PREDICTING EDIT OUTCOMES: THE STRATEGIC USE OF IMPUTATION IN ESTIMATING CORPORATE INCOME STATISTICS

John L. Czajka, Mathematica Policy Research, Inc.

The Statistics of Income (SOI) Division of the Internal Revenue Service (IRS) produces extensive income and tax statistics from data extracted from samples of corporation and individual tax returns. The preparation of these data involves extensive editing, which ranges from automatic correction of internal inconsistencies to the review of portions of the taxpayer's original return. Such review can be particularly costly because it involves first securing the return from service center files; reading what is frequently handwritten information supplied in a nonstandard format; making a determination as to the appropriate correction; entering the change; and then retesting and, if necessary, re-correcting the edited fields. The corporation data program provides several examples where the taxpayer's return contains information that can be used to refine items in the SOI data base, but where the cost of obtaining this additional information for all records would be prohibitive, given other needs and pressure on resources.

To address this problem in the context of one set of items, IRS devised a strategy that entails editing only some of the sampled returns and using the results of these edits to impute edit outcomes to the remainder. The strategy utilizes elements of double sampling in conjunction with the substitution method of imputation [1], but it departs from traditional applications of these techniques, as will be explained. The specific procedures implemented in tax years 1981 and 1982 have been described in a number of papers presented in earlier years at these meetings [2,1,3] and elsewhere [4]. Several changes are being introduced for tax year 1985, however. This paper departs from the earlier discussions to:

(1) consider the potential impact of these procedures upon estimation error at small levels of aggregation

(2) discuss several modifications being introduced into the imputation procedures, in part to reduce this estimation error

(3) speculate about the more general utility of this type of approach to reducing bias in administrative and survey data

Before taking up these three objectives we begin with an overview of the editing and imputation problem as it relates to one of the items in question: Other Income.

COMPENSATING FOR NON-EDITS: AN OVERVIEW OF THE PROBLEM

If a firm reports Other Income, it provides a supplementary schedule detailing the sources. SOI editors reviewing these schedules often find that portions of what is reported as Other Income can be reclassified as more specific kinds of amounts--which are more informative to analysts--e.g., Gross Receipts or Other Dividends. They edit the data fields, moving out some of what was originally reported as Other Income and adding it to one or more of these other fields. For this presentation we ignore the impact upon these other fields and focus on Other Income.

The Other Income estimator for an individual sample firm is thus:

$$(1) \qquad Y_i = B_i - C_i$$

where Y_i represents the final amount, B_i represents the beginning amount, and C_i represents the change. (It is possible for C_i to be negative, but this happens very rarely, and we ignore such cases in this presentation.) For an aggregate of sample firms, the Other Income estimator is:

$$(2) \qquad \sum Y_i = \sum B_i - \sum C_i$$

The aggregate estimate consists of the total original amount less the total amount removed. In the absence of editing, therefore, $\sum C_i$ is the aggregate bias.

Rather than review all of the Other Income schedules associated with the sampled returns, IRS does the following to correct Other Income:

(1) reviews the schedules for all "large" returns

(2) reviews the schedules for selected smaller returns--those with a high likelihood of being changed by an edit

(3) reviews the schedules for a random subsample of the remaining returns (20% of those in financial industries; 10% of those in nonfinancial industries)

(4) imputes edit outcomes to the remaining returns, using the random subsample as donors

This paper focuses on the fourth step. Note that if step (2) could be executed to perfection, the returns left unedited after that point would require no changes, and imputation would be unnecessary.

To impute edit outcomes IRS uses the substitution method. Each record to be imputed is matched to a donor. Change information from the donor is then substituted into the empty change fields. The matching is accomplished by a variation on the hot deck. Within an adjustment cell, described in the next section, a donor is selected from the top of a once shuffled deck. That donor is then recycled to the bottom of the deck. This process is repeated until all returns within the adjustment cell have been imputed. Generally this requires nine cycles through the deck for nonfinancial returns and four cycles for financial returns.

For returns not edited, the C_i term in equation (1) is not observed. An imputed change C_i^* must be substituted. If we let the variable E_i indicate whether a return was edited (E_i=1) or not edited (E_i=0), the general estimator of Other

Income for a sample firm is given by:

$$(3) \quad Y_i = B_i - E_iC_i - (1 - E_i)C_i^*$$

that is, the original amount less either the observed change or an imputed change.

IRS does not provide micro-level corporation data to users, so error at the micro-level is important primarily insofar as it affects the error for reported aggregates. However, individual records do have to satisfy an extensive battery of consistency tests. For Other Income the test with the most direct bearing on the imputations is the requirement that the amount removed from Other Income not exceed the original amount. Thus equation (3) must satisfy the constraint that C_i^* not exceed B_i.

To meet this constraint, IRS imputes the change to Other Income as a proportion of the original amount. In other words, what is substituted from the donor is the ratio of the donor's change to the donor's original amount. The imputed change is thus defined as:

$$(4) \quad C_i^* \equiv B_iR_i^* \quad \text{where} \quad 0 \leq R_i^* \leq 1.$$

For records with imputed changes, the final amount is given by:

$$(5) \quad Y_i = B_i - B_iR_i^* = B_i(1 - R_i^*) ,$$

which satisfies the consistency test regarding the relative magnitudes of the original amount and the imputed change.

THE IMPACT OF IMPUTATION

The objective of IRS's strategic use of imputation in the corporation SOI program is to substantially reduce the number of fields manually edited while minimizing the increase in the mean square error (MSE) of reported estimates. The reduction in the volume of editing has indeed been sizable; in 1982 imputation replaced manual editing for over 170,000 schedules [3]. Generally the associated increase in MSE was negligible at high levels of aggregation [3,4]. Here we consider how the design of the imputation procedure may affect MSE at small levels of aggregation.

Levels of Aggregation

IRS publishes corporation income and tax estimates for several levels of aggregation. The most widely distributed tabulations include aggregates over all firms plus subaggregates by 10 industrial divisions, 58 major industrial classes, and 185 minor classes. Estimates are published by 12 size classes as well. The less widely distributed but highly important Source Book of Statistics of Income on Corporation Income Tax Returns reports item totals for a cross-classification of firms by all 185 minor industrial classes and 12 size classes--yielding 2,220 potential reporting cells. Many of these cells are in fact empty, and cells in the top size classes are often represented by 100 percent of their respective firms. For most of the cells, however, the published aggregates depend on sample data, with sample sizes varying widely around a mean of about 100 returns per cell.

The impact of imputation on the MSE of reported aggregates of final amounts has two components: the variance of the aggregate imputed change and its (squared) bias. The principal source of variance is sampling error in the information taken from the donors. Because of double sampling, the aggregate contribution of sampling error is much more significant than in more typical examples of imputation by substitution. The imputation procedures themselves introduce variance through the selection of donors (some donors will be used more often than others) and the pairing of donors with imputes (recall that the imputed change is the product of a proportionate change substituted from the donor and the original amount present on the return being imputed). The potential for bias exists in the possibility that the donors used in a reporting cell may not be representative of that cell--e.g., if reporting cells are subsets of adjustment cells or cut across adjustment cells. Bias is also introduced by the imputation of ratios rather than actual amounts, although it should be noted that imputing actual amounts would necessitate subsequent editing, which would generate bias of another kind.

Three major factors differentiate high and low levels of aggregation (by industry and size) with respect to the impact of imputation on MSE. First, in most industries the distribution of returns by the dollar amount of a given financial item is heavily skewed, with a small number of very large returns accounting for most of the total dollars. Because returns in the highest size classes are not subject to imputation, totals aggregated across all size classes will reflect predominantly edited rather than imputed outcomes. Second, at high levels of aggregation the net imputed change is based on a much larger sample of donors and hence is characterized by much smaller sampling error than is true at low levels of aggregation. Third, at high levels of aggregation the included returns encompass entire imputation adjustment cells. Since one component of bias washes out at the adjustment cell level (see below), aggregates of adjustment cells exhibit significantly less imputation bias than subaggregates.

For these reasons, estimates of small impacts of imputation on the MSE of high level aggregates in 1982 [3,4] are not at all surprising and do not rule out the possibility that much larger impacts on MSE may exist among detailed reporting cells. Empirical estimates of the impact of imputation upon reported totals at this level of aggregation are still in preparation, and they will be subject to high variation. However, certain theoretical inferences can be drawn from an understanding of the design of the imputation procedures, knowledge of the corporation sample, and empirical information on the distribution of edit outcomes. These inferences have contributed to the redesign of the imputation procedures for use in the production of 1985 corporation SOI data.

Because the variance and bias of the imputations are affected by the adjustment cell definitions and by the substitution of ratios for actual donor amounts, we examine the theoretical implications of these elements of the imputation procedures in detail below.

Adjustment Cell Definition

Equating the adjustment cells with the most detailed reporting cells would minimize the imputation bias in these reporting cells. However, this would be accomplished at a cost of very high imputation variance and MSE. There are approximately 3,000 donors for Other Income. Even if one third of the detailed reporting cells in the first nine size classes were empty, nearly 1,000 would remain, implying an average of about three donors per cell. The mean imputed R_i* in the average reporting cell, therefore, would be based on a sample of size three.

One way to achieve a reduction in this variance is to define the adjustment cells as combinations of reporting cells with similar distributions of change ratios. This increases the sample base for the imputed ratios in each reporting cell, thus reducing the variance around this expectation. Combining similar cells minimizes the resultant increase in bias.

The adjustment cells used by IRS in 1981 and 1982 were defined principally by industrial classification and size of return. To provide adequate numbers of donors per adjustment cell, both classifications were collapsed:

(1) industrial classification was reduced from the 185 minor industries to 10 "major" industry groupings

(2) size of return was reduced from nine classes to three

The net effect was a reduction in the number of potential adjustment cells from 1,665 to only 30, implying an average of about 100 donors per cell. With this configuration the sampling error component of imputation variance would be substantially reduced, but the bias at the level of the reporting cell could be greatly increased.

This bias can be reduced by adding one or more covariates that cut across reporting cells (as opposed to being coterminous with them). Such a strategy maintains much of the variance reduction gained by collapsing the industry and size classifications because the potential pool of donors for each reporting cell is still one of the 30 large aggregates. Within each of these aggregates, however, returns are matched to the donors they most resemble, as defined by the covariates. Thus if the index j refers to one of the 30 aggregates, the expected change imputed to a return is not R_j* but $R_{jk}*$, where k denotes a covariate. A reduction in bias is achieved to the extent that $R_{jk}*$ is closer than R_j* to the true change for each return. This will occur if there is a positive covariance between k and the expected change ratio.

The addition of one or more covariates is likely to reduce the effective sample size somewhat for each reporting cell because donors with certain characteristics will be drawn more often than donors with other characteristics. However, strong covariates can reduce another component of imputation variance in a manner roughly analogous to stratified sampling. Consider the extreme case where there are two strata: one in which all of the donors exhibit the same proportionate magnitude of change and another in which no donor exhibits change. In this case there will be no variance associated with the pairing of donors and imputes. There will still be variability in the relative numbers of donors selected from the two strata. The net effect of adding the covariates, however, could be a reduction in the overall imputation variance.

In more traditional applications of the substitution method, where complete records may outnumber incomplete records by more than nine to one, it is possible to employ multiple covariates in order to obtain a high degree of resemblance between the donor and each record to be imputed. In the present application, where the relative numbers of donors and imputes are reversed, close matches are rarely possible. Minimizing the MSE introduced by imputation requires judicious selection of covariates.

IRS did add a covariate to the adjustment cell definitions in 1981 and 1982. However, this was done primarily to differentiate among records requiring imputation for alternative combinations of items. Other Income is only one of seven miscellaneous items to which the imputation procedures described here were applied. The selective editing described as step two of the four-step sequence above was applied on an item by item basis. Therefore, a return might require imputation for one to seven items. To make it possible to carry out all of the imputations to a given return with a single donor, IRS defined a third adjustment cell dimension reflecting the pattern of items to be imputed. Since six of the seven were always edited (or imputed) in pairs, there were 15 possible patterns, eight of them pertinent to any given item (in other words, eight of these patterns include Other Income, for example).

The introduction of this "pattern" variable increased the number of possible adjustment cells from 30 mutually exclusive and exhaustive groupings of reporting cells to 240 exhaustive but not mutually exclusive groupings.

Figure 1 depicts the relationships between the detailed reporting cells and the imputation adjustment cells. Each row represents a combination of minor industry and size--i.e., a detailed reporting cell. Each column represents one of the eight patterns applicable to, let us say, Other Income. Collectively, the rows are intended to represent one of the 30 groupings of reporting cells by major industry and the collapsed size classification alluded to earlier. Therefore, a given column in this figure corresponds to one of the 240 adjustment cells. The figure is thus a cross-classification of detailed reporting cells by imputation adjustment cells within one combination of major industry and broad size class. A given row and column location includes all returns falling into the same detailed reporting cell and imputation adjustment cell.

The figure indicates that returns in S_{111} will be matched with donors drawn at random from that entire column. Returns in S_{112} will be matched with donors drawn from column two, and so on.

Figure 1--Depiction of Relationship between Adjustment Cells and Reporting Cells for Other Income in 1982 Imputations

Industry (m=1,M) and Size (n=1,3)	"Pattern" (k=1,8)							
	1	2	3	4	5	6	7	8
1,1	S_{111}	S_{112}	S_{113}	S_{114}	S_{115}	S_{116}	S_{117}	S_{118}
1,2	S_{121}							
1,3	S_{131}							
2,1	S_{211}							
2,2	S_{221}			S_{mnk}				
2,3	S_{231}							
.	.							
.	.							
.	.							
M,3	S_{M31}							S_{M38}

NOTE: Each entry represents a set of returns with the same minor industry, size and pattern of items to be imputed. A row constitutes a Source Book reporting cell. The M minor industries depicted here constitute one major industry. A column represents an adjustment cell in the 1982 imputation procedure.

Thus a given reporting cell may have its unedited returns imputed from donors drawn from all of the reporting cells represented in the figure. However, the donors will be drawn from different columns in proportion to the numbers of unedited returns that the reporting cell includes from those columns. If all of the returns in a reporting cell fall into two patterns, then donors will be drawn from only two columns. Different reporting cells may draw donors from the eight columns according to different mixes of probabilities.

As was noted above, the pattern variable was introduced into the adjustment cell definition primarily to make it possible to impute all of the unedited fields on a given record from a single donor. There was some expectation that the mean amounts of changes do vary across the categories of the pattern variable, but substantial reductions in bias and variance were not anticipated. The empirical evidence that we have examined suggests that there is little association between the categories of the pattern variable and the magnitudes of changes observed among the donor records. The pattern variable is not a strong covariate, therefore.

Imputation of Ratios Rather than Amounts

Imputing the edit outcome as a proportion of the original amount, rather than as the donor's actual change, has implications for both the bias and the variance introduced into the final estimates by the imputation procedure. The imputation of ratios rather than absolute changes may increase the bias but lower the variance of the imputation procedure. We consider the variance impact first.

Within an adjustment cell, if the original amount and the change are correlated, the ratio of the latter to the former will exhibit a smaller coefficient of variation--i.e., its values will be distributed more tightly about their adjustment cell mean--than will the change itself. If the correlation is high enough, then the variance of $B_i R_i*$ over all possible imputations to a given return will be smaller than the variance of C_i* if substituted directly from the donor. In the extreme case where B_i and C_i are perfectly correlated, R_i (and therefore R_i*) is constant over all returns. In this case the imputation variance will be due solely to the sampling error of R_i; there will be no additional contribution from the donor-impute pairing.

While reducing the variance, the imputation of ratios rather than absolute amounts may increase the bias of the imputations. The aggregate final amount within a reporting cell, j, is given by:

$$(6) \qquad \sum_i Y_{ij} = N_j \bar{B}_{.j}(1 - \bar{R}_{.j}*) - N_j \sigma_{B_{ij}R_{ij}}$$

The covariance term results from the multiplication of B_i by R_i* to obtain the imputed change. The covariance term is relevant to the bias in the following way. If a reporting cell coincided with an adjustment cell, then drawing C_i* at random from the donors would yield an expected value of C_i* equal to the true mean change in that reporting cell, and the imputed change in the reporting cell would be unbiased. Drawing, instead, R_i* from the donors will yield an expected value of R_i* equal to the true mean proportion, but the covariance of B_i and R_i* is constrained to an expectation of zero by the random draw. Unless the covariance between B_i and true R_i in the reporting cell is also zero, the imputed change will be biased, with the bias having an expected value equal in magnitude to the true covariance.

Adding one or more covariates to the adjustment cell definition can introduce covariance between B_i and R_i* at the reporting cell level and thus reduce this component of bias. Referring back to Figure 1, the imputations will now incorporate the covariance between the adjustment cell (column) mean change ratios and the corresponding mean original amounts for each reporting cell. In 1981 and 1982 the covariance generated by the inclusion of the pattern variable is not likely to be very large, however.

MODIFICATIONS INTRODUCED

Following the last use of the imputation procedures in 1982, SOI staff identified several desired improvements. Taking into account these suggestions as well as our own assessment of the strengths and weaknesses of the earlier imputation procedures, Mathematica Policy Research (MPR) has developed and coded a new imputation program. In addition to a number of features that increase the flexibility of the imputation process (e.g., in allowing the use of prior year donors), the new program incorporates several innovations that we believe promise significant reductions in MSE for estimates at the detailed reporting cell level. These modifications build upon observations presented in the preceding section. The key changes are:

(1) elimination of "pattern" from the adjustment cell definition

(2) separation of the imputation of change/no change from the imputation of the conditional amount of change

(3) expansion of the number of adjustment cell industries to 23

(4) introduction of a new covariate

(5) imputation of change/no change from a probability matrix with smoothed cell values

(6) matching on the original amount when imputing the conditional amount of change

(7) application of consistency tests within the imputation program

These changes are discussed below.

Elimination of the "pattern" variable. The earlier pattern variable is no longer relevant because IRS is limiting imputation to three of the former seven schedules. Rather than imputing all three from the same return, we are separating the imputation of Other Income from that of the pair, Other Deductions and Cost of Goods. Independent imputation of the two sets of items allows us to specify set-specific adjustment cells. For Other Deductions and Cost of Goods, the adjustment cell specification will distinguish returns with nonzero amounts in only one or both items. This corrects what was actually an oversight in the design of the earlier imputation system, which allowed a donor with edit information on Other Deductions but not Cost of Goods to be matched to a return requiring imputation to both items (in which case no changes would be imputed to Cost of Goods). Failure to match on the presence of both items produced a downward bias in the imputation of changes to both items--particularly to Cost of Goods [3].

Separate imputation of change and amount. Among the subsampled returns, edits of the fields in question frequently produce no changes. Consequently, the relevant donors for the imputation of change amounts are a fraction of the full donor subsample. Furthermore, there is evidence that the occurrence of a change and the magnitude of the change display different degrees of covariation with size, industrial class, and other candidates for inclusion in the adjustment cell definition. In short, an adjustment cell specification that produces good imputations of change versus no change (i.e., whether R^* is nonzero or zero) may not produce particularly good imputations of amounts (i.e., the nonzero value of R^*). Separating the two steps of the imputation process gives us the flexibility to address the two problems in potentially quite different ways.

Expansion of the industrial classes. As we noted above, IRS collapsed the 185 minor industries to 10 classes in defining the adjustment cells in the earlier imputation program. These classes varied substantially in the number of firms they encompassed and therefore the number of returns selected as donors. Since the most direct way to attempt improvement in the aggregate estimates for detailed reporting cells is to increase the degree of correspondence between the adjustment cells and reporting cells, this approach merited investigation. The large size of some of the 10 industry classes permitted further disaggregation without generating excessively small donor samples. Based on analyses with the 1982 data we were able to identify subclasses with differential edit outcomes.

Introduction of a new covariate. As we demonstrated earlier, adding a third covariate to the adjustment cell definition provides a means to reduce both the bias and variance of the imputation procedure. A strong candidate was available in the variable IRS uses to determine which returns outside of the largest size classes should be edited with certainty rather than left for possible imputation. For Other Income this variable is the ratio of Other Income to Total Income. This ratio covaries with the probability that editing Other Income will produce a change. Currently IRS edits the Other Income field if the value of this ratio exceeds an industry-specific level (a separate determination is made for Other Deductions and Cost of Goods). It seemed logical to extend the use of this ratio to the definition of adjustment cells, and we confirmed that the ratio continues to predict the probability of change among the returns subsampled for use as donors. Only for Other Income, however, did the ratio also appear to predict the conditional magnitude of change.

Imputation from a probability matrix. Separating the imputation of change from the imputation of amounts makes it possible to impute changes from a probability matrix, as there are only two alternative outcomes to be imputed. With a probability matrix we are no longer forced to rely on the collapsing of industry, size and other categories to achieved desired sample sizes for adjustment cells. Instead, we can smooth the observed probabilities in small cells to increase the effective sample sizes in these cells. In the 1985 imputations we are employing a simple, model-based smoothing algorithm to obtain the probabilities required for the first stage prediction of edit outcomes: namely, whether or not a change is to be made to the recorded values. This makes it possible to reduce the imputation bias at small levels of aggregation by maintaining a large number of adjustment cells.

Matching on original amounts. As was explained above, the random pairing of donors and imputes within adjustment cells is a potential source of bias, in that it does not account for any covariation that may exist (within the adjustment cell) between the original amount and the change ratio. In addition to its effect on imputation bias, this random pairing creates a possibility that a donor with a proportionately large change from a small original amount could be matched to a return with a very large original amount, yielding an imputed change much greater than an editor would be likely to make. To avoid this likelihood and to ensure that the imputed changes are consistent with the distribution of observed changes, we incorporated a suggestion from the IRS staff that when imputing magnitudes of changes each return be matched to the donor with the closest original amount. This "nearest neighbor" match is being carried out within adjustment cells defined more broadly than those used to impute change versus no change, as the sample base is smaller. Furthermore, the cells are to be collapsed as necessary to provide a specified minimum cell size. This modification also addresses the problem of the potential underestimation of the covariance between the original amounts and the imputed changes.

Application of consistency tests. Returns passed through the imputation procedures must be subjected to consistency tests because the imputed changes could conceivably produce violations. Subsequent corrections obscure the imputation "audit trail" created by the imputation program and may generate new but untested inconsistencies. Such masking of the audit trail interferes with the subsequent evaluation of the imputation results. To reduce such risks we have incorporated into the imputation software all of the consistency tests and corrections applicable to the imputed fields. We have also included indicators to denote where corrections were made to imputed values in order to satisfy particular tests. Evaluation of the frequency of such corrections will facilitate future improvement of the imputation and automatic editing routines.

DISCUSSION: APPLICATIONS TO OTHER DATA

The procedures employed by IRS to correct assorted items in the corporation SOI data base--employing a combination of double sampling and imputation--are applicable to other settings where the raw data are known to be biased but where the collection of supplementary information is prohibitively expensive or otherwise not possible. The many federal surveys and administrative record extractions provide numerous examples of the production and publication of statistics that the producers and users acknowledge to be biased. For example, household income is significantly underreported in the most widely used income surveys, yet editing and imputation procedures have been limited, generally, to correcting for nonresponse. (The Census Bureau's research on the undercounting of population in the decennial census provides a notable exception, although the Bureau has not yet developed a method it is willing to use to adjust population estimates below the national level.)

The procedures discussed here and the results that IRS has obtained with them suggest a possible approach to correcting for underreporting of amounts in surveys and administrative data. The procedures involve collecting additional information for a strategically selected subsample of respondents or records, and then extrapolating from this subsample information to the full sample by means of imputation to individual records. The strategic sampling procedures would appear to have particular merit for developing improved aggregate estimates of quantities with skewed distributions, since most of the unreported information is attributable to an identifiable, small portion of the sample.

Expanded applications of these and other techniques for correcting survey and administrative estimates for bias may first require significant innovations in methods of obtaining unreported information from individual units or estimating the magnitude of bias more broadly. In many cases the difficulties associated with such measurement are enormous. Where it can be accomplished, however, the techniques discussed in this paper provide a means to make use of whatever information can be collected.

ACKNOWLEDGMENTS

The author wishes to thank Roderick Little and Donald Rubin for their collaboration on the project under which the research reported herein was conducted; the SOI Division for its support of this work, and especially Susan Hinkins and Fritz Scheuren for their extensive contributions to my understanding of the earlier procedures and the need for modifications; and Patrice Turner and Lena Cunningham for their preparation of the final manuscript.

REFERENCES

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

[2] 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.

[3] Czajka, J. (1986), Imputation of Selected Items in Corporate Tax Data: Improving upon the Earlier Hot Deck, American Statistical Association, Proceedings of the Section on Survey Research Methods.

[4] Hinkins, S., and F. Scheuren (1986), Hot Deck Imputation Procedure Applied to a Double Sampling Design, Survey Methodology, Vol. 12, No. 2 (December), pp. 181-196.

[5] Internal Revenue Service (1987), Source Book, Statistics of Income, Active Corporation Income Tax Returns, July 1984-June 1985, Publication 1053, Department of the Treasury, U.S. Government Printing Office, Washington, D.C.