87-02

ESTIMATION OF COUNTY CROP ACREAGES USING LANDSAT DATA AS AUXILIARY INFORMATION

Raj S. Chhikara
and
James J. McKeon

University of Houston - Clear Lake
2700 Bay Area Boulevard
Houston, Texas 77059

June 1987

ABSTRACT

Certain regression estimators of county crop acreages have been proposed based on a nested error model with county component as random (Battese and Fuller, 1981 ASA Proceedings of Survey Research Methods Section). An empirical evaluation was made showing the suitability of some of the estimators (Walker and Sigman, Commun. Statist., 13, No. 23, 1984). In the present study, these and other alternative estimators are compared. Estimators investigated here are based on different possible models and shrinkage criteria. Numerical evaluations of their bias and mean square errors were made using a simulation study. Although no single estimator was uniformly better, some shrinkage estimators reduced the mean square error consistently across all counties.

1. INTRODUCTION

The U.S. Department of Agriculture makes county crop acreage estimates based on its annual June Enumerative Survey (JES) data. Furthermore, these ground survey estimates are improved upon for several states by utilizing the remotely sensed LANDSAT data. The estimation procedure involves classification of LANDSAT measurements into different ground categories (crops, etc.) and estimation of individual crop sizes. The crop size estimates are utilized as regressors to the ground survey estimates and the regression estimates of crop acreages are obtained at the county and higher levels. The USDA methodology is presented in details in Hanuschak, et. al. (1979).

Often, ground observations are available for a small number of sample segments in a county. As such reliable regression estimates of crop acreages for individual counties may not be feasible unless a combined regression model approach is adopted. Considering a nested error random component model, Battese and Fuller (1981) suggested a prediction approach to estimation of county crop acreages. Walker and Sigman (1984) extended it to the stratified case and evaluated empirically its performance as well as of some others previously developed by the U.S. Department of Agriculture. Fuller (1986) considered the county acreage estimation as a measurement error problem and discussed a multivariate version of crop acreage estimation.

In general, this is a small area estimation problem. Harter (1984) described some of the small area estimation approaches previously available in the literature. Rao (1986) provides an excellent review on the topic.

In this paper we consider the county crop acreage estimation methodology as adopted by the U.S. Department of Agriculture and discuss several estimators including those proposed by Fuller and his collaborators. All these are regression estimators, but they vary in model assumptions and/or slope estimation. The regression models and the estimators of the county mean crop acreage per area segment are described in Sections 2-4. The various estimators considered here fall into two major categories, regular regression and those with a shrinkage component. The estimators are evaluated for their bias and mean square error. A simulation study was conducted for this evaluation and its results are presented in Section 5. It is shown that no single estimator is uniformly better. The regular regression estimators are unbiased but have higher mean square error as compared to the shrinkage type estimators which are biased yet have smaller mean square error. The conclusions are summarized in Section 6.

2. REGRESSION MODELS AND ESTIMATORS

Let y_{ij} be the actual crop acreage and x_{ij} be its estimate obtained from LANDSAT data for the jth segment in ith county in an analysis district (stratum). Assume that there are N_i segments in the ith county and of these n_i are sampled, $i=1,2,\ldots,K$. Let

$$\overline{Y}_i = \sum^{N_i} y_{ij}/N_i$$

and

$$\overline{Y} = \sum^{K} \sum^{N_i} y_{ij}/N$$

where

$$N = \sum^{K} Ni, \tag{1}$$

be the mean crops acreages for the ith county and the analysis district, respectively. Suppose the x_{ij} are obtained for all segments in all counties. Denote the county and analysis district mean crop acreage estimates by

$$\overline{X}_i = \sum^{N_i} x_{ij}/N_i$$

and

$$\overline{X} = \sum^{K} \sum^{N_i} x_{ij}/N \qquad (2)$$

We address estimation of $\overline{Y}_i$ given the ground observed y_{ij}, $j=1,2,\ldots,n_i$, and the x_{ij}, $j=1,2,\ldots,N_i$ and $i=1,2,\ldots,K$. Four different models are considered; three of these are fixed effect models and one assumes the county effects to be random as proposed by Battese and Fuller (1981) who have considered the county crop acreage estimation as a prediction problem.

2.1 Model 1

Let

$$y_{ij} = \alpha + \beta\, x_{ij} + e_{ij} \qquad (3)$$

where the errors e_{ij} are independent with zero mean and common variance, say σ^2. This model does not assume any county effects. Considering the least square estimators for the parameters α and β, an estimator of the mean crop acreage for ith county can be obtained as

$$\hat{\overline{Y}}_i = \overline{y} + b\,(\overline{X}_i - \overline{x}) \qquad (4)$$

where $\overline{y}$ and $\overline{x}$ are the sample mean for the stratum and b is the estimated regression coefficient given by

$$b = \sum^{K} \sum^{n_i} (y_{ij} - \overline{y})\,(x_{ij} - \overline{x}) / \sum^{K} \sum^{n_i} (x_{ij} - \overline{x})^2. \qquad (5)$$

Walker and Sigman (1984) attributed this estimator to Huddleston and Ray who initially suggested its use by the U.S. Department of Agriculture. Since no distinction is made in the regression across counties, the estimator in (4) is likely to have large bias, though its variance may be small, whenever the county effect is significant.

For the stratum, the mean crop acreage estimate is

$$\hat{Y} = \bar{y} + b(\bar{X} - \bar{x}). \tag{6}$$

The estimator given in (4) will be known as <u>E1</u>.

2.2 <u>Model 2</u>

Let

$$y_{ij} = \mu + \nu_i + \beta x_{ij} + e_{ij} \tag{7}$$

where ν_i denotes the <u>ith</u> county effect and μ is the overall mean for the stratum. It is assumed that the ν_i are fixed effects so that $\Sigma\, \nu_i = 0$ and that

$$\sum^{K} n_i\, \nu_i(\bar{x}_i - \bar{x}) = 0. \tag{8}$$

The assumption in (8) implies that the regression for the county means $(\bar{y}_i, \bar{x}_i)$ also has slope β .

For estimating the mean crop acreage in <u>ith</u> county, we consider the following estimator, to be known as <u>E2</u>,

$$\hat{\bar{Y}}_i = \bar{y}_i + \hat{\beta}(\bar{X}_i - \bar{x}_i) \tag{9}$$

where

$$\hat{\beta} = \sum^{K} \sum^{n_i} (y_{ij} - \bar{y})(x_{ij} - \bar{x}) / \sum^{K} \sum^{n_i} (x_{ij} - \bar{x})^2. \tag{10}$$

Since β is a common slope for both within and between county regressions, its estimator in (10) is considered on the basis of total variation for the stratum. Clearly, the estimator in (9) is unbiased. The estimators of mean μ and effect ν_i given next are also unbiased.

$$\hat{\mu} = \bar{y}_w - \hat{\beta}\,\bar{x}_w$$

$$\hat{\nu}_i = \bar{y}_i - \hat{\mu} - \hat{\beta}\,\bar{x}_i \tag{11}$$

where $\bar{y}_w$ and $\bar{x}_w$ are the weighted sample means,

$$\bar{y}_w = \sum^{K} N_i y_i / N$$

$$\bar{x}_w = \sum^{K} N_i x_i / N. \tag{12}$$

For the stratum, the corresponding mean crop acreage estimator is

$$\hat{\bar{Y}} = \hat{\mu} + \hat{\beta}\,\bar{X}. \tag{13}$$

When the assumption in (8) is not considered, it is more appropriate to estimate the slope β by using the pooled sum of squares and cross products, that is

$$\hat{\beta} = \sum^{K} \sum^{n_i} (y_{ij} - \bar{y}_i)(x_{ij} - \bar{x}_i) / \sum^{K} \sum^{n_i} (x_{ij} - \bar{x}_i)^2, \tag{14}$$

instead of the total sum of squares and cross products as in (10).

2.3 Model 3

Consider the model form as in (7), but assume that the county effect is random. In specific, let

$$y_{ij} = \mu + v_i + \beta x_{ij} + e_{ij} \tag{15}$$

where

$$E(y_{ij}) = \mu + \beta x_{ij}$$

and

$$\mathrm{Cov}(y_{ij}, y_{i'j'}) = V_i = \begin{cases} \sigma_v^2 + \sigma_e^2, \text{ if } i=i', j=j' \\ \sigma_v^2, \text{ if } i=i', j\neq j' \\ o, \text{ otherwise} \end{cases} \tag{16}$$

This model was first considered by Battese and Fuller (1981) and later on by others as discussed in Section 1. It can be rewritten in a matrix form as described by these authors. We omit further details and give the following estimator for the county crop acreage mean:

$$\hat{\bar{Y}}_i = \bar{y}_i + \hat{\beta}\ (\bar{X}_i - \bar{x}_i) \tag{17}$$

where

$$\hat{\beta} = S_{yx}/S_{xx} \tag{18}$$

with

$$S_{yx} = \sum^{K} \sum^{n_i} (y_{ij} - \bar{y}_i)(x_{ij} - \bar{x}_i) + \sum^{K} w_i(\bar{y}_i - \bar{y}_w)(\bar{x}_i - \bar{x}_w)$$

$$S_{xx} = \sum^{K} \sum^{n_i} (x_{ij} - \bar{x}_i)^2 + \sum^{K} w_i(\bar{x}_i - \bar{x}_w)^2$$

$$y_w = \sum^{K} w_i \, \bar{y}_i / \sum^{K} w_i$$

$$\bar{x}_w = \sum^{K} w_i \, x_i / \sum^{K} w_i$$

$$w_i = n_i/[1 + n_i \, \sigma_v^2 / \sigma_e^2] \tag{19}$$

The estimator in (17) will be known as E3.

It may be noted that $w_i = n_i$ and $\hat{\beta}$ is the same as in (10) when σ_v^2=o, but w_i=o and $\hat{\beta}$ becomes the pooled within county slope as σ_v^2/σ_e^2 becomes infinite.

Similarly, estimators of μ and v_i are as in (11) but $\hat{\beta}$, $\bar{y}_w$ and $\bar{x}_w$ are those given here in (18) and (19).

Another estimator of $\bar{Y}_i$ as currently used by the U.S. Department of Agriculture is obtained by estimating the regression coefficient β as follows:

$$\hat{\beta}_c = \sum^{K} w_i' \sum^{n_i} (y_{ij} - \bar{y}_i)(x_{ij} - \bar{x}_i) / \sum^{K} w_i' \sum^{n_i} (x_{ij} - \bar{x}_i)^2 \tag{20}$$

where

$$w_i' = (N_i/N)^2 / n_i(n_i - 1). \tag{21}$$

Here $\hat{\beta}_c$ is the same as the combined estimator of slope given in Cochran (1977). Its use leads to the following estimator, to be known as E4, for the county crop acreage mean:

$$\hat{\bar{Y}}_i = \bar{y}_i + \hat{\beta}_c(\bar{X}_i - \bar{x}_i). \tag{22}$$

2.4 Model 4

Consider the following model:

$$y_{ij} = \mu + \nu_i + \beta_o X_i + \beta_1 x_{ij} + e_{ij} \tag{23}$$

This includes an additional regression term involving the county means. A rationale for this is to recognize that the county effect is perhaps linearly related to the LANDSAT estimated county crop acreage mean. Since a positive correlation between y_{ij} and x_{ij} or $\overline{X}_i$ is expected, estimates of β_o and β_1 are constrained to be nonnegatives.

It can be shown that the least square estimators of β_o and β_1 are given by

$$\hat{\beta}_o = (S_{\bar{y}\bar{X}} - \hat{\beta}_1 S_{\bar{x}\bar{X}})/S_{\bar{X}\bar{X}} \tag{24}$$

and

$$\hat{\beta}_1 = (S_{yx} + S_{\bar{y}\bar{x}} - \hat{\beta}_o S_{\bar{x}\bar{X}})/S_{xx} + S_{\bar{x}\bar{x}} \tag{25}$$

where

$$S_{yx} = \sum^{K} \sum^{n_i} (y_{ij} - \bar{y}_i)(x_{ij} - \bar{x}_i)$$

$$S_{\bar{y}\bar{x}} = \sum^{K} n_i (\bar{y}_i - \bar{y})(\bar{x}_i - \bar{x})$$

$$S_{yx} = \sum^{K} n_i (\bar{y}_i - \bar{y})(\overline{X}_i - \overline{X}.) \tag{26}$$

and similarly others, S_{xx}, $S_{\bar{x}\bar{x}}$, $S_{\bar{x}\bar{X}}$ and $S_{\bar{X}\bar{X}}$. In (26), $\overline{X}.$ is defined by

$$\overline{X}. = \sum^{K} n_i X_i / n \tag{27}$$

where

$$n = \sum^{K} n_i .$$

Note that the estimates $\hat{\beta}_o$ and $\hat{\beta}_1$ are to be obtained iteratively.

The estimators of the other model parameters are

$$\hat{\mu} = \bar{y}_w - \hat{\beta}\,\bar{x}_w$$

$$\hat{\nu}_i = \bar{y}_i - \hat{\mu} - \hat{\beta}_o\,\bar{X}_i - \hat{\beta}_1\,\bar{x}_i \tag{28}$$

where $\bar{y}_w$ and $\bar{x}_w$ are the weighted sample means as defined in (12).

The county mean crop acreage estimators are

$$\hat{\bar{Y}}_i = \hat{\mu} + \hat{\nu}_i + \hat{\beta}_o\,\bar{X}_i + \hat{\beta}_1\,\bar{x}_i$$

Equivalently,

$$\hat{\bar{Y}}_i = \bar{y}_i + \hat{\beta}_1\,(\bar{X}_i - \bar{x}_i) \tag{29}$$

This estimator of county means will be called <u>E5</u>. Again, the stratum estimator is

$$\hat{\bar{Y}} = \hat{\mu} + \hat{\beta}_1\,\bar{X} \tag{30}$$

3. CERTAIN SHRINKAGE ESTIMATORS

Each county is regarded as a separate population under models 2, 3 and 4. The estimators given in Section 2 for these K-population models are essentially unbiased but may have large variance and hence, large mean square error (mse). On the other hand, model 1 assumes a single population structure for all K counties, thus reducing the number of parameters to be estimated. As such, the variance of a county mean estimator is likely to be reduced under model 1, but the bias may increase substantially.

Battese and Fuller (1981) constructed a shrinkage estimator for county effects which attempts to minimize~~s~~ the mse of the county mean crop acreage estimate. Although the shrinkage approach was applied only in the case of model 3, one can easily develop shrinkage estimators with respect to other models, as well. The general idea is to shrink the county effect estimate $\hat{\nu}_i$ by considering $\delta_i \hat{\nu}_i$, where $o \leq \delta_i \leq 1$. Clearly the shrinkage estimator $\delta_i \hat{\nu}_i$ has bias $(1-\delta_i)\ \nu_i$ and

$$\text{mse}\ (\delta_i \hat{\nu}_i) = (1-\delta_i)^2\ \nu_i^2 + \delta_i^2\ \sigma^2/m_i \tag{31}$$

where σ^2/m_i denotes the variance of $\hat{\nu}_i$. When the n_i are equal, we take $m_i = n_i$, but in general,

$$m_i = n_i(K-1)/(K - 2 + n_i\bar{n}_H^{-1}) \tag{32}$$

where $\bar{n}_H$ is the harmonic mean of n_i, $i=1,2,\ldots,K$.

The mean square error in (31) is minimum for

$$\delta_i = \nu_i^2/(\nu_i^2 + \sigma^2/m_i). \tag{33}$$

The corresponding expression for δ_i in the case of Battese-Fuller shrinkage estimator is obtained by replacing ν_i^2 by σ_v^2 and m_i by n_i so that

$$\delta_i = \sigma_v^2/(\sigma_v^2 + \sigma^2/n_i). \tag{34}$$

This reduces all county effects proportionally while the δ_i in (33) reduces the smaller county effects by a proportionally greater amount. The resulting shrinkage estimator of the county mean is

$$\hat{\bar{Y}}_i - (1-\hat{\delta}_i)\,\hat{\nu}_i \tag{35}$$

in the case of fixed effect model (Subsection 2.2) and

$$\hat{\bar{Y}}_i - (1-\hat{\delta}_i)\,\hat{v}_i \tag{36}$$

in the case of random model (Subsection 2.3). The estimator in (35) will be designated as E6 and that in (36) as E7. Another estimator of $\bar{Y}_i$ is obtained by applying the fixed effect shrinkage given in (33) to the estimator in (36). The resulting estimator will be designated as E8.

Next, by applying the shrinkage δ_i in (33) to the estimator given in Subsection 2.4, one obtains another shrinkage estimator of $\bar{Y}_i$ given by

$$\hat{\bar{Y}}_i + (1 - \hat{\delta}_i)\,\hat{\nu}_i \tag{37}$$

where $\hat{\nu}_i$ is as in (28) and $\hat{\bar{Y}}_i$ as in (29). This estimator of county mean will be called E9.

Another estimator similar to (37) will be considered with the slope estimate given in (25) replaced by that obtained using the within sum of squares and cross products. It will be designated as E10.

For the random model, another shrinkage estimator was constructed by Harter (1983). This differs slightly from the Battese-Fuller estimator given above in (36). Their major difference lies in the determination of shrinkage coefficient δ_i. We skip the mathematical details of the Harter estimator because of space limitation. However, we have included it in our evaluations discussed later on in section 5. This estimator will be designated as E11.

All these shrinkage estimators (E6 - E11) are consistent; that is, $\hat{\bar{Y}}_i$ approaches $\bar{Y}_i$ as n_i becomes infinite.

4. FULLER'S ESTIMATOR

An approach due to Fuller (1986) is based on an error-in-variable model and does not attempt directly to reduce the mean square error as was the case in Section 3. It involves a two stage estimation procedure: At the first stage, the county mean $\overline{Y}_i$ is estimated as described in Subsection 2.2 with $\hat{\beta}$ given in (14). At the second stage, this estimator is modified by taking into account the residual (for the ith county) obtained from the regression of the county mean $\overline{X}_i$ onto the first stage estimator $\hat{\overline{Y}}_i$, i = 1, 2, . . ., K. Suppose this residual is Z_i. Then the Fuller estimator, to be called E12, is of the form

$$\hat{\hat{\overline{Y}}}_i = \hat{\overline{Y}}_i - A_i Z_i \tag{38}$$

where A_i is a shrinkage factor involving the regression slope at the second stage and the estimation error variances for the both stage regressions. Again, mathematical details are skipped due to space limitation.

The estimator in (38) is a consistent estimator of $\overline{Y}_i$. Note that this estimator essentially is a shrinkage estimator.

5. NUMERICAL EVALUATIONS

A simulation study was conducted to evaluate the bias and mean square errors of the county mean estimators discussed in Sections 2-4. For input, we utilized the data acquired for the 33 JES segments in Northern Missouri during 1979. The data consisted of both ground observations of the land tracks and LANDSAT acquired MSS pixel measurements for the segments. This information was used to simulate (in a boot-strap manner) a set of 200 area segments and their MSS data in four spectral bands for each of the seven cover types which combinedly covered more than 99 percent of the total area in the 33 JES segments. The maximum likelihood rule was applied to the simulated spectral data for pixel classification. For each of the 200 segments, the actual number of pixels (y) and the corresponding number of pixels classified (x) into a cover type were determined. The statistical properties of the bivariate (x, y) data for the 200 segments are discussed in Chhikara (1987). A description of the simulation program can be found in Chhikara, et. al. (1986). Moreover, a simulation package is being developed at the University of Houston - Clear Lake which will allow one to simulate the agricultural and LANDSAT data for an area of interest and to conduct ~~evaluations~~ evaluation studies for estimation of crop acreages.

The 200 segments were subdivided into 5 groups each containing 40 segments. These five groups were treated as five counties. The subdivision was carried in two ways: (a) ordered groups where the segment size for the corn and soybean acreages was used to order the segments and (b) random groups. Listed in Table 1 are the county means and standard deviations for the cover types in case (a). In both cases, the individual county estimates were made for each cover type using each estimator. Both equal and unequal sample sizes were considered. In the equal sample size cases, $n_i = 4$ and 10 ($i = 1, 2, 3, 4, 5$), whereas $n_1 = 8$, $n_2 = 8$, $n_3 = 6$, $n_4 = 4$ and $n_5 = 4$ in the case of unequal sample sizes. The estimation was repeated 500 times and the bias and mean square error were calculated from the 500 estimates of a county crop acreage mean in each case.

For a comparision of the estimators, we examined the maximum bias and median mse values obtained across five counties for each estimator . The reason for considering maximum bias was to safeguard against an overly biased county mean estimate. We computed relative bias for each estimator and the ratio of its mse to that of the sample mean for each county. Tables 2(a) - 2(i) present the performance levels of the estimators for the three major cover types, pasture,

TABLE 1: County means and standard deviations for (y,x) in case (a)

Cover type and size	Statistic	County mean (standard deviation)				
		1	2	2	4	5
Pasture 29.6%	$\bar{Y}_i$	133.1 (84.8)	193.2 (95.1)	210.5 (88.1)	203.9 (107.7)	178.4 (87.2)
	$\bar{X}_i$	141.4 (67.2)	153.8 (65.0)	137.4 (51.2)	153.0 (61.0)	121.2 (51.8)
Soybeans 24.5%	$\bar{Y}_i$	381.8 (195.8)	233.0 (102.5)	163.3 (75.9)	141.1 (95.4)	88.2 (58.6)
	$\bar{X}_i$	354.0 (136.9)	234.5 (64.4)	169.3 (49.8)	139.9 (49.1)	86.4 (40.3)
Corn 12.8%	$\bar{Y}_i$	140.6 (179.0)	115.9 (105.6)	109.5 (81.5)	70.2 (62.7)	77.2 (61.3)
	$\bar{X}_i$	162.8 (114.5)	156.9 (67.8)	153.4 (54.8)	136.4 (46.0)	108.0 (37.5)
Waste 13.3%	$\bar{Y}_i$	103.2 (93.8)	95.2 (92.5)	78.1 (77.9)	114.3 (113.1)	144.8 (137.2)
	$\bar{X}_i$	58.9 (30.7)	50.2 (27.8)	60.6 (31.5)	59.2 (26.1)	85.1 (58.6)
Wood 8.4%	$\bar{Y}_i$	54.3 (106.1)	49.4 (57.5)	106.2 (125.4)	77.9 (98.2)	57.2 (64.5)
	$\bar{X}_i$	53.8 (44.1)	56.0 (37.8)	69.9 (49.8)	64.5 (43.9)	54.9 (41.9)
Hay 7.2%	$\bar{Y}_i$	81.7 (109.7)	94.0 (140.7)	66.3 (92.4)	123.2 (157.2)	97.6 (123.0)
	$\bar{X}_i$	117.7 (61.4)	125.4 (45.8)	143.8 (61.2)	177.8 (76.8)	187.7 (88.9)
Winter Wheat 3.4%	$\bar{Y}_i$	12.1 (20.1)	13.5 (25.9)	23.7 (38.6)	17.2 (28.0)	22.4 (34.7)
	$\bar{X}_i$	19.7 (29.6)	19.4 (16.9)	27.6 (23.1)	25.3 (19.7)	32.0 (26.8)

soybeans, and corn. Similar results were made for other cover types, but these are omitted. Depicted in these tables are the distributions of the estimators into various categories in terms of maximum relative bias and median mse ratio. An estimator is indicated by its numerical figure with letter E omitted (Recall we previously numbered these estimators but each preceeded by letter E.)

Based on these results, the estimators may be viewed to fall into three categories, (i) E2, E3, E4, E5, (ii) E6, E8, E9 - E12, and (iii) E1, E7. None of the category (i) estimators are shrinkage estimators and these are essentially unbiased but have the largest mse, whereas those in category (iii) have the smallest mse but involve the largest bias. It seems the estimators in category (ii) are the most desirable since they generally fall in between the other two categories showing moderate bias and mse. Category (ii) consists of all shrinkage estimators, except E7 which belongs to category (iii). Since E1 is based on the no county effects model, it is clear that E7 shrinks the county estimates toward the stratum estimate much more than any other shrinkage estimator. So it seems there are some advantages in the use of a shrinkage estimator (excluding E7) over a regular regression estimator.

For a comparison of category (i) estimators, one finds that the underlying model and how the slope parameter is estimated, influence the estimation of county mean crop acreage. For example, E2 uses the total slope accounting both the within and the between county variation and appears to be more stable than others. However, if the county effects are large and not symmetrically located about the regression line, the total slope provides an inferior estimator. Under the variance component (random) model the slope "adjusts" between the total and within estimates depending upon the county effects. As the between county variation increases, the slope estimation shifts toward the within part making the estimators in (17) and (22) behave similarly. Model 4 includes regression on

TABLE 2 (a): Performance distribution of estimators for cover type Pasture and sample size $n_1 = 4$.

Case (a): Ordered Groups

Max Bias

		0 - 2.5%	2.5 - 5.0 %	5.0 - 10.0%	10 + %
	0 - .2				
mse ratio	.2 - .4				1
	.4 - .6			9	7
	.6 - .8			6, 8, 10	
	.8 - 1.0	2, 3, 5	12		
	1.0 +	4			

Case (b): Random Groups

Max Bias

		0 - 2.5%	2.5 - 5.0%	5.0 - 10.0%	10 + %
	0 - .2				
mse ratio	.2 - .4				1
	.4 - .6			7, 9	
	.6 - .8			6, 8, 10	
	.8 - 1.0	2, 3, 5			
	1.0 +	4		12	

TABLE 2 (b): Performance distribution of estimators for cover type Soybeans and sample size $n_i = 4$.

Case (a): Ordered Groups

Max Bias

mse ratio	0 - 2.5%	2.5 - 5.0 %	5.0 - 10.0%	10 + %
0 - .2				
.2 - .4		9, 10	12	
.4 - .6			6, 8	1, 7
.6 - .8	5	3, 4		
.8 - 1.0	2			
1.0 +				

Case (b): Random Groups

Max Bias

mse ratio	0 - 2.5%	2.5 - 5.0%	5.0 - 10.0%	10 + %
0 - .2	12	2 - 10	1	
.2 - .4				
.4 - .6				
.6 - .8				
.8 - 1.0				
1.0 +				

TABLE 2(c): Performance distribution of estimators for cover type Corn and sample size $n_i = 4$.

Case (a): Ordered Groups

Max Bias

		0 - 2.5%	2.5 - 5.0 %	5.0 - 10.0%	10 + %
	0 - .2				
mse ratio	.2 - .4			9	1, 7
	.4 - .6			6, 8, 10, 12	
	.6 - .8	2, 3, 4, 5			
	.8 - 1.0				
	1.0 +				

Case (b): Random Groups

Max Bias

		0 - 2.5%	2.5 - 5.0%	5.0 - 10.0%	10 + %
	0 - .2			1, 7	
mse ratio	.2 - .4	10	6, 8, 9, 12		
	.4 - .6	2, 3, 4, 5			
	.6 - .8				
	.8 - 1.0				
	1.0 +				

TABLE 2 (d): Performance distribution of estimators for cover type Pasture and sample size unequal.

Case (a): Ordered Groups

Max Bias

		0 - 2.5%	2.5 - 5.0 %	5.0 - 10.0%	10 + %
	0 - .2				
mse ratio	.2 - .4				1
	.4 - .6				7
	.6 - .8		10	6, 8, 9, 11	
	.8 - 1.0	2, 3, 4	12		
	1.0 +	5			

Case (b): Random Groups

Max Bias

		0 - 2.5%	2.5 - 5.0%	5.0 - 10.0%	10 + %
	0 - .2				
mse ratio	.2 - .4				
	.4 - .6				1, 7
	.6 - .8		6, 8, 10, 11	9	
	.8 - 1.0	2, 3, 5		12	
	1.0 +	4			

TABLE 2 (e): Performance distribution of estimators for cover type Soybeans and sample size unequal

Case (a): Ordered Groups

Max Bias

mse ratio		0 - 2.5%	2.5 - 5.0 %	5.0 - 10.0%	10 + %
	0 - .2				
	.2 - .4			9	12
	.4 - .6			10	6, 7, 8
	.6 - .8	3, 4, 5		11	1
	.8 - 1.0	2			
	1.0 +				

Case (b): Random Groups

Max Bias

mse ratio		0 - 2.5%	2.5 - 5.0%	5.0 - 10.0%	10 + %
	0 - .2	2-6, 8, 10-12	1, 7, 9		
	.2 - .4				
	.4 - .6				
	.6 - .8				
	.8 - 1.0				
	1.0 +				

TABLE 2 (f): Performance distribution of estimators for cover type Corn and sample size unequal.

Case (a): Ordered Groups

Max Bias

		0 - 2.5%	2.5 - 5.0 %	5.0 - 10.0%	10 + %
	0 - .2				
mse ratio	.2 - .4			6, 8-10, 12	1, 7
	.4 - .6			11	
	.6 - .8	2, 3, 4, 5			
	.8 - 1.0				
	1.0 +				

Case (b): Random Groups

Max Bias

		0 - 2.5%	2.5 - 5.0%	5.0 - 10.0%	10 + %
	0 - .2			1, 7	
mse ratio	.2 - .4	10, 11	6, 8, 9, 12		
	.4 - .6	2, 3, 4, 5			
	.6 - .8				
	.8 - 1.0				
	1.0 +				

TABLE 2 (g): Performance distribution of estimators for cover type Pasture and sample size $n_i = 10$.

Case (a): Ordered Groups

Max Bias

mse ratio	0 - 2.5%	2.5 - 5.0 %	5.0 - 10.0%	10 + %
0 - .2				
.2 - .4				
.4 - .6				1, 7
.6 - .8		10	6, 8, 9	
.8 - 1.0	2,3,4,5,12			
1.0 +				

Case (b): Random Groups

Max Bias

mse ratio	0 - 2.5%	2.5 - 5.0%	5.0 - 10.0%	10 + %
0 - .2				
.2 - .4				
.4 - .6				1, 7
.6 - .8		6, 8, 10	9	
.8 - 1.0	2, 3, 4, 5	12		
1.0 +				

TABLE 2 (h): Performance distribution of estimators for cover type Soybeans and sample size n_i = 10

Case (a): Ordered Groups

Max Bias

mse ratio	0 - 2.5%	2.5 - 5.0 %	5.0 - 10.0%	10 + %
0 - .2				
.2 - .4				
.4 - .6		9, 10	12	7
.6 - .8	3, 4, 5	8	6	
.8 - 1.0	2			
1.0 +				1

Case (b): Random Groups

Max Bias

mse ratio	0 - 2.5%	2.5 - 5.0%	5.0 - 10.0%	10 + %
0 - .2	2-6, 8, 12	7, 9, 11	1	
.2 - .4				
.4 - .6				
.6 - .8				
.8 - 1.0				
1.0 +				

TABLE 2 (i): Performance distribution of estimators for cover type Corn and sample size $n_i = 10$

Case (a): Ordered Groups

Max Bias

mse ratio	0 - 2.5%	2.5 - 5.0 %	5.0 - 10.0%	10 + %
0 - .2				
.2 - .4				1, 7
.4 - .6		6, 8, 10	9, 12	
.6 - .8	2, 3, 4, 5			
.8 - 1.0				
1.0 +				

Case (b): Random Groups

Max Bias

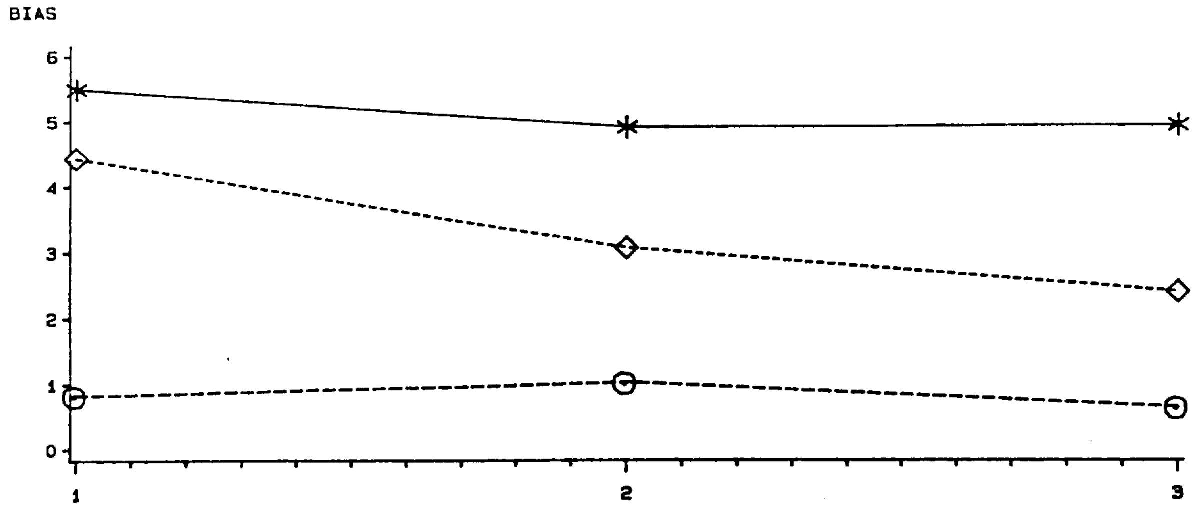
mse ratio	0 - 2.5%	2.5 - 5.0%	5.0 - 10.0%	10 + %
0 - .2			1	
.2 - .4		6, 8, 9, 10	7, 12	
.4 - .6	2, 3, 4, 5			
.6 - .8				
.8 - 1.0				
1.0 +				

the county mean (for the LANDSAT crop acreages) and may improve the county mean estimates, particularly when county effects are large and the correlation between x and y is substantial. Although none of the shrinkage estimators is uniformly better than others in the group, the performance of E10 seems to be slightly better.

For further comparison, we have plotted both the bias and the mse ratios obtained for the three estimators, E3, E10, and E12. The bias plots are shown in Figures 1(a) - 1(c) for the case of ordered groups and Figure 2(a) - 2(c) for the unordered group case. Clearly, E3 has the smallest bias, and E10 tends to be less biased than E12. Next, the mse ratios are plotted in Figure 3(a) - 3(c) for the ordered group case and in Figure 4(a) - 4(c) for the unordered group case. Here, E3 has the largest mse (with one exception) and E10 has mse generally smaller than E12. Hence, E10 may be viewed as the best choice amongst those considered in this study.

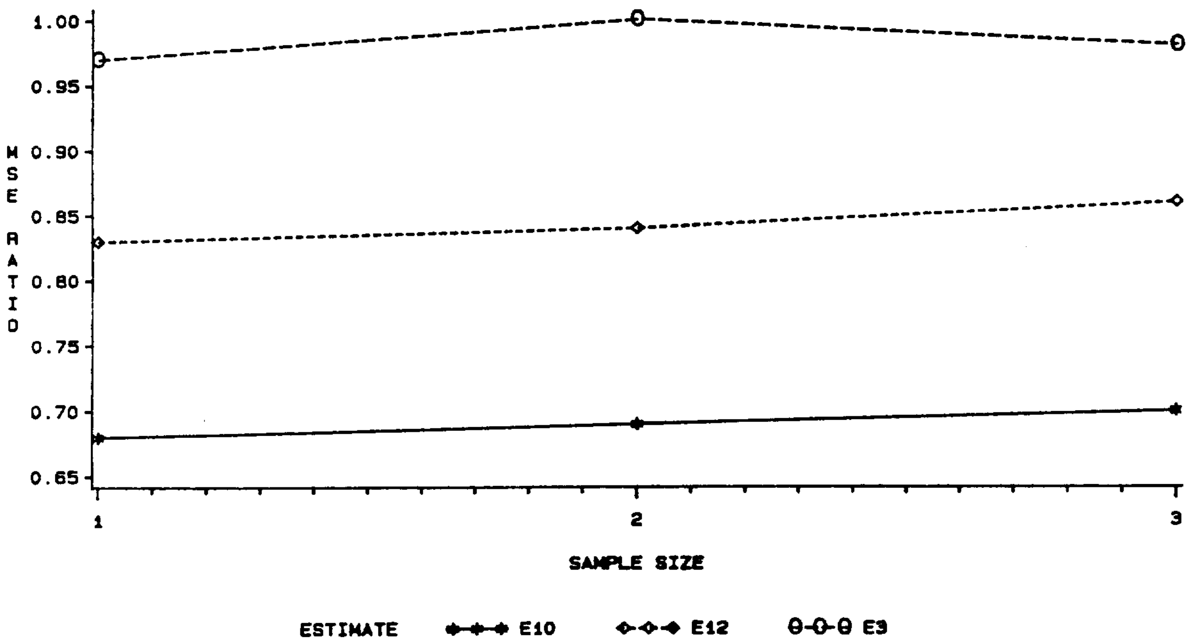
BIAS VS SAMPLE SIZE

CROP-PASTURE



MSE RATIO VS SAMPLE SIZE

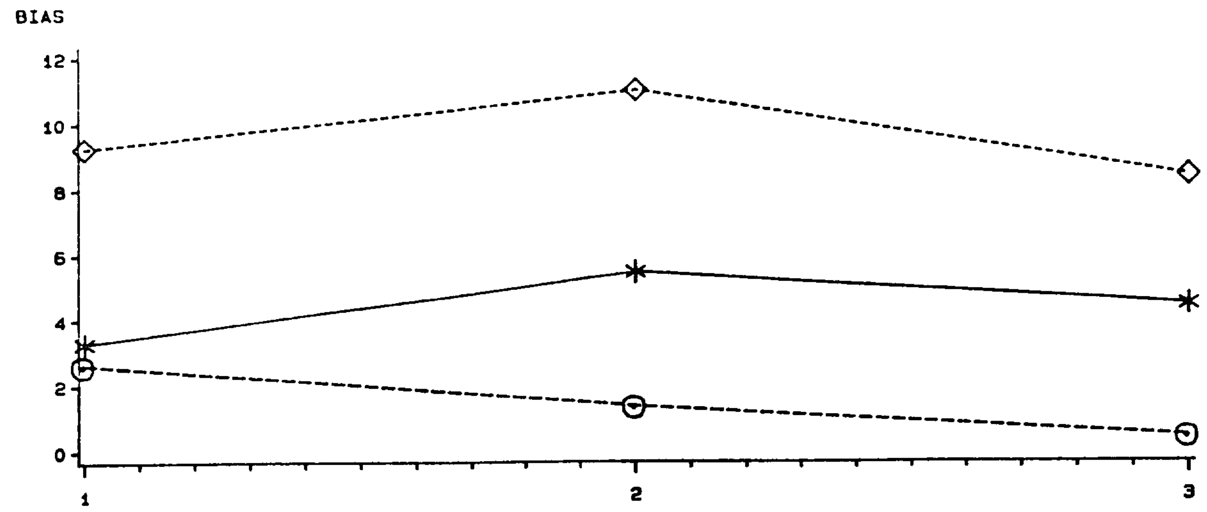
CROP-PASTURE



COUNTY MEANS ORDERED

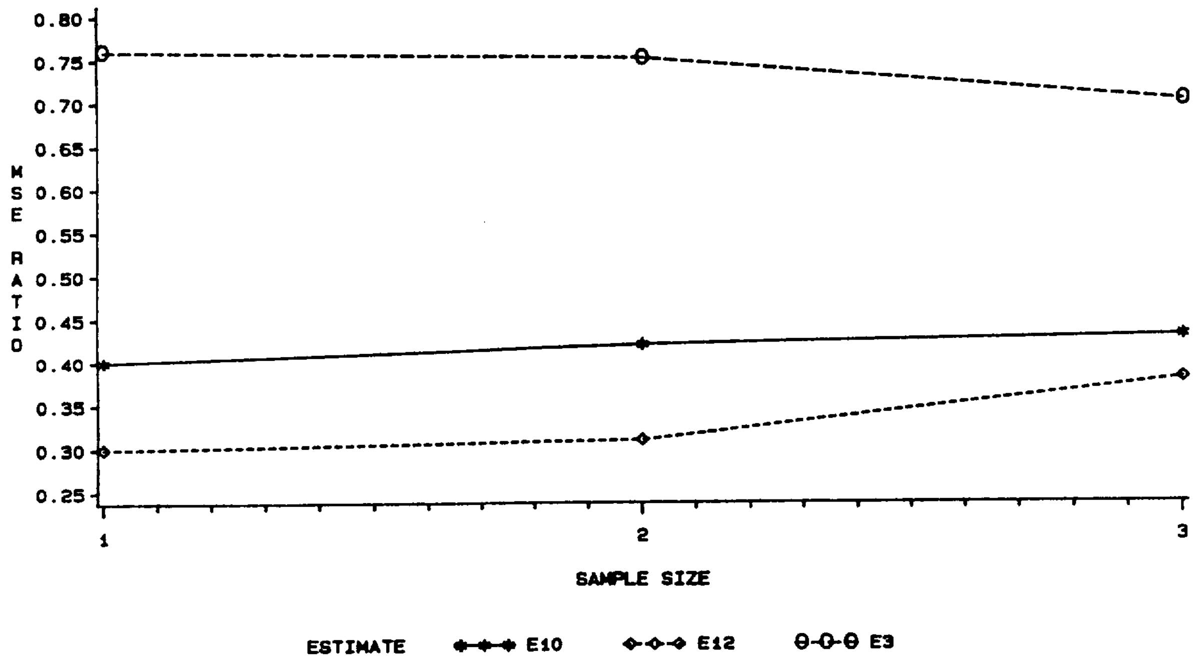
BIAS VS SAMPLE SIZE

CROP=SOYBEANS



MSE RATIO VS SAMPLE SIZE

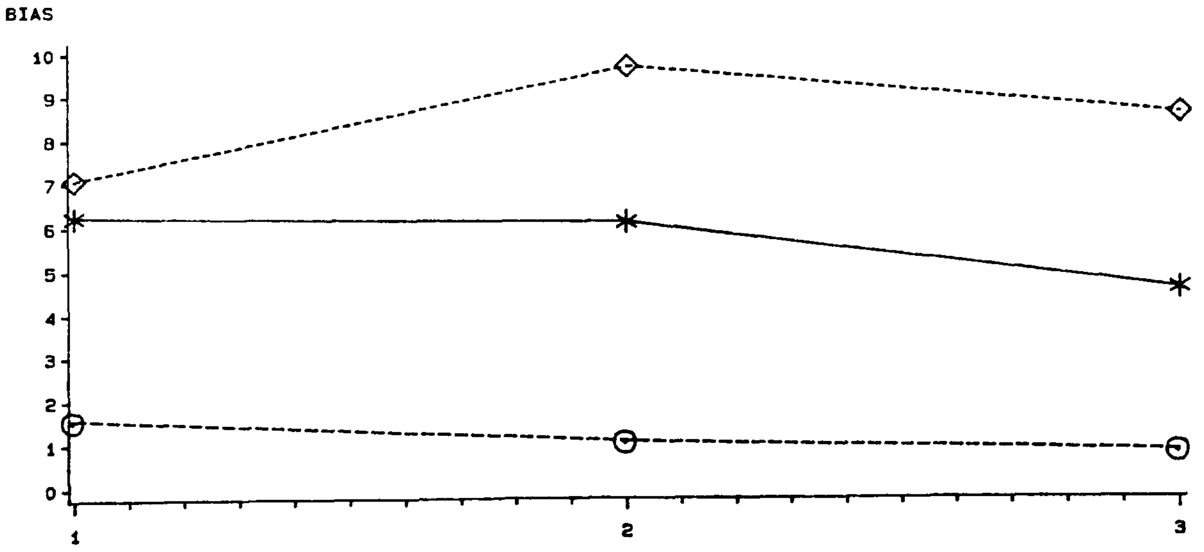
CROP=SOYBEANS



COUNTY MEANS ORDERED

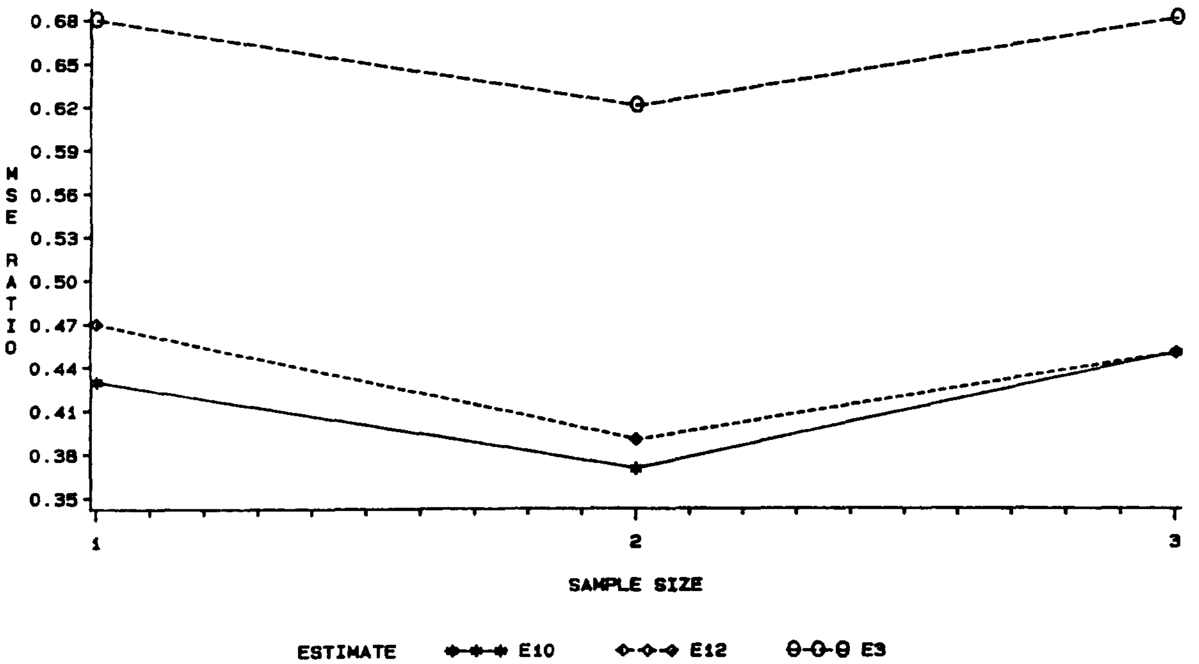
BIAS VS SAMPLE SIZE

CROP=CORN



MSE RATIO VS SAMPLE SIZE

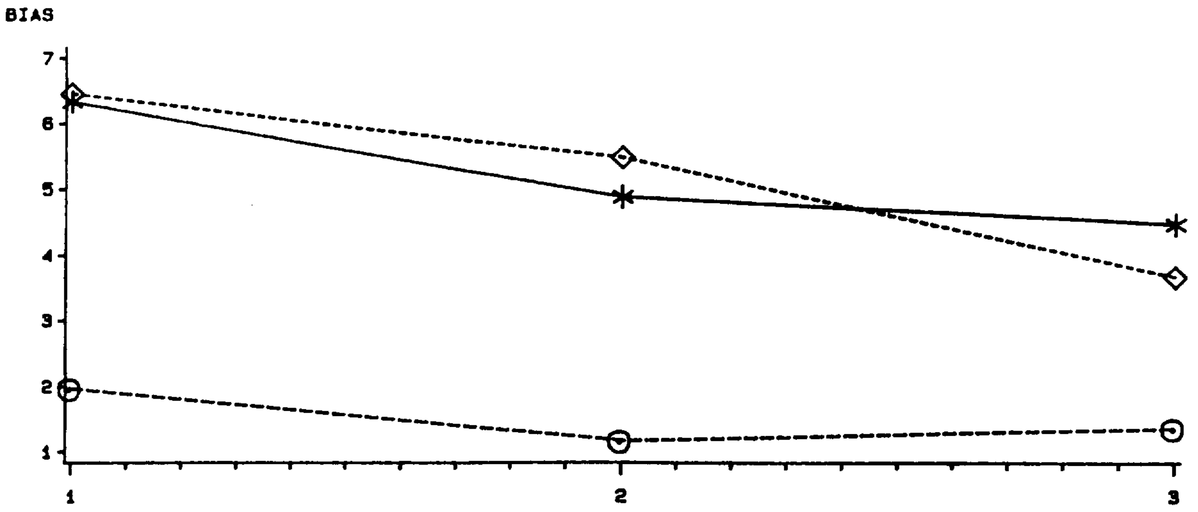
CROP=CORN



COUNTY MEANS ORDERED

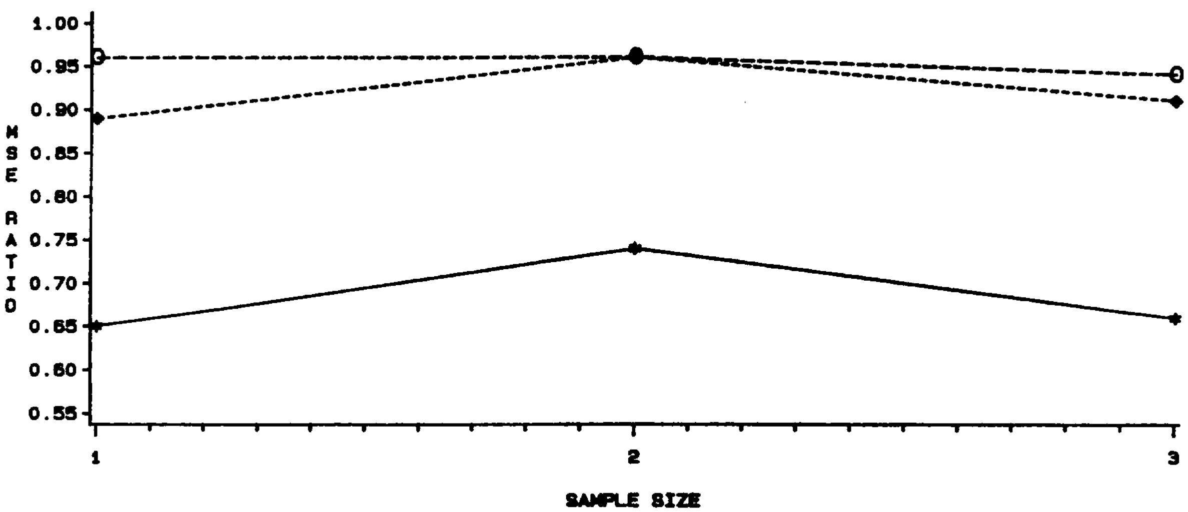
BIAS VS SAMPLE SIZE

CROP-PASTURE



MSE RATIO VS SAMPLE SIZE

CROP-PASTURE

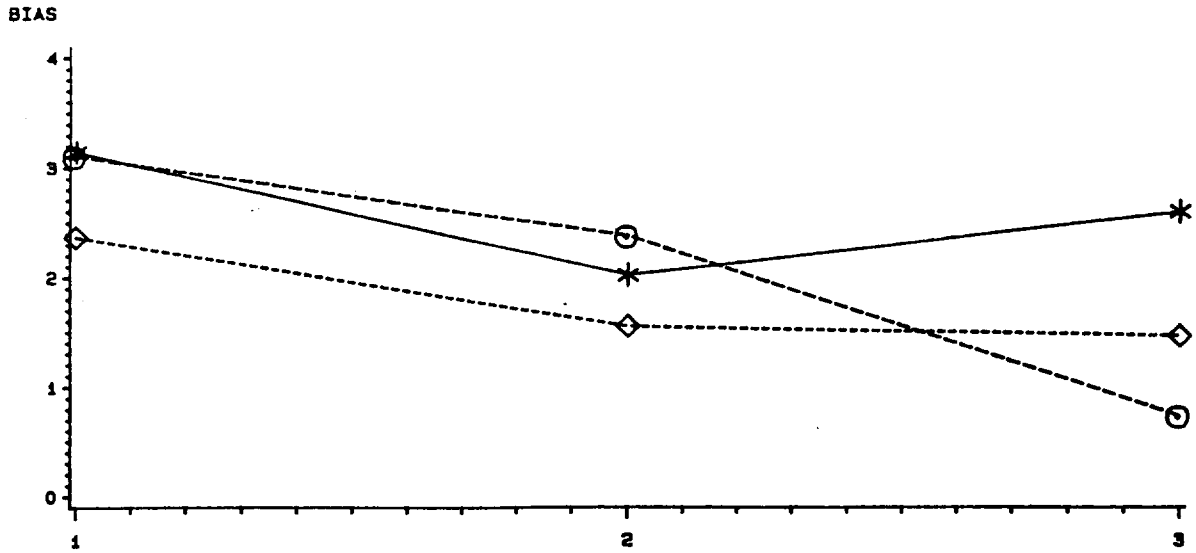


ESTIMATE E10 E12 E3

COUNTY MEANS UNORDERED

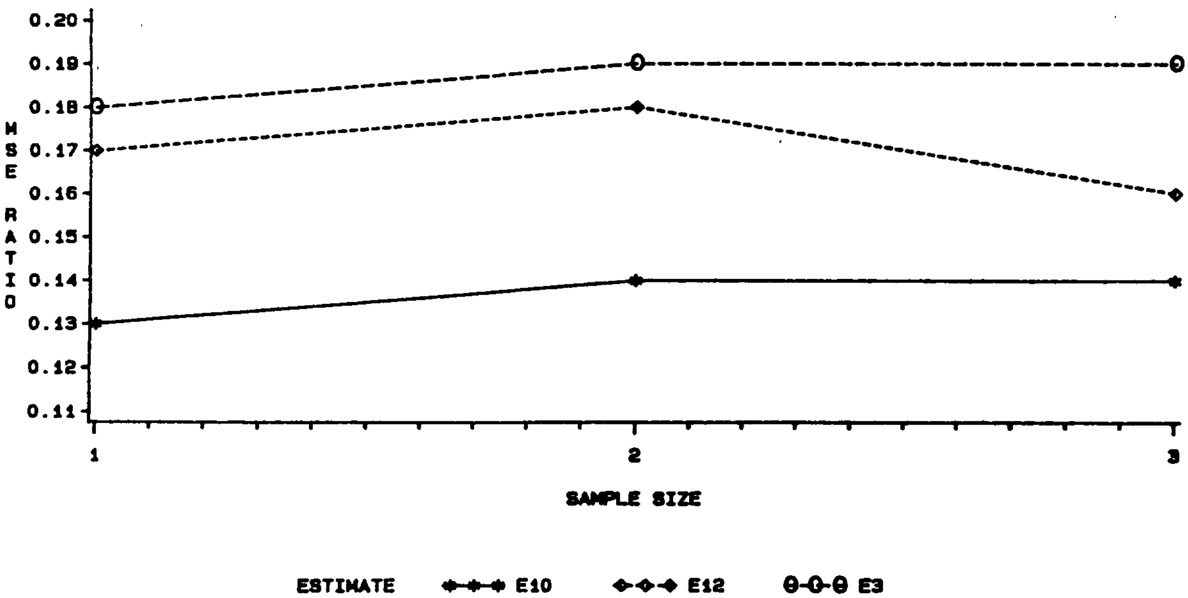
BIAS VS SAMPLE SIZE

CROP=SOYBEANS



MSE RATIO VS SAMPLE SIZE

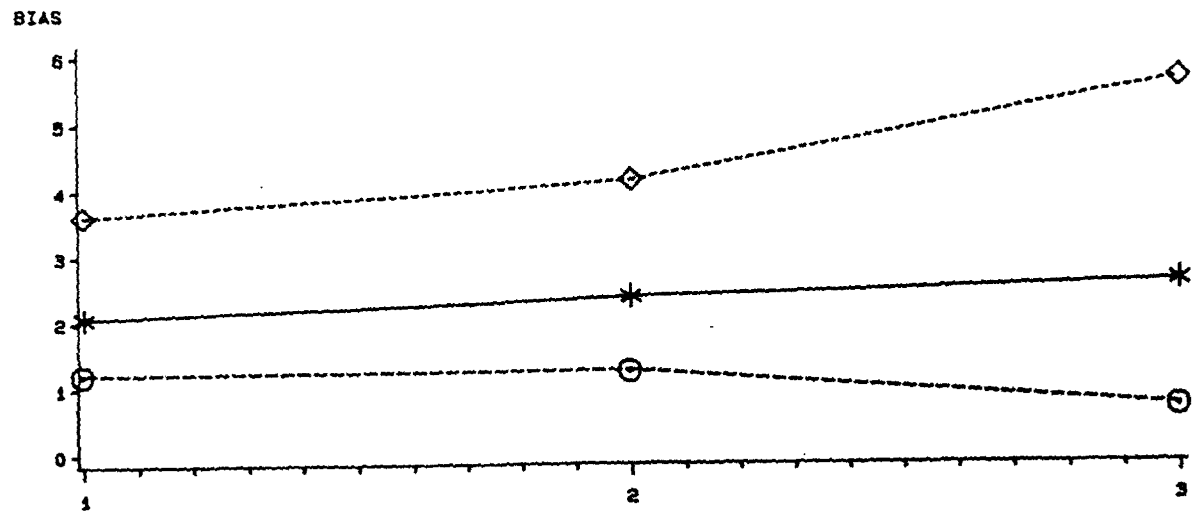
CROP=SOYBEANS



COUNTY MEANS UNORDERED

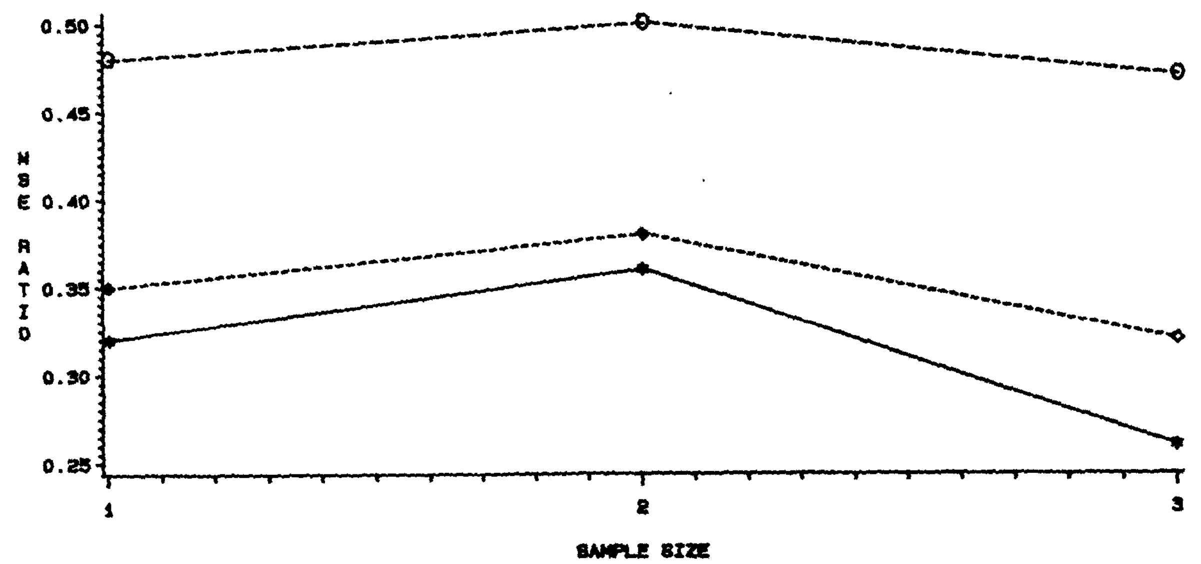
BIAS VS SAMPLE SIZE

CROP=CORN



MSE RATIO VS SAMPLE SIZE

CROP=CORN



ESTIMATE E10 E12 E3

COUNTY MEANS UNORDERED

6. CONCLUSIONS

* Estimator E1 (which is based on no county effect model) is biased.

* Regression estimators E2-E5 are essentially unbiased.

* All shrinkage estimators E6-E11 introduce some bias.

* Fuller estimator E12 is biased and behaves as a shrinkage estimator.

* E10 and E12 have more bias and smaller MSE than E3.

* Estimator E10 has smaller MSE than E12 when county effect is not large, but their MSE's are about equal in the presence of significant county effect.

* Decrease in MSE by shrinkage is accompanied by an increase in bias.

* Slope using the pooled within county variation is preferable to the slope using the total variation when county effects are large.

REFERENCES

Battese, G.E. and Fuller, W.A. (1981). Prediction of county crop areas using survey and satellite data. The 1981 Proceedings of the Section on Survey Research Methods, American Statistical Association, Washington, D.C., pp. 500-505.

Chhikara, Raj S. (1987). A simulation study for an evaluation of USDA LANDSAT derived crop acreage estimation. Technical Report, University of Houston - Clear Lake. Prepared for NASS/USDA, February 1987.

Chhikara, R.S., Lundren, J.C. and Houston, A.G. (1986). Crop acreage estimation using a LANDSAT based estimator as an auxiliary variable. *IEEE Transactions on Geoscience and Remote Sensing*, GE-24, No. 1, pp. 157-168.

Cochran, W.G. (1977). *Sampling Techniques*, Third Edition, John Wiley, New York.

Fuller, Wayne A. (1986). Small area estimation as a measurement error problem. Proceedings of the Small Area Estimation Conference held in June, 1986.

Hanuschak, G., Sigman, R., Craig, M., Ozga, M., Luebbe, R., Cook, P., Kleweno, D., and Miller, C. (1979). Obtaining timely crop area estimates using ground-gathered and LANDSAT data. Technical Bulletin No. 1609, Economics, Statistics, and Cooperatives Services, U.S. Department of Agriculture.

Harter, Rachel M. (1983). Small area estimation using nested-error models and auxiliary data. Ph.D. Thesis, Iowa State University, Ames, Iowa.

Rao, J.N.K., (1986) Small area estimation--an overview. Proceedings of the Small Area Estimation Conference held in June, 1986.

Walker, G. and Sigman, R. (1984). The use of LANDSAT for county estimates of crop areas: evaluation of the Huddleston-Ray and the Battese-Fuller estimators for the case of stratified sampling. *Communications in Statistics*, *13*, No. 23, pp. 2975 - 2996