INVESTIGATION OF STRATIFICATION RULES FOR ESTIMATING TOTAL MARKET VALUE OF TAXABLE PROPERTY IN A DISTRICT

John J. Chai, Syracuse University
Sherman Chottiner, Syracuse University
Michel Legault, Syracuse University

Many local governments allocate the budget to subtaxing districts based on their share of total market value of taxable property. In Westchester County, New York, the authors are involved in estimating total market value using stratified random sampling methods for each of its districts. The data from these surveys form the basis of this empirical study.

The estimation procedure used by the authors, briefly stated, is as follows: First, for the ith property class sampled (e.g., residential, commercial), the ratio of assessed to market values, R_i, is estimated from a sample of parcels, where the assessed value of a parcel in the sample is obtained from the district assessment roll and the market value of the parcel is appraised by a qualified appraiser. Next, the market value of each property class is estimated using $\hat{R}_i$ and the total assessed value of the property class A_i taken from the assessment roll, i.e., $\hat{M}_i = A_i/\hat{R}_i$. Then, the survey ratio (R_s) of the district is estimated by dividing the sum of the total assessed values for the sampled property classes in the district ($A_s = \Sigma A_i$) by the sum of the estimated market values for the sampled property classes in the district ($\hat{M}_s = \Sigma\hat{M}_i$), i.e., $\hat{R}_s = A_s/\hat{M}_s$. Finally, the total taxing district market value is estimated by dividing the district total assessed value (for sampled and unsampled classes) by $\hat{R}_s$, or $\hat{M} = A/\hat{R}_s$. More detailed sample design and estimation procedures are given in the appendix.

Conscientious property appraisal is a costly process; hence the sample size used for this purpose is usually small, and the estimates are subject to relatively large sampling errors. Stratification by class of property (residential, commercial, industrial, vacant land) is effective in reducing sampling error, since wide variations in tax assessment ratios exist among these classes. Because tax assessment ratios often vary with the level of assessment, a second stratification by assessment within a property class can also serve to reduce sampling error. The question which then arises is how to make this stratification, and this is the subject of this paper.

There is considerable literature on stratification methods [e.g., 1, 5, 10, 12], but most of the results have been based on restricted assumptions on the distribution of stratification variable and estimation variable. Furthermore, there are no exact rules for defining optimum stratum boundaries. In recent years, there has been some work reported on the determination of approximately optimum stratification [6, 7, 8, 9, 11]. This paper has much in common with the work by Hess, et al. [7] in that investigation of some well known theoretical stratification rules using "real world" data has been made, though the scope of this paper is much narrower than that of Hess, et al.

The purpose of this paper is to investigate the effect of two stratum-boundary rules (equal assessment value (EAV) and equal number of parcels (ENP)) and 3 different number of strata (two, three, and six) on the coefficient of variation, the bias, and the relative mean square error of the estimated ratio of assessed to market values for selected property classes in several taxing districts. The allocation of the sample is made according to the Neyman optimum rule. For study of the effect of the two stratum boundary rules, sample size, number of strata, sample parcels, base roll used for assessment of properties and the estimator used (ratio-estimator) are fixed. For study of the effect of different number of strata, all the factors listed above are fixed, except that "number of strata" is replaced by a given boundary rule.

The results of our investigation of the two main effects are summarized in Tables 1 and 2. The major findings for each effect are now presented.

For comparison of equal assessed value (EAV) and equal number of parcels (ENP) rules (Table 1):

(1) The bias of the estimator for both rules are negligible for all taxing districts considered except for Peekskill. Similar results were reported in our previous report [2]. The reason for this is that the regression of assessed values on market values in most cases considered is approximately a straight line through the origin.

(2) The EAV rule is more efficient than the ENP rule in five out of seven districts considered as measured by the ratio of relative mean square error for ENP divided by relative mean square error for EAV ($R.MSE_{EAV}/R.MSE_{ENP}$). In fact, the ratios range from 1.4542 to 4.5372 for these five districts.

As the results in Table 1 indicate, there is no clear cut conclusion one can draw as to which rule is uniformly better. Further research with larger samples and other districts is necessary before such conclusion can be made.

For the second investigation on the effect of the number of strata (Table 2), our results conform with the general pattern of decrease in variance as the number of strata increases, as reported in many previous works [e.g., 7], though there is no definite pattern observed in the results as far as the rate of decrease is concerned. The range of $R.MSE_{L_i}/R.MSE_{L_i-1}$ (L_i is number of strata for the ith property class) for $L_i = 3$ is .7026 to .9599 and the range of $R.MSE_{L_i}/R.MSE_{L_i-3}$ for $L_i = 6$ is .4948 to .7476. This rate of decrease is larger than that for the cases reported by Hess,

TABLE 1

COEFFICIENT OF VARIATION, RELATIVE BIAS, AND RELATIVE ROOT MEAN SQUARE ERROR OF ESTIMATED RATIO OF ASSESSED TO MARKET VALUES FOR "EQUAL ASSESSED VALUE" AND "EQUAL NUMBER OF PARCEL" BOUNDARY RULES

Taxing District	No. of Strata	Sample Size	C.V.*		Bias/S.D.		$R\sqrt{MSE}$		$\frac{R.MSE_{ENP}}{R.MSE_{EAV}}$
			EAV	ENP	EAV	ENP	EAV	ENP	
White Plains	3	14	.0382	.0344	.0301	.0124	.0382	.0344	.8109
City of Rye	4	18	.0174	.0251	.0302	.0498	.0174	.0251	2.0809
Poundridge	3	19	.0204	.0246	-.0029	.0055	.0204	.0246	1.4542
Peekskill	3	8	.0602	.1292	.2078	.1674	.0615	.1310	4.5372
North Castle	3	31	.0324	.0404	.0333	.0330	.0324	.0404	1.5548
Lewisboro	5	17	.0349	.0247	.0249	.0148	.0349	.0247	.5009

*Coefficient of Variation

TABLE 2

COEFFICIENT OF VARIATION, RELATIVE BIAS, AND RELATIVE ROOT MEAN SQUARE ERROR OF ESTIMATED RATIO OF ASSESSED TO MARKET VALUES FOR DIFFERENT NUMBER OF STRATA

Taxing District	Sample Size	Coefficient of Variation			Bias/S.D.			$R\sqrt{MSE}$			$R.MSE_{L_i}/R.MSE_{L_i-1}$	
		2	3	6	2	3	6	2	3	6	$\frac{R.MSE_3}{R.MSE_1}$	$\frac{R.MSE_6}{R.MSE_2}$
White Plains	22	.0476	.0398	.0345	-.0052	.0744	.0372	.0476	.7026	.7476	.0399	.0345
Poundridge	19	.0586	.0554	.0410	.0111	.0310	.0315	.0586	.0544	.0410	.8938	.5477
New Rochelle	22	.0296	.0290	.0204	-.0002	.0045	.0127	.0296	.0290	.0204	.9599	.4948
Lewisboro	17	.0435	.0377	.0290	-.0065	.0075	.0259	.0435	.0377	.0290	.7511	.5917

Remark: The data base and the boundary rules used for this table are different from those for Table 1.

Appendix

SAMPLE DESIGN, ESTIMATOR, MEAN SQUARE ERROR OF ESTIMATOR, AND ESTIMATOR OF MEAN SQUARE ERROR

Sample design:

A population of N taxable properties in a taxing district is stratified first by property type (i.e., residential, commercial, vacant land, and industrial) and second by a specified stratification rule within a property class. Let N_i be the number of taxable properties for the i-th property class, N_{hi} be the number of taxable properties for the h-th stratum of the i-th property class, and L_i be the number of strata for the ith property class. Hence

$N = \sum_{i=1}^{K} N_i = \sum_{i=1}^{K} \sum_{h=1}^{L_i} N_{hi}$, where K is the number of sampled property classes. We define x_{hij} and y_{hij} respectively by the assessed and market values of the j-th property in the h-th stratum for the i-th property class. Furthermore, we define n, n_i, and n_{hi} similarly to N, N_i, and N_{hi} for the sample. Hence, $n = \sum_{i=1}^{K} n_i = \sum_{i=1}^{K} \sum_{h=1}^{L_i} n_{hi}$.

Determination of n_i is made in general considering the cost of appraising the properties and the precision of sample outcome.

The sampling procedure applied within strata is simple random sampling without replacement.

Estimation of the total market value:

We let R_s be the survey ratio of assessed to market values for a district and R_i be the ratio of assessed to market values for the i-th property class. We define

$$x_i = \sum_{h=1}^{L_i} N_{hi}\bar{x}_{hi}\ ,\ \text{where } \bar{x}_{hi} = \sum_{j=1}^{n_{hi}} x_{hij}/n_{hi}$$

$$y_i = \sum_{h=1}^{L_i} N_{hi}\bar{y}_{hi}\ ,\ \text{where } \bar{y}_{hi} = \sum_{j=1}^{n_{hi}} y_{hij}/n_{hi}$$

$$\hat{R}_i = x_i/y_i$$

Now, let A_i be the assessed value in the assessment roll for the i-th property class and M_i be the market value for the i-th property class. We define the estimator of M_i as $\hat{M}_i = A_i/\hat{R}_i$. Furthermore let A_s and $\hat{M}_s$ respectively be defined as

$$A_s = \sum_{i=1}^{K} A_i$$

$$\hat{M}_s = \sum_{i=1}^{K} \hat{M}_i$$

Then the survey ratio $\hat{R}_s$ is estimated by dividing A_s by $\hat{M}_s$, i.e.

$$\hat{R}_s = A_s/\hat{M}_s$$

Finally the estimator of the total market value of a taxing district is given by

$\hat{M} = A/\hat{R}_s$ where A is the total assessed value for a taxing district obtained from the assessment roll.

Variance, bias, and mean square error:

The variance, bias, and mean square error of $\hat{R}_i$ and $\hat{R}_s$ shown below are based on the assumption that the terms beyond the product moments about the means of second power are negligible. This assumption in most cases is justified for the districts we studied [2].

Variance of $\hat{R}_i$:

$$\mathrm{Var}(\hat{R}_i) \doteq R_i^2 \sum_h N_{hi}^2 \left(\frac{1-f_{hi}}{n_{hi}}\right)\left(\frac{\mu_{20hi}}{\mu_{10i}^2} + \frac{\mu_{02hi}}{\mu_{01i}^2} - \frac{2\mu_{11hi}}{\mu_{10i}\mu_{11i}}\right)$$

where: $\mu_{10i} = E(x_i)$, $\mu_{01i} = E(y_i)$

$\mu_{20i} = E\{x_{hij} - E(x_{hij})\}^2$

$\mu_{11i} = E\{x_{hij} - E(x_{hij})\}\{y_{hij} - E(y_{hij})\}$

$\mu_{02i} = E\{y_{hij} - E(y_{hij})\}^2$

$\vdots$

In general,

$$\mu_{k\ell i} = E\{x_{hij} - E(x_{hij})\}^k\{y_{hij} - E(y_{hij})\}^\ell$$

and

$$f_{hi} = n_{hi}/N_{hi}.$$

Bias of $\hat{R}_i$:

$$B(\hat{R}_i) \doteq R_i \sum_h N_{hi}^2 \left(\frac{1-f_{hi}}{n_{hi}}\right)\left(\frac{\mu_{02hi}}{\mu_{01i}^2} - \frac{\mu_{11hi}}{\mu_{10i}\mu_{11i}}\right)$$

Mean Square Error of $\hat{R}_i$:

$$\mathrm{MSE}(\hat{R}_i) \doteq \mathrm{Var}(\hat{R}_i) + \{B(\hat{R}_i)\}^2$$

Relative Root Mean Square Error of $\hat{R}_i$:

$$R\sqrt{\mathrm{MSE}(\hat{R}_i)} = \sqrt{\mathrm{MSE}(\hat{R}_i)}/R_i$$

Variance of $\hat{R}_s$:

$$\mathrm{Var}(\hat{R}_s) \doteq (R_s/M_s)^2 \sum_i^K A_i^2 \cdot \mathrm{Var}(\hat{R}_i^{-1})$$

$$= (R_s/M_s)^2 \sum_i^K (A_i^2/R_i^2)\{V^2(R_i)\}$$

where: $V^2(\hat{R}_i) = \mathrm{Var}(\hat{R}_i)/R_i^2$, the rel-variance of $\hat{R}_i$

Bias of $\hat{R}_s$:

$$B(\hat{R}_s) \doteq (R_s/M_s)\left[\sum_i^K (A_i^2/R_i^2) \cdot V^2(\hat{R}_i)/M_s - \sum_i^K M_i\{V^2(x_i) - \rho_{x_i y_i} V(x_i) \cdot V(y_i)\}\right]$$

where: $V^2(x_i) = \mathrm{Var}(x_i)/\{E(x_i)\}^2$

$V(x_i) = \sqrt{V^2(x_i)}$

$V^2(y_i)$ and $V(y_i)$ respectively are similarly defined.

$\rho_{x_i y_i}$ is the coefficient of correlation between x_i and y_i.

Mean Square Error of $\hat{R}_s$:

$$\mathrm{MSE}(\hat{R}_s) = \mathrm{Var}(\hat{R}_s) + \{B(\hat{R}_s)\}^2$$

Estimation of variance, bias, and mean square error from a sample:

The estimators used for estimating the various parameters in the variance, bias, and mean square error formulas are summarized in the following table:

Parameter to be estimated	Estimator
μ_{10}	$x_i = \sum_h N_{hi}\bar{x}_{hi}$
μ_{01}	$y_i = \sum_h N_{hi}\bar{y}_{hi}$
μ_{20h}	$s^2_{xhi} = \sum_j (x_{hij}-\bar{x}_{hi})^2/(n_{hi}-1)$
μ_{02h}	$s^2_{yhi} = \sum_j (y_{hij}-\bar{y}_{hi})^2/(n_{hi}-1)$
μ_{11h}	$s_{xyhi} = \dfrac{\sum_j (x_{hij}-\bar{x}_{hi})(y_{hij}-\bar{y}_{hi})}{(n_{hi}-1)}$
μ_{12h}	$\hat{\mu}_{12hi} = \dfrac{\sum_j (x_{hij}-\bar{x}_{hi})(y_{hij}-\bar{y}_{hi})^2}{(n_{hi}-1)}$
$\vdots$	$\vdots$
In general:	
$\mu_{k\ell hi}$	$\hat{\mu}_{k\ell hi} = \dfrac{\sum(x_{hij}-\bar{x}_{hi})^k(y_{hij}-\bar{y}_{hi})^\ell}{(n_{hi}-1)}$

Neyman Allocation of n_i:

$$n_{hi} = (N_{hi}S_{hiz}/\sum_h N_{hi}S_{hiz}) \cdot n_i$$

where: $S_{hiz} = \mu_{20i} + R_i^2\mu_{02i} - 2R_i\mu_{11i}$

References

[1] Anderson, D. W., Kish, L., and Cornell, R., "Quantifying Gains From Stratification for Optimum and Approximately Optimum Strata Using A Bivariate Normal Model", a paper presented at the 1975 American Statistical Association Meeting in Atlanta, Georgia.

[2] Chai, John J. and Sherman Chottiner, "Estimating Mean Square Error of Estimator of the Property Class Ratio of Assessed to Market Values From a Small Sample -- An Experimental Study", Proceedings of the American Institute for Decision Sciences, Western Regional Meeting, March 1975.

[3] Cochran, W. G., Sampling Techniques, 2nd ed. New York: John Wiley and Sons, Inc., 1963.

[4] Cochran, W. G. "Comparison of Methods for Determining Stratum Boundaries," Bulletin of the International Statistical Institute, 38 (2), Tokyo (1961), pp. 345-58.

[5] Dalenius, T., "The Problem of Optimum Stratification," Skandinavisk Aktuarietidskrift, 33 (1950), pp. 203-13.

[6] Dalenius, T. and Hodges, J. L., Jr., "Minimum Variance Stratification," Journal of the American Statistical Association, 54 (1959), pp. 88-101.

[7] Hess, I., Sethi, V. K., and Balakrishnan, T., "Stratification: A Practical Investigation," Journal of the American Statistical Association, 61 (1966).

[8] Seafling, R. J., "Approximately Optimal Stratification," Journal of the American Statistical Association, 63 (1968), pp. 1298-1309.

[9] Sethi, V. K., "A Note on Optimum Stratification of Populations for Estimating the Population Means," Australian Journal of Statistics, 5 (1963), pp. 20-33.

[10] Singh, R. and Sukhatme, B. V., "Optimum Stratification," Annals of the Institute of Statistical Mathematics, 21 (1969), pp. 515-528.

[11] Singh, R., "Approximately Optimum Stratification on the Auxiliary Variable," Journal of the American Statistical Association, 66 (1971), pp. 829-833.

[12] Singh, R. and Sukhatme, B. V., A Note on Optimum Stratification," Journal of the Indian Society of Agricultural Statistics, XXIV (1972), pp. 91-98.