THE EFFECT OF A LARGE CHANGE ON TWO RATIO REGRESSION ESTIMATORS

George Kattakkuzhy, Health Care Financing Administration

1. INTRODUCTION: In sample survey we faced the dilemma of choosing one of several estimators available. Here we are comparing a combined estimator and a separate estimator to see which one is more stable under a large change in one finding. In Medicaid Quality Control (MQC) program, the State takes a sample of size n to review. Out of these n cases the Federal takes a subsample of size n' and re-reviews them. The results are combined using a regression. An error payment percentage is calculated from the regression results as the ratio of the error amount to the total payment amount. The MQC program has cases in three strata - MAO, AFDC and SSI. For the combined estimator the error payment percentage is derived as the ratio of a combined error amount of three strata to the combined total payment of three strata. For the separate estimator, an error payment percentage is calculated for each stratum and the rates are joined together using the weight of each stratum.

2. NOTATION: The combined estimator $r_c = \sum w_i \bar{y}" / \sum w_i \bar{m}"$ where w_i is the stratum case weight (1), $\bar{y}"_i$ is the stratum average error amount (regression), $\bar{m}"_i$ is the stratum average total payment amount (regression);
The separate estimator $r_s = \sum w_i r_i$ where w_i is the stratum payment weight (2);
$r_i = \bar{y}"_i / \bar{m}"_i$ is the stratum error rate and,
$\bar{y}"_i = \bar{y}_i + b(\bar{x}_i - \bar{x}'_i)$,
$\bar{m}"_i = \bar{m} + c(\bar{u}_i - \bar{u}"_i)$ (3);
where $\bar{y}_i$, $\bar{m}_i$ are subsample average amounts according to Federal re-reviews; b, c are regression coefficients; $\bar{x}_i$, $\bar{u}_i$ are State averages for full sample and $\bar{x}_i'$, $\bar{u}_i'$ are State averages for the subsample part. X and Y are error amounts while m and u are total payment amounts.

3. SIMULATION: This is an empirical study of the two estimators given above. In actual practice, this occurs when either the Regional Office finds an entry error or the State finds an error. For the study purposes we consider a state with full sample sizes of 428 in MAO, 984 in AFDC, and 1217 in SSI and subsample sizes of 137 in MAO, 193 in AFDC, and 134 in SSI. We assume that there is no change in the total payment amount. The change is in the error amount designation.

One case could be in the following situations. (1) The case is in subsample and only y_{ij} changed. (2) The case is in subsample and only x_{ij} changed. (3) The case is in subsample and both x_{ij} and y_{ij} changed. (4) The case is not in subsample but is in full sample and x_{ij} changed. The only changing error amount is assumed $669 for one case in MAO stratum. (669 to 0). The effect is same if the change occurred in any other stratum. (We did study changes under a larger and smaller amounts but results are similar.)

4. OBSERVATIONS: Table 1 gives the full sample mean and subsample means under the changes.

Table 1

MAO stratum means...	$\bar{X}$	$\bar{X}'$	$\bar{Y}'$
No change	13.379	12.978	24.168
Only changed	13.379	12.978	19.285
Only X, S' changed	11.816	8.095	24.168
All changed	11.816	8.095	19.285
Only X changed	11.816	12.978	24.168

The regression coefficient b has to be calculated for the changing conditions. The results are given below.

Table 2

Regression Coefficient

	MAO separate	AFDS separate	Both combined
No change	.97863	1.0095	.97867
Y only	.50914	1.0095	.51779
X,X' only	.96429	1.0095	.96574
X, X & Y	.97509	1.0095	.97619
X only	.97863	1.0095	.97867

There was no change assumed for the total payment. The $\bar{m}"$ for MAO is 307.33, for AFDC is 101.91 and the combined is 188.54 for the two strata together. Table 3 gives the MAO error rate along with the 2 stratum Combined and Separate estimates.

Table 3

Error Rates

	(separate) MAO only	r_c%	r_s%
No change	7.9915_5	5.9903	6.2319
X only	6.3414_{4572}	4.8990	5.0794
X and X' only	9.0314_{37445}	6.7311	7.0014
X, X' and Y	7.4556_{0469}	5.6170	5.8366
X only	7.4938_{4938}	5.6484	5.8697

It can be seen from the above table that r_c is more closer together than r_s. This is the true even if we look at the percentage of change in number r_s. However, the variance of these estimates show much variation. Variance of combined estimate is almost double that of separate estimate. The variances are given in Table 4.

Table 4

Variance and Standard Deviation

	Combined	Separate	Combined %	Separate %
No change	.0005137	.0002616	2.26%	1.62%
Y only	.0005843	.0002722	2.42%	1.65%
X and X only	.0007445	.0003490	2.73%	1.87%
X, S' and 4	.0006036	.0002647	2.46%	1.63%

This shows that separate estimate has interval estimate shorter than the combined estimate. But if we look at the point estimate only then the combined estimate looks better under the condition we are checking.