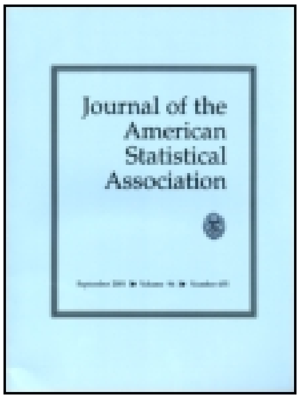

Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uasa20

Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data

Robert E. Fay III [a] & Roger A. Herriot [b]
[a] Statistical Methods Division, U.S. Bureau of the Census, Washington, DC, 20233, USA
[b] Population Division, U.S. Bureau of the Census, Washington, DC, 20233, USA
Published online: 06 Apr 2012.

To cite this article: Robert E. Fay III & Roger A. Herriot (1979): Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data, Journal of the American Statistical Association, 74:366a, 269-277

To link to this article: http://dx.doi.org/10.1080/01621459.1979.10482505

Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data

ROBERT E. FAY III and ROGER A. HERRIOT*

An adaptation of the James-Stein estimator is applied to sample estimates of income for small places (i.e., population less than 1,000) from the 1970 Census of Population and Housing. The adaptation incorporates linear regression in the context of unequal variances. Evidence is presented that the resulting estimates have smaller average error than either the sample estimates or an alternate procedure of using county averages. The new estimates for these small places now form the basis for the Census Bureau's updated estimates of per capita income for the General Revenue Sharing Program.

KEY WORDS: Biased estimation; Small-area statistics; James-Stein; Income; Revenue sharing.

1. INTRODUCTION

The State and Local Fiscal Assistance Act of 1972 specifies the distribution of funds to states and units of general-purpose local government for operational or capital expenditures. The resulting General Revenue Sharing Program, administered by the Treasury Department, allocates monies to state and local governments on the basis of interdependent formulas: Funds are distributed to approximately 39,000 units of local government by dividing state allocations. Statistics on population, per capita income (PCI), and adjusted taxes are used to determine the allocations within states.

The Census Bureau provides the Treasury Department with current estimates of these statistics for the states and local jurisdictions receiving funds under the General Revenue Sharing Program. Separate methodologies are used to update the population counts and the income figures from the 1970 Census of Population and Housing. Data from the Internal Revenue Service (IRS) and the Bureau of Economic Analysis form the basis for updating the census estimates of income. In general, the 1970 census values of PCI in 1969 are multiplied by ratios of an administrative estimate of PCI in the current year and a similarly derived estimate for 1969. Herriot (1977) described this methodology in greater detail.

The 1970 census thus constitutes the foundation for the current estimates of PCI, but a significant problem arises in this regard. Of the estimates required, more than one-third, or approximately 15,000, are for places with population of fewer than 500 persons in 1970. Because income was collected on the basis of a 20 percent sample in the 1970 census, the sampling error for the estimates for such small places is an important consideration. For a place of 500 persons, the coefficient of variation (relative standard error) for the 1970 census estimate of PCI is about 13 percent; for a place of 100 persons, 30 percent. The magnitude of these sampling errors initially led the Census Bureau and the Treasury Department to agree to set aside the census figures for these places and to substitute the respective county average figures instead.

This substitution of the county figures for the census estimates for places with fewer than 500 persons would seem to be based on the following statistical reasoning: For larger places the sampling errors of the census sample estimates are sufficiently small so that they might be chosen as the best estimates, but for smaller places substituting biased estimates with negligible sampling error (the county values) for estimates with large sampling error is preferable. This sort of reasoning is of course present, formally or informally, in a great deal of statistical practice. Aspects of this particular problem suggested, however, that this initial solution might be improved considerably: The dividing line of 500 persons was essentially an arbitrary choice; the census estimates for a significant number of small places were many standard errors removed from the county values that had been substituted, thus suggesting a failure of the county values to represent adequately the true values for these places; and auxiliary data related to PCI from the IRS and the 1970 census had not been incorporated in the estimation. In this article we shall describe the application of procedures adapted from the original James-Stein estimator to the problem of estimating 1969 PCI for these small places by addressing each of the deficiencies of the original choice. The revised estimator consisted of the following elements:

1. Fitting a regression equation to the census sample estimates, using as independent variables the

* Robert E. Fay III is a Staff Assistant, Statistical Methods Division, and Roger A. Herriot is Assistant Chief for Socio-economic Programs, Population Division, both at the U.S. Bureau of the Census, Washington, DC 20233. The authors wish to thank staff members, particularly Emmet Spiers, for their help on the project and to acknowledge the support of Daniel B. Levine, Associate Director for Demographic Fields, and of Harold Nisselson, Associate Director for Statistical Standards and Methodology. The authors also wish to thank Carl Morris for helpful comments on the research and to express appreciation to an associate editor and a referee for their suggestions on the exposition of this material.

June 1979, Volume 74, Number 366
Applications Section

county values, tax-return data for 1969, and data on housing from the 1970 census;
2. Measuring the goodness of fit between the regression equation and the sample data, taking into consideration the expected contribution of sampling error to the observed differences, and deriving an estimated measure of average lack of fit between the regression estimates and the underlying true values for the places;
3. Forming a weighted average of the sample and the regression estimate for each place, adjusting the weights to reflect the relative magnitudes of the average lack of fit of the regression and the variance of the sample estimate; and
4. Constraining each such weighted average to be within one standard error of the sample estimate, thus preventing severe disagreement between the sample and final estimate.

Because of the mathematical and logical consistency of the revised procedures, and on the basis of independent empirical evidence, the Census Bureau has used this methodology in forming the estimates for 1974 and subsequent years. To our knowledge, the Census Bureau's use is the largest application of James-Stein procedures in a federal statistical program.

2. THE JAMES-STEIN ESTIMATOR AND ITS DESCENDANTS

In order to describe the nature of the estimator that we developed for this problem, we will briefly review some of its predecessors. Other authors, for example, Efron and Morris (1973a, 1975), have given a more comprehensive presentation of much of the material summarized in this section.

Suppose that we have a single observation $\mathbf{Y} = (Y_1, \ldots, Y_k)^T$ from a k-dimensional multivariate normal distribution with mean $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)^T$ and covariance matrix $D\mathbf{I}$, where D is a known scalar constant. Equivalently, the Y_i's are assumed to be independent and identically distributed according to normal distributions with means θ_i and variance D, that is, $Y_i \sim_{\text{ind}} N(\theta_i, D)$. The maximum likelihood estimator of $\boldsymbol{\theta}$ is $\mathbf{Y}$; each Y_i is the obvious estimate of its respective θ_i. Stein (1955) showed that for $k \geq 3$, $\mathbf{Y}$ is not admissible under the usual loss function defined for an estimator $\boldsymbol{\theta}^* = (\theta_1^*, \ldots, \theta_k^*)^T$ by

$$R(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = EL(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = \sum_i E_{\theta_i}(\theta_i - \theta_i^*)^2 . \quad (2.1)$$

We have, of course, $R(\boldsymbol{\theta}, \mathbf{Y}) = kD$. For $k \geq 3$, James and Stein (1961) exhibited the estimator $\boldsymbol{\delta}' = (\delta_1', \ldots, \delta_k')^T$ defined by

$$\delta_i' = (1 - ((k-2)D/S))Y_i , \quad (2.2)$$

where

$$S = \sum_i Y_i^2 \quad (2.3)$$

with risk $R(\boldsymbol{\theta}, \boldsymbol{\delta}') < kD$ for all $\boldsymbol{\theta}$. Consequently, $\boldsymbol{\delta}'$ dominates the maximum likelihood estimator $\mathbf{Y}$ with respect to the loss function (2.1).

The result is far from obvious: The Y_i's estimate the respective θ_i's, which in turn need to have no specific relationship to each other; yet by combining information from apparently unrelated estimation problems, the expected total loss (2.1) may be reduced. To do this, $\boldsymbol{\delta}'$ in effect shrinks $\mathbf{Y}$ towards $\mathbf{0}$; that is, each component of $\mathbf{Y}$ is proportionally reduced by the same factor. The amount of shrinkage depends on the relative closeness of $\mathbf{Y}$ to $\mathbf{0}$; for $\mathbf{Y}$ near $\mathbf{0}$, the shrinkage is substantial, while for $\mathbf{Y}$ far from $\mathbf{0}$, $\boldsymbol{\delta}'$ becomes essentially $\mathbf{Y}$. Roughly speaking, to the extent that $\boldsymbol{\theta}$ lies close to $\mathbf{0}$, $\mathbf{Y}$ is also in a sense an estimate of $\mathbf{0}$, and $\boldsymbol{\delta}'$ incorporates this information in estimating $\boldsymbol{\theta}$.

James and Stein noted that (2.2) could be uniformly improved for all $\boldsymbol{\theta}$ by restricting $(k-2)D/S$ to $[0, 1]$, replacing this term by 1 in cases in which it was greater. This restriction prevents $\mathbf{Y}$ from being partially reflected through the origin and is routinely incorporated in applications of (2.2).

The estimator $\boldsymbol{\delta}'$ has inspired a number of important variations. The link between $\boldsymbol{\delta}'$ and many of the subsequent adaptations can be traced most easily through the correspondence between (2.2) and a classical Bayes estimator. Suppose that we assume that $\boldsymbol{\theta}$ has a prior distribution $\theta_i \sim_{\text{ind}} N(0, A)$, that is, normal with variance A. Then the Bayes estimator $\boldsymbol{\theta}_B^*$ of $\boldsymbol{\theta}$ is given by

$$\boldsymbol{\theta}_B^* = (1 - (D/(A + D)))\mathbf{Y} . \quad (2.4)$$

Thus, the Bayes estimator in this situation also shrinks $\mathbf{Y}$ towards $\mathbf{0}$.

The James-Stein estimator (2.2) mimics the Bayes estimator in the following manner: Under the given prior distribution $\theta_i \sim_{\text{ind}} N(0, A)$, the expectation of $(k-2)D/S$, taken over the joint distribution of $\boldsymbol{\theta}$ and $\mathbf{Y}$, is $D/(D + A)$, showing the correspondence between (2.2) and (2.4). In the Bayesian context, regardless of the value of A, (2.2) approximates the Bayes estimator (2.4) by in effect estimating A on the basis of $\mathbf{Y}$. This principle forms the basis from which the other estimators discussed here are derived. In each instance, an estimate A^* of A is obtained, providing both a notion of the average variation of θ_i about some prior estimate and an indication of how much weight should be given to the prior and sample estimates in order to estimate θ_i.

An immediate generalization of (2.2) follows in the case in which a p-dimensional row vector $\mathbf{X}_i$ is available for each θ_i, representing auxiliary information about θ_i. For $Y_i \sim_{\text{ind}} N(\theta_i, D)$ and $\theta_i \sim_{\text{ind}} N(\mathbf{X}_i\boldsymbol{\beta}, A)$, with a uniform (improper) prior distribution on $\boldsymbol{\beta}$, the regression estimate (for $\mathbf{X}^T\mathbf{X}$ of full rank p),

$$Y_i^* = \mathbf{X}_i(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \quad (2.5)$$

may be combined with the sample estimate Y_i to form the Bayes estimator

$$\theta_{B_i}^* = Y_i^* + (1 - (D/(D + A)))(Y_i - Y_i^*) \quad (2.6)$$

in a form similar to (2.4). Equivalently,

$$\theta_{Bi}^* = (D/(D + A))Y_i^* + (A/(D + A))Y_i \quad (2.7)$$

expresses the estimator as a weighted average of Y_i^* and Y_i. The James-Stein analogue of (2.6) and (2.7) for $p < k - 2$ is

$$\delta_i' = Y_i^* + (1 - ((k - p - 2)D/S))(Y_i - Y_i^*) \quad (2.8)$$

$$= ((k - p - 2)D/S)Y_i^* + (1 - ((k - p - 2)D/S))Y_i \quad (2.9)$$

where

$$S = \sum_i (Y_i - Y_i^*)^2 . \quad (2.10)$$

A special case of (2.5) and (2.8) through (2.10) is for $p = 1$, $X_i = 1$: (2.5) makes each Y_i^* the mean of all of the Y_i's, and (2.9) averages each Y_i with the mean. (The estimator (2.5) and (2.8) through (2.10) in the general case in fact follows directly from (2.2) and (2.3) without requiring a Bayesian formulation, but the intent of the estimator is more clearly illustrated in the Bayesian context.)

Efron and Morris (1971, 1972) remarked that both Bayes estimators such as (2.4) and empirical Bayes estimators such as (2.2) may perform well overall but poorly on individual components. In these instances the shrinkage of (2.2) or (2.4), which benefits most components of $\mathbf{Y}$, is singularly inappropriate for the particular θ_i. For the Bayes case (2.4), θ_i may be unusual relative to the prior distribution, while for the empirical Bayes case (2.2), θ_i may lie much further from 0 than the other components of $\boldsymbol{\theta}$. Efron and Morris suggested a straightforward compromise, which consists of restricting the amount by which δ_i' differs from Y_i by some multiple of the standard error of Y_i. With this restriction, (2.2) becomes

$$\delta_i'' = \delta_i' \quad \text{if} \quad Y_i - c \le \delta_i' \le Y_i + c \quad (2.11)$$

$$= Y_i - c \quad \text{if} \quad \delta_i' < Y_i - c \quad (2.12)$$

$$= Y_i + c \quad \text{if} \quad \delta_i' > Y_i + c \quad (2.13)$$

The estimator (2.11) through (2.13) compromises between limiting the maximum possible risk to any component and preserving the average gains of $\boldsymbol{\delta}'$. The choice $c = D^{\frac{1}{2}}$, for example, ensures that $E(\delta_i'' - \theta_i)^2 < 2D$, while retaining more than 80 percent of the average gain of $\boldsymbol{\delta}'$ over $\mathbf{Y}$.

For $Y_i \sim_{\text{ind}} N(\theta_i, D_i)$, the possible strategies for extending the James-Stein estimator are numerous but more theoretically difficult if the D_i's are known but not all equal. The most simple extension of (2.2) may be derived by assuming a Bayes prior $\theta_i \sim_{\text{ind}} N(0, AD_i)$. This problem may be solved by transforming $\mathbf{Y}$, applying (2.2) to the vector of elements $Y_i/D_i^{\frac{1}{2}}$, which have the common variance $D = 1$. The resulting $\boldsymbol{\delta}'$ from (2.2) may be transformed back to the original scale by computing $\delta_i' D_i^{\frac{1}{2}}$. Even outside the Bayesian formulation, this estimator dominates the maximum likelihood estimator $\mathbf{Y}$ with respect to the loss function

$$R(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = \sum_i E_{\theta_i}(\theta_i - \theta_i^*)^2/D_i \quad (2.14)$$

for all $\boldsymbol{\theta}$. (A similar approach may be used to extend (2.8).) This estimator will be most effective against a Bayes prior in which the variance of the prior distribution is proportional to the sampling variance. The resulting estimator applies an equal amount of shrinkage to each component of $\mathbf{Y}$.

In many applications, however, the linkage between the sampling variance of Y_i about θ_i and the Bayes variance of θ_i about 0 is less direct. An alternate approach is to develop an estimator that more closely parallels the Bayes estimator for the prior distribution $\theta_i \sim_{\text{ind}} N(0, A)$, that is, with constant prior variance regardless of D_i. Efron and Morris (1973a) first proposed an extension of (2.2) under this second assumption. The estimator that we used in this application, however, more closely resembled one suggested by Carter and Rolph (1974). In considering the situation $Y_i \sim_{\text{ind}} N(\theta_i, D_i)$ and $\theta_i \sim_{\text{ind}} N(\nu, A)$, with known D_i but unknown ν and θ_i, they observed for the weighted sample mean

$$\nu^* = \sum_i Y_i/(A + D_i)/\sum_i 1/(A + D_i) \quad (2.15)$$

that

$$E\left(\sum_i \frac{(Y_i - \nu^*)^2}{A + D_i}\right) = k - 1 \quad (2.16)$$

for the joint expectation over both $\mathbf{Y}$ and $\boldsymbol{\theta}$, when A is a known constant. They suggested estimating A as the unique solution $A^* \ge 0$ such that (2.15) and (2.16) are simultaneously satisfied when the expectation operator is omitted from (2.16),

$$\sum_i \frac{(Y_i - \nu^*)^2}{A^* + D_i} = k - 1 . \quad (2.17)$$

They set $A^* = 0$ if no positive joint solution of (2.15) and (2.17) exists. Each θ_i is estimated by a weighted average of Y_i and ν^*,

$$\delta_i' = (A^*/(A^* + D_i))Y_i + (D_i/(A^* + D_i))\nu^* . \quad (2.18)$$

The estimator that we applied to the 1970 census estimates of PCI is an extension of (2.15), (2.17), and (2.18) to the linear regression case. We considered $Y_i \sim_{\text{ind}} N(\theta_i, D_i)$ and $\theta_i \sim_{\text{ind}} N(\mathbf{X}_i\boldsymbol{\beta}, A)$ for a p-dimensional row vector $\mathbf{X}_i$ and regression coefficients $\boldsymbol{\beta}$ with an (improper) uniform prior distribution. The row vectors $\mathbf{X}_i$ and sampling variances D_i were known, but $\boldsymbol{\beta}$ and A were both to be estimated from the data.

To derive the estimator, we first considered relationships when A was known. Over the joint distribution of $\mathbf{Y}$ and $\boldsymbol{\theta}$ in this case, the weighted regression estimates

$$Y_i^* = \mathbf{X}_i(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y} \quad (2.19)$$

where $\mathbf{V}$ is a diagonal matrix with $V_{ii} = D_i + A$ give the minimum variance unbiased estimates of $\mathbf{X}_i\boldsymbol{\beta}$, the prior means of θ_i. (These estimates are also the posterior means of $\mathbf{X}_i\boldsymbol{\beta}$.) Over this same joint distribution with known A,

$$E\left(\sum_i \frac{(Y_i - Y_i^*)^2}{A + D_i}\right) = k - p \ . \qquad (2.20)$$

Equation (2.20) is a standard result in weighted least squares under the preceding assumptions and may be found in texts by Rao (1965, pp. 187–188) and by Draper and Smith (1966, pp. 77–81).

Following the program of Carter and Rolph, we estimated A from the data by removing the expectation operator from (2.20)

$$\sum_i \frac{(Y_i - Y_i^*)^2}{A^* + D_i} = k - p \qquad (2.21)$$

and found the unique $A^* \geq 0$ solving both (2.21) and (2.19), using $A^* = 0$ when no positive solution could be found. The estimator was then

$$\delta_i' = (A^*/(A^* + D_i))Y_i + (D_i/(A^* + D_i))Y_i^* \ . \qquad (2.22)$$

This weighted average of the sample and regression estimate would be the classical Bayes estimator in the case that A were known. The restrictions (2.11) through (2.13) were then imposed on each component δ_i' with $c = D_i^{\frac{1}{2}}$. The actual numerical operations used to solve equations (2.19) and (2.21) simultaneously are described in the Appendix. (We also discuss there an alternate estimator for this problem based on a maximum likelihood approach to fitting the model $\theta_i \sim_{\text{ind}} N(\mathbf{X}_i\boldsymbol{\beta}, AD_i^{\alpha})$, where $\boldsymbol{\beta}$, A, and α may be jointly estimated from the data.)

We have traced the development of this estimator here through its relation to general results for the James-Stein estimator; yet parallel research in estimation for local areas also precedes these results. Ericksen's work (1973, 1974) explored use of sample data to determine regression estimates for small areas, and Madow (see Madow and Hansen 1975) first remarked on the merit of forming a weighted average of the sample and regression estimates. The estimator presented here represents a further development of these basic ideas.

3. APPLICATION TO ESTIMATION OF INCOME FOR SMALL PLACES

This section will describe the steps used to apply the preceding theory to the estimation of PCI in 1969. The elements of the approach consisted of

1. A division of the total problem into a set of separate estimation problems;
2. Logarithmic transformation of the census values to a scale in which the sampling variances could be considered known;
3. Identification and similar transformation of auxiliary variables available for each place;
4. Derivation of a regression estimate for each place, which was combined with the sample estimate by using (2.19), (2.21) and (2.22), and (2.11) through (2.13);
5. Retransformation of the resulting estimates back to the original scale; and
6. A final proportional readjustment of the resulting estimates to sum to sample estimates of total income at the state and county level.

The following discussion treats each of these points in detail.

Although the initial substitution of the county values of PCI had been carried out only for places of population less than 500 before this investigation, we extended the problem to all places with 20 percent sample estimates of population less than 1,000. (The 20 percent sample count, which is approximately proportional to the number of sample persons in the place, is often in minor disagreement with the complete count for places of this size.) We divided the overall problem into 100 separate estimation problems along two dimensions: a division between places with 20 percent sample estimates of population less than 500 and those between 500 and 999, and an independent consideration of each state. An average of 200 to 300 places with population less than 500 in a given state were thus treated as a joint estimation problem, although there was considerable variation in the size of this group. In some states only 10 or 20 cases were involved. (In addition, some states required estimates for two kinds of geography, places and townships. For simplicity we will discuss the problem for places only, although parallel procedures were applied separately to obtain estimates for the townships.)

For almost all places, a sample estimate Z_i and a weighted 20 percent sample count N_i were available. As part of the processing of the 1970 census, variance computations were performed in eight states and the findings generalized to the rest of the country (U.S. Bureau of the Census 1976, pp. 11-8–11-9). An unpublished finding of this generalization was the approximation of the coefficient of variation of Z_i as $3.0/N_i^{\frac{1}{2}}$. Because the coefficient of variation does not depend on the expected value, the standard deviation increases in direct proportion to the expected value. Hence, the log transformation stabilizes this variance, and the variance of $Y_i = \ln(Z_i)$, the natural logarithm of Z_i, is approximately $9.0/N_i$ and does not depend on the expected value of Z_i. This procedure of stabilizing the variances has appeared in some other applications of the James-Stein estimator (e.g., Carter and Rolph 1974).

Each place, without exception, has an associated county value of PCI from the 1970 census. (With a handful of exceptions, places do not cross county lines.) We computed the natural logarithms of these county figures for use as an independent variable in the regression model. Because of the considerably larger county populations, this variable has typically negligible sampling error.

Two other important sources of data are available for these places: the value of owner-occupied housing from the 1970 census, and the average adjusted gross income per exemption from the 1969 IRS returns for 1969. Both variables are free from sampling error, but each has other limitations. The value of owner-occupied housing was collected in the 1970 census only for nonfarm dwellings; we consequently chose to omit this variable from the analysis for places with a substantial proportion of farm residences. The IRS data, on the other hand, are affected by errors in coding tax returns to Census Bureau geography on the basis of mailing address. Some places more than others are affected by substantial ambiguity between the mailing addresses and place boundaries. Places thus affected were identified on the basis of unusual ratios between the number of exemptions coded to the place and the 100 percent population count, and in such cases the IRS results were omitted from the analysis. The IRS results were also dropped for places with significant boundary changes since 1970.

After editing the IRS and housing data in the preceding fashion, the natural logarithms of each of these variables were taken, whenever the case met the criteria for inclusion, and matched to logarithms of the respective county values for these variables. Four separate regressions were possible:

1. A constant term and the logarithm of PCI for the county (with $p = 2$ in the notation of the preceding section);
2. A constant term, the logarithm of PCI for the county, and logarithms of the value of housing for both the place and the county ($p = 4$);
3. A constant term, the logarithm of PCI for the county, and logarithms of IRS-adjusted gross income per exemption for both the place and the county ($p = 4$); and
4. A constant term, the logarithm of PCI for the county, the logarithms of the value of housing for both the place and the county, and the logarithms of IRS-adjusted gross income per exemption for both the place and the county ($p = 6$).

Inclusion of both the logarithms of the county and place values for either the housing or tax data is mathematically equivalent to inclusion of both the logarithm of the place value and the logarithm of the ratio of the place to county values. Thus, the regression was able to use the data for the places on an absolute scale, across the entire state, and in relation to the county values.

Our strategy consisted of computing each of the four regressions for those Y_i's with the necessary independent variables for the particular regression by solving (2.19) and (2.21). Using the regression equation corresponding to all the available variables for each place, we computed (2.22) subject to a constraint of the form (2.11) through (2.13). For states with only a few small places, the number of regressions fitted was restricted by insufficient data. Places without any census sample estimate were estimated directly from the regression (2.19).

The preceding estimates developed on the logarithmic scale were transformed back to the original scale. A final two-dimensional iterative proportional adjustment (raking) was applied to all places in each state, including those with population more than 1,000, to force two constraints: the addition of total estimated income (PCI times population) for places belonging to the classes of places with less than 500, 500 to 999, and more than 1,000, to the sample estimates of these totals at the state level; and the addition of the estimates for all places, disregarding size, within a county to the sample estimate of the total for all places in the county. These adjustments, on the order of 1 or 2 percent, were quite small relative to the other aspects of this estimation problem, but they imposed a logical consistency on the outcome and ensured that the analysis of the data on the logarithmic scale did not induce systematic bias across all small places.

The values of A^* provide a measure of the average fit of the regression models to the sample data, after allowance is made for sampling error in Y_i. Table 1 shows the values of A^* obtained for the states with the largest number of places of size less than 500. In a sense, a value for A^* of .045 indicates an average level of accuracy equivalent to the accuracy of a sample estimate for a place of size 200 ($9.0/200 = .045$, from the formula for the approximate coefficient of variation noted earlier), because (2.22) weights the sample and regression estimates equally in this case. (The value .045 for A^* may be thought to correspond to an average—in the sense of root mean square—error of prediction by the regression of the true value of PCI of approximately 21 percent, because $.045 = 0.21^2$.) In turn, the expected improve-

1. Estimated A^* for Places With 20 Percent Sample Estimates of Population Less Than 500

	Regression Equation			
States	County	County and Tax	County and Housing	County, Tax, and Housing
States With More Than 500 Places in Class				
Illinois	.036	.032	.019	.017
Iowa	.029	.011	.017	.000
Kansas	.064	.048	.016	.020
Minnesota	.063	.055	.014	.019
Missouri	.061	.033	.034	.017
Nebraska	.065	.041	.019	.000
North Dakota	.072	.081	.020	.004
South Dakota	.138	.138	.014	—
Wisconsin	.042	.025	.025	.004
States With 200 to 500 Places in Class				
Arkansas	.074	.036	.039	.018
Georgia	.056	.081	.067	.114
Indiana	.040	.012	.003	.000
Maine	.052	.015	—	—
Michigan	.040	.032	.028	.023
Ohio	.034	.015	.004	.004
Oklahoma	.063	.027	.049	.036
Pennsylvania	.020	.018	.016	.011
Texas	.092	.048	.056	.040

NOTE: A dash (—) indicates that the regression was not fitted because of too few observations.

2. Estimated A^* for Places With 20 Percent Sample Estimates of Population 500 to 999

	Regression Equation			
States	County	County and Tax	County and Housing	County, Tax, and Housing
States With More Than 250 Places in Class				
Illinois	.032	.023	.012	.008
Indiana	.017	.014	.007	.009
Michigan	.019	.014	.005	.008
Minnesota	.056	.040	.021	.007
New York	.052	.015	.028	.006
Ohio	.024	.010	.005	.000
Pennsylvania	.035	.025	.015	.026
Wisconsin	.039	.030	.014	—
States With 100 to 250 Places in Class				
Iowa	.017	.005	.016	.004
Kansas	.025	.010	.014	.008
Maine	.022	.021	—	—
Missouri	.042	.019	.011	.013
Nebraska	.027	.007	.008	.008
Texas	.050	.017	.013	.012

NOTE: A dash (—) indicates that the regression was not fitted because of too few observations.

ment of an equal weighting of two estimates with equivalent estimates of error would be to reduce the variance by one-half, or to give, on the average, the combined estimate an accuracy that would be achieved by a sample estimate alone for a place of 400 persons, that is, a relative error of about 15 percent.

In fact, for the regression equation based on county values alone, more than half the states in Table 1 have values for A^* greater than .045, suggesting that the county value is often not so good a prediction of the true value as the sample estimate for places with more than 200. Parenthetically, this finding suggests that the original decision that had preceded this investigation, namely, to replace the sample estimates with the county values for places of size less than 500, actually exaggerated the ability of the county values alone to serve as a good prediction for these places. If no James-Stein estimation was to have been done, it would have been better as a rule to use sample estimates down to a population of approximately 200, instead of 500. The James-Stein procedures here, however, allow a combination of the two estimates to achieve an improvement in the average accuracy of prediction.

Table 1 also shows that regressions, involving either IRS or housing data, but especially those including both, are significantly more effective in estimating the true values than the regression on the county values alone. The fit for these other regressions is particularly good among states in the North Central Region. (One large value of A^* for Georgia is based on a relatively small number of cases.)

Before processing the entire set of estimates, we experimented with alternative forms for the regression equations, using the value of A^* as the criterion. Surprisingly, we did not find any appreciable improvement through further transformation of the independent variables.

Table 2 displays values of A^* obtained for places between 500 and 999. The values in the table tend to be somewhat less than those in the first table, indicating slightly better fit for larger places. The differences between Tables 1 and 2, however, are less than the difference between the average sampling errors of these two groups of places. Roughly speaking, places with less than 500 would have an average size of 250, while places between 500 and 999 would have an average size of about 750. Thus, the average sampling variances might differ by a factor of up to 3 between the two groups, while the ratios between the average estimated A^*'s are about 1.5. Thus, the assumption that the prior variance A^* is independent of D_i seems to hold reasonably, although not perfectly. Furthermore, possible inadequacies in the approximation used to give the sampling variances may affect the estimates in Table 1. In general, overestimates of the sampling variances will lead to underestimates of A^*.

For cases in which there may be some linkage between the sampling variance D_i and the variation of the true values about the predicted values, we include in the Appendix a procedure to fit the assumption $\theta_i \sim_{\text{ind}} N(X_i\beta,\ AD_i^{\alpha})$. Use of the procedure would be encouraged, however, only if many cases, perhaps on the order of hundreds, were available and the true values of D_i were known to almost complete accuracy.

4. EVALUATION OF THE ESTIMATOR

The values of A^* indicated that the revised estimator would be superior to the county values. In some applications, these statistics may constitute the only available assessment of the improvement achieved by the James-Stein estimator, where small values of A^* relative to the sampling variances D_i point to substantial overall gains. For this problem, however, we devised two additional demonstrations of characteristics of the revised estimator: one based on a limited number of special censuses taken in 1973, and the other derived from the 1970 data used in the estimation.

As a general verification of the methodology to update the 1970 census estimates of population and income on the basis of changes in administrative data, the Census Bureau conducted complete censuses of a random sample of places and townships in 1973, collecting income for 1972 on a 100 percent basis. (The difference in years here is the same as for the 1970 census collecting income for 1969; in general, Census Bureau income questions are asked for income during the preceding calendar year.) Of these special censuses, 17 were for places of size less than 500 in 1970, and 7 fell into the interval 500 to 999. In general, the methodology to update the estimates produced for each place a factor f_i used to multiply a base figure for 1969. By keeping this updating factor f_i constant, three separate estimates of PCI for 1972 were

3. Comparison of Selected 1972 PCI Estimates With 1973 Special Census Values of 1972 PCI

		1972 PCI Estimates and Percentage Difference From Special Census PCI					
		Using 1970 Sample Base		Using Revised Base (James-Stein)		Using County Base	
Special Census Areas	1973 Special Census 1972 PCI	1972 Estimate	Percentage Difference	1972 Estimate	Percentage Difference	1972 Estimate	Percentage Difference
	1970 Census Weighted Sample Population Less Than 500						
Newington, Ga.	$2,019	$2,225	10.2	$2,302	14.0	$2,279	12.9
Foosland Village, Ill.	2,899	2,771	4.4	3,199	10.3	3,796	30.9
Bonaparte, Iowa	2,331	3,126	34.1	2,942	26.2	2,542	9.1
McNary, La.	2,333	2,303	1.3	2,527	8.3	2,908	24.6
Freeborn Village, Minn.	2,741	3,693	34.7	3,338	21.8	2,922	6.6
Spruce Valley Twp., Minn.	2,430	1,894	22.1	1,949	19.8	2,076	14.6
Jacksonville, Mo.	2,723	2,338	14.1	2,611	4.1	3,233	18.7
Thayer, Nebr.	2,742	2,245	18.1	2,870	4.7	3,452	25.9
Benton Town, N.H.	1,788	2,874	60.7	3,284	78.7	3,570	99.7
Nora Twp., N.Dak.	1,780	2,629	47.7	2,754	54.7	3,476	95.3
Riga Twp., N.Dak.	1,454	2,749	89.1	2,411	65.8	2,711	86.5
Deer Creek, Okla.	2,451	2,493	1.7	2,673	9.1	2,762	12.7
Dudley Borough, Pa.	2,446	2,168	11.4	2,411	1.4	2,608	6.6
Brookings Twp., S.Dak.	3,132	3,400	8.6	3,309	5.7	2,395	23.5
Valley Twp., S.Dak.	1,574	1,946	23.6	1,972	25.3	2,114	34.3
Bryant Twp., S.Dak.	2,412	1,120	53.6	2,158	10.5	2,695	11.7
Parrish Town, Wis.	3,567	5,399	51.4	4,079	14.4	2,721	23.7
Average Percentage Difference	—	—	28.6	—	22.0	—	31.6
	1970 Census Weighted Sample Population Between 500 and 999						
Caswell Plantation, Maine	$1,946	$2,656	36.5	$2,490	28.0	$2,646	36.0
Sugar Creek Twp., Mo.	2,224	2,035	8.5	2,315	4.1	2,018	9.3
Jeromesville, Ohio	3,329	3,081	7.4	3,418	2.7	3,072	7.7
Rush Twp., Ohio	2,241	2,545	13.6	2,619	16.9	2,546	13.6
Dennison Twp., Pa.	3,521	4,411	25.3	4,095	16.3	4,430	25.8
Manor, Tex.	2,062	2,746	33.2	2,765	34.1	2,740	32.9
Derby Center, Vt.	2,968	2,694	9.2	2,754	7.2	2,675	9.9
Average Percentage Difference	—	—	19.1	—	15.6	—	19.3

possible: multiplying the census sample estimate by f_i; multiplying the revised James-Stein estimate for 1969 by f_i; or multiplying the county values by f_i. The last, of course, was the original choice for the 1972 Revenue Sharing estimates. Comparison of the three sets of 1972 estimates with the special census results provides an indirect assessment of three sets of estimates for 1969, because each set is affected by errors both in the bases and in the updating factors f_i. Table 3 presents the results. The revised James-Stein estimator shows smaller average errors and, to a lesser extent, a lower incidence of extreme error than either the sample estimates or the county values. (The reader may note, however, that the estimates, particularly for the revised James-Stein base, run consistently higher than the special census values. The explanation lies with the special censuses themselves. Approximately 60 additional special censuses not included in this table were taken at the same time for places with population greater than 1,000, where the 1970 census sample estimates are used as base figures. There too, the estimates fall slightly above the special census results. One factor possibly involved is that missing income was not imputed in the processing of the special censuses, while it was in the 1970 census. The special censuses estimates, which are based on only complete cases, may be subject to a downward bias for this reason.)

A second test illustrates the manner in which the revised estimates, far more than the county values,

4. Relation of 1969 Revised Estimates and 1969 County Averages to 1970 Census Sample Estimates for Groups of 10

	1969 Revised Estimates		1969 County Averages	
Relation to 1969 Sample Estimates	Number	Percentage	Number	Percentage
Total Groups	212	100.0	212	100.0
Within 10% of Sample PCI	172	81.1	111	52.4
Outside 10% of Sample PCI	40	18.9	101	47.6
Within One Standard Error	149	70.3	61	28.8
Between One and Two Standard Errors	28	13.2	60	28.3
Outside Two Standard Errors	35	16.5	91	42.9
Closer to Sample PCI	154	72.6	58	27.4

NOTE: For places with the ratio of 1969 IRS exemptions to 1970 census population between .8 and 1.1.

preserve much of the underlying dispersion of the true values for PCI among places. The logic of the test was simple: Although the estimates for places with population less than 500 have large sampling errors in the 1970 census, if they were pooled into a suitably large group, the sampling error of the resulting estimated sum would be relatively less. We selected a sample of places with valid IRS estimates of adjusted gross income per exemption and assembled them into groups of 10 after sorting them by the IRS values. (This grouping is legitimate in the sense that the groups are defined independently of the sample selection in the census.) Table 4 shows that the sum of the revised James-Stein estimates for 1969 more often falls closer to the sum of the census estimates than does the sum of the county values. This demonstration illustrates how the James-Stein estimates capture more of the true differences in income among these places than does the substitution of county values.

5. DISCUSSION

The General Revenue Sharing Program is one of a number of important federal programs that allocates funds according to formulas using statistical counts or estimates. A general study of critical considerations in the design and administration of these programs has recently appeared, *Statistical Policy Working Paper 1: Report on Statistics for Allocation of Funds* (U.S. Department of Commerce 1978). Although this study principally concerns policy issues beyond the scope of this article, the report reaffirms the need for accurate and timely data in these allocation programs.

The use of a James-Stein estimator in this instance does not form the precedent for its wholesale application by the Census Bureau to all other estimation problems involving small-area data, and we should emphasize the special circumstances in this application. Planning for the 1970 Census of Population and Housing did not anticipate the requirements of the State and Local Fiscal Assistance Act of 1972, which mandated the General Revenue Sharing Program. Consequently, the legislation forced a provisional program of estimation falling short of the ideal in this case, a 100 percent census for small places. The requirements of the act are in consideration in the planning of the 1980 census. In the meantime, however, the lack of sample estimates with acceptable statistical reliability forces a choice among alternatives, and in this instance James-Stein procedures provide an attractive solution.

Future applications may be required under certain conditions. For example, if the 1980 census is conducted with a 50 percent sample in small places, the sampling reliability of the estimates will be much greater than in 1970. Nonetheless, the sampling errors for places of size 200 or less may suggest a repetition of the same methodology. Parallel situations may arise in which the sampling errors of census or survey estimates may require the consideration of alternate estimators. In these instances, the James-Stein procedures may be viewed as a way to maximize the use of the data rather than as a means to replace them.

Although the theory is now sufficient to form the basis for many applications of the James-Stein estimator to practical problems, this article illustrates directions for further research. The general problem of unequal variances will require more investigation to produce estimators with good properties for practical applications. The question of how independent estimation problems may be grouped for joint consideration was partially addressed by Efron and Morris (1973b), but we conjecture that unequal variances introduce further complications in this question and that larger groupings than Efron and Morris suggested may have merit. (In retrospect, we would now contemplate further dividing the estimation problem along the dimension of population size in states with large numbers of small places, but we hesitated doing this initially in an effort to keep the procedures as uniform across states as possible.) The use of prior distributions other than the normal to motivate an empirical Bayes estimator may produce somewhat better results for practical problems in which only a few of the observations lie far from the general tendencies. Full solution of these problems may encourage further application of these techniques to practical problems.

APPENDIX

A1. Iterative Solution of (2.19) and (2.21)

Equation (2.19) for any specific value of A is, of course, simply weighted linear regression. To denote the dependence of (2.21) on the value of A, we will write $Y_i^*(A)$. Using the functions

$$f(A_n) = \sum_i \frac{(Y_i - Y_i^*(A_n))^2}{A_n + D_i} \tag{A.1}$$

$$g(A_n) = -\sum_i \frac{(Y_i - Y_i^*(A_n))^2}{(A_n + D_i)^2} \tag{A.2}$$

we started with $A_0 = 0$ and defined $A_{n+1} = A_n + (k - p - f(A_n))/g(A)$, constraining $A_{n+1} \geq 0$. The function g is an approximation to the derivative of f. Convergence is rapid, generally requiring less than 10 steps.

A2. Alternate Estimators

Carl Morris suggested to us a maximum likelihood approach to estimating A, which requires the simultaneous solution of (2.19) and

$$\sum_i \frac{(Y_i - Y_i^*)^2}{(A + D_i)^2} = \sum_i \frac{1}{A + D_i}. \tag{A.3}$$

Equations (A.3) and (2.21) weight the significance of the deviations $(Y_i - Y_i)^2$ differently: (A.3) places relatively more importance on the observations with small D_i than does (2.21). The maximum likelihood approach improves the efficiency of the estimation of A in the full Bayes setting. We preferred in this application, however, to balance out the estimation of A over the sample to ensure that A was representative of all places in the class less than 500 rather than of just the larger places.

A more complex model may be fitted for cases in which both the assumed Bayes variance of the population and the sampling variance are related to a measure of size. In such cases, $Y_i \sim_{\text{ind}} N(\theta_i, D_i)$ and $\theta_i \sim_{\text{ind}} N(\mathbf{X}_i\boldsymbol{\beta}, AD_i^{\alpha})$ may be a reasonable model. Maximum likelihood techniques may be used to estimate $\boldsymbol{\beta}$, A, and α jointly, requiring solution of (2.19) and

$$\sum_i \frac{(Y_i - Y_i^*)^2 D_i^{\alpha}}{(D_i + AD_i^{\alpha})^2} = \sum_i \frac{D_i^{\alpha}}{D_i + AD_i^{\alpha}} \qquad (A.4)$$

$$\sum_i \frac{(Y_i - Y_i^*)^2 D_i^{\alpha} \ln D_i}{(D_i + AD_i^{\alpha})^2} = \sum_i \frac{D_i^{\alpha} \ln D_i}{(D_i + AD_i^{\alpha})^2}. \qquad (A.5)$$

We would conjecture, however, that the sampling error of α is too large to make this estimator preferable to simpler versions unless many, possibly several hundred, observations were involved.

A3. Details of the Implementation

For each place, the data were edited according to the following rules:

1. The census sample estimates were considered to be missing if the sample estimate of the number of persons was zero, or if the estimated PCI was less than $200. The latter situation can arise from losses, particularly on farm income, but the difficulty of assigning a reasonable standard error to the estimate in this instance led us to exclude such cases.
2. The housing data were considered missing if more than 20 percent of the owner-occupied units were farm dwellings, or if the data were otherwise unavailable.
3. The IRS data (originally prepared according to 1972 geography) were considered missing if the boundary changes between 1969 and 1972 had involved more than a 10 percent change in population, or if the number of exemptions was less than 70 percent or more than 100 percent of the 100 percent census count.

The equations incorporating county, tax, and housing values were calculated only for states with 16 or more complete cases. Similarly, regressions with either tax or housing data only were fitted for 12 or more valid cases and the county-only regressions required at least 8.

[*Received December 1977. Revised December 1978.*]

REFERENCES

Carter, Grace M., and Rolph, John E. (1974), "Empirical Bayes Methods Applied to Estimating Fire Alarm Probabilities," *Journal of the American Statistical Association*, 69, 880–885.

Draper, N.R., and Smith H. (1966), *Applied Regression Analysis*, New York: John Wiley & Sons.

Efron, Bradley, and Morris, Carl (1971), "Limiting the Risk of Bayes and Empirical Bayes Estimators—Part I: The Bayes Case," *Journal of the American Statistical Association*, 66, 807–815.

——— (1972), "Limiting the Risk of Bayes and Empirical Bayes Estimators—Part II: The Empirical Bayes Case," *Journal of the American Statistical Association*, 67, 130–139.

——— (1973a), "Stein's Estimation Rule and Its Competitors—An Empirical Bayes Approach," *Journal of the American Statistical Association*, 68, 117–130.

——— (1973b), "Combining Possibly Related Estimation Problems," *Journal of the Royal Statistical Society*, Ser. B, 35, 379–421.

——— (1975), "Data Analysis Using Stein's Estimator and Its Generalizations," *Journal of the American Statistical Association*, 70, 311–319.

Ericksen, Eugene P. (1973), "A Method of Combining Sample Survey Data and Symptomatic Indicators to Obtain Population Estimates for Local Areas," *Demography*, 10, 137–160.

——— (1974), "A Regression Method for Estimating Population Changes for Local Areas," *Journal of the American Statistical Association*, 69, 867–875.

Herriot, Roger A. (1977), "Updating Per Capita Income for General Revenue Sharing," in *Small Area Statistics Papers, Series GE-41, No. 4*, U.S. Bureau of the Census.

James, W., and Stein, Charles (1961), "Estimation With Quadratic Loss," in *Proceedings of the Fourth Berkeley Symposium of Mathematical Statistics and Probability, Vol. 1*, Berkeley: University of California Press, 361–379.

Madow, W.G., and Hansen, M.H. (1975), "On Statistical Models and Estimation in Sample Surveys," in *Contributed Papers, 40th Session of the International Statistical Institute*, Warsaw, Poland, 554–557.

Rao, C. Radhakrishna (1965), *Linear Statistical Inference and Its Applications*, New York: John Wiley & Sons.

Stein, Charles (1955), "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1*, Berkeley: University of California Press, 197–202.

U.S. Bureau of the Census (1976), *U.S. Census of Population and Housing: 1970, Procedural History PHC(R)-1*, Washington, D.C.

U.S. Department of Commerce (1978), *Statistical Policy Working Paper 1: Report on Statistics for Allocation of Funds*, prepared by the Subcommittee on Statistics for Allocation of Funds, Federal Committee on Statistical Methodology.