

BUREAU OF THE CENSUS
STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR-90/08

AN EVALUATION OF IMPUTED DATA
FOR NON-RESPONDENTS TO THE 1987
ECONOMIC CENSUSES-SINGLE-UNIT
ESTABLISHMENTS

by

Leroy Bailey, A. Elizabeth Jansto, Charlene Smith
Statistical Research Division
U.S. Bureau of the Census
Washington, D.C. 20233

This series contains research reports, written by or in cooperation with staff members of the Statistical Research Division, whose content may be of interest to the general statistical research community. The views reflected in these reports are not necessarily those of the Census Bureau nor do they necessarily represent Census Bureau statistical policy or practice. Inquiries may be addressed to the author(s) or the SRD Report Series Coordinator, Statistical Research Division, Bureau of the Census, Washington, D.C. 20233.

Report issued: April 24, 1991

An Evaluation of Imputed Data for Nonrespondents to the 1987 Economic Censuses-Single-Unit Establishments

Leroy Bailey, A. Elizabeth Jansto, Charlene Smith
U.S. Bureau of the Census

1. Introduction and Summary

The quinquennial censuses of business establishments conducted by the Bureau of the Census are subject to nonresponse, for which imputation procedures are used to compensate for the missing data and possibly reduce related biases. For this report imputation refers to an estimation procedure leading to the derivation of values for missing census data for nonresponding business establishments. The compensatory procedures that are used in the censuses for nonresponse at the establishment level vary according to the census and item under consideration. However, they usually involve the direct use of administrative data for missing census items or imputation based on previous census and survey data and estimates of plausible period-to-period changes in the activity of the establishments. Administrative data refer to data for the designated census items that have been compiled from Internal Revenue Service (IRS) and Social Security Administration (SSA) files and stored in the Standard Statistical Establishment List (SSEL) database maintained by the Census Bureau's Economic Programming Division. The imputed data cited in the report are derived from the SSEL, but are revised, if necessary, and maintained by the Business Division. These data reflect the results of imputation for item and establishment nonresponse and the editing or correction of data that were considered erroneously reported. The values from the file containing the imputed data are used for census tabulations.

The Evaluation of Imputed Data for Nonrespondents to the 1987 Economic Censuses (EID) was designed to assess the accuracy of imputed census data from selected standard industrial classification (SIC) groups within the Censuses of Wholesale Trade, Retail Trade and Service Industries. A study with the same principal objective as that of the EID was conducted for the 1977 Censuses of Wholesale Trade, Retail Trade and Service Industries (Dyke, 1984). Although the report concluded that at the trade area level there were statistically significant differences between the reported and the imputed data for most of the survey items, potentially

sizeable response errors diminished the credibility of those results. The timely follow-up of edit failures substantially reduced the potential for the occurrence of such errors in the EID. King and Trager, (1980), presented the results of an evaluation of administrative data for below cutoff establishments in the 1977 Retail and Service Censuses. The overall results of this evaluation indicated a general pattern in which the reported values from the sample of below cutoff cases were generally larger than the administrative data.

For the EID, the four primary items of interest were first quarter employment (the number of paid employees on March 12, 1987), first quarter payroll, annual payroll, and sales and receipts (revenue for tax-exempt establishments). Secondary objectives of the study included efforts to (1) facilitate the identification of misclassification problems among the nonrespondents encountered in the designated trade areas and (2) contribute to the further development of census imputation methodology.

Planning for the project began late in 1987. A sample of nonrespondent establishments was selected from the designated censuses; during the period extending from November, 1988 through July, 1989, efforts were made to interview them by telephone regarding key census items. The resultant data were used as a standard by which the imputed data for nonrespondents to the census were compared.

From the outset of the study there was concern about limitations on the utility of the results in the absence of plausible explanations for disparity between data reported in the follow-up study and the corresponding imputed data. Consequently, a small number of supplemental questions were also asked during the telephone interviews. These questions related to (1) the compatibility of the reference periods for the census data and the administrative data, (2) the process by which census forms are completed and returned; (3) perceived complexity of census forms; and (4) the comprehension of the meaning of terms used for the census items under investigation. However, response to all but the question on reference periods was not sufficient to conduct any useful analyses. Consequently there were a small number of survey outliers for which differences in the reported and imputed values for census items remained unexplained.

Table 1.1, presented below, provides single-unit imputation correction ratios for each trade area; that is, the estimates for the ratio of reported to imputed totals for the four major items. The imputed totals are from the Business Detail File (the values used in the census tabulations).

The reported totals are calculated by summing over the entire data set of survey respondents and imputes for the EID nonrespondents. The ratios in the table are the result of averaging the imputation correction ratios from five complete data sets consisting of the EID respondents and five separate imputations for the EID nonrespondents. A description of the imputation procedures is presented in section 5.3.

Ratios for the table were calculated using the SIC codes reported by the establishments. Standard deviations are given in italics in the lower right hand corner of each box, with p-values in the upper right hand corner based on the t-tests of the hypothesis that the reported and imputed data are equal for the nonrespondent population.

Table 1.1 Imputation Correction Ratios
- Trade area level

	Wholesale	Retail	Services
First Quarter Employment	0.985 <i>0.69</i> <i>0.035</i>	1.016 <i>0.73</i> <i>0.044</i>	1.095 <i>0.23</i> <i>0.069</i>
First Quarter Payroll	0.992 <i>0.89</i> <i>0.055</i>	1.077 <i>0.26</i> <i>0.067</i>	1.233 <i>0.02</i> <i>0.092</i>
Annual Payroll	1.015 <i>0.59</i> <i>0.027</i>	1.159 <i>0.12</i> <i>0.092</i>	1.261 <i>0.04</i> <i>0.122</i>
Sales and Receipts	1.075 <i>0.42</i> <i>0.083</i>	1.141 <i>0.18</i> <i>0.103</i>	1.154 <i>0.06</i> <i>0.074</i>

Overall, the ratios indicate that census imputes are reasonably accurate, with only three of the twelve trade area ratios differing from 1.00 at a 10% significance level, as indicated by p-values that are less than 0.1. However, it is interesting to note that all three of the significant ratios are in the service are, which seemingly was not just a change occurrence. Moreover, we can observe that ten of the trade area level ratios and three-fourths of the SIC level ratios are greater than 1, suggesting a tendency toward underimputation in the censuses.

2. Sample Selection

A stratified systematic sample of about 6,000 establishments, roughly equally divided between single and multi-unit establishments, was selected to facilitate the objectives of the study. The principal stratification was based on type of unit (single or multi), trade area and SIC code. The 1987 SIC codes on which the evaluation is based are at the three - digit level for wholesale establishments and the two - digit level for retail trade and services. In addition, the SIC groups were stratified by establishment payroll categories to ensure that the precision of the desired estimates was within an acceptable range.

The sampling frame for the evaluation study was the set of approximately 340,290 mail nonrespondents to the censuses as of August 31, 1988. In September, 1988 the Economic Programming Division provided us with an abbreviated version of the economic censuses database containing only key identifying variables, from which a sample was selected from the designated nonrespondents. Establishment records were deleted from the file if they were associated with establishments that had not been mailed a questionnaire, did not belong to any of the three trade areas selected for the study, or were for inactive businesses. To ensure that the desired sample sizes were achieved for the smallest unit of analysis, the SIC group, adjustments were made to the respective sampling rates to allow for these contingencies and for the possibility of late respondents to the censuses. A listing of census file numbers (CFN's) for the establishments that responded after the cutoff date was provided by Business Division. This list was compared with the list of sampled CFN's and all matches (indicating a census respondent in the sample) were removed. The remaining establishments constituted the evaluation sample. The portion of their records containing the variables required for the study were stripped from the database, forming the initial administrative data file.

The remaining discussion of this report will focus on the collection, processing, and analysis of data from the single-unit portion of the evaluation sample.

3. Data Collection

Establishments in the sample for the EID were contacted initially by mail, reminded of the nonreceipt of their census form, and informed that they would be contacted by telephone and

asked to respond to questions relating to several census items. The telephone interviews began on an average of five to ten working days after the introductory letters were mailed. Wording for the questions asked during the telephone interview followed that of the mailed census questionnaires whenever possible. A copy of one of the five interview forms is included in Appendix A. All interviews were conducted from the Census Bureau's headquarters by statistical assistants from the Statistical Research Division and a temporary staff. From the administrative data file, labels were computer generated, containing the establishment name, address, census file number, and the survey control number. These labels were affixed to the questionnaire and call record forms. After the requested data were collected from the specified establishments, transcription reviews and preliminary edit checks were performed. Data which passed these checks were then compared to the corresponding administrative data in lieu of the then inaccessible imputed census data. Establishments for which the ratio of the data reported in the EID to the administrative data fell outside of the range 0.5-2.0 were identified and investigated promptly. Large discrepancies that could not be ascribed to data processing procedures, or otherwise explained, were noted and reconciliation interviews were conducted. These interviews were handled by the interviewing coordinator. Establishments were told that the collected data were being checked for consistency and possible interviewer recording error, and they were given the opportunity to confirm or revise the reported data in light of the perceived inconsistencies. If an establishment insisted that the original information was correct, the interviewing coordinator concluded the interview and the data were left unchanged.

Like any survey of comparable size and complexity, the EID was subject to measurement errors relating to nonresponse, the accuracy of respondent reporting and the job performance of interviewers and other survey personnel. In an effort to preclude or rectify and reduce the effects of any interviewer behavior that might have compromised the quality of the survey data, the work of the telephone interviewers was reviewed regularly. All interviewers underwent a three day training program on the survey questionnaires and were provided with an interviewer reference manual. Moreover, extensive administrative data and interviewer performance statistics (Appendix B) were collected weekly to monitor the progress of the data collection activities. Table B1 present distributions of interview outcomes. Note that over all trade areas about 69 percent of the single-unit interview attempts resulted in completed interviews or gleaned enough

principal item data to be termed partial interviews. About 11 percent of the establishments were refusals. Desired data were not available for roughly three percent of the cases due to problems relating to administrative matters, such as changes in ownership of businesses, establishment mergers or dissolutions and inept record keeping. We were unable to obtain correct telephone numbers or addresses from another 16 percent of the establishments. A sizable portion of these cases were thought to have gone out of business, while others might have relocated or effected changes in their organizational structures and/or operations that encumbered efforts to contact them using only the information that was available for the evaluation. The interview outcome distributions for the individual trade areas followed a pattern similar to that observed for all respondents. However, the response rate for wholesale establishments was a little larger than the overall average, while the rates for services and retail sales were somewhat lower.

Regarding other operational matters, we can observe that although a substantial percentage of our successful interviews were conducted within ten minutes and required less than five calls, it took three weeks or more to make productive contact with about 40 percent of the sample (Tables B2-B4). Moreover, about 10 percent of the sample required a reconciliation interview for at least one survey item.

4. Data Processing

Following the transcription review, preliminary edit checks and any required reconciliation interviews, the resultant data were coded, keyed and verified. This reported data file then underwent a computer edit before other data adjustments were made. All edit failures were checked for transcription errors and corrections were made to the file as required. This preliminary edit included the following:

- The reported SIC codes were compared to the SIC group from the administrative data file. When discrepancies were encountered, the accuracy of the coding and keying of the reported SIC codes was verified.
- Sales values for wholesale broker/agent establishments were adjusted for possible commission value instead of gross sales, similar to the procedure used during the census. Sales were compared to annual payroll for wholesale establishments, with a flag indicating broker/agent. If the

resulting ratio was less than 10, the reported sales value was inflated by an SIC commission rate factor.

- The payroll and sales and receipts items for the respective establishments were compared for inconsistencies, such as annual payroll entries that were less than the corresponding first quarter payroll and sales less than either payroll item.
- The sales and revenue variables were compared to insure that there was only one receipts entry per establishment.
- Reference period codes were compared with reference period dates for consistency.

• Following the computer editing of the data file containing the principal survey items (reported data file), editing procedures were performed on the interview monitoring file, which contained the operational statistics on the data collection. These editing procedures included a review of the starting and ending interview times reported by the interviewers to identify recording errors leading to interviews of unreasonable length. At the completion of this editing, the establishment reported data file and the interview monitoring file were merged with a file containing the corresponding imputed data, the Business Detail File (BDF). This merge comprised the survey analysis file.

5. Estimation

5.1 Notation

For any census item of interest, Y_{hij} will denote its value for the j th census nonrespondent of the i th payroll category and h th SIC group. Let π_{hij} be the selection probability for the establishment with this value. Y_T will be the generic representation for the respective trade area totals for the item, that is for retail trade, wholesale trade and service industries. The value of Y_{hij} reported in the census follow-up study and the corresponding census impute will be given by $Y_{hij}^{(r)}$ and $Y_{hij}^{(e)}$, respectively. The number of payroll categories associated with the h th SIC

group will be given by M_h ; the number of nonrespondent establishments belonging to the i th payroll group will be denoted by N_{hi} , with the corresponding EID sample size given by n_{hi} .

5.2 Ratio Estimators

The major objective of the follow-up study of nonrespondents in the designated census areas was to effect comparisons between item totals based on data reported during the follow-up and corresponding totals derived from census imputes. The population value associated with the desired comparison, denoted by R , is the ratio of the reported total to the corresponding total based on imputation. This ratio will be referred to subsequently as the "imputation correction ratio". Its estimator at the SIC and trade area levels are:

$$\hat{R}_h = \frac{\sum_{i=1}^{M_h} \sum_{j=1}^{n_{hi}} y_{hij}^{(r)} \pi_{hij}^{-1}}{\sum_{i=1}^{M_h} \sum_{j=1}^{n_{hi}} y_{hij}^{(c)} \pi_{hij}^{-1}}$$

$$= \frac{\hat{Y}_h^{(r)}}{\hat{Y}_h^{(c)}} \quad (5.2.1)$$

and

$$\hat{R}_T = \frac{\sum_{h \in T} \hat{Y}_h^{(r)}}{\sum_{h \in T} \hat{Y}_h^{(c)}} \quad (5.2.2)$$

where the $y_{hij}^{(r)}$ and $y_{hij}^{(c)}$ are sample measurements from establishments in the SIC's that comprise the given trade area.

For estimates of the fraction of an SIC's establishments that are misclassified, the

following estimator is used.

$$\hat{\theta} = \frac{\sum_{i=1}^{M_h} \sum_{j=1}^{n_{hi}} x_{hij} \pi_{hij}^{-1}}{\sum_{i=1}^{M_h} \sum_{j=1}^{n_{hi}} z_{hij} \pi_{hij}^{-1}} = \frac{\hat{X}_h}{\hat{Z}_h} \quad (5.2.3)$$

where $x_{hij} = \begin{cases} 0 & \text{if establishment is classified correctly} \\ 1 & \text{if establishment is misclassified} \end{cases}$

and $z_{hij} = 1$.

- A discussion of the variance estimators is included in Appendix C.

5.3 Imputation Procedures

Nonresponse to the EID survey occurred at both the item and establishment level. Two different methods of imputation were used to handle the two types of missing data. For cases where item nonresponse occurred and an appropriate set of respondents could be fitted to linear regression models, those models were used to derive imputes for the item or partial nonrespondents. For the other item nonresponse and establishment nonresponse, a multiple imputation procedure was utilized. These methods are discussed in more detail below.

5.3.1 Regression Imputation Procedure

In order to develop the regression equations to impute for item nonresponse, partial respondents were classified by their response pattern over the four major survey items. Table 5.1 summarizes these cases by response pattern, with 1 indicating a response to the item and 0 indicating no response. Partial and complete responses were partitioned by SIC group. Then for each missing item in each response pattern, the complete respondents within each SIC group were

used to estimate parameters for "the best" regression equations based on all combinations of reported items within that response pattern. For example, for response pattern 1100, complete respondents by SIC group were used to develop two sets of equations, the first to estimate annual payroll and the second to estimate sales. The resultant equations were linear functions of first quarter employment, first quarter payroll or both items. Using the derived regression equations, values for the missing items were imputed. Missing items for which a viable regression equation did not exist were combined with establishment nonresponse cases by item and imputed for through the multiple imputation procedure.

Table 5.1 Summary of Partial Respondents by Response Pattern

Response Pattern				Number of Cases
Emp	QPI	AP	S/R	
0	0	0	1	1
0	0	1	1	3
0	1	0	1	1
0	1	1	1	14
1	0	0	0	38
1	0	0	1	2
1	0	1	0	2
1	0	1	1	12
1	1	0	0	3
1	1	0	1	3
1	1	1	0	22
Total				101

Key: Emp = Employment AP = Annual Payroll
 QPI = First Quarter Payroll S/R = Sales or Revenue

5.3.2 Multiple Imputation Procedure

Multiple imputation is a method in which a single imputation procedure is repeated a specified number of times, generating multiple sets of complete data, each with its own survey estimates. These estimates can then be combined to provide a single estimate for each item of interest. The process is designed to minimize the potential effects of any one impute and to reflect the uncertainty in the imputation procedure. Details on the multiple imputation procedure used for the EID are discussed in Appendix D.

After a value was imputed, through a hot deck procedure, for each missing value encountered in the evaluation, estimates for the imputation correction ratios at the SIC group and trade area level were generated as detailed in Section 5.2. This process was repeated five times and the resulting set of estimates were averaged into one set of imputation correction ratio estimates, presented in Table 6.1.

6. Analysis

6.1 Ratio Analysis

Initially we note that although the imputation correction ratios presented in this section were computed after the application of the EID imputation procedures, the ratios and general results are very similar to those without those data adjustments.

Tables 6.1-6.3 present the imputation correction ratios at the trade area and SIC group levels. At the trade area level, imputation correction ratios for service items are among the most extreme (all three significant ratios are for service items), while the ratios for wholesale items exhibit the least deviation from 1. At the SIC group level, for each item ratios for all three trade areas vary tremendously between groups. SIC group level ratios for the sales item within wholesale span the greatest range of values, varying from 0.776 to 2.206. The second greatest range of ratio values occurs in the annual payroll item for services, with ratios ranging from 1.042 to 1.560. Although services and wholesale trade both exhibit large ranges of SIC level ratios for all four items, the ranges for wholesale ratio are centered around 1, while the centers of the ranges for service ratios are greater than 1. Of the 25 SIC group level ratios less than 1,

seventeen are within wholesale trade, six are within retail trade, and only two are within services. These differences are evident in the ratio values at the trade area level.

A discussion of the imputation correction ratios by trade area follows Tables 6.1-6.3.

Table 6.1 Imputation Correction Ratios
- Wholesale

	Employment		First Quarter Payroll		Annual Payroll		Sales/Receipts	
	Ratio	Std Error	Ratio	Std Error	Ratio	Std Error	Ratio	Std Error
WHOLESALE	0.985	0.035	0.992	0.055	1.015	0.027	1.075	0.083
501 - Motor Vehicles/Parts	1.075	0.052	1.057	0.106	1.036	0.073	1.037	0.130
502 - Home Furnishings	1.036	0.060	1.067	0.063	1.011	0.024	0.899	0.064
503&505 - Lumber & Metals	1.026	0.029	1.159	0.156	1.069	0.217	2.206	1.199
504&508 - Equip./Machinery	1.045	0.098	0.845	0.152	0.985	0.063	1.072	0.142
506&507 - Electrical/Hardware	0.916	0.116	1.056	0.048	1.043	0.029	0.776	0.164
509 - Misc. Durable Goods	0.989	0.167	1.018	0.119	1.034	0.040	1.175	0.107
511&512 - Paper/Drugs	0.920	0.606	1.151	0.136	1.106	0.068	1.077	0.059
513 - Apparel/Piece Goods	1.070	0.042	1.173	0.060	1.130	0.078	1.072	00.91
514&518 - Groc./Beer & Wine	0.858	0.065	0.920	0.057	0.938	0.062	0.989	0.136
515&519 - Farm/Misc. Nondurable	0.979	0.031	1.192	0.113	0.984	0.050	0.978	0.068
516&517 - Chem./Petro	0.914	0.102	0.939	0.056	0.989	0.057	1.144	0.149

Table 6.2 Imputation Correction Ratios
- Retail

	Employment		First Quarter Payroll		Annual Payroll		Sales/Receipts	
	Ratio	Std Error	Ratio	Std Error	Ratio	Std Error	Ratio	Std Error
RETAIL	1.016	0.044	1.077	0.067	1.159	0.094	1.141	0.103
52 - Hardware Materials	0.975	0.040	1.094	0.057	1.084	0.044	1.037	0.062
53&56 - Merchandise	1.041	0.077	1.074	0.053	1.007	0.056	0.836	0.159
54 - Food Stores	0.973	0.068	1.125	0.158	1.185	0.133	1.335	0.236
55 - Car Dealers/Gas Stations	1.049	0.070	1.100	0.043	1.083	0.109	1.083	0.059
57 - Home Furnishings	0.998	0.061	1.034	0.061	1.303	0.250	1.336	0.385
58 - Eating & Drinking Places	1.041	0.085	1.062	0.137	1.204	0.230	0.989	0.057
59 - Misc. Retail	0.922	0.048	1.055	0.086	1.139	0.083	1.302	0.384

Table 6.3 Imputation Correction Ratios
- Services

	Employment		First Quarter Payroll		Annual Report		Sales/Receipts	
	Ratio	Std Error	Ratio	Std Error	Ratio	Std Error	Ratio	Std Error
SERVICES	1.095	0.069	1.233	0.092	1.261	0.123	1.154	0.074
70 - Hotels/Motels	1.044	0.078	1.239	0.147	1.224	0.121	1.158	0.123
72 - Personal Services	1.018	0.085	1.160	0.099	1.125	0.103	1.163	0.075
73 - Business Services	1.246	0.256	1.490	0.334	1.560	0.347	1.194	0.133
75 - Automotive Services	1.098	0.107	1.156	0.141	1.131	0.090	1.249	0.153
76 - Misc Repair Services	1.160	0.086	1.066	0.083	1.072	0.044	1.110	0.124
78,79&84 - Recreation Services	1.027	0.099	1.025	0.116	1.092	0.110	1.145	0.296
81 - Legal Services	0.972	0.045	1.143	0.070	1.052	0.065	0.970	0.103
82&83 - Educational & Social Services	1.054	0.056	1.036	0.027	1.079	0.041	1.106	0.118
86 - Membership Organizations	1.007	0.055	1.056	0.044	1.042	0.049	1.224	0.097

Wholesale:

Item	Ratio	Standard deviation	p-value
Employment	0.985	0.035	0.69
First Quarter Payroll	0.992	0.055	0.89
Annual Payroll	1.015	0.027	0.59
Sales/Receipts	1.075	0.083	0.42

At the trade area level, the imputation correction ratios for the four items are not significantly different from 1. As illustrated in Appendix E, for the two payroll items, approximately half of all respondent establishments exhibit an individual ratio between 0.99 and

1.01. The individual ratios for employment are distributed more evenly over the three categories, with 37% less than 0.99, 32% between 0.99 and 1.01, and 31% greater than 1.01. The establishment level sales ratios are the most disperse, with only 25% falling between 0.99 and 1.01. Overall, the percentage of individual ratios less than 0.99 and greater than 1.01 are closer (indicating a more symmetrical distribution) than for retail or services. These data indicate that, at this level of aggregation, census imputes are reasonable approximations for wholesale nonrespondents.

At the SIC group level, imputation correction ratios are more varied, ranging from 0.776 to 2.206. However, with larger standard errors for the more extreme ratios, only three of the 44 ratios are significantly different from 1 at the 10% level. These three are the employment ratio for SIC group 514&518, groceries, beer and wine, with a ratio value of 0.858, and the first quarter payroll ratios for SIC group 513, apparel, piece goods and notions, and SIC group 515&519, farm goods and miscellaneous nondurable goods, with ratio values of 1.173 and 1.192, respectively. The first quarter payroll ratio for SIC group 513 is influenced by five outliers, three of which have an imputed value of \$0 and reported values ranging from less than \$500 to over \$100,000. An establishment ratio was considered to be an outlier if the individual ratio value was greater than 7.5 or less than 1/7.5. These cutoffs were determined by observing the natural breaks in the data over all trade areas. Follow-up investigations of the outlier establishments in SIC group 513 mentioned above provided no definitive evidence of erroneous reporting of the EII data or in its processing. In general, the stability of the trade area level payroll ratios is evident at the SIC group level as well; the percentage of individual ratios between 0.99 and 1.01 is between 40% and 70% for all SIC groups. The two most extreme ratio values are for sales, contributing to the greater standard error for that item at the trade area level. Across all SIC groups, only 17% to 33% of individual sales ratios fall between 0.99 and 1.01. Overall, with 39% of SIC level imputation correction ratios between 0.95 and 1.05, and 70% between 0.9 and 1.1, the data indicate that census imputes are reasonably good approximations for wholesale nonrespondents at the SIC group level. However, for SIC groups and items with more extreme ratios, particularly for the three mentioned above, the observance of their imputation patterns for future censuses and surveys may prove useful in effecting imputation improvements.

Retail:

Item	Ratio	Standard deviation	p-value
Employment	1.016	0.044	0.73
First Quarter Payroll	1.077	0.067	0.26
Annual Payroll	1.159	0.091	0.12
Sales/Receipts	1.141	0.103	0.18

Similar to wholesale, the imputation correction ratios for the four retail items are not significantly different from 1 at the trade area level for a 10% significance level. However, for payroll and sales items note the dramatic reduction in the p-values, where the ratios are considered significantly different from 1 at the 12 and 18 percent significance levels, respectively. For both payroll items, 42-44% of respondent establishments exhibit an individual ratio between 0.99 and 1.01. The individual ratios for sales are distributed roughly evenly over the three categories, with 36% less than 0.99, 30% between 0.99 and 1.01, and 34% greater than 1.01. For employment, the 40% of respondent establishments with individual ratios less than 0.99 are counterbalanced by six outliers with an imputed value of 0 and reported values ranging from less than 5 to about 500. Seemingly the census imputation procedures were quite appropriate at this level of aggregation for nonrespondents to the retail items; although, as noted above, these ratios are less stable than those for wholesale.

Again, at the SIC group level, imputation correction ratios are more varied for retail items, ranging from 0.836 to 1.336. Only six of 28 ratios for retail SIC groups are less than 1, compared with 17 out of 44 for wholesale. None of the imputation correction ratios for either payroll item are less than 1. Despite the range in ratio values, only two are considered significantly different from 1. One is the first quarter payroll ratio for SIC 55, car dealers and gasoline stations, with a ratio value of 1.084. The distribution of individual first quarter payroll ratios for SIC 55 is skewed, with 17% less than 0.99 and 32% greater than 1.01. This ratio is further influenced by two outliers, each with an imputed value of \$0 and reported values of about \$30,000 and more than \$2,000,000. Similar to wholesale, the two most extreme retail ratios are for sales items, contributing to the high standard deviation for the trade area level sales ratio. Overall, 36% of retail SIC level imputation correction ratios are between 0.95 and 1.05, and 68%

are between 0.90 and 1.10. However, further investigation is warranted for the groups and items with the more extreme ratio values, and in particular for the two significant ratios mentioned above.

Services:

Item	Ratio	Standard deviation	p-value
Employment	1.095	0.069	0.23
First Quarter Payroll	1.233	0.092	0.02
Annual Payroll	1.261	0.123	0.05
Sales/Receipts	1.154	0.074	0.06

At the trade area level, three of the four services imputation correction ratios are considered significantly different from 1 for at least the 10% significance level. For all four items, the imputation ratios are larger for services than for either wholesale or retail. For the annual payroll ratio, 32% of individual ratios are greater than 1.01 while only 25% are less than 0.99. For first quarter payroll and sales, the distribution of individual ratios is more evenly divided between those less than 0.99 and those greater than 1.01, but for sales, only 24% of the establishment level ratios are between 0.99 and 1.01. This greater dispersion contributes to the instability of the estimate. All four trade area level ratios are greater than 1, suggesting various levels of underestimation of the services nonrespondent values by the census imputes.

Similar to wholesale and retail, at the SIC group level, the service imputation correction ratios are more varied, ranging from 0.970 to 1.560. However, only two of 36 ratios are less than 1, reemphasizing the underimputation indicated at the trade area level. Contrary to wholesale and retail, the largest ratio is not in sales, but in annual payroll. Six of the SIC group level ratios are significantly different from 1, including the employment ratio for SIC group 76, miscellaneous repair services, with a ratio value of 1.160 and the first quarter payroll ratio for SIC group 81, legal services, with a ratio value of 1.143. Two annual payroll ratios are significant - SIC group 70, hotels and motels, with a ratio value of 1.224 and SIC group 82&83, educational and social services, with a ratio value of 1.079. The remaining two significant ratios are in sales - SIC group 72, personal services, with a ratio value of 1.163 and SIC group 86,

membership organizations, with a ratio value of 1.224. None of the significant ratios are strongly influenced by outliers. Overall, only 25% of SIC group level ratios fall between 0.95 and 1.05, and only 44% fall between 0.90 and 1.10. With the largest number of significant and extreme ratios, the service industries appear to require the greatest focus for future imputation improvements.

6.2 Source Flag Analysis

One of the more interesting analyses of the EID study has been the comparison of the distribution of establishments by imputation method for the outliers with the distribution of all establishments included in the sample. The census impute for each item is marked with a code indicating the method used for the impute. Table 6.4 below lists the imputation methods and codes.

Table 6.4 Imputation Methods by Code

Code	Source Description
A	Administrative data
C	Corrected by problem solving clerk
D	Derived from other reported data
H	Ratio imputation based on 1982 census data for same establishment
I	Ratio imputation based on 1982 industry averages (cold deck)
J	Ratio imputation based on 1987 industry averages (warm deck)
M	Midpoint of range that will pass all edit checks
P	Ratio imputation based on prior year (1986) administrative data
R	Reported
Z	Zero imputed from blank

The vast majority of imputation method codes are "A", indicating the use of the administrative value. However, for the outliers, the percentage of establishments imputed with administrative values is much less than for all sampled establishments. Imputing with a zero ("Z") or a ratio adjustment based on 1987 industry averages ("J") is far more prevalent among the outliers, suggesting that these imputation methods might warrant further investigation. In previous results it was observed that "sales and receipts" is the item for which the largest disparity is likely to occur between the reported and imputed data. Moreover, we can observe that for all cases the extent to which administrative data were used as census imputes for the sales and receipts item (48%) is far less than that for the other three items. This certainly suggests a relationship between deviation from the use of administrative data as imputes and the size of the imputation correction ratios. Table 6.5 presents the comparative distributions of source codes by item. The "All cases" category of the table excludes establishments selected for the EID that were subsequently indentified as census respondents and those cases whose trade area designations were other than the four included in the study.

Table 6.5 Comparative Distributions of Establishments
By Imputation Method and Item

Employment

Code	A	C	D	H	J	P	R	Z	m
All cases	2754	53	2	1	3	6	24	117	11
%	92.7	1.8	0.1	0.0	0.1	0.2	0.8	3.9	0.4
Outliers	13	1	0	0	0	2	0	22	2
%	32.5	2.5	0.0	0.0	0.0	5.0	0.0	55.0	5.0

First Quarter Payroll

Code	A	C	H	I	J	P	R	Z	m
All cases	2557	47	22	4	105	85	23	117	11
%	86.1	1.6	0.7	0.1	3.5	2.9	0.8	3.9	0.4
Outliers	12	1	1	0	7	5	4	15	2
%	25.5	2.1	2.1	0.0	14.9	10.6	8.5	31.9	4.2

Annual Payroll

Code	A	C	H	I	J	M	P	R	m
All cases	2537	55	23	3	161	5	149	27	11
%	85.4	1.9	0.7	0.1	5.4	0.2	5.0	0.9	0.4
Outliers	11	1	0	2	10	0	3	0	2
%	37.9	3.4	0.0	6.9	34.5	0.0	10.3	0.0	6.9

Sales and Receipts

Code	A	C	H	I	J	M	P	R	m
All cases	1426	200	201	21	624	2	459	27	10
%	48.0	6.7	6.7	0.7	21.0	0.0	15.5	0.9	0.3
Outliers	7	2	2	1	33	0	4	0	2
%	13.7	3.9	3.9	2.0	64.7	0.0	7.8	0.0	3.9

"m" indicates imputation code was missing

6.3 Misclassification Analysis

For the economic censuses, questionnaires were designed using the 1972/77 version of the SIC codes. After the conclusion of the data collection, these codes were updated to reflect the 1987 revision and this version was used for census publication purpose. At the time of sample selection for the EID, the 1972/77 version of the codes was the most current and was used in the selection of the establishments for the survey. Verbal descriptions of the establishment's principal line of merchandise or service provided, which were obtained during the interview, were coded to both the 1972/77 and the 1987 versions and verified manually. Any ambiguous descriptions were coded to the original SIC code present on the SSEL file and then translated (using the four digit SIC code on the SSEL) to the 1987 version. Descriptions were coded to three digit SIC codes for wholesale and to two digit codes for retail and services. An establishment was considered misclassified if the SIC code imputed by Business Division differed from the classification based on the reported data. Misclassification rates were generated for the evaluation sample and estimates of the misclassification rates were produced for the population of nonrespondents.

Both methods of comparison, by mailout basis (the 1972/77 version) or publication basis

(the 1987 version) of the SIC codes, generated comparable results. Table F1.1 presents the survey sample misclassification rates for the trade areas and SIC groups for the 1972/77 codes and Table F1.2 presents the misclassification rates for the 1987 codes. Estimates for the population misclassification rates were calculated using the sample selection weights and are present in Table F2. For the purposes of this discussion, the results from the comparison based on the 1987 SIC codes will be used.

At the trade area level, estimates of misclassification for the nonrespondent population varied between 6.2% for services and 20.1% for wholesale. At the SIC group level, estimated misclassification rates ranged from 0% to 30.4%. Each trade area had at least one SIC group with an estimated misclassification rate greater than 20%. Five of the eleven SIC groups in wholesale were included in that category. At the sample level, for the 32 establishments coded in the EID as '504/508' (professional equipment, machinery and supplies) that were misclassified, '509' (miscellaneous durable goods), was the most commonly imputed code, followed by '506/507' (electrical goods, hardware and plumbing). The most commonly imputed codes for misclassified establishments in group '57' (home furnishings and equipment) were '502' (wholesale furnishings) and '53/56' (general merchandise and apparel stores); 64% were imputed as either wholesale or service. Code '57' constitutes the primary imputed classification for establishments in '73' (business services) that were misclassified. A total of 48% of the misclassified establishments in group '73' were imputed as wholesale or retail. It is interesting to note that two-thirds of misclassified establishments from group '82/83' (educational and social services) were classified as '86' (membership organizations). Also of note is that legal services has a misclassification rate of 0. Tables F3.1-3.3 contain establishment counts cross tabulated by report and imputed SIC groups.

In the previous sections, all ratio estimates have been calculated and presented using the survey reported SIC code. It was of some concern that the presence of misclassification would influence the ratio values. To investigate this claim, imputation correction ratio estimates were calculated at both the SIC group and trade area levels using only "identicals", establishments whose imputed SIC code matched the SIC code reported in the survey. These ratio estimates were then compared with the ratio estimates based on all respondents, grouped by reported SIC code. Results at the trade area level are presented in Table 6.6 below.

Except for wholesale items, standard errors for ratios based on identicals are larger than those based on all respondents due to the smaller sample size. Comparisons at the SIC group level are similar, with a few exhibiting larger discrepancies. In general, the confounding effect of misclassification on the values of the ratio estimates appears to be minimal.

Table 6.6 Comparison of Ratio Estimates
Based on Identicals Only and
on All Respondents

	First Quarter Employment	First Quarter Payroll	Annual Payroll	Sales & Receipts
Wholesale Identicals	1.024	0.945	0.993	1.006
All respondents	1.027	0.975	1.014	1.089
Retail Identicals	1.046	1.056	1.054	1.092
All respondents	1.042	1.026	1.033	1.069
Service Identicals	1.174	1.206	1.129	1.097
All respondents	1.160	1.207	1.129	1.101

7. Conclusions

As stated earlier, the trade area imputation correction ratios are indications that at this level of aggregation, the imputation for the selected censuses is "reasonably" good for wholesale and retail trade. However, for three of the items included in the study, the trade area imputes for the service industries appear to underestimate the item value. In addition for retail trade, trade area estimates for annual payroll and sales and receipts differed from 1.00 at 12 and 18 percent significance levels, respectively.

As was expected, the discussion of tabular results and associated discussion of previous sections suggested a relationship between the use of imputation procedures, other than the

substitution of administrative data, and the size of the corresponding imputation ratio adjustments. Specifically, increased use of procedures based on adjustments to alternative data sources to produce census imputes tends to generally produce increases in the size of the estimated imputation correction ratios.

Relative to census misclassification, for the nonrespondents the trade area rates are between six and twenty percent; moreover, at the SIC level, the corresponding rates vary considerably. In order to effect a reduction in the level of misclassification among census nonrespondents, perhaps census follow-up interview attempts for which only data required to classify the establishment could be considered.

Regarding potential imputation biases relating to the differences in the effectiveness of the currently-used imputation procedures, there is a need to seek alternative adjustment methodology appropriate for census items for which administrative data are more likely to be unavailable. Both theoretical and empirical research in this area seem warranted. Moreover, it is suggested that census planners consider the estimation of correction ratios from nonrespondent studies similar to the EID, conducted as an integral part of the economic censuses.

APPENDIX A -- INTERVIEW FORM

**EVALUATION OF IMPUTED DATA FOR NONRESPONDENTS
TO THE 1987 CENSUS OF RETAIL TRADES**

Item 1 - Employer Identification *(verify and correct if necessary)*

Control File Number : _____
 Census File Number : _____
 Name of Business : _____
 Address : _____

 Contact : _____
 Phone Number : _____

FOR OFFICE
USE ONLY

Item 2 - Multi-unit vs Single-unit

Does the Census File Number above begin with a zero? If not, skip to item 3.

1. Is this business owned or controlled by a company or person who controls at least one other business? (1) yes - name of company or person _____
 address _____
 _____ zip _____
 phone # _____
Ask question 4.
 (0) no - Ask question 2.

2. Is your business located at two or more physical locations?
 (1) yes - Ask question 3. (0) no - Skip to item 3.

3. Do the separate locations of your business operate independently or are they sites for a central operating unit in the same metropolitan area (for example, collection sites for a dry cleaning plant)?
 (1) independently - Ask question 4.
 (2) sites - Skip to item 3.

4. What are the names, addresses, and principal activity or line of merchandise based on receipts for each location or business?

2.1
 2.2
 2.3
 2.4

Item 3 - Classification

1. Would you classify your business, based on majority of your sales and receipts, as a
 (1) retail trade? (2) wholesale trade? (3) service industry? *Skip to item 4.*
 (4) other? _____

2. What product line or service area constitutes the primary activity of your business, based on sales and receipts? _____ *Skip to item 5.*

3.1
 3.2

Item 4 - Kind of Business

1. Describe the principal line of merchandise of your business, based on your sales and receipts. _____

Building Materials, Hardware, Garden Supply, and Mobile Home Dealers (52) Apparel and Accessory Stores (56)
 General Merchandise Stores (53) Furniture, Home Furnishings, and Equipment Stores (57)
 Food Stores (54) Eating and Drinking Places (58)
 Automotive Dealers and Gasoline Service Stations (55) Miscellaneous Retail (59)
 Other (777) code = _____

4.1

Item 5 - Employment

1. How many paid employees were on your payroll during the pay period that included March 12, 1987? Include both full- and part-time employees. _____

Item 6 - Payroll and Receipts

(Value figures may be reported to the nearest dollar or rounded to thousands.)

- 1. What was your total payroll for the first quarter of 1987 (Jan - Mar)?
- 2. What was your total annual payroll in 1987, before deductions?
- 3. What was your total sales and operating receipts in 1987, excluding sales or other taxes collected and paid directly to a local, state, or Federal tax agency?
- 4. *Applicable for STE form only*
- 5. What is the approximate percentage of your 1987 sales to household consumers?

Mil.	Thous.	Dol.

- 6. (If SIC 52 or 777) What is the approximate percentage of your 1987 sales to construction contractors? _____
- 7. (If SIC 55 or 777) What percentage of your receipts are from the sale of merchandise, excluding the sale of parts and supplies associated with repair service conducted at your establishment? _____
- 8. Is your information for annual payroll and annual receipts based on a fiscal or a calendar year?

Annual Payroll	Annual Receipts
<input type="checkbox"/> (1) fiscal (give dates) _____	<input type="checkbox"/> (1) fiscal (give dates) _____
<input type="checkbox"/> (2) calendar (January 1 - December 31)	<input type="checkbox"/> (2) calendar (January 1 - December 31)

6.5 [] [] []

6.6 [] [] []

6.7 [] [] []

6.8 payroll []

receipts [] [] [] [] []

Item 7 - Nonresponse Status

1. We mentioned that we had not received a 1987 Census of Retail Trade form from you. We are trying to improve the form and could use your help. What difficulties did you have in completing it? _____

if none and no explanation - Did you mail it in then? _____

2. What is the position title or titles of the person(s) responsible for completing Census forms (for example, owner or accountant)? _____

Is this person also responsible for completing tax-related forms?

(1) yes

(0) no

(2) other _____

7.1 [] []

7.2 [] []

APPENDIX B -- INTERVIEW MONITORING RESULTS

Table B1. Distribution of Interview Outcomes
- All Trade Areas

Outcome Category	Number in Category	Percentage of all Single-unit Establishments
Completed Interview	2021	66.046
Partial Interview	97	3.170
Refusal	330	10.784
Data Not Available	81	2.647
No Contact Made	494	16.144
Out-of-Scope	37	1.209
	—	—
Total	3060	100.000

Table B2. Distribution of Length of Call Resulting in Final Interview Outcome
- All Trade Areas

Length of Call	Count	Percentage of All Single-unit Establishments
1-5 minutes	2076	67.843
6-10 minutes	724	23.660
11-15 minutes	187	6.111
over 15 minutes	73	2.386
	—	—
Total	3060	100.000

Table B3. Distribution of Elapsed Time Between
Initial Contact and Final Call
- All Trade Areas

Interview Processing Period	Count	Percentage of All Single-unit Establishments
Same Day	615	20.098
1 week	871	28.464
2 weeks	347	11.340
3 weeks	270	8.824
4 weeks	227	7.418
over 1 month	730	23.856
TOTAL	3060	100.000

Table B4. Distribution of Interview Outcomes
By Number of calls Required for Final Disposition
- All Trade Areas

Number of Calls Required	Count	Percentage of All Single-unit Establishments
Less than 5	2459	80.359
6-10	545	17.810
Over 10	56	1.830
Total	3060	100.000

APPENDIX C. -- VARIANCE ESTIMATION

Variance Estimators

Since our principal estimators are ratios of random variables, there variances can be approximated by the following:

$$\text{Var}(\hat{R}_h) = (\hat{R}_h)^2 \left[\frac{\text{Var}(\hat{Y}_h^{(r)})}{(\hat{Y}_h^{(r)})^2} + \frac{\text{Var}(\hat{Y}_h^{(c)})}{(\hat{Y}_h^{(c)})^2} - 2 \frac{\text{Cov}(\hat{Y}_h^{(r)}, \hat{Y}_h^{(c)})}{\hat{Y}_h^{(r)} \hat{Y}_h^{(c)}} \right] \quad (\text{C.1})$$

$$= \frac{1}{(\hat{Y}_h^{(c)})^2} \left[\text{Var} \hat{Y}_h^{(r)} + (\hat{R}_h)^2 \text{Var} \hat{Y}_h^{(c)} - 2 \hat{R}_h \text{Cov}(\hat{Y}_h^{(r)}, \hat{Y}_h^{(c)}) \right] \quad (\text{C.2})$$

and similarly,

$$\text{Var}(\hat{\theta}_h) = \frac{1}{(\hat{Z}_h)^2} \left[\text{Var}(\hat{X}_h) + (\hat{\theta}_h)^2 \text{Var}(\hat{Z}_h) - 2 \hat{\theta}_h \text{Cov}(\hat{X}_h, \hat{Z}_h) \right] \quad (\text{C.3})$$

Now since independent systematic samples were selected from each payroll group within a given SIC, the variances of the components of the respective ratios for the SIC can be obtained by summing the individual payroll category variances. For example,

$$\text{Var}(\hat{Y}_h^{(r)}) = \sum_{i=1}^{M_h} \text{Var}(\hat{Y}_{hi}^{(r)}) \quad (\text{C.4})$$

From the theory of systematic sampling we know the $\text{Var}(\hat{Y}_{hi}^{(r)})$ can be estimated by

$$\text{Var}(\hat{Y}_{hi}^{(r)}) = \left(1 - n_{hi}^{-1} \sum_{i=1}^{n_{hi}} \pi_{hij} \right) \frac{\sum_{j=2}^{n_{hi}} n_{hi} \left(\frac{y_{hi(j+1)}^{(r)}}{\pi_{hi(j+1)}} - \frac{y_{hij}^{(r)}}{\pi_{hij}} \right)^2}{2(n_{hi}-1)} \quad (\text{C.5})$$

The corresponding expressing for $\text{Var}(\hat{Y}_{hi}^{(c)})$ obviously results from replacing every occurrence of the superscript r with c in equation C.5. Similarly the associated covariance term is estimated

by

$$\text{Cov}(\hat{Y}_{hi}^{(r)}, \hat{Y}_{hi}^{(c)}) = \left(1 - n_{hi}^{-1} \sum_{i=1}^{n_{hi}} \pi_{hij} \right) \sum_{j=2}^{n_{hi}} \left[\frac{n_{hi} \left(\frac{y_{hi(j+1)}^{(r)}}{\pi_{hi(j+1)}} - \frac{y_{hij}^{(r)}}{\pi_{hij}} \right)}{2(n_{hi} - 1)} \right] \quad (\text{C.6})$$

Therefore, the desired estimator for the variance of \hat{R}_h results from summing (C.5), its counterpart for the imputed data and the expression given by (C.6) over the M_h payroll groups and then combining the sums as indicated in C.2.

Finally, the variances of $\hat{Y}_T^{(r)}$ and $\hat{Y}_T^{(c)}$ and $\text{Cov}(\hat{Y}_T^{(r)}, \hat{Y}_T^{(c)})$ are obtained by summing the respective variances of the affected SIC groups, from which an estimator of the variance of \hat{R}_T is obtained by appropriately combining the resultant sums. Expressions for estimators of $\text{Var}(\hat{\theta}_h)$ and $\text{Var}(\hat{\theta}_T)$ are similarly determined.

APPENDIX D. - MULTIPLE IMPUTATION PROCEDURE

Multiple Imputation Procedure

For the EID, the multiple imputation procedure utilized five separate hot deck imputation results for each item, yielding five sets of imputation correction ratio estimates by SIC group and trade area, which were then averaged to produce one set of estimates. The variability of the five imputation runs is reflected in the standard error for the averaged estimates.

Each hot deck imputation was based on the same criteria. For each item, EID nonrespondents and respondents were sorted by SIC group and payroll category. All respondents whose individual ratio of reported value to census impute exceeded 7.5 or was less than 1/7.5 were removed from the donor selection pool. Payroll levels within SIC groups were combined as necessary to insure at least two respondents per category and that the respondents provided a suitable set of donors for the nonrespondents.

Both respondents and nonrespondents were represented with a pair notation where the first member of each pair represents the value reported in the EID and the second member denotes the census imputed value. The "." indicates a missing reported value and N_{dr} indicates a nonzero reported value for a donor. The nonzero census imputed values are designated by N_{nc} for a nonrespondent and N_{dc} for a respondent. Thus nonrespondents were divided into two types - (\cdot , N_{nc}) and (\cdot , 0), and respondents were classified as one of four types - (N_{dr} , N_{dc}), (N_{dr} , 0), (0, N_{dc}) and (0,0). For each of the multiple imputation data sets a set of donors equal in number to the number of sample respondents in the SIC group and payroll category was selected at random and with replacement from the same respondents. From this set, a donor was selected at random and with replacement for each nonrespondent. A value was then imputed for the missing reported value according to the rules in Table D.1. To complete the description of the tabular entries let $V_{nc} = N_{nc}$ or 0; $V_{dr} = N_{dr}$ or 0 and $V_{dc} = N_{dc}$ or 0.

Table D.1 Rules for Imputing Missing Value From Donor

Nonrespondent	Donor	Imputation Rule
1. (\cdot, V_{nc})	(V_{dr}, V_{dc}) Except $(V_{dr} = N_{dr}, V_{dc} = 0)$	$\bullet = \left(\frac{V_{dr}}{V_{dc}} \right) V_{nc}$
2. $(\cdot, 0)$	$(N_{dr}, 0)$	$\bullet = 0$
3. (\cdot, N_{nc})	$(N_{dr}, 0)$	$\bullet = \left(\frac{\sum_{l \in D} V_{(dr)_l}}{\sum_{l \in D} V_{(dc)_l}} \right) N_{nc}$ where the summation is over the donor group D.

In the application of rule 1, when both V_{dr} and V_{dc} were zero, the ratio of the reported donor value to the corresponding census impute $\left(\frac{0}{0} \right)$ was considered 1.00; therefore, the impute for the donor becomes the impute for the nonrespondent. As was indicated earlier, we were unable to locate telephone numbers or addresses for about 16 percent of the establishments in our survey, the majority of which were thought to be out of business during the census reference period. Rule 2 was adopted to accommodate specific cases which were included in this "out of business" category.

APPENDIX E. -- DISTRIBUTIONS OF ESTABLISHMENT LEVEL RATIOS
FOR RESPONDENTS

Figure E 1: Distributions of Establishment Ratios for Wholesale, Outliers Capped

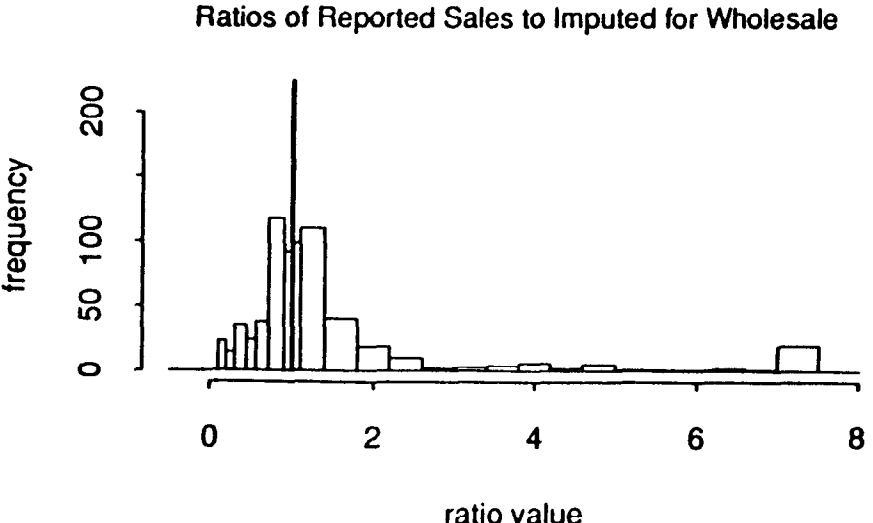
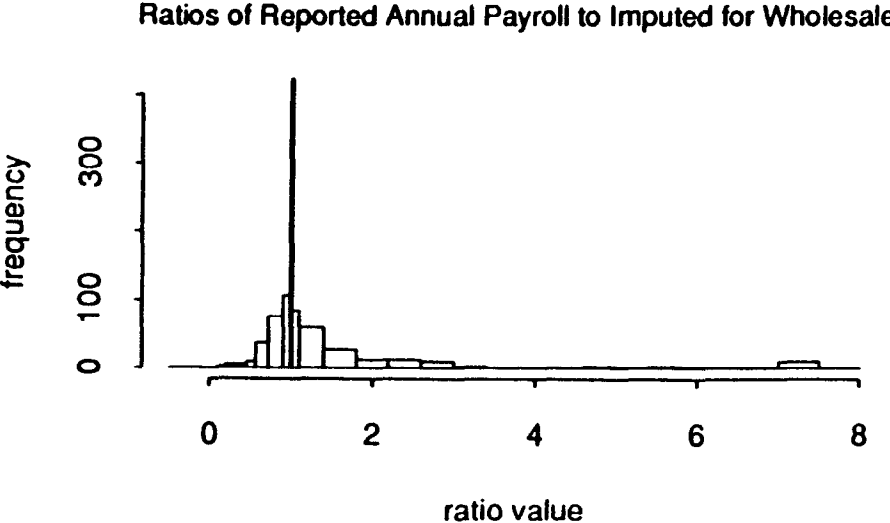
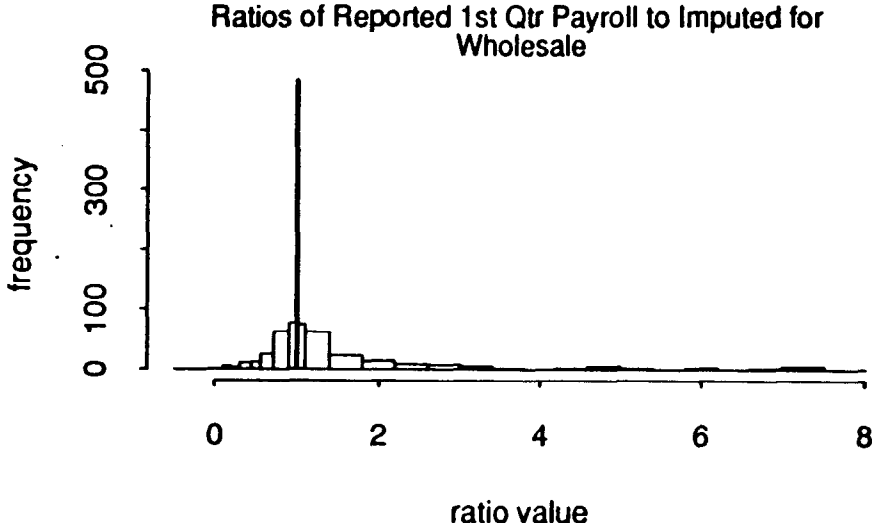
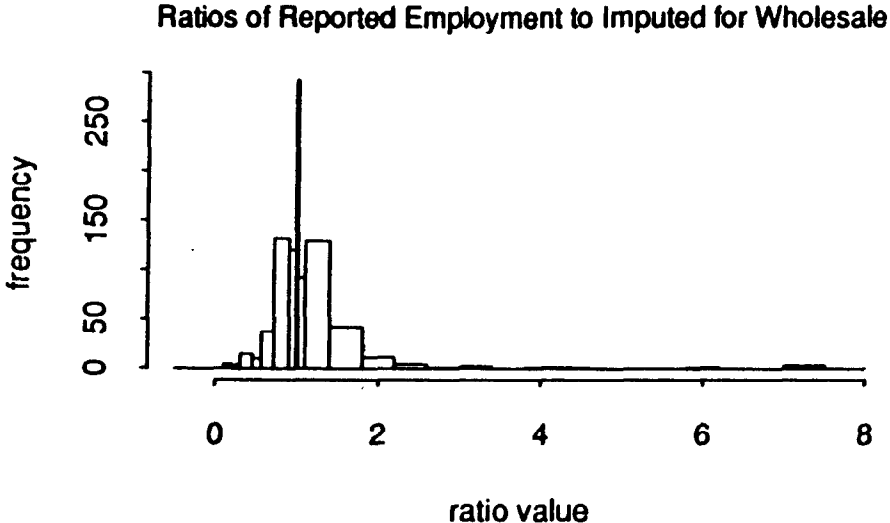


Figure E2: Distributions of Establishment Ratios for Retail, Outliers Capped

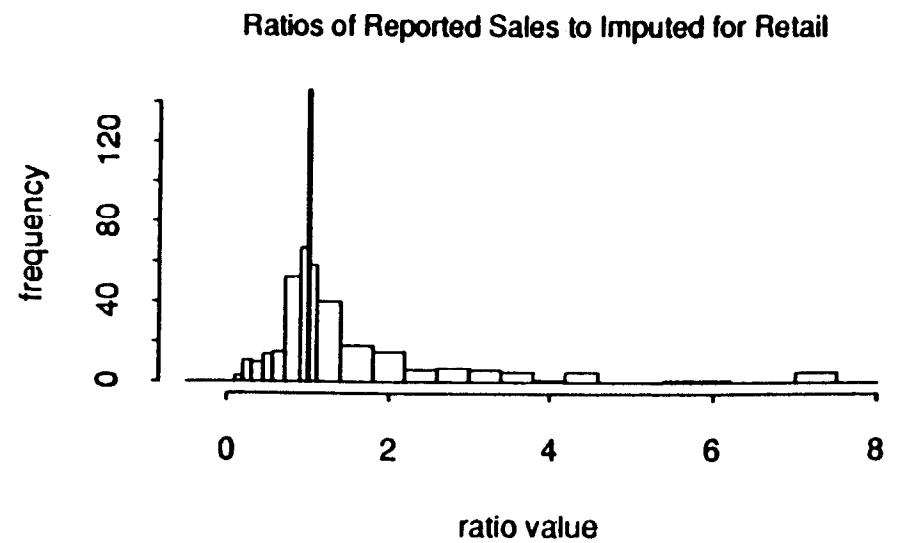
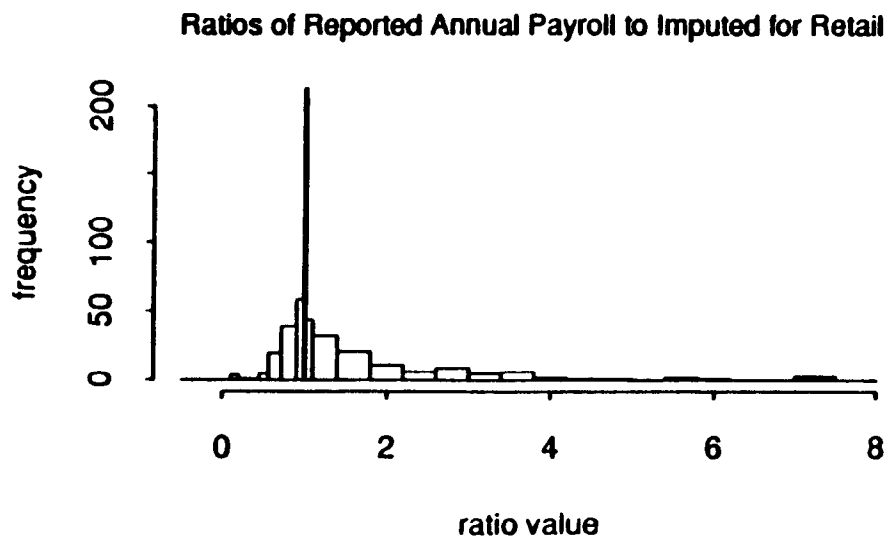
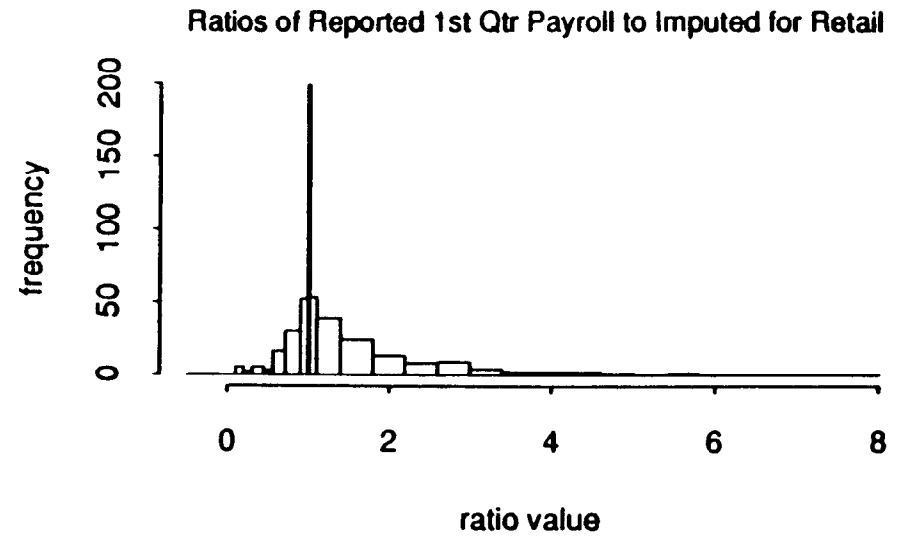
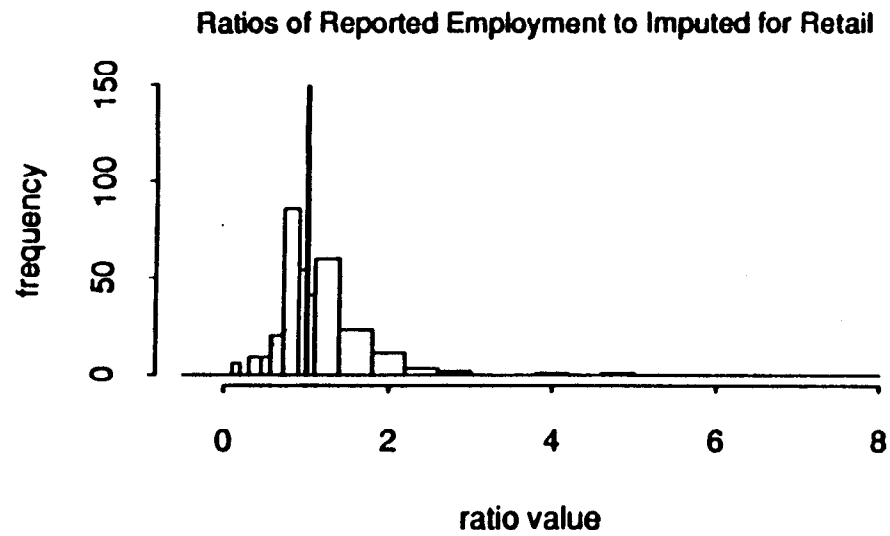


Figure E3: Distributions of Establishment Ratios for Services, Outliers Capped

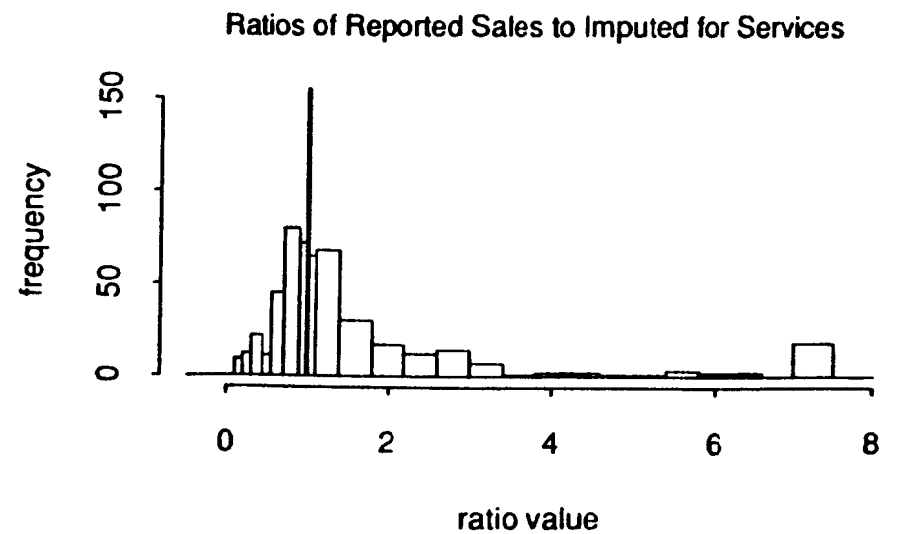
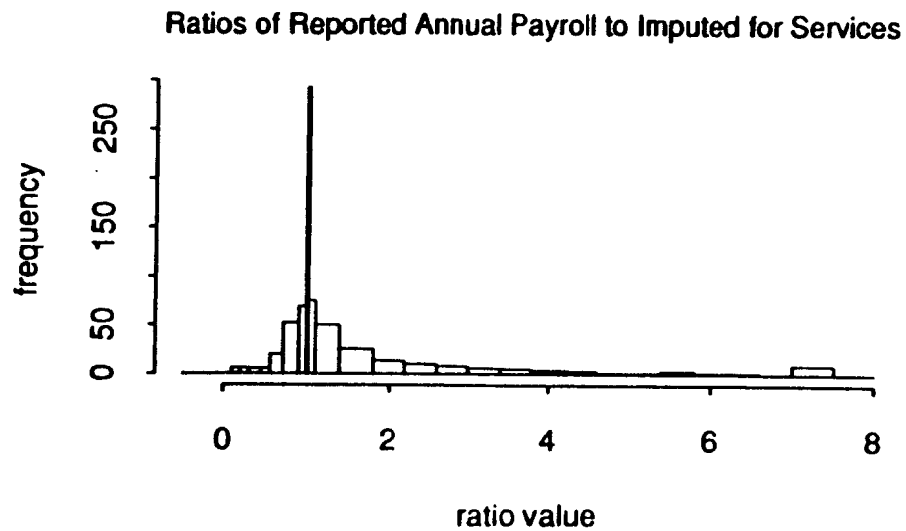
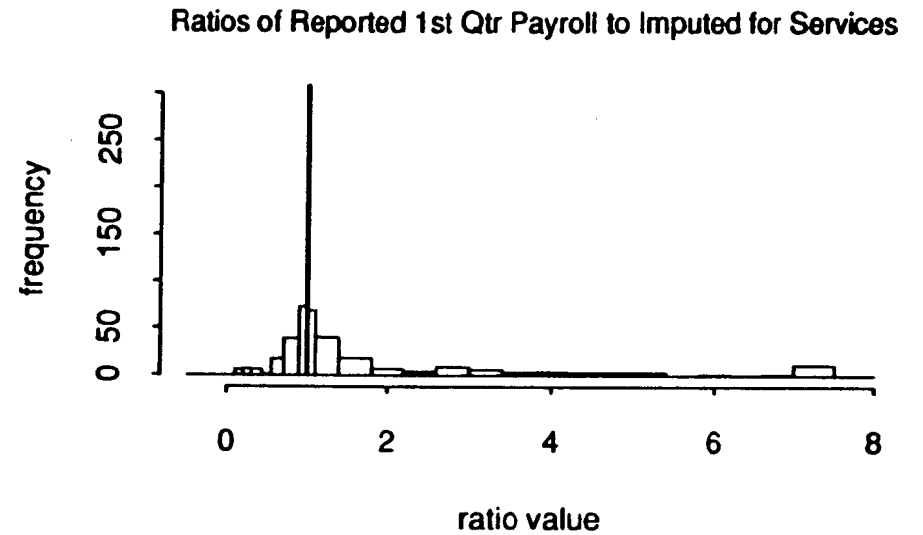
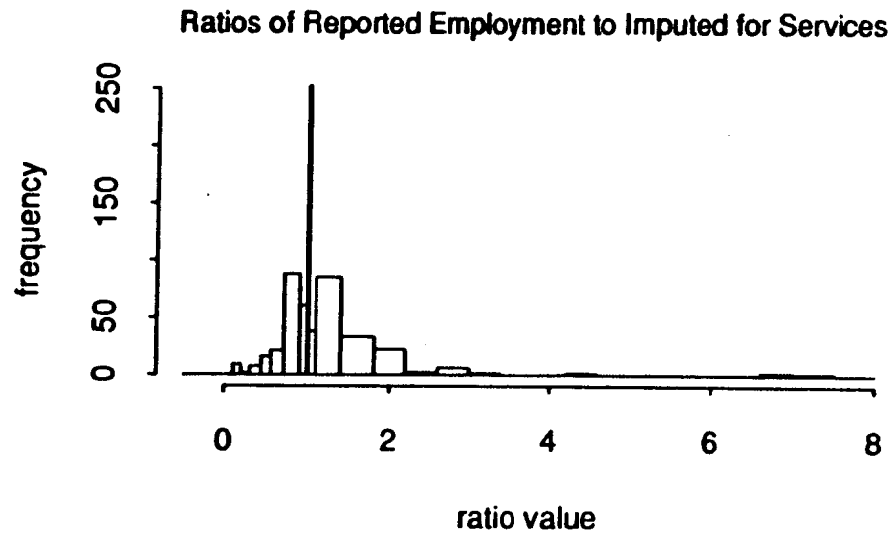


Figure E4. Distributions of Establishment-Level Ratios by SIC Group

	Percentages of Establishment-level Ratios											
	Employment			First Quarter Payroll			Annual Payroll			Sales/Receipts		
	<0.99	>0.99 &<1.01	>1.01	<0.99	>0.99 &<1.01	>1.01	<0.99	>0.99 &<1.01	>1.01	<0.99	>0.99 &<1.01	>1.01
Wholesale	36.6	31.9	31.5	23.0	53.7	23.3	27.0	47.3	25.7	38.3	25.0	36.6
1	35.4	38.0	26.6	35.4	41.8	22.8	35.4	45.6	19.0	36.7	22.8	40.5
2	35.1	29.9	35.1	27.6	47.4	25.0	24.3	41.9	33.8	35.5	26.3	38.2
3	35.0	35.8	29.2	20.9	60.9	18.3	27.0	48.7	24.3	36.8	21.9	41.2
4	42.0	26.1	31.9	26.1	39.1	34.8	26.5	44.1	29.4	40.6	17.4	42.0
5	45.8	24.1	30.1	22.2	51.9	25.9	27.2	49.4	23.5	49.4	24.1	26.6
6	34.2	31.6	34.2	21.5	49.4	29.1	23.1	51.3	25.6	33.3	29.5	37.2
7	35.4	27.3	37.7	17.6	63.5	18.9	25.3	48.0	26.7	32.0	33.3	34.7
8	32.6	29.1	38.4	22.9	47.0	30.1	23.5	39.5	37.0	46.3	18.3	35.4
9	31.3	38.5	30.2	18.6	69.1	12.4	29.6	52.0	18.4	30.2	31.3	38.5
10	39.2	35.4	25.3	18.2	59.7	22.1	23.4	53.2	23.4	40.3	24.7	35.1
11	39.7	32.4	27.9	25.0	54.4	20.6	32.4	44.1	23.5	44.1	26.5	29.4
Retail	39.7	30.9	29.4	25.8	41.9	32.4	25.8	44.2	30.0	35.7	30.1	34.2
12	39.0	30.5	30.5	23.8	42.5	33.7	25.3	43.0	31.6	41.8	25.3	32.9
13	40.0	29.1	30.9	22.2	48.1	29.6	21.8	49.1	29.1	41.8	30.9	27.3
14	44.8	28.4	26.9	36.4	42.4	21.2	34.3	35.8	29.9	23.9	37.3	38.8
15	47.3	20.9	31.9	16.9	51.7	31.5	22.5	47.2	30.3	39.3	27.0	33.7
16	29.1	44.2	26.7	22.7	35.2	42.0	15.9	51.1	33.0	28.4	34.1	37.5
17	45.6	28.1	26.3	40.1	29.1	30.9	42.6	38.9	18.5	35.2	35.2	29.6
18	41.5	30.2	28.3	31.5	33.3	35.2	29.6	38.9	31.5	38.9	20.4	4.07
Services	34.9	37.3	27.8	27.4	45.8	26.9	25.0	43.4	31.7	38.7	23.5	37.8
19	35.0	35.0	30.0	41.7	28.3	30.0	41.7	28.3	30.0	27.1	28.8	44.1
20	41.1	41.1	17.9	32.7	32.7	34.5	25.0	37.5	37.5	25.9	33.3	40.7
21	38.5	35.4	26.2	18.8	43.7	37.5	12.5	37.5	50.0	37.1	19.4	43.5
22	37.7	40.3	22.1	27.0	47.3	25.7	25.7	55.4	18.9	37.0	31.5	31.5
23	26.7	33.3	40.0	31.6	45.6	22.8	16.9	52.5	30.5	38.6	24.6	36.8
24	29.0	40.6	30.4	39.7	38.2	22.1	30.9	39.7	29.4	47.0	15.2	37.9
25	34.8	41.3	23.9	19.6	57.6	22.8	24.7	44.1	31.2	35.6	32.2	32.2
26	41.6	29.2	29.2	15.6	61.1	23.3	15.6	50.0	34.4	51.1	13.6	35.2
27	31.0	38.0	31.0	28.2	43.7	28.2	33.3	38.9	27.8	42.0	15.9	42.0

APPENDIX F. -- MISCLASSIFICATION RATES

Table F1.1 Survey Misclassification Rates By Trade Area and SIC Group
- Mailout Basis (1972/77 codes)

Trade Area	# misclassified	%	Reported SIC Group	# misclassified	%	Business Detail File SIC Codes
wholesale	112	13.3	501	8	10.3	55, 59, 504/508, 509, 75
			502	7	9.1	57, 503/505, 504/508, 513, 72, 78/79/84
			503/505	13	10.8	509, 52, 515/519, 57, 76
			504/508	32	36.8	506/507, 509, 501, 59, 502, 503/505, 511/512, 513, 516/517, 57, 76, 515/519, 52, 73
			506/507	13	15.1	503/505, 509, 504/508, 502, 511/512, 516/517, 57
			509	14	22.6	504/508, 502, 59, 503/505, 511/512, 515/519, 516/517, 76, 78/79/84
			511/512	3	3.9	504/508, 515/519, 59
			513	11	12.6	53/56, 504/508, 502, 509
			514/518	6	6.1	515/519, 54, 58
			515/519	10	12.5	52, 502, 503/505, 504/508, 509, 513, 516/517, 53/56, 59
			516/517	5	7.4	501, 502, 509, 513, 515/519
			retail	71	14.1	52
53/56	4	7.1				513, 52, 82/83
54	8	11.6				53/56, 58, 514/518, 516/517
55	12	12.9				501, 516/517, 52, 53/56, 59, 75
57	22	25.0				53/56, 502, 504/508, 76, 52, 506/507, 509, 72, 75
58	7	12.3				59, 54, 70, 73, 78/79/84, 86
59	10	17.5				53/56, 54, 57, 501, 504/508, 509, 514/518
service	76	11.3	70	5	8.1	86
			72	6	9.0	78/79/84, 73, 53/56
			73	27	31.4	57, 72, 78/79/84, 82/83, 76, 504/508, 506/507, 75, 502, 52, 53/56
			75	8	10.4	76, 501, 73, 72, 509
			76	5	8.3	72, 73, 75, 501, 504/508, 57
			78/79/84	5	8.2	72, 57, 82/83, 509
			81	0	0	-
			82/83	18	19.8	86, 78/79/84, 81, 87
86	2	2.7	73, 82/83			

Table F1.2 Survey Misclassification Rates By Trade Area and SIC Group
- Publications Basis (1987 codes)

Trade Area	# misclassified	%	Reported SIC Group	# misclassified	%	Business Detail File SIC Codes
wholesale	121	13.2	501	8	10.1	55, 509, 59, 75
			502	7	9.1	503/505, 57, 504/508, 513, 72, 78/79/84
			503/505	13	10.7	52, 509, 515/519, 57
			504/508	33	47.8	509, 506/507, 501, 502, 503/505, 76, 511/512, 513, 516/517, 52, 57, 59, 73
			506/507	13	15.3	504/508, 509, 502, 503/505, 511/512, 516/517, 57, 73, 87
			509	12	15.0	502, 59, 515/519, 516/517, 76
			511/512	3	3.9	515/519, 59
			513	11	12.6	53/56, 504/508, 502, 509
			514/518	6	6.1	515/519, 54, 58
			515/519	10	12.7	52, 502, 503/505, 504/508, 509, 513, 516/517, 53/56, 59
			516/517	5	7.4	501, 502, 509, 513, 515/519
retail	68	13.5	52	8	9.8	59, 53/56, 54, 55, 503/505, 515/519, 76
			53/56	4	6.9	513, 52, 82/83
			54	8	11.6	53/56, 58, 514/518, 516/517
			55	12	12.9	501, 516/517, 52, 53/56, 59, 75
			57	22	24.7	53/56, 502, 504/508, 76, 52, 506/507, 509, 72, 75
			58	7	12.2	59, 54, 70, 73, 78/79/84, 86
			59	7	13.0	53/56, 54, 57, 501, 504/508, 509, 514/518
service	75	11.4	70	5	8.1	86
			72	3	5.1	78/79/84, 73, 53/56
			73	25	35.2	57, 72, 78/79/84, 82/83, 76, 504/508, 506/507, 75, 502, 52, 53/56
			75	8	10.4	76, 501, 73, 72, 509
			76	5	8.2	72, 73, 75, 501, 504/508, 57
			78/79/84	5	13.0	72, 57, 82/83, 509
			81	0	0	-
			82/83	18	19.8	86, 78/79/84, 81, 87
			86	2	2.7	73, 82/83

Table F2. Comparison of Survey Misclassification Rates
and Estimates of Population Misclassification Rates
- Publication Basis (1987 codes)

Wholesale:

Group	501	502	503/5	504/8	506/7	509	511/2	513	514/8	515/9	516/7	wholesale
Survey misclassification rate	1.01	9.1	10.7	47.8	15.3	15.0	3.9	12.6	6.1	12.7	7.4	13.2
Population misclassification rate estimate	23.6	20.7	19.8	29.3	23.2	30.4	8.9	5.7	3.3	4.1	12.3	20.1

Retail:

Group	52	53/6	54	55	57	58	59	retail
Survey misclassification rate	9.8	6.9	11.6	12.9	24.7	12.2	13.0	13.5
Population misclassification rate estimate	26.1	11.6	20.2	2.5	20.3	8.0	5.7	11.1

Service:

Group	70	72	73	75	76	78/79/84	81	82/3	86	services
Survey misclassification rate	8.1	5.1	35.2	10.4	8.2	13.0	0.0	19.8	2.7	11.4
Population misclassification rate estimate	0.1	0.2	9.0	5.2	8.0	13.2	0.0	23.2	0.4	6.2

Table F3.1 Cross Tabulations of Reported SIC Group
By Imputed SIC Group for Wholesale Trades
- Publication Basis (1987 codes)

Imputed SIC Group	Reported SIC Group										
	1 501	2 502	3 503/505	4 504/508	5 506/507	6 509	7 511/512	8 513	9 514/518	10 515/519	11 516/517
1 - 501	71	0	0	3	0	0	0	0	0	0	1
2 - 502	0	70	0	2	1	2	0	1	0	1	1
3 - 503/505	0	1	108	2	2	1	0	0	0	1	0
4 - 504/508	0	1	0	36	3	0	1	0	0	0	0
5 - 506/507	0	0	0	5	72	0	0	0	0	0	0
6 - 509	2	0	5	7	3	68	0	4	0	2	1
7 - 511/512	0	0	0	2	1	1	74	0	0	0	0
8 - 513	0	1	0	2	0	0	0	76	0	1	1
9 - 514/518	0	0	0	0	0	0	0	0	92	0	0
10 - 515/519	0	0	1	0	0	2	1	0	3	69	1
11 - 516/517	0	0	0	2	1	1	0	0	0	1	63
Retail	4	2	6	5	1	3	0	6	3	4	0
Services	1	2	1	3	0	2	0	0	0	0	0
Other	1	0	0	0	1	0	0	0	0	0	0
TOTAL	79	77	121	69	85	80	77	87	98	79	68

Table F3.2 Cross Tabulations of Reported SIC Group
By Imputed SIC Group for Retail Trades
- Publication Basis (1987 codes)

Imputed SIC Group	Reported SIC Group						
	12-52	13-53/56	14-54	15-55	16-57	17-58	18-59
12-52	74	1	0	1	1	0	0
13-53/56	1	54	4	1	5	0	2
14-54	1	0	61	0	0	1	1
15-55	1	0	0	81	0	0	0
16-57	0	0	0	0	67	0	0
17-58	0	0	2	0	0	50	0
18-59	2	0	0	1	2	2	47
Wholesale	2	2	2	8	9	0	3
Services	1	1	0	1	5	3	0
Other	0	0	0	0	0	1	1
TOTAL	82	58	69	93	89	57	54

Table F3.3 Cross Tabulations of Reported SIC Group
By Imputed SIC Group for Service Industries
- Publication Basis (1987 codes)

Imputed Sic Group	Reported SIC Group								
	19-70	20-72	21-73	22-75	23-76	24-78/79/84	25-81	26-82/83	27-867
19-70	57	0	0	0	0	1	0	0	0
20-72	0	56	3	1	0	3	0	0	0
21-73	0	1	46	1	1	0	0	1	1
22-75	0	0	1	69	1	0	0	0	0
23-76	0	0	1	3	56	0	0	0	0
24-78/79/84	0	0	1	0	0	60	0	1	0
25-81	0	0	0	0	0	0	94	1	0
26-82/83	0	0	2	0	0	1	0	73	1
27-86	2	0	0	0	0	0	0	12	71
Wholesale	0	0	5	4	2	1	0	0	0
Retail	0	1	7	0	1	3	0	0	0
Other	3	1	5	0	0	0	0	3	0
TOTAL	62	59	71	77	61	69	94	91	73