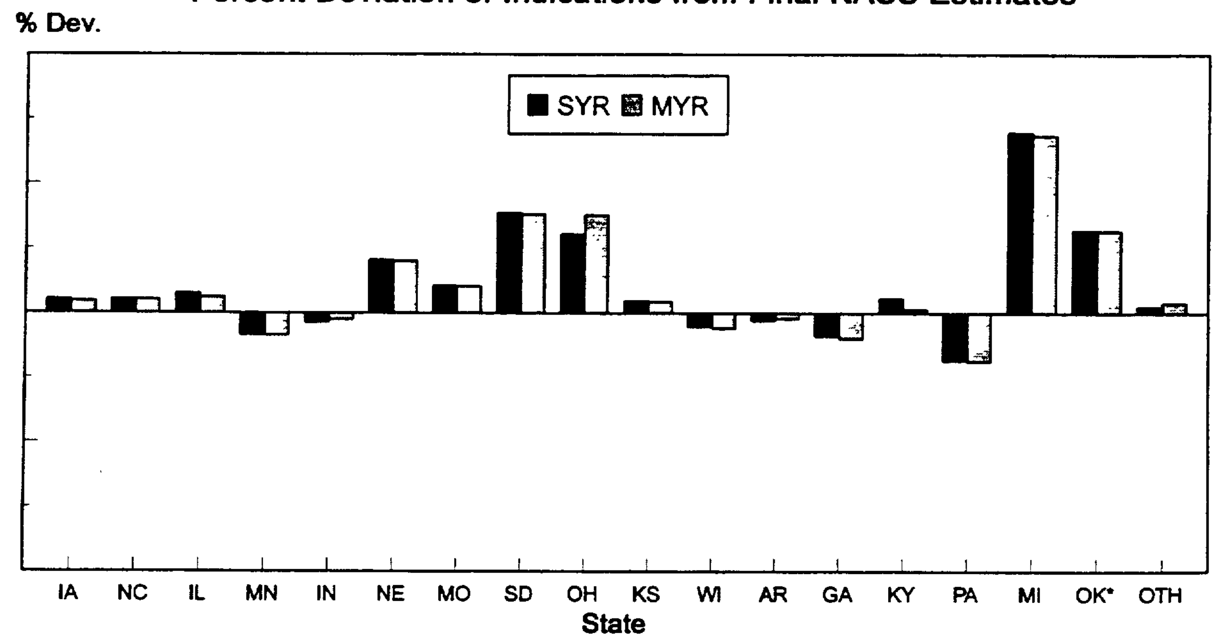


* - state received new area frame between 1992-95
SYR - single year, MYR - multiyear

Figure 10. State Level Estimator Comparison for Sows Farrowed (Dec.-Feb.)

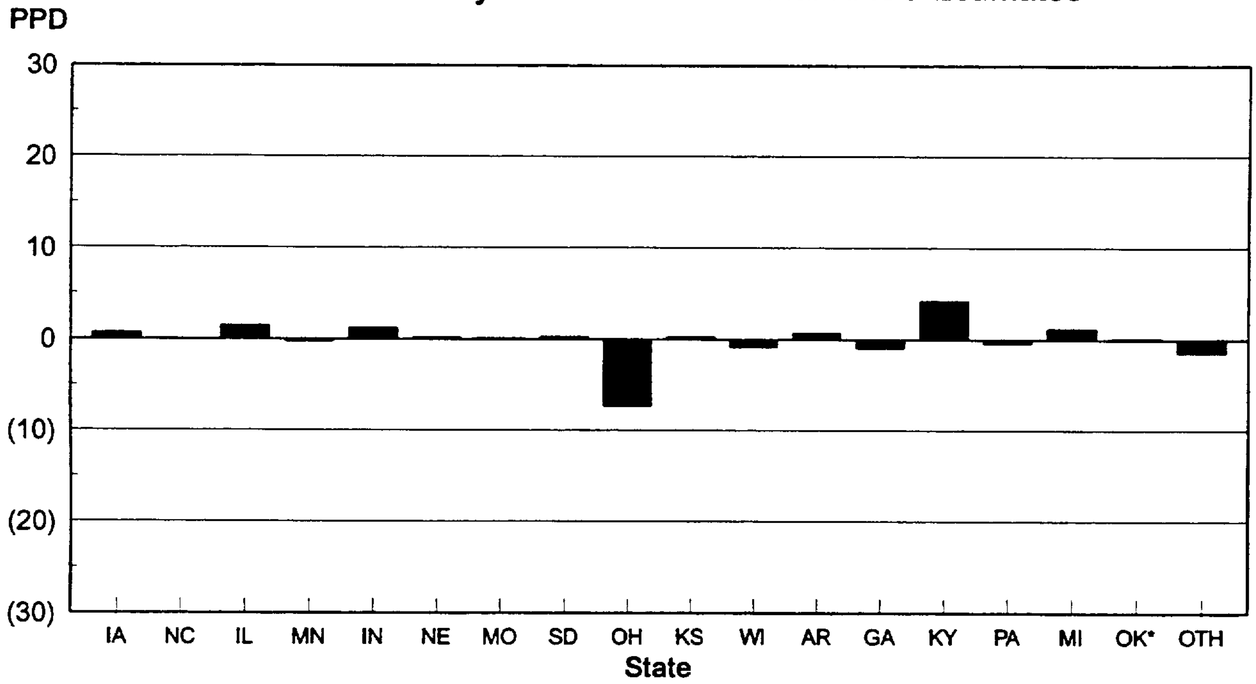


Percent Deviation of Indications from Final NASS Estimates



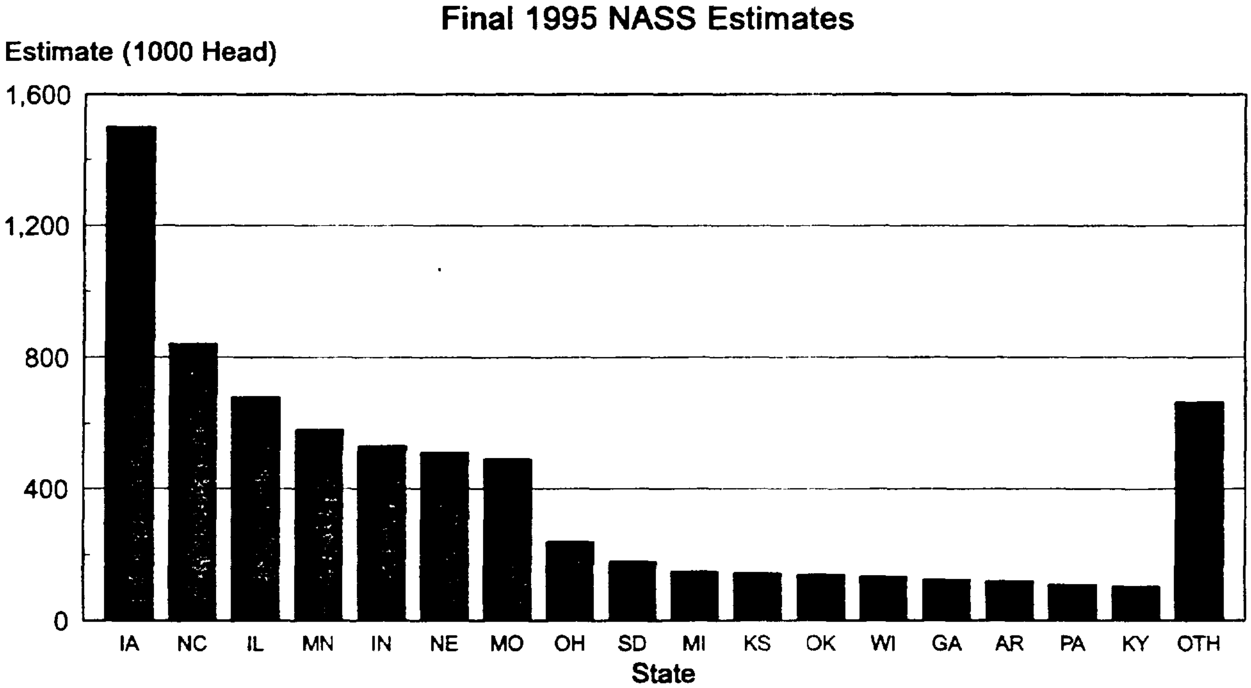
(scale suppressed as adminstratively confidential)

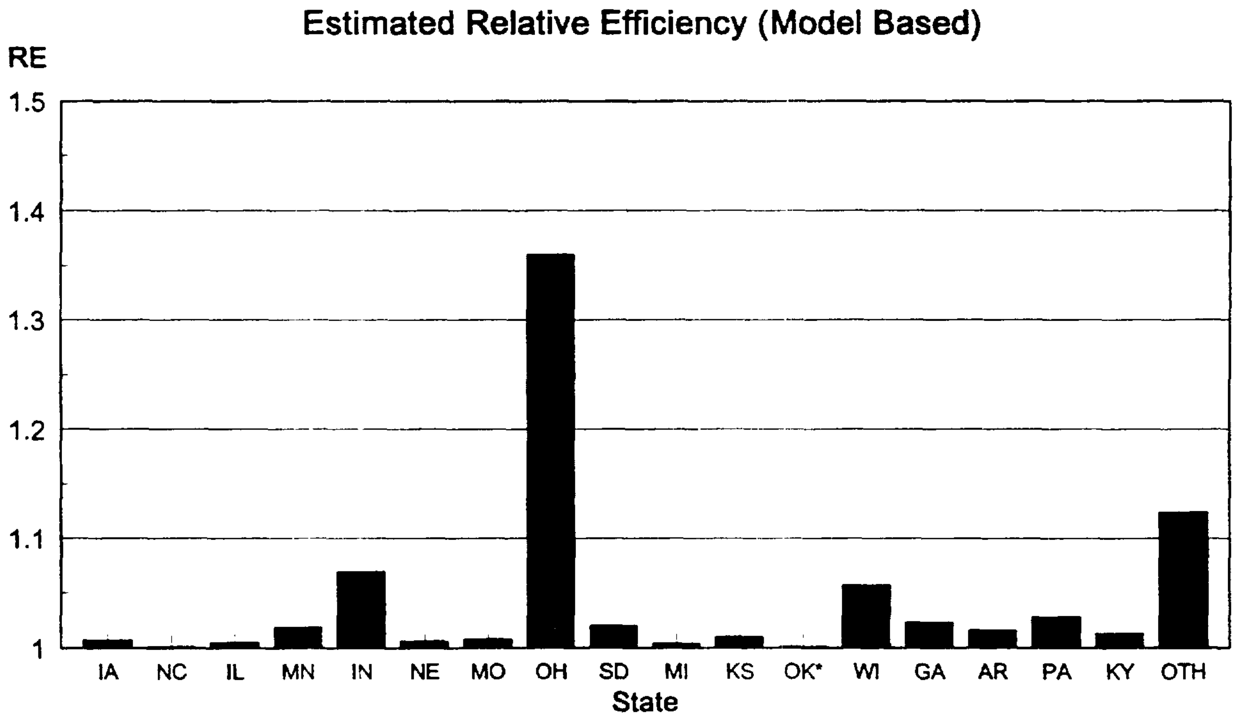
Percent Proximity Difference from Final NASS Estimates



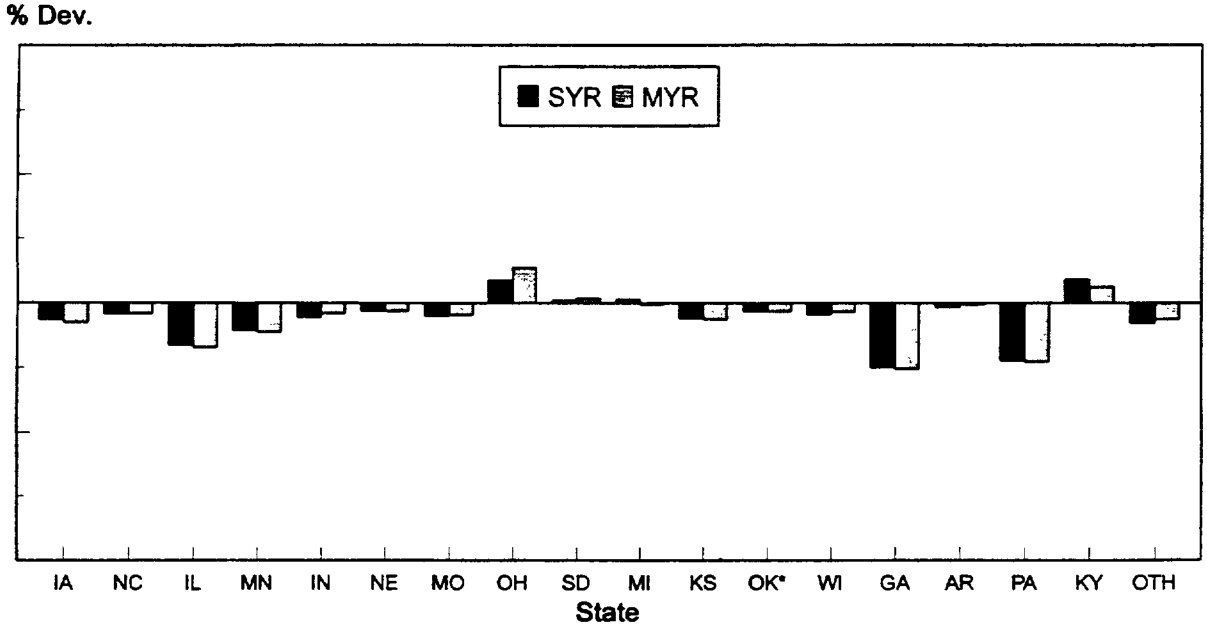
* - state received new area frame between 1992-95
SYR - single year, MYR - multiyear

Figure 11. State Level Estimator Comparison for Total Breeding Stock (Hogs)



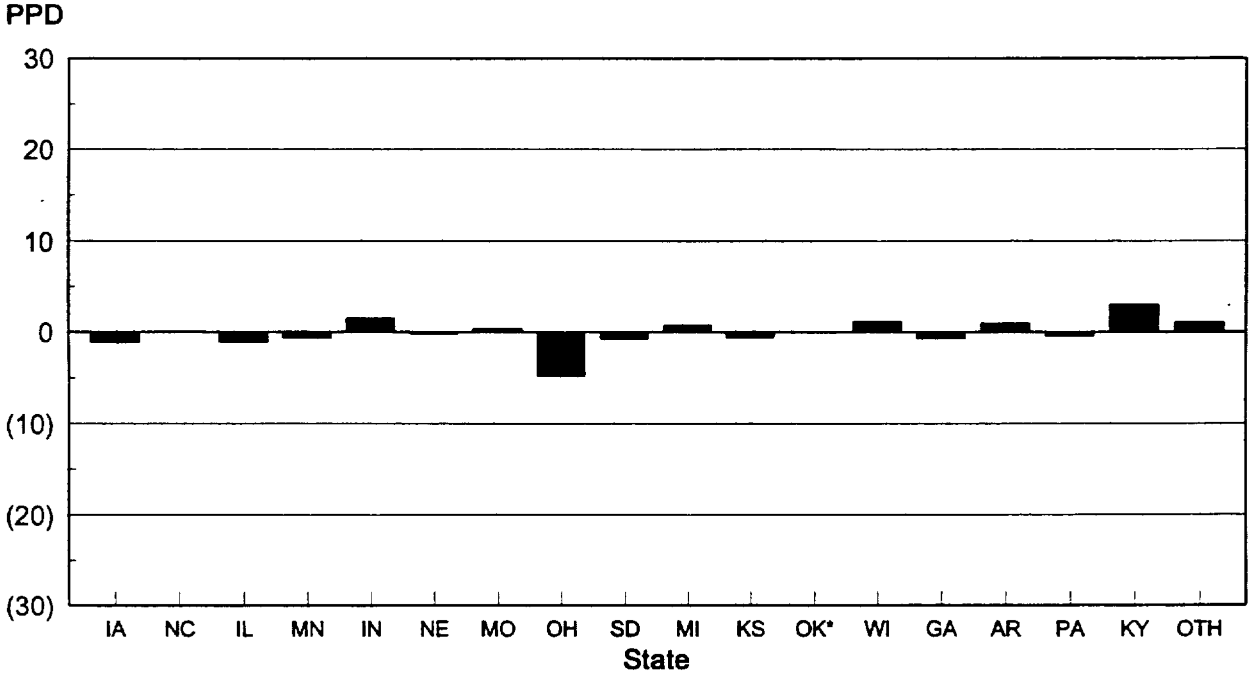


Percent Deviation of Indications from Final NASS Estimates



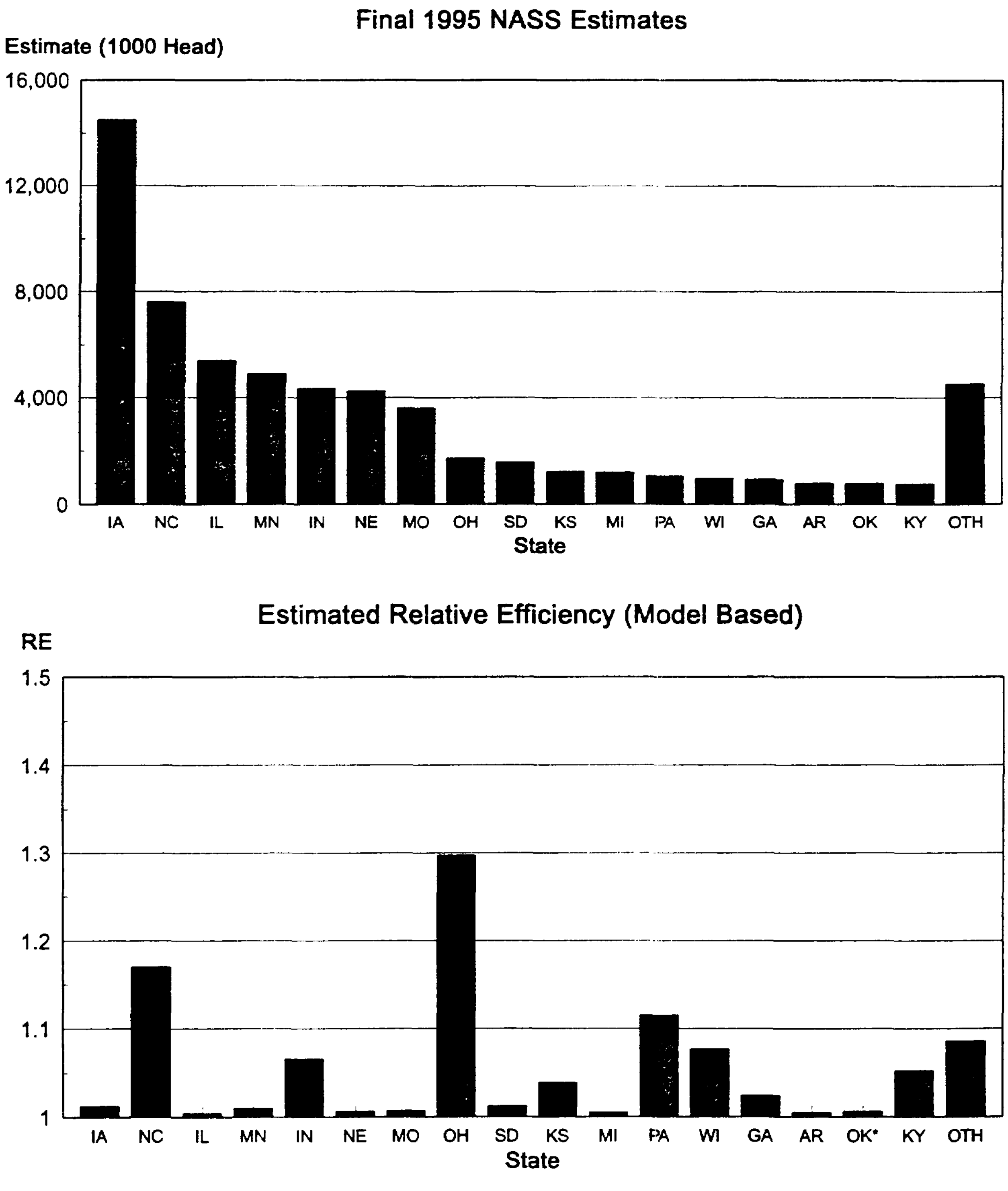
(scale suppressed as adminstratively confidential)

Percent Proximity Difference from Final NASS Estimates

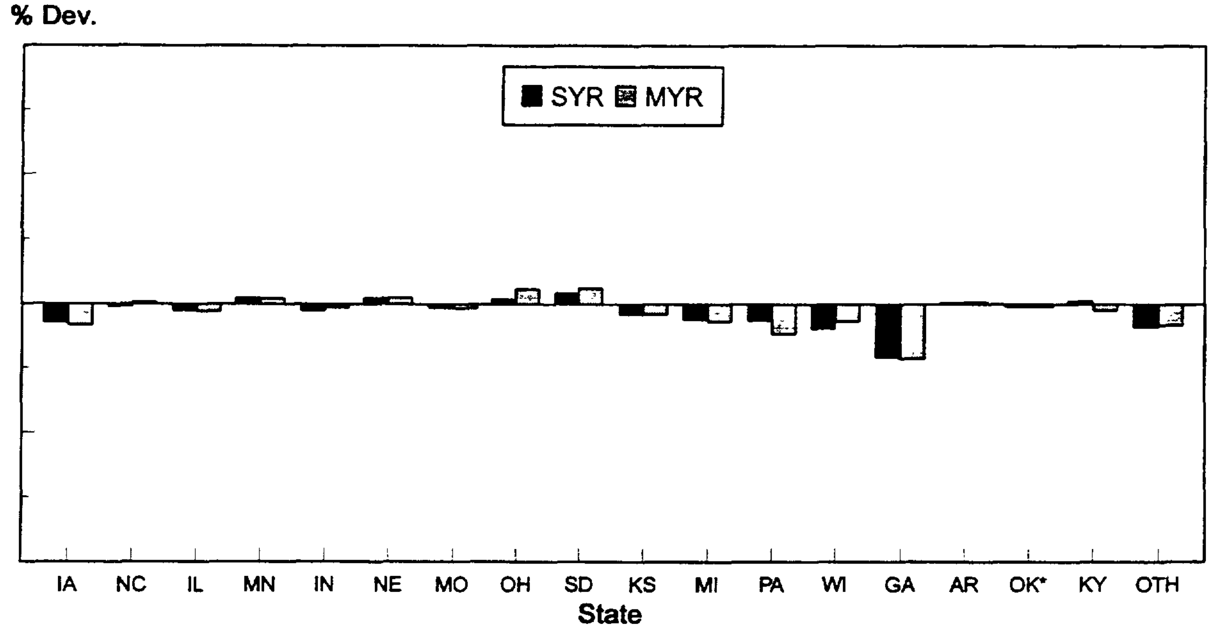


* - state received new area frame between 1992-95
SYR - single year, MYR - multiyear

Figure 12. State Level Estimator Comparison for Total Hogs and Pigs

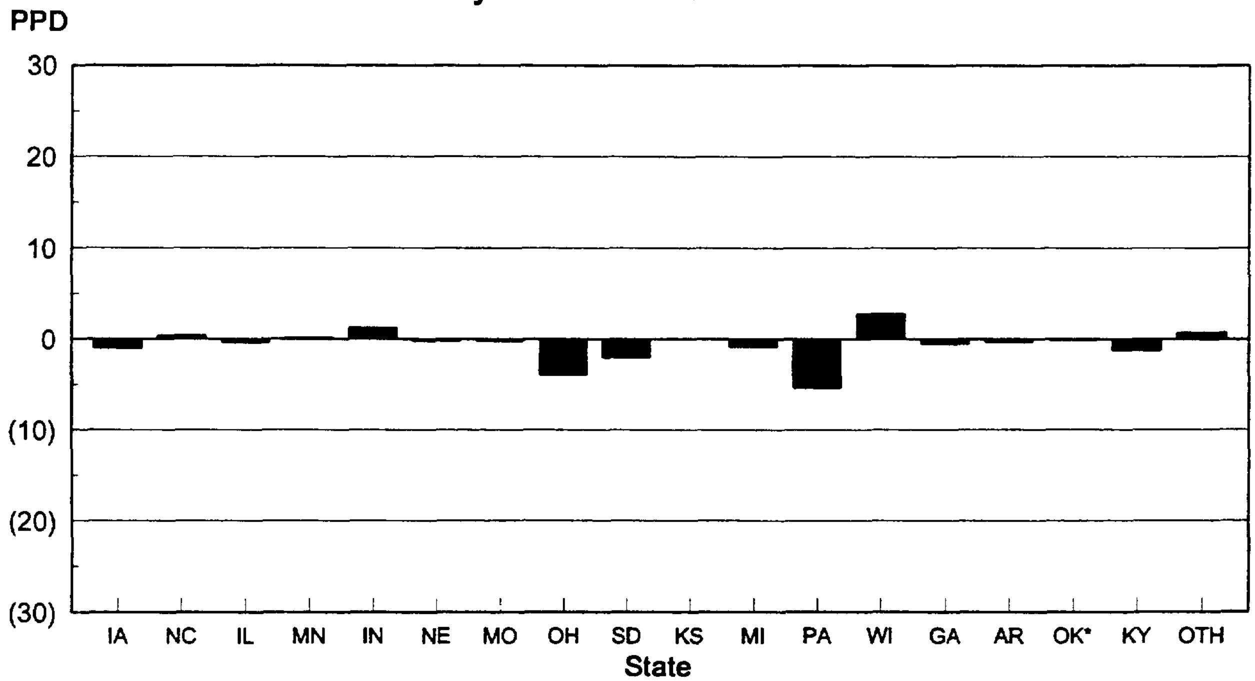


Percent Deviation of Indications from Final NASS Estimates



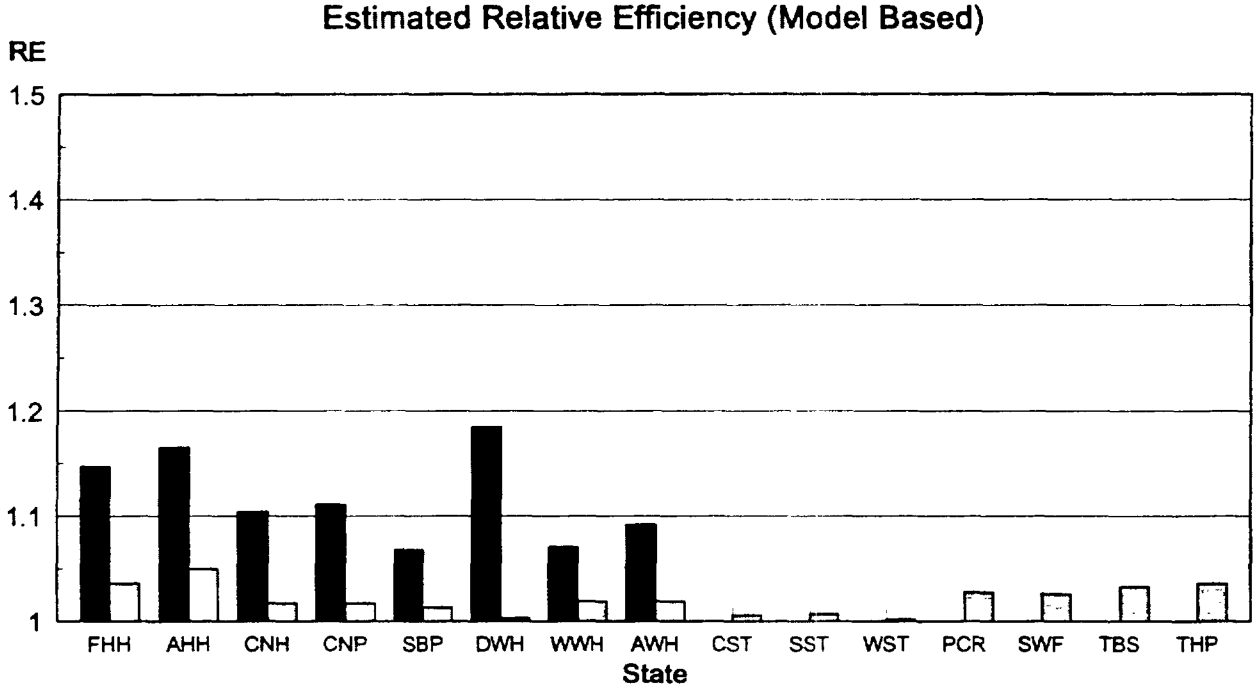
(scale suppressed as adminstratively confidential)

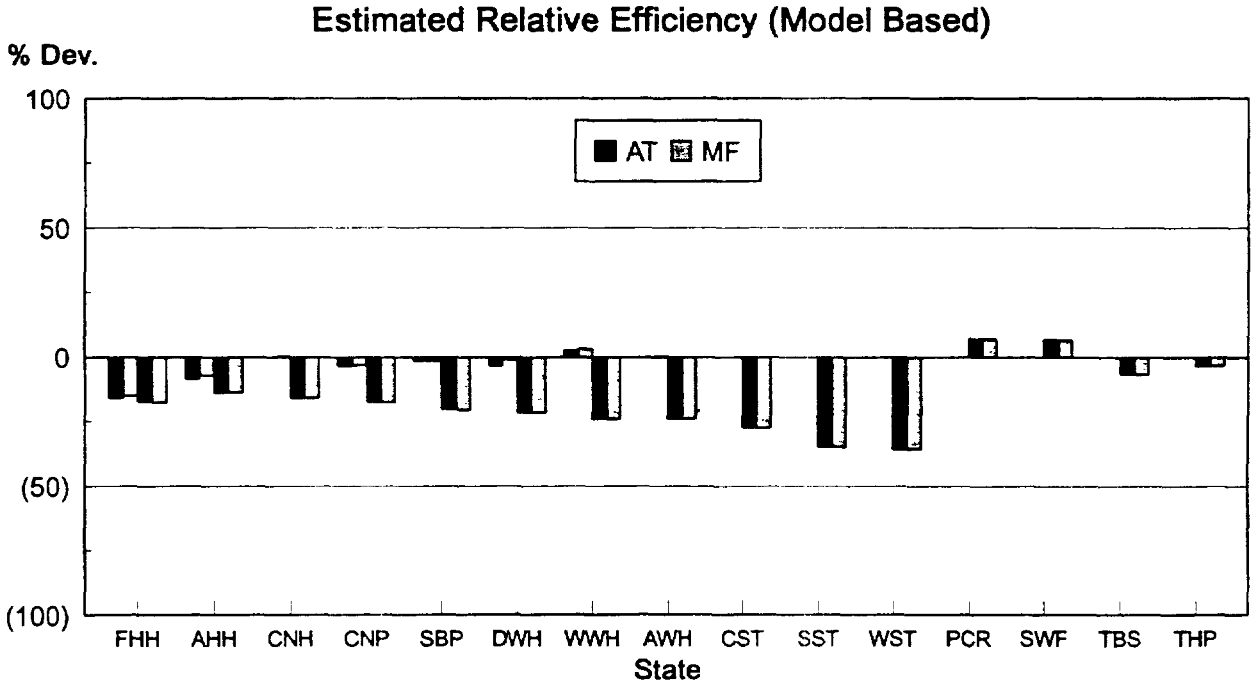
Percent Proximity Difference from Final NASS Estimates



* - state received new area frame between 1992-95
SYR - single year, MYR - multiyear

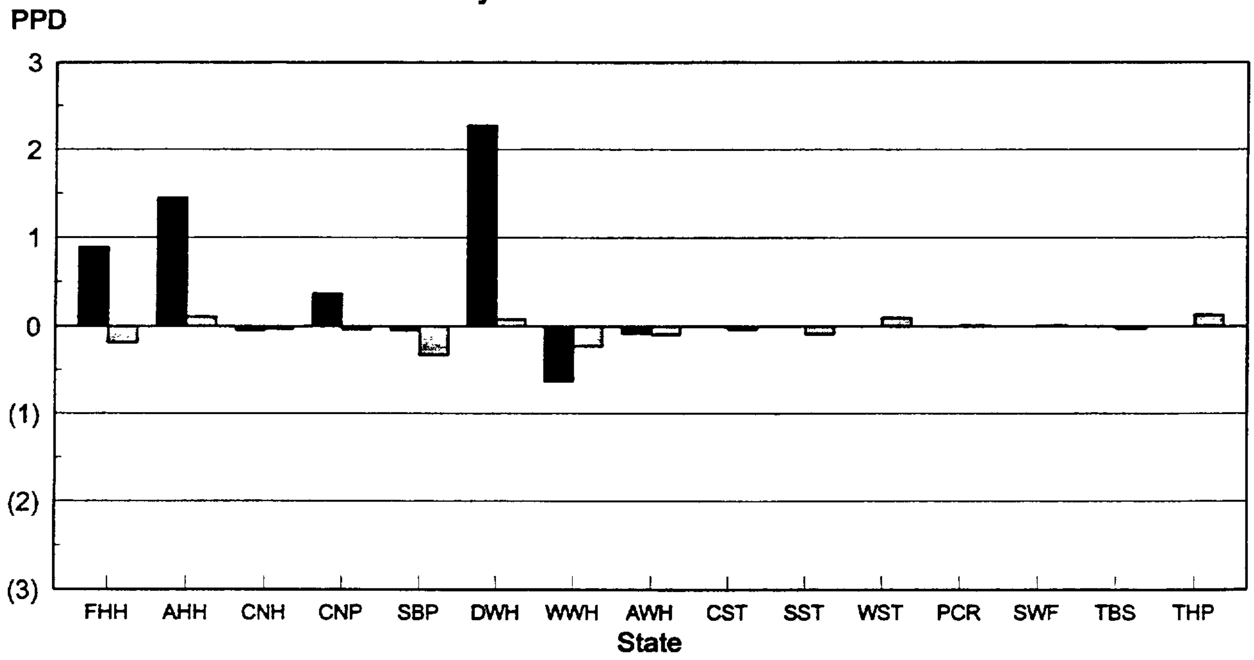
Figure 13. U.S. Level Estimator Comparison for all Items





(scale suppressed as adminstratively confidential)

Percent Proximity Difference from Final NASS Estimates



FHH - alfalfa hay harvested, AHH - all hay harvested, CNH - corn harvested, CNP-corn planted, SBP - soybean planted, DWH - durum wheat harvested, WWH - winter wheat harvested, AWH - all wheat harvested, CST - corn stocks (on farm), SST - soybean stocks (on farm), WST- wheat stocks (on farm) , PCR - pig crop (Dec.-Feb.), SWF - sows farrowed (Dec.-Feb.), TBS - total breeding stock (hogs), THP - total hogs and pigs.

APPENDIX - Derivation of Variance Formula for Bootstrap Study

In Section 4, it was stated that $\mathbf{f}'\mathbf{f}$ is an unbiased estimator of the random error variance σ_e^2, in the transformed model. The following is proof of that statement.

The unadjusted residual vector of the transformed model $\mathbf{z} = \mathbf{B}\alpha + \mathbf{f}$ is:

$$\hat{\mathbf{f}} = \mathbf{z} - \mathbf{B}\hat{\alpha}$$

The expected sum of squared unadjusted residuals is given by:

$$\begin{aligned}
E[\hat{\mathbf{f}}'\hat{\mathbf{f}}] &= E[(\mathbf{z} - \mathbf{B}\hat{\alpha})'(\mathbf{z} - \mathbf{B}\hat{\alpha})] \\
&= E[\mathbf{z}'\mathbf{z} - \hat{\alpha}'\mathbf{B}'\mathbf{z} - \mathbf{z}'\mathbf{B}\hat{\alpha} + \hat{\alpha}'\mathbf{B}'\mathbf{B}\hat{\alpha}] \\
&= E[(\mathbf{B}\alpha + \mathbf{f})'(\mathbf{B}\alpha + \mathbf{f})] - E(\hat{\alpha}'\mathbf{B}'\mathbf{z} + \mathbf{z}'\mathbf{B}\hat{\alpha}) + E(\hat{\alpha}'\mathbf{B}'\mathbf{B}\hat{\alpha}) \\
&= \alpha'\mathbf{B}'\mathbf{B}\alpha + E(\mathbf{f}'\mathbf{f}) - E[(\mathbf{z}'\mathbf{B}\hat{\alpha})' + \mathbf{z}'\mathbf{B}\hat{\alpha}] + E(\hat{\alpha}'\mathbf{B}'\mathbf{B}\hat{\alpha}) \\
&= \alpha'\mathbf{B}'\mathbf{B}\alpha + N\sigma_e^2 - 2E(\mathbf{z}'\mathbf{B}\hat{\alpha}) + E(\hat{\alpha}'\mathbf{B}'\mathbf{B}\hat{\alpha})
\end{aligned}$$

Using the ordinary least squares formula for $\hat{\alpha}$ in the transformed model and an expression for the expectation of a quadratic form (Seber, 1977, p. 13), we have:

$$\begin{aligned}
E(\mathbf{z}'\mathbf{B}\hat{\alpha}) &= E[\mathbf{z}'\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{z}] \\
&= tr[\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\sigma_e^2 I] + [E(\mathbf{z})]'\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'E(\mathbf{z}) \\
&= \sigma_e^2 tr[\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'] + (\mathbf{B}\alpha)'\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'(\mathbf{B}\alpha) \\
&= \sigma_e^2 tr[\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'] + \alpha'\mathbf{B}'\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}(\mathbf{B}'\mathbf{B})\alpha \\
&= \sigma_e^2 tr[\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'] + \alpha'\mathbf{B}'\mathbf{B}\alpha
\end{aligned}$$

$$\begin{aligned}
E(\hat{\alpha}'\mathbf{B}'\mathbf{B}\hat{\alpha}) &= E[\hat{\alpha}'(\mathbf{B}'\mathbf{B})(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{z}] \\
&= E(\hat{\alpha}'\mathbf{B}'\mathbf{z}) \\
&= E(\mathbf{z}'\mathbf{B}\hat{\alpha})
\end{aligned}$$

Therefore:

$$\begin{aligned}
E[\hat{\mathbf{f}}'\hat{\mathbf{f}}] &= \alpha'\mathbf{B}'\mathbf{B}\alpha + N\sigma_e^2 - E(\mathbf{z}'\mathbf{B}\hat{\alpha}) \\
&= \alpha'\mathbf{B}'\mathbf{B}\alpha + N\sigma_e^2 - \sigma_e^2 tr[\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'] - \alpha'\mathbf{B}'\mathbf{B}\alpha \\
&= \sigma_e^2[N - tr(\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}')]
\end{aligned}$$

Hence:

$$E[\hat{\mathbf{f}}_a{}'\hat{\mathbf{f}}_a] = N\sigma_e^2$$

where:

$$\hat{\mathbf{f}}_a = [N/(N-tr(\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'))]^{1/2}$$

so $(\hat{\mathbf{f}}_a{}'\hat{\mathbf{f}}_a)/N$ is an unbiased estimator of σ_e^2.