# MULTIVARIATE RAKING RATIO ESTIMATION IN THE 1973 EXACT MATCH STUDY

H. Lock Oh and Fritz Scheuren
Social Security Administration

Raking ratio estimation was first proposed in a paper by Deming and Stephan in 1940. Since then, additions to their work have been many and varied. In this paper we will describe a multivariate extension made at the Social Security Administration as part of the 1973 Exact Match Study.

Organizationally, the material is divided into four sections. We begin with some background on how the need to do raking arose in the Match Study (Section 1). This is followed by a brief discussion of the basic theory of raking ratio estimation (Section 2). The third section of the paper describes a multivariate extension which we made during the project. Section 4 concludes with a few remarks that relate this paper to the companion presentation on "Some Unresolved Application Issues in Raking Ratio Estimation."

(Space limitations do not permit the inclusion of the extensive numerical material prepared for the meetings. This is, however, available upon request.)

## 1. BACKGROUND ON THE 1973 EXACT MATCH STUDY

Serious gaps exist in our general knowledge of the overall income distribution of persons and families in the United States [e.g., 38].* This is so despite several major Federal statistical programs which collect detailed income information. Periodically, in an attempt to fill some of these gaps, the Bureau of the Census and the Social Security Administration (SSA), in partnership with the Internal Revenue Service (IRS), have engaged in interagency data linkages for statistical purposes. The 1973 Exact Match Study is the most recent such effort to be completed and the largest to date. Its starting point was the March 1973 Current Population Survey (CPS). For the sampled individuals a match was made between the CPS and social security benefit and earnings records. As part of the project, a limited set of tax items from 1972 Federal income tax returns was also furnished to the Census Bureau by IRS.

Need for raking.--The CPS data linkage studies conducted prior to that for March 1973 all suffered from significant survey undercoverage errors and errors arising from failures to complete all the required matching to administrative records [42]. Major changes in the Current Population Survey and in administrative procedures at Social Security and IRS would have been needed to avoid such an outcome in the 1973 Study. The Census Bureau and Social Security personnel who worked on the project, however, had to operate within the existing systems. Instead of expecting to eliminate virtually all coverage and matching errors, adjustment procedures had to be developed which would mitigate their effect on the subsequent analyses. The particular strategy adopted to deal with both these problems was to reweight the matched CPS cases employing raking ratio estimation techniques.

Nature of reweighting.--In order to prepare the Exact Match Study weights, an extensive research and development program was required. Four activities were important in this:

1. studying the nature of the coverage errors in the CPS, including an examination of the Census Bureau's adjustments which partially "correct" for such errors;

2. studying the nature of the nonmatching errors in the project;

3. developing the required estimates from administrative and other outside sources to adjust for the Study's coverage and matching errors; and,

4. developing the appropriate raking ratio estimation procedures and computational algorithms for reweighting the sample to "correct" for undercoverage and nonmatching.

It is the last of these which, of course, is the primary focus of the current paper. Readers interested in the other aspects of the reweighting may consult [44, 47 and 49] for details.

## 2. UNIVARIATE RAKING RATIO ESTIMATION

"Raking" is a procedure for iteratively ratioing sample data to known (outside) marginal totals. The technique was first proposed by Deming and Stephan in 1940 [6] and has been rediscovered and renamed [e.g., 14, 24] several times since.

One way to specify the raking problem is to set up a series of condition (or constraint) equations. Consider, for example, a three-way table of (weighted) sample counts $\{n_{ijk}\}$. Assume that we know the "true" row $\{m_{i..}\}$, column $\{m_{.j.}\}$, and "layer" $\{m_{..k}\}$ marginals and that we wish to obtain adjusted sample counts $\{\tilde{n}_{ijk}\}$ such that

$$(1)\quad \begin{cases} \sum_{j,k} \tilde{n}_{ijk} = m_{i..} \\ \sum_{i,k} \tilde{n}_{ijk} = m_{.j.} \\ \sum_{i,j} \tilde{n}_{ijk} = m_{..k}. \end{cases}$$

---

*Bracketed numbers pertain to the bibliography found at the end of the paper. These references are also cited in the companion paper provided next in these Proceedings.

In a simple ratio estimation problem, we must solve only one of these "condition equations," say

(2) $$\sum_{j,k} \tilde{n}_{ijk} = m_{i..}$$

where the subscripts "j" and "k" represent all of the uncontrolled dimensions in the sample and where the subscript "i" stands for the "controlled" (or constrained) marginal.

To solve expression 2 using a simple ratio estimator, we first find

(3) $$\sum_{j,k} n_{ijk} = n_{i..}$$

from the sample and then (if $n_{i..} \neq 0$) choose

(4) $$\tilde{n}_{ijk} = \{F_i\}\, n_{ijk}$$

where the

(5) $$F_i = \left\{\frac{m_{i..}}{n_{i..}}\right\}$$

are the ratio adjustment factors needed to reweight the sample so that it will agree with the known outside totals. The process is non-iterative, unlike raking.

To derive the $\{\tilde{n}_{ijk}\}$ in the more general situation where all the condition equations in expression 1 must be satisfied, we proceed by proportionately adjusting the cell values $\{n_{ijk}\}$ so that every one of the equations is satisfied in turn. Each step begins with the results of the previous step, the process terminating when all the equations are <u>simultaneously</u> satisfied to the degree of closeness desired.

<u>Iterative adjustment process and weighting factors</u>.--Specifically in the present case, one could begin with a proportionate ratio adjustment by rows

(6) $$n_{ijk}^{(1)} = \left\{\frac{m_{i..}}{n_{i..}}\right\} n_{ijk} = a_i^{(1)}\, n_{ijk}$$

followed by a column adjustment

(7) $$n_{ijk}^{(2)} = \left\{\frac{m_{.j.}}{n_{.j.}^{(1)}}\right\} n_{ijk}^{(1)} = b_j^{(1)}\, n_{ijk}^{(1)}$$

and a layer adjustment

(8) $$n_{ijk}^{(3)} = \left\{\frac{m_{..k}}{n_{..k}^{(2)}}\right\} n_{ijk}^{(2)} = d_k^{(1)}\, n_{ijk}^{(2)} \quad .$$

These three successive ratio adjustments constitute a cycle that is then repeated in whole or in part until the individual cell entries cease to change appreciably.

The weighting factors $\{F_{ijk}\}$ from the raking process can be derived in either of two ways. Since the $\{\tilde{n}_{ijk}\}$ are calculated as an integral part of the procedure, we can simply let

(9) $$F_{ijk} = \begin{cases} \tilde{n}_{ijk}/n_{ijk} & n_{ijk} \neq 0 \\ 0 & n_{ijk} = 0 \end{cases} \quad .$$

Alternatively, we can keep track of the cumulative products over all the cycles "c" required, that is, of

(10) $$\begin{cases} \tilde{a}_i = \prod_{c=1} a_i^{(c)} \\ \tilde{b}_j = \prod_{c=1} b_j^{(c)} \\ \tilde{d}_k = \prod_{c=1} d_k^{(c)} \end{cases} \quad .$$

The factors $F_{ijk}$ could then be expressed as

$$F_{ijk} = \tilde{a}_i\, \tilde{b}_j\, \tilde{d}_k \quad .$$

This second formulation is preferred since it displays the underlying structure of the adjustment and allows one to compare more readily the sampling properties of simple and raking ratio estimators.

<u>Statistical properties of raking ratio estimators</u>.--In discussing the statistical properties of raking ratio estimators, it is of some value to distinguish between three types of variables. These are--

1. <u>Sample marginals</u>.--First, we have the row $\{\tilde{n}_{i..}\}$, column $\{\tilde{n}_{.j.}\}$, and layer $\{\tilde{n}_{..k}\}$ totals which are constrained directly and, assuming convergence, can be brought as close as desired to the "true" marginal totals $\{m_{i..}\}$, $\{m_{.j.}\}$, and $\{m_{..k}\}$. In many applications, however, including several in the Exact Match Study, for cost reasons the iterations are allowed only to proceed a few cycles. Good, or at least acceptable, results seem to have been achieved in U.S. and Canadian census studies when raking from two to four cycles [19-20, 36]. Our own work, in a setting where bias was an issue (unlike the census studies), suggests that if there is any substantial disagreement between the <u>expected</u> sample totals and the outside "controls" then sizable differences can persist for a very long time before the raked sample marginals begin to settle down.

2. <u>Sample cell frequencies</u>.--The second set of statistics to consider are the weighted cell counts $\{\tilde{n}_{ijk}\}$ which make up the interior of the multi-way table whose marginals are constrained in the raking. In simple random sampling, assuming convergence, certain optimality properties exist for the raked cell counts $\{\tilde{n}_{ijk}\}$; for example, they can be shown to be BAN estimators [3, 11]. In more general sampling settings, the raked weighted cell frequencies still possess certain desirable properties [12]. Alternative adjustment

procedures [e.g., 17], though, may yield estimators with smaller variances, especially in nonself-weighting designs.

3. Other statistics.--Finally, and usually most important, are the remaining variables $\{Y_{ijkh}\}$ about which information was collected in the survey. For instance, we might construct an estimator of, say, a particular population total Y from

$$\tilde{Y} = \sum_{i,j,k,h} W_{ijkh} \; Y_{ijkh}$$

before raking; and from

$$\approx{Y} = \sum_{i,j,k,h} F_{ijk} \; W_{ijkh} \; Y_{ijkh}$$

after raking, where the $\{F_{ijk}\}$ and the subscripts "i," "j," and "k" have the same meaning as before. The subscript "h" is added to allow us to identify any particular sample observation in the ijk[th] cell. The $\{W_{ijkh}\}$ are the sample weights.

In general, for estimators like $\approx{Y}$, the results which exist suggest that raking can yield appreciable reductions in mean square error over the corresponding unraked statistics, i.e., $\tilde{Y}$, but this outcome is by no means assured. The situation seems to be similar to that which pertains when one is engaged in ordinary separate (univariate or multivariate) ratio estimation [e.g., 39].

The above discussion of statistical properties assumes basically that the raking algorithm will converge. However, that is not always the case. Much theoretical work has been done on this subject. As yet, though, there does not seem to be any easily verifiable set of necessary and sufficient conditions which will allow us to determine when convergence will in fact occur.

The convergence proofs which exist for the raking algorithm make strong assumptions about cell counts--for example, that they are all present [11] or that some particular combination is present [3]. For many practical problems, therefore, we have found that the best method of checking for convergence is simply to attempt to carry out the raking adjustment. In all the applications we made in the 1973 Study, once we set up the statistical problem "sensibly," convergence always occurred. The main things we had to guard against were trying to impose too many constraints on the sample, imposing constraints that were themselves contradictory, or controlling, either explicitly or implicitly, groups with very small expected sample sizes. (An example of one "sensible" rule we tried to follow was not to rake a weighted survey total unless its effective sample size was greater than, say, 20 or 25 times the number of dimensions being constrained. 1/)

## 3. MULTIVARIATE RAKING RATIO ESTIMATION

So far we have described what might be called univariate raking ratio estimation. As part of the 1973 Exact Match Study, we had to develop a multivariate version of raking. To see what our particular multivariate extension consists of, consider the three-way table $\{\underline{n}_{ijk}\}$, each cell of which is a three-component vector

$$\underline{n}_{ijk} = \begin{pmatrix} n_{1ijk} \\ n_{2ijk} \\ n_{3ijk} \end{pmatrix} . \tag{11}$$

Assume we wish to obtain adjusted cell entries

$$\tilde{\underline{n}}_{ijk} = \begin{pmatrix} \tilde{n}_{1ijk} \\ \tilde{n}_{2ijk} \\ \tilde{n}_{3ijk} \end{pmatrix} , \tag{12}$$

subject to marginal constraints on each component of the $\{\underline{n}_{ijk}\}$ such that

$$\begin{cases} \sum_{j,k} \tilde{n}_{1ijk} = m_{1i..} \\ \sum_{i,k} \tilde{n}_{2ijk} = m_{2.j.} \\ \sum_{i,j} \tilde{n}_{3ijk} = m_{3..k} \end{cases} . \tag{13}$$

Rather than fit each of these constraints separately, the requirement is made that all the counts in a cell must receive the same adjustment. The $\{\tilde{\underline{n}}_{ijk}\}$, therefore, have to be of the form

$$\tilde{\underline{n}}_{ijk} = a_i b_j d_k \, \underline{n}_{ijk} \tag{14}$$

where the row, column, and layer adjustment factors, $a_i$, $b_j$, and $d_k$, respectively, are scalars.

The raking ratio procedure for deriving the $\{\tilde{\underline{n}}_{ijk}\}$ still retains the straight-forwardness of the univariate case. For the three-way table we are looking at, the process could begin with

$$\underline{n}^{(1)}_{ijk} = \left\{\frac{m_{1i..}}{n_{1i..}}\right\} \underline{n}_{ijk} = a^{(1)}_i \; \underline{n}_{ijk} \tag{15}$$

followed by

$$(16) \quad \underline{n}_{ijk}^{(2)} = \left\{ \frac{m_{2.j.}}{n_{2.j.}^{(1)}} \right\} \underline{n}_{ijk}^{(1)} = b_j^{(1)} \; \underline{n}_{ijk}^{(1)}$$

and

$$(17) \quad \underline{n}_{ijk}^{(3)} = \left\{ \frac{m_{3..k}}{n_{3..k}^{(2)}} \right\} \underline{n}_{ijk}^{(2)} = d_k^{(1)} \; \underline{n}_{ijk}^{(2)} \; .$$

These three successive steps constitute the first cycle of the adjustment procedure and, just as in the univariate iteration, would be repeated in whole or in part until the individual cell entries cease to change appreciably.

Need for multivariate raking.--At least three types of units are important in the statistical literature on surveys. These are the "sampling (or design) unit," the "unit of analysis," and what we have called the "unit of estimation." 2/ To discuss why we developed a multivariate extension of raking, it may be helpful to examine each of these units in the context of the March 1973 Current Population Survey [41] since, as has been mentioned, that sample was the basis for the Exact Match Study:

1. CPS design units.--In a multi-stage (household) probability sample, like the CPS, different sampling or design units are defined at each stage. For example, for selections at the first stage, the unit is a large area (county, city, SMSA, etc.). At subsequent stages, smaller areas (e.g., enumeration districts) are defined. The last stage, or ultimate sampling unit (USU), is a particular address or set of addresses, all of whose members are to be interviewed. In the March 1973 CPS, the USU's primarily consisted of four housing units.

2. CPS units of analysis.--In the CPS, as in most large multi-purpose surveys, there are many different units of analysis. CPS income statistics, for instance, are prepared for persons, families, and various other aggregations of housing unit or household members (couples, STATS units [45], subfamilies, etc.).

3. CPS units of estimation.--In general, this is simply the unit used in constructing the sample weights. The choice of the unit of estimation depends (among other things) on the design or sampling unit, the unit of analysis desired, the nature of the outside information (if any) which is to be incorporated, and the model one has for the structure and severity of the nonsampling errors to which the survey measurement process was subjected.

   In the standard weighting of the CPS there are several stages of estimation, each of which is carried out with a different unit. For example, the first step employs the (ultimate) sampling unit as the unit of estimation. It is at this step that the inverse of the probability of selection is developed and assigned. In the next phase of the weighting, the unit of estimation is an occupied housing unit or household. (Here the noninterview nonresponse is adjusted.) Later stages of the estimation use outside information from the previous decennial census (first-stage ratio), up-to-date independent U.S. population data (second-stage ratio), and data from previous CPS's (composite weighting). At all of these three latter stages the estimation unit is basically an adult civilian. Adult civilians, of course, are the units of analysis in the monthly CPS labor force series.

   Because of the differential undercoverage between men and women in the CPS, the standard (person) weighting of the sample can lead to major inconsistencies in the weighted counts of the number of men estimated to be living with their wives and the number of women estimated to be living with their husbands. Therefore, in March of each year, when the CPS is used to study income and family status, the weighting procedure is modified. Discrepancies between counts of "married-spouse-present" males and females are avoided by first ratioing the sampled females up to their controls. All males coded as "married-spouse-present" are then assigned the newly-derived weights of their wives. Finally, the remaining men are ratioed to new "controls" obtained by subtracting from the overall male totals the adjusted counts for men who are "married-spouse-present."

The March Supplement weighting is an attempt to introduce outside "control" totals about persons in order to improve statistics for a unit of analysis (families) about whose numbers we have no such independent knowledge. Implicit in the procedure is the notion that the wife can be used to represent both members of the couple, i.e., that the wife is the unit of estimation for the couple.

The Supplement weighting has two serious deficiencies which could have been avoided if a multivariate raking procedure had been carried out instead. First, in some cases, particularly for nonwhites, the procedure yields estimators which can be quite unstable. The subtraction step, moreover, can lead to the absurdity of negative "controls" for men who are not "married-spouse-present." A second objection to the method is that it leaves other inconsistencies in family status still unresolved (for example, counts of children obtained using the weights assigned to each child versus counts derived by using the weights of their parents). 3/

As part of the 1973 Exact Match Study, we decided to replace the Supplement weight with one derived using multivariate raking. This would allow us to consider simultaneously the differential coverage of both men and women as well as that of the children living with them. This change in procedure also allowed us to adjust the sample not only to population totals by age, race, and sex (as in the regular CPS), but also to totals derived from administrative sources for taxfilers, social security covered workers, and OASDI beneficiaries.

Statistical properties of multivariate extension.--To the extent we have studied them so far, it seems that there are many similarities in the statistical properties of univariate and multivariate raking ratio estimators:

1. Sample marginals.--As in the univariate case, sample totals for the components constrained in the raking can be brought as close as desired to the corresponding known outside marginal totals. In the very limited Monte Carlo work we have done, large simple random samples were drawn from a synthetic CPS-type population [34]. For the cases studied, the sample marginals being adjusted seemed to settle down "fairly soon" to values close to their expectations. We might add, however, that convergence was much slower in applications where there appeared to be a substantial disagreement between the expected sample totals and the outside "controls." [35]

2. Sample cell frequencies.--The weighted cell vectors, e.g., the $\{\tilde{n}_{ijk}\}$, possess at least some of the optimality properties of the corresponding univariate setting. In particular, because of the nature of the iteration, all the cross-product ratios [12] in all the planes of the table are preserved. Also, it should be possible to show, arguing as in [11], that the cell vectors $\{\tilde{n}_{ijk}\}$ are BAN estimators in simple random sampling. This suggests that, ceteris paribus, one should choose the "unit of estimation" in raking to be the same as the "unit of analysis." 4/

3. Other statistics.--Multivariate raking estimators of other statistics have exactly the same form as do univariate estimators. Our limited Monte Carlo results (and more extensive applications) confirm what this similarity in form suggests. Multivariate raking, like its univariate counterpart (see the companion paper [32]), can lead, sometimes, to a reduction in the mean square error over that of an unraked estimator.

Of course, the above discussion of statistical properties assumes convergence. Unlike univariate raking, there has been almost no theoretical work done on convergence for the multivariate extension. Identifying some of the necessary conditions for convergence is fairly easy. For example:

1. Each component of the vector of counts must be such that the univariate raking of it would converge to the controls associated with that component.

2. The set of controls imposed on the various components must be compatible (i.e., noncontradictory).

It must be added, however, that obtaining conveniently checkable sufficient conditions has eluded us so far.

## 4. CONCLUDING REMARKS

There are several aspects of both univariate and multivariate raking which need more study. In order to examine some of our concerns, therefore, we have begun a series of Monte Carlo experiments. A few early results from these are described in the companion paper [32] on "Some Unresolved Application Issues in Raking Ratio Estimation."

In the present paper, we have described a multivariate extension of raking ratio estimation. The paper as published here makes only brief mention of a number of the application issues which arise in practice. There was more discussion of these in the longer version of this paper which was distributed at the meetings and which is available upon request.


## ACKNOWLEDGEMENTS

The help of C. Terrence Ireland of The George Washington University is gratefully acknowledged. He was an active participant in the early stages of the work described in this paper and continues in a consultant role. Editorial, computational and other assistance was provided by Linda DelBene, Clarise Lancaster, and Beth Kilss. Joan Reynolds typed the present draft.


## FOOTNOTES

1/ This is an extension of the various rules of thumb used in simple ratio estimation. See for instance [48, p. 53] or more generally [40, p. 194].

2/ "Elementary units" are also commonly discussed in survey sampling texts, sometimes as synonyms for analysis units [40, p. 238; 43, p. 6] and sometimes simply as the units of the population which cannot be decomposed further [46, p. 11].

3/ In recent years, negative and very unstable "controls" have been eliminated by collapsing the categories within which the ratio adjustment is carried out. The second objection to the Supplement weighting still continues, however, in full force.

4/ In the longer version of this paper, available by writing to us , there is an extended discussion of the degree to which univariate and multivariate raking have "equivalent" properties.

# PARTIAL BIBLIOGRAPHY ON RAKING RATIO ESTIMATION AND OTHER SELECTED REFERENCES CITED IN RAKING PAPERS


### Some Historical References on Raking and Related Techniques (1940-1972)

[1] Bacharach, M. Estimating nonnegative matrices from marginal data, International Economic Review, vol. 6, No. 3, 294-310, 1965.

[2] Bishop, Y. M. Multidimensional contingency tables: cell estimates. Ph.D. Dissertation, Harvard, 1967.

[3] Bishop, Y. M. and Fienberg, S.E. Incomplete two-dimensional contingency tables. Biometrics, vol. 25, 383-400, 1969.

[4] Causey, B. D. Sensitivity of raked contingency table totals to changes in problem conditions, Annals Math. Stat., vol. 43, 656-8, 1972.

[5] Deming, W. E. Statistical Adjustment of Data. Wiley, 1943.

[6] Deming, W. E. and Stephan, F. F. On a least squares adjustment of a sampled frequency table when the expected marginal tables are known, Annals Math. Stat., vol. 11, 427-44, 1940.

[7] El-Badry, M. A. and Stephan, F. F. On adjusting sample tabulations to census counts, J. Amer. Stat. Assn., vol. 50, 738-62, 1955.

[8] Fienberg, S. E. An iterative procedure for estimation in contingency tables, Annals Math. Stat., vol. 41, 907-17, 1970.

[9] Friedlander, D. A technique for estimating a contingency table, given the marginal totals and some supplementary data, J. Royal Stat. Soc. - A, vol. 124, 412-420, 1961.

[10] Goodman, L.A. The analysis of cross-classified data: independence, quasi-independence and interactions in contingency tables with or without missing entries, J. Amer. Stat. Assn., vol. 63, 1091-131, 1968.

[11] Ireland, C. T. and Kullback, S. Contingency tables with given marginals, Biometrika, vol. 55, 179-88, 1968.

[12] Mosteller, F. Association and estimation in contingency tables, J. Amer. Stat. Assn., vol. 63, 1-28, 1968.

[13] Nargundkar, M. S. and Arora, H. R. The raking-ratio estimation procedure for the 1971 Census, Census Division Memorandum, Statistics Canada, 1971 (unpublished).

[14] Rosenblatt, H. M. Notes on a study of proposed ratio estimators for the 1970 Census. Census Bureau Working Paper (unpublished).

[15] Sinkhorn, R. A relationship between arbitrary positive matrices and doubly stochastic matrices, Annals Math. Stat., vol. 35, 876-9, 1964.

[16] Smith, J. H. Estimation of linear functions of cell proportions, Annals Math. Stat., vol. 18, 231-254, 1947.

[17] Stephan, F. F. Iterative method of adjusting sample frequency tables when expected margins are known, Annals Math. Stat., vol. 13, 166-78, 1942.

### Recent Canadian References on Raking (1973-1978)

[18] Arora, H. R. and Brackstone, G. J. The bias and variance of the raking ratio estimator, Technical Report, Statistics Canada, July 1975 (unpublished).

[19] Arora, H. R. and Brackstone, G. J. An investigation of the properties of raking ratio estimators: I. with random sampling, Survey Methodology, vol. 3, No. 1, 62-83, 1977.

[20] Arora, H. R. and Brackstone, G. J. An investigation of the properties of raking ratio estimators: II. with cluster sampling, Survey Methodology, vol. 3, No. 2, 232-252, 1977.

[21] Bankier, M. D. An estimate of the efficiency of raking ratio estimators under simple random sampling, Survey Methodology, vol. 4, No. 1, 115-124, 1978.

[22] Brackstone, G. J. and Rao, J. N. K. Raking ratio estimators, Survey Methodology, vol. 2, No. 1, 63-69, 1976.

[23] Brackstone, G. J. and Rao, J. N. K. An investigation of raking ratio estimators, 1978 (draft working paper).

[24] Rao, J. N. K. Raking ratio estimators, Technical Report, Statistics Canada, January 1974 (unpublished).

[25] Rao, J. N. K. Raking ratio estimators, II, Technical Report, Statistics Canada, March 1974 (unpublished).

[26] Rao, J. N. K. Raking ratio estimators, III, Technical Report, Statistics Canada, October 1974 (unpublished).

Other Recent References on Raking
(1973-1978)

[27] Causey, B. D. Variance of raked-table entries. U.S. Census Bureau Memorandum, 1973.

[28] Ireland, C. T. and Scheuren, F. J. The rake's progress, The George Washington University, Computer Programs for Contingency Table Analysis, 155-216, 1975.

[29] Killian, R. A., Singh, R., Thompson, J. and Chapman, D. Study Plan for the 1980 Census Estimator Investigation (unpublished, U.S. Census Bureau working paper), 1978.

[30] Koch, G. G. and Freeman, D. H., Jr. The asymptotic covariance structure of estimated parameters from marginal adjustment (raking) of contingency tables. Presented to the Washington Statistical Society, January 6, 1976.

[31] Oh, H. L. and Scheuren, F. J. Multivariate raking ratio estimation in the 1973 Exact Match Study, American Statistical Association Proceedings, Section on Survey Research Methods, 1978.

[32] Oh, H. L. and Scheuren, F. J. Some unresolved application issues in raking ratio estimation, American Statistical Association Proceedings, Section on Survey Research Methods, 1978.

[33] Pugh, R. E., Tyler, B. S. and George, S. Computer-based procedure for the N-dimensional adjustment of data--NJUST, U.S. Social Security Administration, Staff Paper No. 24, 1976.

[34] Scheuren, F. J., et al. Working notes on social security applications of multivariate raking--MULTIRAKE (unpublished SSA working paper), 1976.

[35] Scheuren, F. J. Methods of estimation in 1973 CPS-IRS-SSA Exact Match Study, 1977 (unpublished SSA working paper to appear in Report No. 10 in the series Studies from Interagency Data Linkages).

[36] U.S. Bureau of the Census, Procedural History, Chapter 11: Sampling and Estimation, PHC(R)-1E, 1974.

[37] Waksberg, J., Hanson, R. and Bounpane, R. Estimation and presentation of sampling errors for sampling data from the 1970 U.S. Census. Paper prepared for the joint meetings of the International Statistical Institute and the International Association of Survey Statisticians, Vienna, Austria, Aug. 20-30, 1973.

Remaining References
Cited in Papers

[38] Budd, E., Radner, D. and Hinrichs, J. Size distribution of family personal income: methodology and estimates for 1964, Bureau of Economic Analysis, Staff Paper No. 21, 1973.

[39] Cochran, W. Sampling techniques, Chapter 6, Wiley, 1976.

[40] Hansen, M., Hurwitz, W. and Madow, W. Sample survey methods and theory, vol. I, Wiley, 1953.

[41] Hanson, R. The Current Population Survey: design and methodology, Technical Paper No. 40, U.S. Bureau of the Census, 1978.

[42] Kilss, B. and Scheuren, F. The 1973 CPS-IRS-SSA Exact Match Study: past, present, and future (presented at the 1978 SSA Workshop on Policy Analysis with Social Security Research Files).

[43] Kish, L. Survey sampling, Wiley, 1976.

[44] Oh, H. and Scheuren, F. Working notes on the derivation of 1972 social security earner, tax return and OASDI beneficiary counts for CPS-eligible U.S. civilians on April 1, 1973 (staff paper, revised 1977).

[45] Projector, D., Millea, M. and Dymond, K. Projections of March Current Population Survey: population, earnings, and property income, March 1972 to March 1976, Studies in Income Distribution, Report No. 1, 1974.

[46] Raj, D. The design of sample surveys, McGraw-Hill, 1972.

[47] Scheuren, F. and Oh, H. Fiddling around with mismatches and nonmatches, American Statistical Association Proceedings, Social Statistics Section, 1975.

[48] U.S. Bureau of the Census, The Current Population Survey: a report on methodology, Technical Paper No. 7, 1963.

[49] Yuskavage, R., Hirschberg, D. and Scheuren, F. The impact on personal and family income of adjusting the Current Population Survey for undercoverage, American Statistical Association Proceedings, Social Statistics Section, 1977.