EMPIRICAL BAYES ESTIMATION FOR MULTIPLE CHARACTERISTICS

Robert E. Fay, U.S. Bureau of the Census[1]
Stat. Methods Div., U.S. Bureau of the Census, Washington, DC 20233

Keywords: Components of variance, small area estimation, small domain estimation.

1. Introduction

Application of linear regression to small domain estimation (Ericksen 1973, 1974) has become well established. This method combines relationships estimated from sample survey data and auxiliary data available at the individual or small domain level. By combining the detail of the auxiliary data with the systematic relationships suggested by the survey data, the method produces estimates of characteristics for small domains for which the available sample survey data, by themselves, are of inadequate precision. Numerous applications of this basic approach have occurred. A recent volume (Platek, Rao, Särndal, and Singh 1987) describes several new developments in the area of small domain estimation, including applications of linear regression.

Typical applications of linear regression involve estimation for a single characteristic. Estimates of the characteristic are often available from sample survey data for all or many of the small domains of interest, but with high sampling variances. Auxiliary data related to the characteristic of interest are typically measured without sampling variance, since they are often derived from a previous census or administrative sources. The relationship between the characteristic and the auxiliary variables is then expressed in the form of a linear regression, with the survey estimates of the characteristic treated as the dependent variable and the auxiliary data employed as the independent variables. The predicted value from this regression fitted to the survey data represents a possible choice as a small domain estimate.

Composites involving both the regression estimate and the sample estimate may also be considered. For example, empirical Bayes estimation (e.g., Fay and Herriot 1979) derives the form of the composite based on the comparison of the fit of the model relative to the variance of the sample estimates.

This paper describes multivariate extensions of the linear regression approach. A multivariate treatment may be motivated under either of the following circumstances:

1. The interest is in more than one characteristic at the small domain level. A multivariate approach may result in improved accuracy and consistency if the characteristics are related.

2. Some of the auxiliary information is itself subject to significant sampling variance. If such auxiliary information is incorporated simply as an independent variable or variables in a linear regression, an inappropriate specification of the model typically results, with an associated cost in the performance of the small domain estimator.

Because of the second reason, the methods described in this paper are of potential relevance to some small domain problems concerning a single characteristic.

This paper will describe preliminary work on the development of a small area estimation strategy to estimate median rents by size of unit for rental units meeting a number of specific characteristics. The small domain estimation procedures are based on components of variance models. There are some precedents for an approach of this sort to small domain estimation (e.g., Fuller and Harter (1987) and several of the papers cited by them.) One such application (Fay 1986, 1987) has already been implemented for the estimation of median incomes for 4-person families by State, based on data from the Current Population Survey and estimates of per capita income from the Bureau of Economic Analysis.

This paper describes preliminary research to apply multivariate methods to the estimation of median rents. The specific requirements and need for these estimates is described in detail in section 2. This section also describes the sources of data for this problem. Section 3 describes the estimation strategy generally. Section 4 discusses the potential applications of this methodology to the problem. The paper does not make a detailed proposal to produce a set of estimates; rather, the emphasis in the paper is on the importance of and approaches to assessing alternative strategies.

2. Estimating Median Rents

The Department of Housing and Urban Development (HUD) is required by law to establish Fair Market Rents (FMR) for rental units. The FMRs are necessary for a government program, the Housing Assistance Payments Programs for Existing Housing. The intent of the FMRs is to reflect local market conditions to the extent possible.

For most of the 1970s, the FMRs were obtained from 1970 census data on rents, updated by changes in the rents and utilities component of the Consumer Price Index. The census data did not provide the specific detail necessary to determine which units were substandard, however, Since the administrative purposes imply or require that the FMRs exclude the downward bias in monthly rents contributed by substandard units, a revision of the methodology in 1979 employed instead data from the American Housing Survey (AHS, formerly called the Annual Housing Survey). This survey provided the necessary substantive detail for a determination of whether sampled rental units were substandard. The criteria reflected several factors related to state of repair, adequacy of plumbing and basic equipment, interruptions in service, and safety. The survey data, while meeting substantive purposes, provided far less geographic detail than the census.

Additional criteria further restrict the subset of rental units considered in determining the FMRs by the current methodology. The complete set of criteria are:

1. the units must not be public housing;

2. the units must be in buildings at least two years old;

3. the units must not be substandard, according to a definition employed by HUD on the basis of AHS characteristics; and

4. the units must be occupied by recent movers who began occupancy within the last two years.

Generally, the purpose of these exclusions is to determine an appropriate universe of rental units to define the FMRs for program purposes.

Size of unit is defined by the number of bedrooms. In the current methodology, however, only two-bedroom units are used in estimating subnational differentials. Local FMRs for units of other sizes are calculated by an adjustment to the local two-bedroom value.

The AHS comprises two component surveys. The AHS - National Sample (AHS-NS) is now conducted every two years, although at one time this part of the survey was conducted annually. The 1983 AHS-NS provides the national estimates used for discussion in this paper. The 1985 AHS-NS has become available for analysis within the last few months. (By implication, estimates from the 1987 AHS-NS will presumably not be ready for some time.) Each year, the AHS - Metropolitan Sample (AHS-MS) includes a subset from 60 Standard Metropolitan Statistical Areas (SMSAs), now Metropolitan Statistical Areas (MSAs), so that periodic estimates become available for each of these SMSAs/MSAs. (After analysis of the 1980 Census, the Office of Management and Budget replaced the SMSAs by MSAs. The latter areas replaced the SMSAs as basis of geographic selection starting with the 1984 AHS-MS.) For purposes of illustration, some results appear in this paper from the 1983 AHS-MS for the thirteen SMSAs included in that year.

Table 1 presents preliminary unit counts from the 1983 AHS-NS. The table shows the effect of the various exclusions. The counts are preliminary, in the sense that they are based on application of the available written documentation and some internal documentation used by HUD. The counts are provided only for purposes of discussion here, and further revision of the implementation of the definition may produce minor revisions.

The first line of table 1 shows the counts for 7453 sample units meeting all criteria. The second line reports that 2321 units are excluded because they are public housing or receive rent subsidies. An additional 325 units are excluded as too new, and 2647 of the remaining units are defined as substandard according to the criteria. The remaining units are excluded because they are not occupied by recent movers. Table 1 divides this remaining group into those who occupied their units within the last five years (60 months or less), 2884 cases, and those who have been tenants longer, 2454 cases.

Table 2 presents median gross rents for the various cells of table 1. Figure 1 displays the relationship between median rent and number of bedrooms for these groups. Median rents for public housing and for substandard units are considerably lower than those of units meeting the complete set of criteria. Units in new buildings have higher rents of than those meeting the criteria. Median rents for those occupying their units more than 2 years but no more than 5 years closely approximate those of recent movers, but rents for long-term tenants fall below those for those meeting all criteria, especially for the larger units.

Table 1 indicates that there are more two-bedroom units meeting the criteria than any other size. The numbers of one- and three-bedroom units are both substantial, however, and their combined number is greater than the number of two-bedroom units. As mentioned earlier, the current estimation methodology implemented by HUD employs the proportional relationship from the AHS-NS between the unit of a specific size and two-bedroom units. For example, the values in the first row of table 2 represent, in principle, the basic components of these ratios in 1983. Estimates for SMSAs/MSAs included in the AHS-MS are produced by proportionally adjusting the sample estimate for two-bedroom units for the SMSA/MSA by the national ratios for units of other sizes. In other words, the current methodology disregards any sample data for units of other sizes, except to the extent that these units contribute to the estimation of the overall national ratios. This exclusion occurs even though the objective is to provide estimates for each size. A more detailed discussion of this issue follows in section 4.1.

Two objectives may be set for this study. One is to determine whether the current methodology, which uses only the data for two-bedroom units at the SMSA/MSA level, can be significantly improved by incorporating the data for units of other sizes, particularly for one- and three-bedroom units. The second is to attempt to determine if there are any reasonable alternative strategies to produce subnational estimates below the regional level for areas not now included among the 60 SMSAs/MSAs covered by the design of the AHS-MS. As noted earlier, the geographic detail available from the decennial censuses is not incorporated in any form under the existing methodology.

3. Components of Variance Models for Small Domains

As noted in the first section, the methodology to be discussed here is related to the use of regression estimation for small domain estimation. Suppose that y_i represents a characteristic of interest for small domain, i, and the vector $\mathbf{y} = \{y_i\}$ denotes these quantities over a set of domains of interest. For example, **y** could denote a vector of median gross rents for two-bedroom units meeting the criteria for FMRs, over a set of domains, such as MSAs. For purposes of illustration, suppose that auxiliary variables X_{i1} and X_{i2} are available for each domain i, and X_{i0} is set identically to the value 1 in order to represent the intercept term. For example, the auxiliary information X_{i1} and X_{i2} may denote census values of related characteristics for the same areas. A linear regression model for y_i may be expressed as

$$y_i = b_0 X_{i0} + b_1 X_{i1} + b_2 X_{i2} + a_i$$

The last term, a_i, represents a model error for domain i, and depends on the specific choices for b_0, b_1, and b_2. A vector representation of the more general situation is

$$\mathbf{y} = \mathbf{Xb} + \mathbf{a}. \qquad (3.1)$$

When the a_i are small for an appropriate choice of **b**, the regression prediction

$$\hat{y}_i = b_0 X_{i0} + b_1 X_{i1} + b_2 X_{i2}$$

or more generally

$$\hat{\mathbf{y}} = \mathbf{Xb} \qquad (3.2)$$

represents a possible strategy to obtain small domain estimates of **y**.

The determination of **b** affects the overall performance of estimator (3.2). Although the form of (3.1) does not require treatment of any of its components as random, distributional assumptions are an effective strategy for motivating small domain estimators. Assuming a stochastic distribution on the a_i may suggest a form of estimator; for example, assuming that the a_i have mean 0 and are uncorrelated with approximately the same variance suggests the use of least squares. More generally, if the a_i are assumed to have a non-singular covariance **A**, then estimation of **b** through generalized least squares, i.e., $\hat{\mathbf{b}} = (\mathbf{X'A}^{-1}\mathbf{X})^{-1}\mathbf{X'A}^{-1}\mathbf{y}$, represents a likely choice for efficient estimation of **b**.

In small domain problems, the y_i are not directly observed, but sample estimates Y_i may be available for some or all i. In this example, Y_i may denote sample estimates of medians for the same areas from the AHS. Suppose that we represent the effect of sampling error on **Y** by the random vector $\mathbf{d} = \mathbf{Y} - \mathbf{y}$, so that $\mathbf{Y} = \mathbf{y} + \mathbf{d}$. In other words, the model for the sample estimates takes the form

$$\mathbf{Y} = \mathbf{Xb} + \mathbf{a} + \mathbf{d}. \qquad (3.3)$$

If we assume that **a** and **d** are independent, that the a_i have a multivariate normal distribution with covariance matrix **A**, that the d_i have a normal distribution with covariance matrix **D**, and that the a_i and d_i have expectations of zero, then, conditional on the fixed values of **X**, the best linear unbiased estimator (BLUE) of **b** is given by

$$\hat{\mathbf{b}} = (\mathbf{X'(A+D)}^{-1}\mathbf{X})^{-1}\mathbf{X'(A+D)}^{-1}\mathbf{Y}. \qquad (3.4)$$

The actual interest, however, is in **y**, the values of the small domains. Since $\mathbf{y} = \mathbf{Xb} + \mathbf{a}$, these values depend both on regression parameters, **b**, that may be regarded as fixed effects, that is, not described by a probability distribution, and random effects, **a**. (Alternative small domain estimators treating **b** as random are possible, but they will not be considered here.) Consequently, estimation of **y** falls under the theory for linear combinations of fixed and random effects. The best linear unbiased estimator, over the distribution of **d** and **a**, is given by Harville (1976)

$$\hat{\mathbf{y}} = \mathbf{X\hat{b}} + \mathbf{A(D+A)}^{-1}(\mathbf{Y}-\mathbf{X\hat{b}}). \qquad (3.5)$$

In other words, the regression estimate, $\mathbf{X\hat{b}}$, is modified by a term involving the residuals, $(\mathbf{Y}-\mathbf{X\hat{b}})$. When **D** is negligible relative to **A**, (3.5) gives essentially $\hat{\mathbf{y}} = \mathbf{Y}$. That is, when the sample estimates provide far more reliable estimates than the model in question, the optimal estimator almost exactly agrees with the sample estimates. Conversely, when **A** is negligible relative to **D**, the estimate approaches $\hat{\mathbf{y}} = \mathbf{X\hat{b}}$. To the extent that **D** is of the same order of magnitude or smaller than $\mathbf{A}$, however, (3.5) represents a weighted average of $\mathbf{Y}$ and $\mathbf{X}\hat{\mathbf{b}}$. Thus, (3.5) may be said to shrink the sample estimates $\mathbf{Y}$ towards the regression estimates $\mathbf{X}\hat{\mathbf{b}}$.

Motivation for (3.5) was stated in the terms of variance component models. Equivalently, it is possible to arrive at the same formula through a Bayesian formulation, by putting the appropriate diffuse prior on $\mathbf{b}$.

In small domain estimation problems, information about $\mathbf{D}$, the sampling error of $\mathbf{Y}$, is typically available or can be estimated with reasonable precision from the sample. The value of $\mathbf{A}$, however, is more likely to be unknown. When an estimate of $\mathbf{A}$ is derived from the sample and substituted into (3.5), the resulting estimator may be motivated through variance component models or as an empirical Bayes procedure (e.g., Morris 1983).

Specific forms of the James-Stein estimator (James and Stein 1961) arise in this manner. In the simplest case, both $\mathbf{D}$ and $\mathbf{A}$ are diagonal matrices, each a scalar multiple of the identity matrix. The James-Stein estimator gives a particular choice for $\mathbf{A}$. In this case, each coordinate is estimated by a weighted combination of the regression estimate for the coordinate and the corresponding sample estimate.

The preceding development is suitable for extensions to multivariate applications. In this instance, the sample vector $\mathbf{Y}$ may denote a series of sample estimates of K related characteristics $Y_{k(i)}$, k=1,...,K, in domain i, arranged in the order $\mathbf{Y} = (Y_{1(1)}, Y_{2(1)}, \ldots, Y_{K(1)}, Y_{1(2)}, \ldots)'$. To anticipate the discussion in the next section, the K different sample estimates may represent log-median gross rents for different types of rental units. Typically, and in the application to the AHS, (especially to the AHS-MS, but presumably to the AHS-NS as well), the sampling covariances between different domains may be taken to be zero, or effectively zero, by design. Within each domain, however, the sample estimates $Y_{1(i)}, Y_{2(i)}, \ldots, Y_{K(i)}$ may have nonzero sampling covariances, and, in some applications, these covariances may be quite important. Consequently, the typical form for the sampling covariance matrix $\mathbf{D}$ will be block diagonal:

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_1 & 0 & 0 & 0 & \ldots \\ 0 & \mathbf{D}_2 & 0 & 0 & \ldots \\ 0 & 0 & \mathbf{D}_3 & 0 & \ldots \\ 0 & 0 & 0 & \mathbf{D}_4 & \ldots \end{pmatrix}$$

The submatrices, $\mathbf{D}_1, \mathbf{D}_2, \ldots$, each K by K positive semidefinite symmetric matrices, are rarely identical, since, among other reasons, the sample sizes will almost always vary among the domains. In some applications, it may be appropriate to assume that the $\mathbf{D}_i$ are diagonal matrices, i.e., that the component characteristics for domain i have independent sampling errors. As a first approximation, sample estimates of median rents for rental units of different sizes could have almost zero covariances, since each sample unit contributes to the estimation of at most one of the components. In practice, however, some modest covariances between the components presumably arise from the clustering of the sample at the last stage of selection.

In some applications, design-based estimation of each $\mathbf{D}_i$ separately may prove entirely satisfactorily. When sample sizes are small within the domains, however, smoothing of design-based estimates or generalized variance formulas may become necessary. Some uncertainty in the estimation of the $\mathbf{D}_i$ is tolerable, but entirely unstable estimates of these matrices would be harmful.

The usual form of $\mathbf{A}$ is block diagonal as well. The general case may be represented:

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & 0 & 0 & 0 & \ldots \\ 0 & \mathbf{A}_2 & 0 & 0 & \ldots \\ 0 & 0 & \mathbf{A}_3 & 0 & \ldots \\ 0 & 0 & 0 & \mathbf{A}_4 & \ldots \end{pmatrix}$$

In practice, however, and in contrast to the $\mathbf{D}_i$, separate estimation of each $\mathbf{A}_i$ is impossible for satisfactory application in (3.4) and (3.5). One possible strategy is to assume that each $\mathbf{A}_i$ is equal to a common $\mathbf{A}^*$ and to estimate $\mathbf{A}^*$ through maximum likelihood, constrained maximum likelihood, or another of the familiar procedures in components of variance models. Another parametric form for the $\mathbf{A}_i$ could be hypothesized, and the corresponding parameters then estimated from the data. Selection among such alternatives may arise on the basis of past experience and data or from comparisons of alternative models fitted to the current sample data. Comments on alternative assumptions for median rents appear in the next section.

A form for $\mathbf{A}$ other than block diagonal is also possible, but generally even more complex. For example, if there were significant geographic correlations among domains that could not be captured entirely by the fixed effects, e.g., $\mathbf{Xb}$, then off-diagonal blocks could express these relationships.

4. Development of Components of Variance Models for Median Rents

4.1 Effect of the Number of Bedrooms

As described in section 2, HUD requires estimates of median rents for different size units but employs sample estimates for only the two-bedroom units to represent subnational variation among domains. The apparent implicit assumption of the method is that the relationship between median rents for two-bedroom units and units of each of the other sizes is captured by the corresponding national ratio. For example, in table 2, the 1983 AHS-NS indicates that, among units meeting all criteria, the median for two-bedroom units is approximately 19 percent larger than for one-bedroom units.

Since table 1 indicates that one- and two-bedroom units are the most numerous sizes available for the comparison, table 3 presents related evidence on the constancy of proportional relationships from two different sources. The thirteen SMSAs included in the 1983 AHS-MS are compared. The first column presents the ratio of median gross rents for two-bedroom relative to one-bedroom units for all contract-rent units from the 1980 census for these SMSAs. These ratios do not reflect the exclusions of units for the current criteria applied in determining the FMRs from AHS. The SMSAs have been ordered by increasing values of the ratios in the first column, although the order in a few cases rests on the third decimal point and may not be statistically significant, in spite of the large sample sizes in the census. The census medians are also interpolated from interval estimates, i.e., rental units were tabulated by intervals of gross rent, e.g., $400-499, and the median interpolated by assuming a distribution within the interval containing the sample median. In spite of these caveats, the census results indicate substantial variation in the ratios of medians among the different SMSAs.

The second column of table 3 presents the ratios of sample medians obtained from the 1983 AHS-MS. The medians used in computing the ratios are exact, that is, calculated from the sample without interpolation, for units meeting the FMR criteria. Sampling variance accounts for a more substantial proportion of the variation in this column than the first. Nonetheless, clear evidence of a persistence of pattern emerges from the comparison of the two columns. Figure 2 displays a plot of the same information. In areas such as the New York SMSA, a comparatively low ratio of median rents observed in the census reappears in the 1983 AHS-MS. Many other SMSAs with high 1980 census ratios continue to reflect these differences from the national average in 1983, even after application of the FMR criteria. The suggested slope appears possibly less than 1.0, perhaps 0.7 to 0.8. In other words, the 1983 values may display some regression toward the mean relative to the 1980 values.

The implications of table 3 and figure 2 are not necessarily welcome from the perspectives of policy. Differences among the medians for different sizes may reflect more than the relative marginal cost of additional space. For example, it is possible that in the New York SMSA the smaller units may be particularly concentrated among more "up-scale" or newer buildings, with many of the larger units remaining in the housing stock concentrated in less expensive areas. In other areas, for example those SMSAs experiencing rapid growth, there may not be such comparable hidden factors differentiating the smaller units, and, indeed, market conditions may place a particularly high premium on larger units.

The apparent spirit behind the FMR determination reflects an attempt to measure the cost of a rental unit of average quality. Possibly, use of national factors may better serve this purpose than an attempt to account more precisely for local variation. Hence, HUD will need to weigh the relative merits of more precise estimates of median rents reflecting local variation by size against the possibility that application of national ratios may provide a more uniform measure of average quality.

From the point of view of small domain estimation, however, the importance of table 3 is clear. The 1983 results indicate that relationships from the 1980 census are indeed predictive for current patterns, and deserve to be incorporated in the model. A specific form for the relationship will be chosen for discussion in the next sections, but full analyses of the data available from the other years of the AHS-MS is clearly warranted in order to select the most appropriate form of the model.

4.2 Models with Fixed Local Effects

The sample size of the AHS-MS generally supports separate analyses for each SMSA/MSA. Hence, although estimators that smooth across small domains are possible, as a matter of practicality a fixed effects model appears the most simple and appropriate for the AHS-MS. Because of the dominant role of the median for two-bedroom units, it is sensible to express the fixed local effect as the expected median for two-bedroom units meeting the FMR criteria.

The recurrent use of ratios throughout the preceding discussion suggests that an appropriate expression of the model is in terms of log-medians. With this reexpression, proportional relationships become additive factors.

In this form, the model is that the true log-median, $\ln(y_{2(i)})$, for two-bedroom units meeting the standard for domain i is a fixed effect, $b_{(i)}$. A possible model for the true log-median, $\ln(y_{k(i)})$ for a unit in size-class k, k=0,1,3,4, is given by

$$\ln(y_{k(i)}) \;=\; b_{(i)} + b_k + b^* x_{k(i)} + a_{k(i)} \tag{4.1}$$

where b_k denotes a fixed effect for size-class k, $x_{k(i)}$ represents the difference of log-medians from the census for size-class k relative to two-bedroom units, and $a_{k(i)}$ represents a random effect. This model is expressed in such a way that $b_2 = 0$, with the remaining b_k's measuring differentials with respect to two-bedroom units. We also have $x_{2(i)} = 0$ and $a_{2(i)} = 0$. More precisely,

$$x_{k(i)} \;=\; \ln(\text{cmed}_{k(i)}/\text{cmed}_{2(i)}),$$

where $\text{cmed}_{k(i)}$ represents the census median for class k in area i. Figure 2 suggests a value of b^* less than 1.0, perhaps 0.7 to 0.8 or so.

Even though the median for two-bedroom units meeting the criteria is represented by a corresponding fixed effect in the preceding model, $b_{(i)}$, application of (3.4) and (3.5) will generally produce an estimate different from the direct sample estimate. If (4.1) had instead been fully expressed in terms of fixed effects, e.g., if each random effect $a_{k(i)}$ were replaced by a fixed effect, $b_{k(i)}$, then (3.4) and (3.5) would reproduce the sample estimate. When the model is in the form (4.1), (3.4) and (3.5) employ sample information about the estimation of medians for units of other sizes in the same domain to the estimation of $\ln(Y_{2(i)})$.

Different strategies are possible for the estimation of (4.1). The model as presented makes the most sense for a single year of the AHS-MS. Derivation of an appropriate form for **A** and its estimation would tend to favor examination of data for several years, however. Additionally, it may be appropriate to further generalize (4.1) by permitting the coefficient b^* to depend on k, giving a series of coefficients b^*_k, again favoring concurrent estimation for a series of years. In generalizing across years, however, it may be necessary to allow some of the fixed effects to depend on year.

A simple model worth trying as a first step would be application of (4.1) to one- and three-bedroom units meeting the criteria. A further possible expansion of the model would be to incorporate the estimates for one-, two-, and three-bedroom units for relatively recent movers (i.e., within the preceding three to five years) as three additional components of the model for each domain, thus doubling the length of **Y**. In principle, the model could be further elaborated; in practice, concerns for estimating sampling variance for cells with small samples combined with possible failures of the model in such instances argues for placing limitations on the scope of the model. Depending on the relative sizes of **D** and **A**, the gains for the SMSA estimates for two-bedroom units could be as large as an effective tripling of the sample size or instead possibly relatively modest. Even if the gains for two-bedroom units were small, the model could produce more effective differentials by size of unit if this were a reasonable objective for purposes of HUD.

4.3 Models with Random Local Effects

Except for the SMSAs/MSAs included in the AHS-MS, HUD now employs only regional relationships in estimation of FMRs. In effect, this approach completely smooths over any information about local variation available from the 1980 census or the AHS-NS. Avoidance of direct use of census data presumably reflects a concern by HUD that the global measures provided by the census do not incorporate the distinctions represented by the FMR criteria applied to the AHS. Avoidance of the use of local data from the AHS-NS presumably stems from the small sample sizes, particularly for two-bedroom units treated separately.

Models with random local effects are more appropriate than models with fixed local effects in instances when there is too little sample data to employ any domain-level sample estimate directly. One simple model with local random effects is of the form:

$$\ln(y_{k(i)}) \;=\; b'_k + b^{*\prime}_k \text{cmed}_{k(i)} + a'_{k(i)} \tag{4.2}$$

where b'_k now represents a fixed effect for size class k, $\text{cmed}_{k(i)}$ the census log-median for this size class in the local area, and $a'_{k(i)}$ a random effect. In this model, the values of coefficients, including $a'_{k(i)}$, are not forced to 0 at k=2, unlike (4.1). This model could be used to form, through (3.4) and (3.5), composite estimates for areas not covered by the AHS-MS. Actual application may depend on the fit of the census data to the AHS sample.

Another possible application of a random effects model would be to attempt to use data from the AHS-NS to update SMSAs/MSAs for which separate components of change were not available from the Consumer Price Index. In this instance, the random effects would indicate differentials between local change and regional change.

[1] This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the author and do not necessarily reflect those of the Census Bureau. The author wishes to thank Michael J. Roebuck for helpful comments on an earlier draft.

REFERENCES

Ericksen, E.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.

Fay, R.E. (1986), "Multivariate Components of Variance Models as Empirical Bayes Procedures for Small Domain Estimation," Proceedings of the Survey Research Methods Section, American Statistical Association, Washington, DC.

______ (1987), "Application of Multivariate Regression to Small Domain Estimation," in Small Area Statistics, R. Platek, J.N.K. Rao, C.E. Sarndal, and M.P. Singh, eds., John Wiley & Sons, New York, 91-102.

______ and Herriot, R. (1979), "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data," Journal of the American Statistical Association, 74, 269-277.

Fuller, W.A., and Harter, R.A. (1987), "The Multivariate Components of Variance Model for Small Area Estimation," in Small Area Statistics, R. Platek, J.N.K. Rao, C.E. Sarndal, and M.P. Singh, eds., John Wiley & Sons, New York, 103-123.

Harville, D.A. (1976), "Extension of the Gauss-Markov Theorem to Include the Estimation of Random Effects," Annals of Statistics, 4, 384-395.

______ (1977), "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems," Journal of the American Statistical Association, 72, 320-338.

James, W., and Stein, C. (1961), "Estimation with Quadratic Loss," in Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, CA, Vol. 1, 361-379.

Morris, C. (1983), "Parametric Empirical Bayes Inference: Theory and Applications," Journal of the American Statistical Association, 78, 47-55.

Platek, R., Rao, J.N.K., Särndal, C.E., and Singh, M.P., eds. (1987), Small Area Statistics, John Wiley & Sons, New York.

Table 1 Preliminary Counts of Rental Units in the 1983 American Housing Survey - National Sample

	Bedrooms					
	0	1	2	3	4+	Total
Meets criteria	199	2180	3196	1526	352	7453
Public housing or unknown status	88	864	852	434	83	2321
New unit	4	104	167	47	3	325
Substandard	219	699	1084	529	116	2647
Tenant 2-5 years	80	878	1273	566	87	2884
Tenant more than 5 years	73	813	1069	433	66	2454

Note: Based on preliminary application of HUD definitions. See text for explanations

Table 2 Preliminary Estimates of Median Gross Rent from the 1983 American Housing Survey - National Sample

			Bedrooms		
	0	1	2	3	4+
Meets criteria	$275	305	364	450	520
Public housing or unknown status	104	138	217	216	235
New unit	335	370	447	537	818
Substandard	160	231	280	308	313
Tenant 2-5 years	255	310	353	412	525
Tenant more than 5 years	251	275	315	330	372

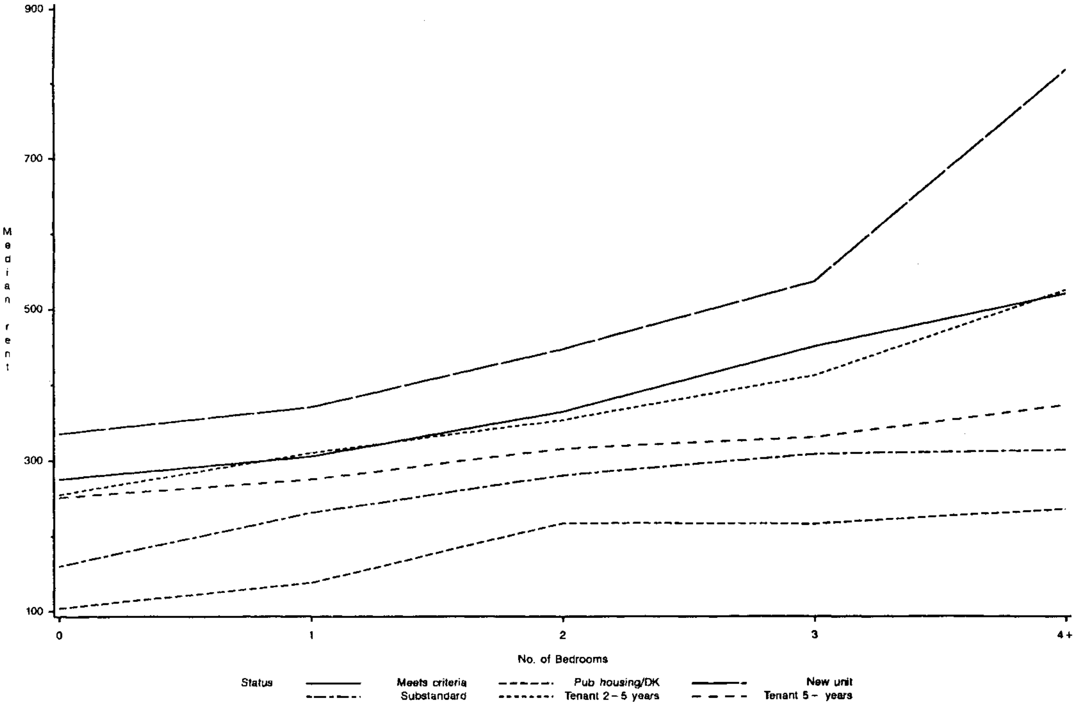
Note: Based on preliminary application of HUD definitions. See text for explanations.

Table 3 Comparison of 1980 Census and 1983 AHS-MS Ratios of Gross Rent, for Two- vs. One-Bedroom Units

SMSA	1980 Census Ratio	1983 AHS-MS Ratio
New York	1.10	1.11
Honolulu	1.15	1.16
Chicago	1.16	1.19
Houston	1.18	1.26
Baltimore	1.23	1.26
Hartford	1.24	1.19
Miami	1.24	1.31
Louisville	1.29	1.33
Portland	1.31	1.24
St. Louis	1.32	1.35
Sacramento	1.33	1.28
Seattle	1.34	1.27
Denver-Boulder	1.36	1.33

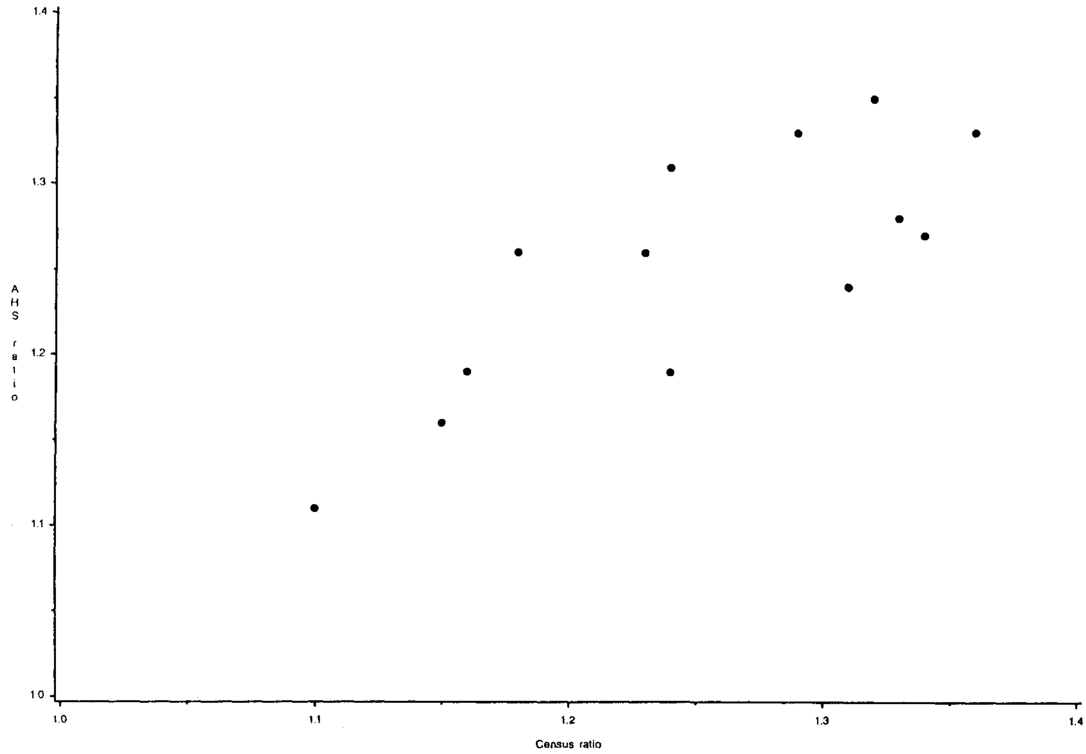
Note: SMSAs are ranked by increasing values of ratios, although not all differences in the ranking are necessarily statistically significant. The 1980 medians are based on all units with contract rent; the 1983 AHS medians are for units meeting the FMR definition of HUD.

Figure 1 Preliminary Estimates of Median Gross Rent from the
1983 American Housing Survey – National Sample



Note: Based on preliminary application of HUD definitions. See text for explanations.

Figure 2 Comparison of Ratios of 2– to 1–Bedroom Units
1980 Census vs. 1983 AHS–MS


AHS ratio
Census ratio