IMPUTATION OF SELECTED ITEMS IN CORPORATE TAX DATA: IMPROVING UPON THE EARLIER HOT DECK

John L. Czajka, Mathematica Policy Research, Inc.

Part of the appeal of administrative records as a data source is their low cost relative to a direct survey and their greater accuracy for particular kinds of data. Yet the extraction of satisfactory data files from administrative records may still entail substantial absolute costs, and the conversion of the reported data fields into consistent and accurate measures for analytic purposes may require extensive editing. Both observations apply to the files of income and tax data produced by the Statistics of Income (SOI) Division of the Internal Revenue Service (IRS) from samples of individual and corporation tax returns. The SOI files serve critical data needs that could not be satisfied by tabulations from the master files produced for revenue processing purposes.

This paper addresses the use of imputation to supplement editing for selected fields on selected returns in the preparation of corporation SOI data. Mathematica Policy Research, Inc. has been involved in an evaluation of this effort, and this paper reports preliminary results. The paper contains three parts:

(1) a review of what IRS sought to accomplish with the imputations and how the objectives were carried out in the 1982 program;
(2) an assessment of the extent to which these procedures achieved their objectives; and
(3) a discussion of improvements planned or being considered for 1985 and later.

IMPUTATION IN THE 1982 CORPORATION SOI DATA

The use of imputation in the preparation of corporation SOI data has been discussed by IRS staff at prior Joint Statistical Meetings [1,2] and elsewhere [3]. Here we provide an overview with some elaboration of points crucial to an understanding of this presentation.

The corporation tax return includes several miscellaneous items where the reporting of a nonzero amount requires the attachment of a supplementary schedule detailing the specific sources of that amount. The information supplied on these schedules often implies that portions of the reported miscellaneous amounts should be redistributed to other, more specific line items on the return. Generally these redistributions have no revenue implications (key totals are not changed), so revenue processing does not routinely make such adjustments, but the corrections are of interest to Treasury Department and Congressional analysts who use the SOI data. Therefore, when a return with one or more of these supplementary schedules falls into the SOI sample, the schedules must be reviewed in order that the final recorded amounts may reflect the distribution suggested by the schedules.

Reviewing the supplementary schedules and then editing the reported line items adds significantly to the cost of preparing the SOI corporation tax data. In 1981 and again in 1982, rather than review all schedules for the more than 90,000 sampled returns, the SOI Division opted to leave some of the schedules unedited on selected returns and use a subset of the edited records as donors to impute changes to the unedited fields.

Seven schedules were designated for such treatment in 1981 and 1982. For editing purposes, the seven are grouped into four sets, as depicted in Figure 1. Redistribution of amounts occurs only within each set, and no item appears in more than one set.

The selection of sample returns for editing versus imputation seeks to achieve two potentially competing objectives: (1) maximize the amount of editing that is averted, and (2) minimize the impact of the imputations upon the estimation error associated with the affected fields. Clearly this suggests restricting imputation to returns with small miscellaneous amounts, taking advantage of the uneven distribution of amounts across returns and the disproportionately large number of returns with relatively small amounts. However, estimation error is a concern not only for amounts aggregated over all corporations, but for subaggregations by detailed industry and asset size, as the SOI Division prepares estimates for hundreds of subclassifications. Consequently, the objectives cannot be fully met by selecting returns for editing or imputation on the basis of the amounts alone.

The selection strategy adopted by IRS includes two steps: (1) designate selected schedules on selected returns for editing only, thereby precluding these from any possibility of imputation; (2) subsample all returns with one or more undesignated schedules to select a group of returns for further editing as donors, to be used in imputing changes to the remaining returns.

The schedules precluded from imputation in 1982 comprised: all schedules on returns from the largest firms in their respective industries and any additional firms with net income or assets in excess of $50 million; all schedules corresponding to reported amounts that failed consistency tests; all schedules corresponding to reported amounts judged likely to be changed by editing. The latter amounts were identified by large values relative to the totals of which they formed parts--e.g., Other Income relative to Total Income, or Other Assets relative to Total Assets. Prior experience has shown that editors are more likely to make a change in such circumstances than if the miscellaneous amount is relatively small.

Perfect implementation of the IRS selection scheme would result in the editing of all fields where such editing would actually result in changes, and the imputation of only fields that, in fact, require no changes. The fields edited to provide donor information would receive no changes, so no changes would be imputed to the remaining unedited fields.

In 1982, records were selected for editing as donors by systematic selection within strata

Figure 1.--Potential Redistribution of Amounts as the Result of Reviewing Seven Miscellaneous Schedules

Miscellaneous Items from Which Amounts May Be Removed:	Set I	Set II		Set III		Set IV	
	Other Income (OI)	Cost of Goods Sold (COGS)	Other Deductions (OD)	Other Current Assets (OCA)	Other Assets (OA)	Other Current Liabilities (OCL)	Other Liabilities (OL)
Selected Items to Which Amounts May Be Added:	Gross Receipts Interest on Govt. Oblig. Rents Royalties	OD COGS Employee Benefits Depreciation Taxes Paid		OA OCA Accounts Receivable Other Investments		OL OCL Short Term Mortgages Long Term Mortgages Accounts Payable	

defined by nine major industries and nine asset classes. Records from nonfinancial industries (the great majority) were selected with a 10 percent probability, while records from financial industries were selected with a 20 percent probability.

The application of the above selection criteria in 1982 produced the following distribution of 86,637 sample returns, the number left after excluding approximately 3,000 Domestic International Sales Corporations and a somewhat greater number of returns filed on special forms:

- 11,340 Returns of firms with $50 million in total assets: no exposure to imputation
- 14,598 Returns of firms with under $50 million in assets but not exposed to imputation:
  - Firms with relatively large total assets in their industries
  - Firms with none of the 7 schedules
  - Firms with all schedules selected for editing
- 60,699 Returns left for subsampling (containing at least one as yet unedited set of schedules):
  - 6,503 Subsampled as donors
  - 54,196 Subsampled as imputes.

Even among the returns subsampled for imputation, most had at least one schedule previously edited.

An editor's review of one of the schedules may produce no changes, or it may result in the movement of some part or all of the miscellaneous amount into one or more other fields, including those listed in the lower part of Figure 1. The left hand portion of Figure 2 illustrates the kind of edits that might result from a review of the Other Income schedule. In the example $40,000 are moved from the original Other Income amount to Gross Receipts ($20,000), Interest on U.S. Government Obligations ($15,000) and Rents ($5,000). Note that this particular transfer adds dollars to one field in which the original amount was zero. Editing can thus change not only the distribution of amounts but the number of firms estimated to have nonzero amounts.

The imputation of changes to unedited fields is accomplished by a hot deck procedure. Donor records are matched to similar records containing one or more unedited sets of items. The matching utilizes 10 industry groups, three asset size classes, and the 15 possible combinations of one to four unedited sets of items. Changes expressed as proportions of the miscellaneous item are imputed from the donor record to the unedited items. This is illustrated on the right in Figure 2. In the example the original Other Income amount on the impute record is half that on the donor record, so the imputed dollar changes are only half those on the donor.

The imputation of changes in the form of ratios, rather than absolute dollars, ensures that no negative miscellaneous amounts will be generated (and then have to be edited out). However, it creates the potential for bias in the imputed changes and, hence, the final amounts, as the average number of dollars added to or subtracted from a given field on an impute

Figure 2.--Illustration of Edit Based on Other Income Schedule and Subsequent Imputation of Changes

Item	Record Edited for Use As Donor			Information Transferred from Donor to Impute	Record with Changes Imputed		
	Original Amount	Change	Final Amount		Original Amount	Imputed Change	Final Amount
Gross Receipts	$180,000	+$20,000	$200,000	+20.0%	$120,000	+$10,000	$130,000
Interest on U.S. Oblig.	0	+15,000	15,000	+15.0	10,000	+7,500	17,500
Rents	30,000	+5,000	35,000	+5.0	0	+2,500	2,500
Other Income	100,000	-40,000	60,000	(-Σ) -40.0	50,000	-20,000	30,000

record is not constrained to equal the average number added to or subtracted from that same field on the donor records. The magnitude of this bias was presumed to be small when IRS introduced the imputation procedure, but empirical estimates have since been prepared. These are related below.

EVALUATION OF THE IMPUTATIONS

This evaluation of the corporation imputations addresses three areas: (1) how much editing was actually averted? (2) what was the price in data quality? and (3) how well did particular elements of the imputation procedure perform?

Volume of Editing Averted

The fact that more than 54,000 records were designated for imputation does not imply that fewer than half of all schedules were reviewed. As was noted earlier, most of the 54,196 records selected for imputation had already had at least one set of items edited. Furthermore, many returns do not have all seven schedules, and the incidence of zero values in the miscellaneous amount fields is likely to be greater among smaller firms than among larger corporations.

Nevertheless, the editing that was saved was substantial. Table 1 reports estimates of the numbers of each of the seven miscellaneous amount fields edited, imputed, or recorded as zero (implying no schedule to be reviewed for editing). The figures exclude several thousand special returns that would not have been subject to imputation. Among the nearly 87,000 returns included in the table, 41.2 percent of all schedules were left for imputation, ranging from 33.3 percent for Other Liabilities to 47.9 percent for Other Current Liabilities.

Table 1.--Frequency of Editing Versus Imputation for Seven Supplementary Schedules: 1982 SOI Sample

Name of Schedule	Zero Amount (No Schedule)	Edit	Impute	Total Schedules
OI	33,188	29,008	24,441	53,449
COGS	27,631	37,715	21,291	59,006
OD	2,741	53,039	30,857	83,896
OCA	26,073	33,711	26,853	60,564
OA	31,669	31,778	23,190	54,968
OCL	17,892	35,843	32,902	68,745
OL	53,674	21,979	10,984	32,963
Total	--	243,073	170,518	413,591
	Percentage of Schedules			
OI	--	54.3%	45.7%	100.0%
COGS	--	63.9	36.1	100.0
OD	--	63.2	36.8	100.0
OCA	--	55.7	44.3	100.0
OA	--	57.8	42.2	100.0
OCL	--	52.1	47.9	100.0
OL	--	66.7	33.3	100.0
Total	--	58.8	41.2	100.0

Note: Tabulations are based on an incomplete file excluding several thousand records.

Price in Data Quality

Because the principal use of corporation SOI data lies in the estimation of aggregates, we measure the reduction in data quality in terms of bias and increased variance. To estimate these impacts we imputed changes to the 6,503 donor records by, first, randomly ordering the records within their hot deck adjustment cells and, then, pairing each record with its successor, the first record in each pair serving as "donor" and the second as "impute." Each record was used as both a donor and an impute. This procedure yielded a set of imputed changes for each set of fields originally edited to provide donor information. The difference between the imputed value and the edited value provides a measure of imputation error.

To calculate aggregate bias, we proceeded as follows. For each donor, i, we computed the imputation error, $Y_i^* - Y_i$, where Y_i^* is the imputed value and Y_i is the edited value. We then doubly weighted each observation, one weight w_i being the inverse of the sample selection probability for that record, and the other weight d_i being the number of times that observation was used as a donor. The sum of the w_i over all donors is the population of firms represented by these 6,503 observations, while the sum of the d_i equals the sample number of records with imputed fields. The sum of the products, $w_i d_i$, is an estimate of the total population of firms represented by the records with imputed fields. An estimate of the bias due to imputation in the estimated aggregate amount over all firms in the population is given by:

$$\Sigma w_i d_i (Y_i^* - Y_i) .$$

To estimate the variance impact of the imputation procedure, we began by calculating the sums and sums of squares of the recorded final amounts within each sampling stratum, using all observations on the SOI file. We then applied the customary formula to obtain the variance of an aggregate, estimated from a stratified sample [4]. This variance estimate includes the impact of imputation error.

Next we used the information contained in the simulated imputation errors to calculate an alternative estimate of variance that excludes the impact of imputation error. The rationale for this tactic begins with the fact that imputation error affects the sums and sums of squares in strata containing imputed records. We cannot determine the magnitude of this effect for the imputed records themselves, but we can calculate sums and sums of squares for edited

versus simulated imputed values over selected aggregates of donor records. While the hot deck procedure does not utilize sampling stratum in matching donor records to imputes, we recall that the donors were subsampled at random from records designated to be used either as donors or imputes. Consequently, we may infer the impact of imputation error on sums and sums of squares in a sample stratum from the observed differences between the statistics calculated for edited versus imputed values among the donor records in that stratum. We did so as follows.

Within each sample stratum we computed sums and sums of squares of the donor Y_i and Y_i^*, weighting in each case by the d_i. We then calculated within each sample stratum the differences:

$$\Sigma d_i Y_i - \Sigma d_i Y_i^* \text{ and}$$

$$\Sigma d_i Y_i^2 - \Sigma d_i (Y_i^*)^2 ,$$

and added these results to the sums and sums of squares calculated from the recorded amounts on the full SOI file. For the reasons noted above, this is approximately equivalent to replacing the sums and sums of squares of the imputed amounts with estimates of their true amounts, subject to sampling error. With these revised sums we then repeated the variance calculation.

The results of these calculations are reported in Table 2 for the seven miscellaneous items and eight additional items. With the exception of four items the estimated biases are below .1 percent and generally well below that amount. The four items with biases above .1 percent include three of the miscellaneous items, with biases estimated to be between .18 and .73 percent.

Comparison of the alternative estimates of the coefficients of variation (C.V.s) of the 15 aggregates reveals even smaller impacts of imputation. Increases due to imputation error were detected for only three items. The largest impact by far was recorded for Depreciation, where we estimate that imputation error increased the coefficient of variation from .336 percent to .343 percent, an increase of 2.08 percent but still a negligible impact. Reductions in the C.V.s due to imputation error were also detected for three items, but these may be attributed to the upwardly biased estimates of total amounts used in the denominators of the C.V.s calculated directly from the SOI file.

It is likely that the differential impacts of imputation across the 15 items can be explained more fully by the average magnitudes of the changes relative to the final amounts than by characteristics of the imputation procedure itself. Recall that the final amount equals the reported original amount plus a change, and that only the change field is affected by the use of imputation versus editing. The impact of imputation is amplified or diminished depending on whether the changes are typically large or small proportions of the final amounts.

Table 3 reports, for the 15 items, the relative frequency of change and the average magnitudes of these changes in the fields edited for use as donors. Net positive changes are expressed as proportions of the final amounts while net reductions are reported as proportions of the original amounts. Both positive and negative changes can occur to six of the seven miscellaneous items, but, except for Other Current Assets and Other Current Liabilities, net increases are rare and are ignored in the discussion. Only increases are recorded for the

Table 2.--Estimated Impact on Aggregate Bias and Coefficient of Variation (C.V) for Selected Items on 1982 Corporation SOI File

(Amounts in Millions of Dollars)

Item	Total Amount Estimated from SOI File	Estimated Bias Due to Imputation	Estimated Percentage Bias	C.V. Estimated from SOI File	C.V. Adjusted to Remove Impact of Imputation	Percentage Increase in C.V. Due to Imputation
OI	73,143	198	0.27%	1.554%	1.559%	-0.32%
Int. on U.S. Oblig.	449,573	-35	-0.01	0.999	0.999	0.00
Gross Receipts	4,647,010	-140	-0.00	0.312	0.312	0.00
Rents	57,843	-4	-0.01	0.808	0.808	0.00
COGS	3,247,710	-660	-0.02	0.377	0.377	0.00
OD	461,237	34	0.01	0.713	0.713	0.00
Bad Debts	20,967	-3	-0.01	0.920	0.919	0.11
Interest Paid	467,843	107	0.02	0.606	0.606	0.00
Depreciation	154,721	539	0.35	0.343	0.336	2.08
Pensions	39,016	25	0.06	0.814	0.814	0.00
Employee Benefits	38,846	29	0.08	0.491	0.488	0.62
OCA	278,698	271	0.10	0.680	0.680	0.00
OA	203,292	1,465	0.73	0.528	0.532	-0.75
OCL	2,381,530	40	0.00	1.199	1.199	0.00
OL	371,174	658	0.18	0.968	0.970	-0.21

NOTE: Estimates are based on an incomplete sample file that excludes several thousand sample firms, many of which are large.

Table 3.--Frequency and Relative Magnitudes of Changes to Selected Fields Edited to Provide Donor Information: 1982 Corporation SOI File

Item	All Edited Fields	Fields with Changes	
	Frequency of Change	Mean Increase As % of Final Amount	Mean Reduction As % of Original Amount
OI	25.5%	--	-70.0
COGS	12.0	78.3*	-7.6
OD	31.2	8.7*	-8.7
OCA	19.9	61.9	-49.8
OA	28.8	65.2*	-76.3
OCL	6.8	41.7	-37.5
OL	4.6	45.0*	-72.1
Interest on Govt. Oblig.	3.9	71.3	--
Gross Receipts	21.2	61.4	--
Rents	1.4	87.7	--
Bad Debts	0.6	66.2	--
Interest Paid	0.6	44.2	--
Depreciation	7.3	62.4	--
Pensions	0.8	50.4	--
Employee Benef.	17.7	91.4	--

*Net increases are rare.

eight remaining items, which receive amounts redistributed from the miscellaneous items.

Among the seven miscellaneous items, rates of change due to editing varied from 4.6 percent for Other Liabilities to 31.2 percent for Other Deductions. Changes were generally much less frequent among the eight other items, although Gross Receipts and Employee Benefits had comparatively high rates of change. Average magnitudes of change show wide variation. For six items the average change exceeded 70 percent of the original or final amount, as the case may be, but for two other items the mean change was less than 10 percent.

Clearly there is some correspondence between the frequency and average magnitude of change (Table 3) and the observation of a comparatively high bias or variance impact (Table 2). The item with the highest bias, Other Assets, had the highest mean change overall (product of frequency and average magnitude). Other Income ranked third in bias and second in mean change. Employee Benefits, generally imputed in its entirety when it was changed at all, showed a relatively high variance impact but negligible bias. However, Gross Receipts exhibited no bias or variance impact, despite a relatively high frequency of change and mean increase, and the high variance impact displayed by Depreciation was not accompanied by an exceptionally high rate or average magnitude of change. This suggests that variation among the items in the effectiveness of the donor-impute matching may have contributed to the results in Table 2.

Effectiveness of Elements of the Procedure

Three elements of the imputation procedure were given careful attention in our evaluation: (1) the imputation of changes in the form of ratios rather than absolute amounts, (2) the selection of records for editing versus imputation, and (3) the matching of donors with the records to be imputed.

Use of ratios. -- Table 4 compares the donor records (weighted by the number of times used) with the imputed records, with respect to the distributional characteristics of selected nonzero change fields. If absolute changes rather than ratios had been imputed, the values reported in the imputed records columns in Table 4 would be identical to those in the donor

Table 4.--Comparison of Distributional Statistics for Selected Nonzero Change Fields Between Donor and Imputed Records, Excluding Financial Firms: 1982 Corporation SOI File

(Amounts in Thousands of Dollars)

Item to Which Amounts Were Added	Item From Which Amounts Were Taken	Donor Records				Imputed Records			
		Median	Mean	Std. Dev.	Number of Sample Records	Median	Mean	Std. Dev.	Number of Sample Records
Receipts	OI	13.8	62.4	157.4	206	5.3	37.3	126.1	1897
Interest	OI	11.0	27.7	45.4	103	3.3	24.6	81.8	971
Rents	OI	4.4	22.4	44.6	29	3.4	22.1	52.3	264
Taxes	COGS	46.4	148.5	274.4	293	39.6	162.3	405.0	2603
Deprec.	COGS	89.3	249.7	534.5	250	77.0	419.2	2228.9	2278
Emp. Ben.	COGS	56.0	145.7	213.0	124	56.5	181.1	469.7	1153
Taxes	OD	1.8	14.0	33.7	420	1.5	29.8	185.3	4095
Emp. Ben.	OD	14.9	41.7	110.4	381	11.6	50.1	180.4	3601
Repairs	OD	7.2	28.7	52.8	85	5.0	28.7	79.9	797
OCA	OA	12.1	54.4	115.5	306	8.5	63.0	194.0	2168
Oth. Inv.	OA	37.9	117.8	246.4	607	19.9	102.3	300.2	4352
Act. Rec.	OCA	25.4	115.9	272.3	247	14.4	84.4	220.4	1988
OCL	OL	35.3	115.2	152.0	46	75.0	271.2	460.0	167
Mortgage	OCL	38.0	167.2	306.6	103	24.2	149.9	329.5	917

records columns. Instead we find the following differences. In almost every case the imputed median change is smaller than the median change among the donors. The imputed mean change, however, is often larger than the mean change among donors. This suggests that ratios yield smaller changes generally but with greater variance. Very large imputations occur more frequently with the ratio method, inflating the mean imputed amount. Improving the donor-impute matching might alter this, however, so we cannot draw a firm conclusion as yet.

Selection of firms. -- Recall that while all schedules from large firms are reviewed, the reviewing of schedules from small firms is selective. If the selection were perfect, then all schedules implying changes to amounts would be reviewed, and all of the remainder would be left for subsampling. (If selection were in fact this perfect, the imputation procedure would be superfluous.) Such perfection may be unattainable, but, if the selection were effective to any degree, the frequency of changes resulting from the review of selected schedules would exceed that from schedules on returns subsampled as donors.

To assess the effectiveness of the selection of schedules for review prior to subsampling, we compared the rates of change to the seven miscellaneous amounts for two sets of edited fields on records subsampled as donors: (1) fields edited prior to subsampling, and (2) fields edited after subsampling (i.e., for the purpose of imputing changes to other records). The comparative rates of change are shown below:

Item	Edits prior to subsampling	Edits after subsampling
OI	45.3%	25.5%
COGS	25.9	19.1
OD	28.2	31.2
OCA	23.3	12.0
OA	42.0	40.9
OCL	11.1	4.9
OL	17.1	13.4

For six of the seven items the rate of change is indeed higher for fields edited prior to rather than after subsampling, but for only three of the six is the difference appreciable.

The cost effectiveness of the editing could be enhanced, and the imprecision introduced by imputation could be reduced, if the selection criteria were improved so as to increase the percentages in the first column and reduce those in the second. The 60,699 sample records that were designated as donors or imputes in 1982 contained 304,503 schedules in support of the seven miscellaneous items. Of this number, 114,137 were selected for review prior to subsampling, with changes resulting in only 27.9 percent of the cases. Thus 82,274 schedules were reviewed unnecessarily, in effect. Another 19,849 schedules were reviewed in order to generate donor information, making it possible to "save" reviewing the remaining 170,517 schedules. The review of donor schedules produced edited changes to 21.1 percent of the 19,849 miscellaneous amounts and imputed changes to 18.3 percent of the 170,517 amounts. The implication is that 31,140 schedules requiring changes were excluded from review and the changes left for imputation. The numbers indicate that there exists a significant potential for reducing the volume of edits and the number of changes left for imputation.

Matching of donor and impute records. -- Direct evidence of the effectiveness of the matching has been difficult to assemble because many of the adjustment cells are so small. For example, an attempt to measure the between- versus within-cell variance of the change ratios was hampered by the absence of changes in many cells and the occurrence of just a single change in many others. Two observations may be made.

The "pattern" variable (combination of sets of schedules not yet edited) accounts for little if any improvement in the matching and, yet, it increases by 15-fold the number of adjustment cells required. This makes a strong case for finding an alternative to the use of pattern in imputing changes.

One factor which the 1982 donor-impute matching did not take into consideration was whether only one or both schedules of a pair required imputation (i.e., had nonzero original amounts). If a donor had a nonzero amount in only one of the fields while the impute record had nonzero amounts in both fields, the field in which the donor amount was zero would have been ignored in the imputation. For both the asset and liability pairs we found markedly lower change rates for at least one of the pair among the impute records than we did among the donor records. We did not see much evidence of this in the COGS item, even though zero fields occurred as often there as with OA. Nevertheless, matching on the schedules present would seem to offer improvement in the imputation results.

FUTURE DIRECTIONS

The results presented here provide strong support for the continued use of imputation in the corporation SOI program. Following a two-year hiatus, imputation will again be used to supplement editing in the 1985 program, but the procedures will include a number of modifications. Imputation will be limited to the OI, COGS and OD schedules (for procedural rather than methodological reasons), and this will greatly reduce the number of adjustment cells required to match on the pattern of schedules to be imputed. In light of the experience in 1982, however, records will be matched on the presence of both COGS and OD amounts. Several additional changes are under consideration for 1985 and later.

The use of donors from prior years could speed up processing and contribute to further cost savings. The availability of prior-year donors at the start of processing could make in-stream imputation feasible; currently imputation does not begin until late in the production cycle, when sufficient donors are available [3]. In addition, the use of prior-year donors would permit a reduction in the number of schedules edited in the "current" year without loss of precision.

Replacement of the hot deck with a group mean imputation would simplify the use of prior-year

donor information and would permit the elimination of pattern entirely from the adjustment cell definition, as each cell would have mean values for changes to all schedules.

Another proposal involves separation of the imputation of change from the conditional imputation of the amount of change. This could be combined with improvement of the selection of schedules for editing prior to subsampling. Both require a model of the probability of change. Results presented here suggest that there remains a substantial payoff to improving the selection.

Finally, there is indirect evidence that the matching of donors to imputes could bear improvement, but there are few candidates for new covariates beyond those already mentioned. Research in the short term is focusing on the ratios used to select records for editing prior to subsampling.

ACKNOWLEDGMENTS

The author wishes to thank Sharon Hirabayashi for generating the empirical results presented herein; Roderick Little and Donald Rubin for their collaboration on the larger project; the SOI Division for its support of this work, and especially Susan Hinkins and Fritz Scheuren for their comments on earlier drafts and continued input to the research; and Lena Cunningham for typing the final manuscript.

REFERENCES

[1] Hinkins, S. (1983), Matrix Sampling and the Related Imputation of Corporate Income Tax Returns, American Statistical Association, Proceedings of the Section on Survey Research Methods.

[2] Hinkins, S. (1984), Matrix Sampling and the Effects of Using Hot Deck Imputation, American Statistical Association, Proceedings of the Section on Survey Research Methods.

[3] Hinkins, S., and F.J. Scheuren (1986), Hot Deck Imputation Procedure Applied to a Matrix Sample Design, paper presented at the Symposium on Nonresponse in Sample Surveys, Ottawa, April 1986 (in this volume).

[4] Cochran, W.G. (3rd ed., 1977), Sampling Techniques, Wiley, New York.