Survey Methodology, December 1987
Vol. 13, No. 2, pp. 209–219
Statistics Canada



# Modified Raking Ratio Estimation

**H. LOCK OH and FRITZ SCHEUREN[1]**

ABSTRACT

A hybrid technique is described that employs both conventional and raking ratio estimation to handle the case when the population frequencies $N_{ij}$ in a two-dimensional table are known, but some of the observed frequencies $n_{ij}$ are small (or zero). Results are provided on the approach taken as it has evolved in the Corporate Statistics of Income Program over the last several years. Changes are still being considered and these will be discussed as well.

KEY WORDS: Raking ratio estimation; Conventional ratio estimation; Conditional bias and variance.

## 1. INTRODUCTION

Raking ratio estimation, or simply "raking," is a widely used technique in sample surveys. Applications differ depending on the nature of the sample design, the extent of the auxiliary information available and the presence of various nonsampling errors (such as might arise because of nonresponse or undercoverage).

Raking was first proposed by Deming and Stephan (1940) as a way of assuring consistency between complete count and sample data from the 1940 U.S. Census of Population. The originators themselves elaborated their ideas early on (Deming 1943; Stephan 1942). Since then, perhaps because of the basic intuitive appeal of the iterative algorithm employed, there have been several wholly independent rediscoveries of the technique (Fienberg 1970).

Advances and modifications have also been numerous. For example, important theoretical work on convergence of the algorithm was done by Ireland and Kullback (1968). As might be expected, practitioners at Statistics Canada, and also at the U.S. Bureau of the Census, have deeply studied the application of raking in census and survey taking, especially in situations where the raking is not allowed to proceed to complete convergence (e.g., Brackstone and Rao 1979; Fan *et al.* 1981). A reasonably complete bibliography of the statistical research on raking prior to 1978 can be found in Oh and Scheuren (1978b).

In many treatments of raking, it is assumed that two (or more) sets of marginal population totals, say $N_{i.}$ and $N_{.j}$, are known, but that the interior of the table $N_{ij}$ can only be estimated from the sample. When the $N_{ij}$ are also known, the usual ratio estimator with weights $N_{ij}/n_{ij}$ would be the natural choice, unless the corresponding sample sizes $n_{ij}$ are "too small."

The present paper describes a hybrid technique that employs both conventional and raking ratio estimations to handle the case when the population cell frequencies $N_{ij}$ are known, but some of the observed frequencies $n_{ij}$ are small (or zero). In Section 2, we describe our approach. Some empirical results from the application of the method to our Corporate Statistics of Income Program are covered in Section 3. In Section 4, we conclude with a brief summary and some plans for the future.

[1] H. Lock Oh and Fritz Scheuren, Statistics of Income Division, Internal Revenue Service, 1111 Constitution Avenue N.W., Washington, D.C. 20224, U.S.A.

## 2. RAKING RATIO ESTIMATION

### 2.1 General Considerations

Raking ratio estimation usually assumes that two (or more) marginal population totals, say, $N_{i.}$ and $N_{.j}$ are known, but that the interior of the table $N_{ij}$ can only be estimated from the sample by, say, $\widetilde{N}_{ij}$, where graphically (Deming 1943) we have

| | 1 | 2 | $\cdots$ | | S | |
|---|---|---|---|---|---|---|
| 1 | $N_{11}$ | $N_{12}$ | $\cdots$ | | $N_{1S}$ | $N_{1.}$ |
| 2 | $N_{21}$ | $N_{22}$ | $\cdots$ | | $N_{2S}$ | $N_{2.}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $N_{ij}$ | $\vdots$ | $\vdots$ $N_{i.}$ |
| $R$ | $N_{R1}$ | $N_{R2}$ | $\cdots$ | | $N_{RS}$ | $N_{R.}$ |
| | $N_{.1}$ | $N_{.2}$ | $\cdots$ | $N_{.j}$ $\cdots$ | $N_{.S}$ | $N$ |

with $i = l, \ldots, R$ and $j = l, \ldots, S$. The corresponding sample count table is

| | 1 | 2 | $\cdots$ | | S | |
|---|---|---|---|---|---|---|
| 1 | $n_{11}$ | $n_{12}$ | $\cdots$ | | $n_{1S}$ | $n_{1.}$ |
| 2 | $n_{21}$ | $n_{22}$ | $\cdots$ | | $n_{2S}$ | $n_{2.}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $n_{ij}$ | $\vdots$ | $\vdots$ $n_{i.}$ |
| $R$ | $n_{R1}$ | $n_{R2}$ | $\cdots$ | | $n_{RS}$ | $n_{R.}$ |
| | $n_{.1}$ | $n_{.2}$ | $\cdots$ | $n_{.j}$ $\cdots$ | $n_{.S}$ | $n$ |

In simple random sampling, the raking algorithm begins by setting

$$\widetilde{N}_{ij} = \frac{N}{n} n_{ij}, \tag{2.1}$$

and then proceeds by proportionately scaling the $\widetilde{N}_{ij}$ such that the relations

$$\sum_{j}^{S} \widetilde{N}_{ij} = N_{i.} \tag{2.2}$$

and

$$\sum_{i}^{R} \widetilde{N}_{ij} = N_{.j} \tag{2.3}$$

are satisfied in turn. Each step in the algorithm begins with the results of the previous step, with the $\tilde{N}_{ij}$ continuing to change; the process terminates either after a fixed number of steps or when expressions (2.2) and (2.3) are simultaneously satisfied to the closeness desired. (See Oh and Scheuren (1983) for further details; see Ireland and Scheuren (1975) for generalizations to multi-way tables and the handling of computational efficiency issues.)

By an application of the theory of minimum discrimination information (Kullback 1968), it can be shown (e.g., Ireland and Kullback 1968) that, under some regularity conditions if only the $N_{i.}$ and $N_{.j}$ are known, the $\tilde{N}_{ij}$ obtained by raking to convergence are asymptotically unbiased, normally distributed and minimum variance (i.e., best asymptotically normal, or BAN, estimators). Theoretical results of this kind are partly what motivates the raking estimator for a general survey characteristic $Y_{ijk}$ (e.g., income or assets), where we are interested in estimating the population total

$$Y = \sum_{i}^{R} \sum_{j}^{S} \sum_{k}^{N_{ij}} Y_{ijk} \tag{2.4}$$

with, say, the statistic

$$\tilde{Y} = \sum_{i}^{R} \sum_{j}^{S} \frac{\tilde{N}_{ij}}{n_{ij}} \left( \sum_{k}^{n_{ij}} Y_{ijk} \right). \tag{2.5}$$

Typically, of course, in survey processing a raking weight

$$\tilde{W}_{ij} = \frac{\tilde{N}_{ij}}{n_{ij}} \tag{2.6}$$

is placed on each individual record on the file for ease of handling. It is important to note that a feature of the raking algorithm is that if $n_{ij} = 0$ then necessarily $\tilde{N}_{ij} = 0$. For convenience, let $\tilde{W}_{ij} = 0$ in such cases as well.

Our interest below will be mainly on the conditional properties of the various estimators being examined. Such an approach has considerable appeal, as advocated by Holt and Smith (1979) and Rao (1985). (As an aside, it may be worth noting that Brackstone and Rao (1979), among others, have looked at the conditional behavior of the raking estimator. They conditioned, however, on the sample marginals $n_{i.}$ and $n_{.j}$.)

### 2.2 Conditional Bias

Following Oh and Scheuren (1983) we focus primarily in this paper on the conditional properties of $\tilde{Y}$, given $\underline{n} = (n_{11}, n_{12} \ldots, n_{RS})$. In particular, let $\bar{Y}_{ij}$ be the population mean for the $ij$-th subgroup. Then the conditional expected value of $\tilde{Y}$ is

$$\mathrm{E}(\tilde{Y} \mid \underline{n}) = \sum_{i}^{R} \sum_{j}^{S} \tilde{N}_{ij} \bar{Y}_{ij} = Y + \sum_{i}^{R} \sum_{j}^{S} (\bar{Y}_{ij} - \bar{Y}) (\tilde{N}_{ij} - N_{ij}). \tag{2.7}$$

Thus $\tilde{Y}$ is conditionally biased with the importance of the bias depending on the structure of the population and whether or not the raking is to convergence. (Of course, when raking to convergence, unconditionally $E(\tilde{N}_{ij}) = N_{ij}$ asymptotically.)

Employing the usual analysis of variance conventions (e.g., Scheffé 1959)

$$(\bar{Y}_{ij} - \bar{Y}) = (\bar{Y}_{i.} - \bar{Y}) + (\bar{Y}_{.j} - \bar{Y}) + (\bar{Y}_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}); \qquad (2.8)$$

hence the conditional bias, given $\underline{n}$, is expressible as

$$\text{Bias}\,(\widetilde{Y} \mid \underline{n}) = \sum_{i}^{R} (\bar{Y}_{i.} - \bar{Y})\,(\widetilde{N}_{i.} - N_{i.}) + \sum_{j}^{S} (\bar{Y}_{.j} - \bar{Y})\,(\widetilde{N}_{.j} - N_{.j})$$

$$+ \sum_{i}^{R} \sum_{j}^{S} (\bar{Y}_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y})\,(\widetilde{N}_{ij} - N_{ij}). \qquad (2.9)$$

If the raking is to convergence, then the first two terms of the conditional bias become zero. For the third term of the conditional bias to be zero for either form of raking, it is sufficient that the $Y_{ij}$ be such that there is no interaction. In large-scale surveying with many variables, this is unrealistic to assume; nonetheless, in practice the interaction is often a minor part of the decomposition of $Y_{ij}$; consequently, the raking ratio estimator may, in many cases, have small biases even in moderate sample sizes.

### 2.3 Conditional Variance

Conditional and unconditional approaches to the variance of the raking ratio estimator have been extensively examined (e.g., Binder 1983; Causey 1972; Bankier 1986; Fan *et al.* 1981; Brackstone and Rao 1979). In our own early work (described in Section 3.2), we have employed replication techniques (e.g., Leszcz, Oh and Scheuren 1983). The replication methods used (which were equivalent to conditioning on the sample marginals) proved expensive, unwieldy, and somewhat unstable, leading us to a simpler attack on the conditional variance estimation problem (albeit the level of conditioning was deeper).

To motivate the approach we are currently taking, consider the conditional variance of $\widetilde{Y}$, given $\underline{n}$. Now it can be shown by a slight extension of Oh and Scheuren (1983) that

$$\text{Var}\,(\widetilde{Y} \mid \underline{n}) = \sum_{i}^{R} \sum_{j}^{S} n_{ij}\,(\widetilde{W}_{ij})^2 \left(1 - \frac{n_{ij}}{N_{ij}}\right) V_{ij} \qquad (2.10)$$

where the $V_{ij}$ are the population variances of the $ij$-th subgroup and if $N_{ij} = 0$ or 1 we define $V_{ij} = 0$. (We are also employing the convention in expression (2.10) that $0/0 = 0$.)

Expression (2.10) holds whether or not the raking goes to convergence. Despite this it has been little studied because it cannot be readily adapted to estimate the conditional variance. The principal difficulty, of course, lies in our inability to calculate stable estimators of the $V_{ij}$ when the $n_{ij}$ are small. To overcome this problem we began looking at collapsing techniques based on the size of the raking weight. First, we let $\widetilde{W}_{ij}$ approximate $N_{ij}/n_{ij}$ which gives us

$$\text{Var}\,(\widetilde{Y} \mid \underline{n}) \cong \sum_{i}^{R} \sum_{j}^{S} n_{ij}\,\widetilde{W}_{ij}\,(\widetilde{W}_{ij} - 1)\,V_{ij}. \qquad (2.11)$$

Now if the $\widetilde{W}_{ij}$ are ordered from smallest to largest and if they vary over a narrow range, then averaging them into (ordered) groups of, say, about $n_g \geq 25$ observations each will alter the value of expression (2.11) very little. It will, however, allow us to calculate collapsed post-stratum variance estimates for the $V_{ij}$. This is the approach we have taken in Section 3.

One final point should be noted. The alternative proposed here is stable and fairly easy to calculate. Our limited empirical work, however, is inconclusive on the method's utility and, while we feel the method is worthy of discussion, we are in no sense advocating its general use at this time.

### 2.4 Modified Raking Estimation

As we have noted, under fairly general conditions the $\tilde{N}_{ij}$ are BAN estimators. This does not mean, however, that $\tilde{Y}$ will share all these properties. Indeed, if the variables used in the raking are not highly correlated with the characteristic Y, the estimator $\tilde{Y}$ may suffer some degradation in variance relative, say, to a simple ratio estimator

$$\tilde{\tilde{\tilde{Y}}} = \sum_{j}^{S} \left(\frac{N_{.j}}{n_{.j}}\right)\left(\sum_{i}^{R} \sum_{k}^{n_{ij}} Y_{ijk}\right). \tag{2.12}$$

Typically, of course, experience has shown that both positive and negative impacts may occur in the same sample. The practitioner's problem is somehow to keep the positive effects while minimizing the negative ones.

There seems to be no general solution to this dilemma but we have had some limited successes, in our application settings, with two techniques that may be of wider interest (see Subsections 3.2 and 3.3 for results).

In most treatments of raking, it is assumed that the marginal population totals $N_{i.}$ and $N_{.j}$ are known; and that the interior of the table $N_{ij}$ can only be estimated from the sample. In our setting we actually have the population values $N_{ij}$ and are employing raking as a way of systematically handling cells in the table where the $n_{ij}$ are small. Conventional collapsing alternatives exist here, of course (e.g., Cochran (1977) Fuller (1966)); but seemed unsuitable for reasons that will be explained later.

It may be possible to agree that raking is a satisfactory way of handling the small cells in this setting; but what about the larger ones? Surely it would be better to use the conventional simple ratio estimator in the large cells. Indeed, if this were done, the conditional bias for these "large" cells would be zero; but what would be the effect on the rest of the cells? This line of reasoning suggested that we employ a hybrid estimation method where, for cells where the $n_{ij}$ was large, the conventional simple ratio estimator is used. These cells are then removed from the population and sample tables, and the remaining sample cells are raked to the adjusted population marginals.

For the remaining smaller cells, a second procedure was introduced to reduce the possible negative impacts of the raking on certain variables. We bounded the raking so that the weights $\tilde{W}_{ij}$ did not vary "too much" from the initial weight. (This kind of constraint is often employed, by the way, in simple ratio estimation, e.g., Hanson 1978.)

The approach to bounded raking ratio estimation is similar to that when "large" sample counts are available in a single cell. That is, it is similar in that, for the cell that is to be constrained, we bound the $\tilde{W}_{ij}$; then take the estimated population total $\tilde{N}_{ij} = \tilde{W}_{ij}\, n_{ij}$ for that cell and the sample $n_{ij}$ for that cell out of the population and out of the sample tables (respectively); and then adjust the remaining observations.

Three problems exist with these partial "solutions." First there is the (uncomfortable) arbitrariness of the definitions of a "large" cell, and of a weighting factor that varies "too much" from its initial value. A related concern was why, if we were willing to use simple ratio estimation for "large" cells, conventional collapsed stratum techniques could not be used for the remaining cells. The third problem has to do with the properties of the raking algorithm's convergence when we employ this hybrid. It is quite clear, for example, from the research that has been done on raking that tables with too many zeros in them will be very unstable and the raking may not converge (e.g., Oh and Scheuren 1978a and 1978b; Ireland and Scheuren 1975). This is of particular concern since the effect of both our modifications is to introduce zeros into the table. If these zeros are strategically placed, or better, *misplaced*, then this could have a very serious detrimental impact on the rate of convergence and, even, on the quality of the estimators. Our recommendation before starting was, therefore, that the number of times that these procedures were employed would have to be fairly small. It is beyond the scope of the present paper to resolve these concerns in general (if indeed that is possible). In Section 3, however, we will consider them further for the applied setting in which we did this work, and also will return to them in Section 4, when discussing areas for future study.

## 3. RAKING IN THE CORPORATE STATISTICS OF INCOME PROGRAM

### 3.1. Background

The U.S. Internal Revenue Service has produced statistics from corporate tax returns annually for over 70 years. Corporate data are, in fact, a mainstay of the so-called Statistics of Income Program, which is the name collectively given to all of the non-administrative statistical series produced by the Internal Revenue Service for public consumption.

Until 1951, corporate statistics were based on a complete census of the returns filed. Since then, a stratified probability sample has been employed, currently running in size at about 90,000 returns annually (from about 3,000,000 returns filed). Assets and income are the principal stratifying variables (Jones and McMahon 1984). Stratification by industry has long been considered, as well, but the quality of the industry coding as self-reported by taxpayers seemed insufficient to justify this step on a wholesale basis. Typically, for example, at the minor industry level perhaps 20 percent or more of the self-reported codes are changed during statistical processing. Nonetheless, because of the importance of industry statistics, efforts to use administrative data by industry to post-stratify the sample still seemed warranted and have been pursued over many years (e.g., Westat, Inc. 1974; Leszcz, Oh, and Scheuren 1983).

In a pilot post-stratification study done by Westat during the early 1970's, substantial improvements in standard errors were achieved for a number of variables, notably Total Receipts (where a reduction of about 12 percent occurred). Some increases in standard errors took place, however, for variables not closely related to industry (e.g., distribution to shareholders), but these were minor. To handle small cells, Westat used conventional collapsed stratum techniques to combine industry post-strata within the then-existing sample strata. Concerns continued to exist about the quality of the administrative industry data, especially for small cells; in any case, due to other operational priorities, the Westat approach was never implemented.

A major series of budget cuts occurred during the 1980-1982 period, and these forced a number of changes in the sample designs and estimation procedures across nearly all the studies that make up the Statistics of Income Program (e.g., Hinkins and Scheuren 1986; Scheuren, Schwartz, and Kilss 1984); in particular, the corporate study experienced sample size cuts during this period which, although later partially rescinded, reopened the issue of post-stratification by industry.

A raking ratio estimation approach to post-stratification seemed to have appeal over what Westat had done. One of the reasons for this was that concerns about the quality of the marginal administrative totals, by industry, were not as great as for the individual cells. The work of implementing a collapsing scheme could be completely avoided, as well.

### 3.2 Early Modified Raking Results

When we implemented a pure raking scheme for the Tax Year 1979 sample, our principal customers expressed concerns about what we had done. They were particularly worried about the potential for large adjustment factors having an adverse effect on certain statistics. We, in turn, having seen the results ourselves, were concerned that we had not done an adequate job for those industry-sample stratum combinations where the number of sample observations were large. As a consequence, these results were never used and the 1979 Tax Year statistics were published employing normal stratified sampling estimation (NORM).

Research continued, however, and in 1983, a paper was given comparing the root mean square errors of six different variations of raking both with each other and with what we had been doing previously (Leszcz, Oh, and Scheuren 1983). Three "pure" raking alternatives were looked at:

PRRE: "Classical" raking ratio estimation to convergence (Deming and Stephan 1940);

PRRE (200): Simple ratio adjustment of cells with samples of 200 returns or more and "classical" raking of the remaining cells to convergence; and

PRRE (400): Simple ratio adjustment of cells with samples of 400 returns or more and "classical" raking of the remaining cells to convergence.

In addition, three versions of bounded raking ratio estimation were examined, all with the bounds set at ($\sqrt{2/3}$, $\sqrt{3/2}$). These were:

BRRE: Bounded raking ratio estimation (2 cycles);

BRRE (200): Simple ratio adjustment of cells with samples of 200 and bounded raking (2 cycles) of the remaining cells; and

BRRE (400): Simple ratio adjustment of cells with samples of 400 and bounded raking (2 cycles) of the remaining cells.

For the bounded raking we were initially not sure that complete convergence was possible; hence, we made an operational simplification and only cycled through the constraint equations, e.g., (2.2) and (2.3), twice.

To make the root mean square error (RMSE) comparison, pseudo-replicate half-samples were drawn, each designed in the same way as the overall sample. The procedure involved: (1) construction of the half-samples; (2) two-way classification - by original sample stratum and major industry (post-stratum) - of sample counts for each half-sample; (3) derivation of a set of weights for each half-sample for each estimator; (4) calculation of estimates of selected items by applying the weight to sample values for each half-sample; and (5) calculation of the RMSE, based on the variations in the estimates that each half-sample produced. For cost reasons only 14 sets of half samples were used.

The resultant summary tabulation presented as Table 1 reveals what one would have expected of the number of returns. Near 100 percent reductions occurred for the PRRE, PRRE(200), and PRRE(400) estimates. Application of the bounding limits $\sqrt{2/3}$ and $\sqrt{3/2}$, and not cycling to convergence, decreased the magnitude of these reductions; however, they were still substantial. As Table 1 also indicates, for Total Receipts, a key variable, there were also improvements, although much less sizable.

**Table 1**

Reduction in Root Mean Square Error (RMSE)
as a Percent of Corresponding Normal Stratified Sampling RMSE

| Estimator | Number of Returns | Total Receipts | Jobs Credit |
|---|---|---|---|
| "Pure" raking ratio estimators: | | | |
| PRRE | 98.6 | 8.3 | −3.09 |
| PRRE (400) | 98.6 | 9.2 | −3.09 |
| PRRE (200) | 98.6 | 11.9 | −3.09 |
| Bounded raking ratio estimators: | | | |
| BRRE | 74.0 | 13.8 | +1.09 |
| BRRE (400) | 73.4 | 15.6 | +1.09 |
| BRRE (200) | 72.3 | 17.4 | +1.09 |

Note: The percentages shown are simple averages of the percent reductions in each of the 56 major industry groups used in the post-stratification. Notice that the percentage improvements for the "number of returns" column are nearly but *not* 100 percent for the PRRE estimators. This occurs because the raking took place for all corporations, with both the $N_{ij}$ and $n_{ij}$ defined on this basis; however, only active corporations (about 90 percent) were tabulated. The BRRE estimators in the "number of returns" column differ from each other and from the PRRE estimator because the cycling was not to convergence. This has subsequently been changed, beginning with Tax Year 1985.

Jobs Credit results in Table 1 are included to illustrate the expected tradeoff that can exist for items not closely related to industry. In particular, we see that in some cases there are (modest) increases in the root mean square errors for this item, due presumably to the fact that this field is less dependent upon the industry groupings utilized in this research.

It should be noted that, for Total Receipts, the decreases shown in the root mean square error, from the initial (NORM) estimate to that utilizing raking ratio estimation, all compare favorably with the Westat pilot study results. While we are encouraged by this comparison, a great deal has changed over the decade between the earlier Westat results and those in Leszcz, Oh and Scheuren (1983). What would really be telling, and what has not been done, is to compare conventional collapsing schemes with our modified approach to raking *on the same data set*.

One final point about Table 1; it reflects improvements in RMSE when tabulating by the administrative industry information which was used in the post-stratification. Because of differences between the administratively and statistically assigned classifications by industry, the figures shown in this table are therefore likely to overstate the improvements being achieved in our published statistics, since so many entities (over 20 percent) are recoded during the indepth processing done of our corporate sample.

### 3.3 Current Modified Raking Results

Beginning with Tax Year 1980, we began to regularly produce and publish our corporate statistics using the bounded raking ratio estimator BRRE(200) (U.S. Department of Treasury 1984). For Tax Years 1983 and later, we made the modifications described in Section 2.3 so that approximate conditional variances could be calculated. These were first published for Tax Year 1984 (U.S. Department of Treasury 1987). Also, in an effort to confirm the earlier results, we undertook for Tax Year 1984 to compare the conditional variance of the modified raking method being employed with the variance that would have been estimated had we used normal stratified sampling estimation. Before discussing the limited comparisons made, it might be worthwhile giving some of the application details on the corporate setting for 1984.

In our earlier work (Leszcz, Oh and Scheuren, 1983), and for 1984, the entire corporate return population of IRS Forms 1120 and 1120S was tallied into 58 major industry groups. For 1984, industry was cross-classified by 14 sample strata in each of the two processing years during which the sample had to be selected. Some of the major industries were so sparse that we immediately collapsed the industry detail to 56 groups. This still left a very large table (of 1568 cells).

It may be of interest to note that there were 414 "natural" zero cells in the population and an additional 125 zero cells arising in the sample. Before raking we removed 96 cells that had 200 or more sample observations; these cells were then each ratio adjusted separately. (In all, 57 percent of the Forms 1120 and 1120S corporate sample were so adjusted.) Finally, there were 73 cells that had to be bounded during the raking itself. This meant that altogether in the raking step there were 708 or 45 percent of the cells being treated as zeroes.

The raking was initiated by introducing the normal stratified estimator into each cell of the table. The marginal constraints imposed were (1) by industry and sampling period, and (2) by sample strata and sampling period. In the published statistics for 1984, and in the comparisons made here, the raking did not go to convergence; it was just carried out for two cycles. (Incidentally, concerns about the conditional bias of this approach have led us to rake our 1985 sample data to convergence.)

The results of the efforts for 1984 were to reduce the overall and industry-by-industry standard errors for frequencies by substantial amounts - only about half as much, however, as is shown in Table 1. Similar dampened improvements occurred for Total Receipts (8.7 percent) with many variables like Jobs Credit and Net Income experiencing little or no change in their standard errors overall (see U.S. Department of Treasury 1987, for details). As already noted, conditioning may be part of the reason for this difference (Holt and Smith 1979). The original results were conditional on the sample marginals $n_{i.}$ and $n_{.j}$; the later figures employed a deeper level of conditioning.

We are still examining other possibilities as to why the improvements are more modest than we found in the earlier work. Some obvious possibilities are the way we grouped the data from the smaller cells, including the consequent averaging of the weighting factors $\widetilde{W}_{ij}$, and the collapsed variance estimation of the $V_{ij}$. Tabulating the data using our statistical industry coding, rather than the administrative coding, as in Table 1, may have been a major factor.

## 4. CONCLUSIONS AND AREAS FOR FURTHER STUDY

### 4.1 General

The modified raking approach for our corporate sample certainly seems to be an improvement over the normal stratified sampling approach taken formerly. There are, however, a number of unsettling *ad hoc* aspects of the method that trouble us. For instance, the connection between conventional collapsed stratum techniques and our modified raking procedure needs more study. Exploring changes in estimation techniques is not enough, however. More work on the basic sample design appears needed too. Finally, the variance approximation being used needs further looking at. We may well have paid a high price for stability and ease of calculation. As noted earlier, the statistical literature is full of good alternatives, and these deserve to be examined in a full-scale comparison with what we are currently doing.

### 4.2 Estimation Issues

There is considerable intuitive appeal in developing a post-stratification method that *smoothly* increases the degree of conditioning from just using marginal totals to using some or all of the interior population counts as well. Our current approach has an embarrassing *ad hoc* flavor. Frankly, we see it just as a stop gap until we can increase the quality of the underlying administrative data by industry. Our main concern is to reduce response variation arising from taxpayer or processing errors. Even if we are unsuccessful in improving the administrative data directly, it may be possible to dampen the response error effects by looking at the tables by industry and sample stratum over several years. This is planned and may allow us to integrate, in a more complete way, raking on the one hand and collapsed post-stratum estimation on the other.

### 4.3 Design Issues

Improved administrative data by industry has obvious uses at the design stage. At the present time, coefficients of variation differ quite widely by industry, with the smaller industries being very poorly represented. No amount of after-the-fact post-stratification can correct for this completely. Improving the balance by industry, and over time, appear to be top priorities (e.g., Hinkins, Jones and Scheuren 1987).

## ACKNOWLEDGMENTS

As is true of nearly all applied work, the authors have many people to thank for the results in this paper. Homer Jones, who has responsibility for the corporate sample, made important contributions, ably assisted by Richard Collins. Nat Shaifer provided illuminating statistics on the degree of comparability of the administrative and statistical coding of industry. Mike Leszcz played the leading role in presenting the earlier results on this topic at the 1983 American Statistical Association meetings in Toronto. The paper profited from conversations with Rod Little and many helpful suggestions were made by the referees. Bettye Jamerson assisted in the manuscript preparation.

## REFERENCES

BANKIER, M.D. (1986). Estimators based on several stratified samples with applications to multiple frame surveys. *Journal of the American Statistical Association,* 81, 1074-1079.

BINDER, D.A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review,* 51, 279-292.

BRACKSTONE, G.J., and RAO, J.N.K. (1979). An investigation of raking ratio estimators. *Sankhyā,* Ser. C, 41, 97-114.

CAUSEY, B.D. (1972). Sensitivity of raked contingency table totals to changes in problem conditions. *Annals of Mathematical Statistics,* 43, 656-658.

COCHRAN, W.G. (1977). *Sampling Techniques, (3rd ed.).* New York: John Wiley.

DEMING, W.E. (1943). *Statistical Adjustment of Data.* New York: Dover.

DEMING, W.E., and STEPHAN, F.F. (1940). On a least square adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics,* 11, 427-444.

FAN, M.C., WOLTMAN, H.F., MISKURA, S.M., and THOMPSON, J.H. (1981). 1980 Census variance estimation procedure. *Proceedings of the Section on Survey Research Methods, American Statistical Association,* 176-181.

FIENBERG, S.E. (1970). An iterative procedure for estimation in contingency tables. *Annals of Mathematical Statistics,* 907-917.

FULLER, W.A. (1966). Estimation employing post strata. *Journal of the American Statistical Association,* 61, 1172-1183.

HANSON, R. (1978). The Current Population Survey: design and methodology. Technical Paper No. 40, U. S. Bureau of the Census.

HINKINS, S., JONES. H., and SCHEUREN, F. (1987). Updating tax return selection probabilities in the corporate Statistics of Income program. A paper presented at the International Symposium on Statistical Uses of Administrative Data, Ottawa, Canada, November 23-25, 1987.

HINKINS, S., and SCHEUREN, F. (1986). Hot deck imputation procedure applied to a double sampling design. *Survey Methodology,* 12, 181-195.

HOLT, D., and SMITH, T. M. F. (1979). Post-stratification. *Journal of the Royal Statistical Society,* Series A, 142, 33-46.

IRELAND, C.T., and KULLBACK, S. (1968). Contingency tables with given marginals. *Biometrika,* 55, 179-188.

IRELAND, C.T., and SCHEUREN, F.J. (1975). The rake's progress. *Computer Programs for Contingency Table Analysis,* Washington, DC: The George Washington University, 155-216.

JONES, H., and MCMAHON, P.B. (1984). Sampling corporation income tax returns for Statistics of Income, 1951 to present. *Proceedings of the Section on Survey Research Methods, American Statistical Association,* 269-274.

KULLBACK, S. (1968). *Information Theory and Statistics.* New York: Dover.

LESZCZ, M., OH, H.L., and SCHEUREN, F. (1983). Modified raking estimation in the corporate SOI program. *Proceedings of the Section on Survey Research Methods, American Statistical Association,* 107-111.

OH, H.L., and SCHEUREN, F.J. (1978a). Multivariate raking ratio estimation in the 1973 Exact Match Study. *Proceedings of the Section on Survey Research Methods, American Statistical Association,* 716-722.

OH, H.L., and SCHEUREN, F.J. (1978b). Some unresolved application issues in raking ratio estimation. *Proceedings of the Section on Survey Research Methods, American Statistical Association,* 723-728.

OH, H.L., and SCHEUREN, F.J. (1983). Weighting adjustments for unit nonresponse. In *Incomplete Data in Sample Surveys, Volume 2, Theory and Bibliographies,* (Eds, W.G. Madow, I. Olkin and D.B. Rubin), New York: Academic Press, 143-184.

RAO, J.N.K. (1985). Conditional inference in survey sampling. *Survey Methodology,* 11, 15-31.

SCHEFFÉ, H. (1959). *The Analysis of Variance.* New York: John Wiley, 106-119.

SCHEUREN, F., SCHWARTZ, O., and KILSS, B. (1984). Statistics from individual income tax returns: quality issues and budget cut impact. *Review of Public Data Use,* 12, 55-67.

STEPHAN, F.F. (1942). Iterative method of adjustment sample frequency tables when expected margins are known. *Annals of Mathematical Statistics,* 13, 166-178.

U. S. DEPARTMENT OF TREASURY (1984). *Statistics of Income – 1980 Corporation Income Tax Returns,* Publication 16, Internal Revenue Service, U.S. Department of Treasury.

U. S. DEPARTMENT OF TREASURY (1987). *Statistics of Income – 1984 Corporation Income Tax Returns,* Publication 16, Internal Revenue Service, U.S. Department of Treasury.

WESTAT, INC. (1974). Results of a study to improve sampling efficiency of statistics of corporation income. Bethesda, Maryland (unpublished).