BUREAU OF THE CENSUS
STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR-86-09

REPORT ON THE IMPUTATION RESEARCH FOR THE MONTHLY RETAIL TRADE SURVEY

by

Elizabeth T. Huang
Statistical Research Division
Bureau of the Census
Room 3524, F.O.B. #3
Washington, D.C. 20233 U.S.A.

Recommended by: Nash Monsour

Report completed: February 26, 1986

Report issued: February 26, 1986

UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, D.C. 20233

February 26, 1986

MEMORANDUM FOR Distribution List

From: Elizabeth T. Huang E.T.H.
Mathematical Statistician
Statistical Research Division

Subject: Report on the Imputation Research for the Monthly Retail Trade Survey

Attached is the final report of the Imputation Research for the Monthly Retail Trade Survey.

Attachment

Distribution List

J. Garrett	(BUS)
C. Konschnik	(BUS)
K. Wolter	(SRD)
N. Monsour	(SRD)
C. Isaki	(SRD)

Report on the Imputation Research for the Monthly Retail Trade Survey

by

Elizabeth T. Huang
Statistical Research Division
Bureau of the Census

February 1986

Abstract

The bias of the estimated totals using different ratio type imputation procedures was examined under models and a Monte Carlo study. In addition the mean square error was computed in the Monte Carlo study.

The current imputation procedure for estimating current month sales is $p\xi$ - unbiased under the ratio model with model variance either proportional to the previous month sales or to the square of the sales.

Under the assumption that the data are missing at random, the bias and MSE for estimating total current month sales from the given data by using different ratio and regression imputation procedures were derived. An optimum ratio imputation procedure was also derived. The reported data for 9 SIC's of the 1982 Monthly Retail Trade Survey were used for the Monte Carlo study.

The results of the Monte Carlo study showed that the current ratio imputation procedure is competitive with the optimum ratio procedure and better than other comparable procedures studied. Under the current ratio imputation procedure, a better set of imputation cells can be formed by using sales quantiles as cutoffs within groups as opposed to the current procedure that uses fixed sales cutoffs. For example, the decrease in MSE in 6 of 9 SIC's ranges from 12% to 59% by using 1/8 quantiles. We recommend that changes in the current imputation cell definitions be considered.

Table of Contents

Elizabeth T. Huang
February 1986

Report on the Imputation Research for the Monthly Retail Trade Survey

I. Introduction

In a previous study (Huang (1984)), we examined and evaluated the current imputation procedures and made recommendations on possible ways to improve them given the current methodology. In this study, we followed up some of the recommendations. We examined different alternative imputation cell definitions and some alternative criteria to evaluate different imputation procedures for the same data set previously used.

II. Examining the Trend, and Error Variance of the Ratio Models Under Alternative Imputation Cell Definitions

In Huang (1984), the Retail Trade Survey data - December 1982 - SIC 562 - were examined under the current imputation cell definition (group size (I, II) x previous month sales size ($50,000 cutoff)). It was concluded that for each imputation cell of the reported data, the error variance of the linear regression model of the current month sales (y) with the previous month sales (x) was approximately $x^2 \sigma^2$ instead of $x\sigma^2$ (where x denotes the previous month sales). In the present study, we investigated six alternative imputation cells definitions to determine whether or not the reported data of each imputation cell conform more closely to imputation model A or to B, where

Model A: $y = R x + e, \quad e \sim (0, x^2\sigma^2)$ (2.1)

Model B: $y = R x + e, \quad e \sim (0, x\sigma^2)$ (2.2) where e_i and e_j are independent for $i \neq j$.

The least squares estimators of R under models A and B using n_r reported units are $(n_r)^{-1} \sum_i y_i/x_i$ and $\sum_i y_i/\sum_i x_i$, respectively; and $(\sum_i w_i)^{-1}(\sum_i w_i(y_i/x_i))$ and $\sum w_i y_i/\sum w_i x_i$, respectively when sampling weights w_i are used. Note that if $1/w_i = c\, x_i/\sum_i x_i$, c is a constant, then the estimator $\sum w_i y_i/\sum w_i x_i$ reduces to $(n_r)^{-1} \sum y_i/x_i$.

The six alternative imputation cells are

1) SIC x group size x sales size (use median as cutoff)

2) SIC x group size x sales size (use 1/4 or 1/8 quantiles as cutoff)

3) SIC x SMSA x group size

4) SIC x geographic division

5) SIC x firm size x sales size (use median as cutoff)

6) SIC x region x sales size (use median as cutoff)

Another imputation cell definition, SIC x firm size x region x sales size, was also considered initially. However, since the number in each cell was not large enough, we omitted this alternative from further consideration. See Table II.11.

The statistics of reported data calculated for each imputation cell are mainly the error variance parameters (λ,ρ), the trend R estimated by four ratio estimation procedures. ($R^{(1)} = \sum w_i y_i/\sum w_i x_i$, $R^{(2)} = \sum y_i/\sum x_i$, $R^{(3)} = \sum w_i(y_i/x_i)/\sum w_i$, $R^{(4)} = (1/n_r)\sum(y_i/x_i)$, and the simple correlation coefficient r between the current month sales y and the previous month sales (x). See Tables II.1-II.10. In preparing the estimation of (λ,ρ), the reported list sample data of SIC 562 of the December 1982 Retail Trade Survey for each imputation cell were sorted by the previous month sales and then grouped with generally 20 units (sometimes 10, 7, or 5 units) in each class. The class variances of y and the means of x were fitted by least squares using a log transformation of the following equation

$$(s_y^2)_i = \lambda \, \bar{x}_i^{\rho} \;, \; i=1,\ldots,k \; . \qquad (2.3)$$

i.e., $\log \; (s_y^2)_i = \log \lambda + \rho \log \bar{x}_i$.

where k is the number of groups.

The outliers from regression and the number of establishments in each group have a definite effect on all statistics calculated. For the alternative imputation cell definition (SIC x group size x sales size) using median sales or 1/4 or 1/8 quantile sales will ensure an approximately equal number of establishments in each imputation cell. To examine the effects of using different quantile sales as sales cutoffs, group II data from SIC 562 - December 1982 were used. The correlation coefficients appear to be decreasing function of the number of quantiles. The correlation coefficients of current month sales with previous month sales using 1/8 quantiles sales as cutoffs are much smaller than using 1/4 quantiles as sales size cutoffs (see Table II.3), and the correlations using 1/4 quantiles sales are smaller than using the median. Also, outliers have more impact on all statistics calculated when fewer establishments are in the imputation cell. In the particular imputation cell of group 2 x sales size [1/8 quantile, 1/4 quantile], there are 12 units with the same reported values after deleting one outlier (x,y) = ($25,043, $162,257). Using 7 units or 10 units to form groups, there is always one group with the same reported values and hence zero group variance. This affects the estimates of λ and ρ . In Table II.3, the parameter ρ calculated after deleting outliers ranges from 0.2028 to 2.7172 when 1/8 quantile sales are used as cutoffs. Using regions, SMSA's or geographic divisions as imputation cells (see Tables II.4, 5, 9), the correlations between current month sales (y) and previous month sales (x) are fairly high for most imputation cells and ρ's are nearer to 2 than to 1. From the criterion we used - the estimated ρ's, it seems that most of the data within the cells conform more closely to model A than B. Using quantiles within Group II, firm sizes and regions, some of the cell's data conform more to model B than A. (See Tables II.2 & 3, Tables II.6, 7, & 8, and Tables II.9 & 10).

III. Examining the Error Variance of the Weighted Ratio Models

As mentioned before, the least squares estimate of R of model B is $\sum_i y_i/\sum_i x_i$, and not $\sum w_i y_i/\sum w_i x_i$, where w_i is the sampling weight. In the following, we study 4 ratio models that use weighted data and which are distinguished by their model variances.

The imputation model is

$$\frac{y_i}{\pi_i} = R\,\frac{x_i}{\pi_i} + e_i \ , \ e_i \sim (0,\ \nu_i) \ , \ i=1,\ldots,n \qquad (2.4)$$

where π_i is the sampling inclusion probability, the sampling weight w_i is the inversion of π_i, and the model variances ν_i are defined as follows:

1. $\nu_i = \sigma^2\, x_i/\pi_i$, (2.5)

2. $\nu_i = \sigma^2\, x_i^2/\pi_i$, (2.6)

3. $\nu_i = \sigma^2 \, x_i^2/\pi_i^2$, (2.7)

4. $\nu_i = \sigma^2 \, x_i/\pi_i^2$. (2.8)

The least squares estimates of R of the above models using n_r reported units are

1. $\hat{R} = (\sum_i \frac{y_i}{\pi_i}) \, / \, (\sum_i \frac{x_i}{\pi_i})$,

2. $\hat{R} = (\sum_i \frac{1}{\pi_i} \, \frac{y_i}{x_i}) \, / \, (\sum_i \frac{1}{\pi_i})$,

3. $\hat{R} = \frac{1}{n_r} \sum_{i=1}^{n_r} (y_i/x_i)$,

4. $\hat{R} = (\sum_i y_i) \, / \, (\sum_i x_i)$

respectively.

The ratio of identicals of the reported data of the current imputation procedure is of the form $(\sum_i \frac{y_i}{\pi_i}) \, / \, (\sum_i \frac{x_i}{\pi_i})$, hence we may argue that the current imputation model is (2.4) with $\nu_i = \frac{x_i}{\pi_i} \sigma^2$ instead of model B in (2.2). That is, the weighted observations of the current month sales (y) have linear relationship with the weighted observations of the previous month sales (x) with model variance $(x_i/\pi_i)\sigma^2$. We can rewrite (2.4) and (2.5) as

$$y' = R \, x' + e \, , \; e \sim (0, \, x' \, \sigma^2) \qquad (2.9)$$

where $y' = y/\pi$, $x' = x/\pi$.

Using the same methodology as before, we sorted the weighted data according to x', and calculated (λ , ρ) using weighted data (y' , x'). Under the current imputation cell definition, the (λ , ρ)'s are tabulated in Table II.12 for each imputation cell of the monthly data of SIC 562 of December 1982 and February 1983. All ρ's are closer to 2 than to 1. If $\pi_i = c\, x_i / \Sigma_i x_i$ for a constant c, model (2.4) with $\nu_i = (x_i^2 / \pi_i^2)\, \sigma^2$ reduces to model A.

IV. Properties of the Estimated Total or Trend Using the Current Imputation Procedure Under Models A or B and the Current Sampling Design

- The current imputation procedure for the missing item y (current month sales) is

$$\hat{y} = \hat{R}\, x .$$

where

$$\hat{R} = \left(\sum_{i=1}^{n_r} y_i/\pi_i \right) / \left(\sum_{i=1}^{n_r} x_i/\pi_i \right),$$

x is the previous month sales, and it is assumed that all the x's are known in the sample,

n_r is the number of cases with reported data in the sample, and

π_i is the sampling inclusion probability. The sampling design for the Monthly Retail Trade Survey is a stratified random sampling design.

We investigated the properties of the estimated total where the missing items are imputed by the current procedures under models A or B using the two criteria below.

(1) $p\xi$ - unbiasedness (see Cassel, Särndal and Wretman (1979)). An estimate $\hat{T}$ is $P\xi$ - unbiased for population parameters T in the nonresponse set up if $E_p E_\xi (\hat{T} - T) = 0$, where E_p is the expectation with respect to the sampling distribution, and E_ξ is the expectation with respect to the model.

(2) Incomplete data bias (Schaible (1979))

When sample data are missing, the incomplete data bias in an estimator $\hat{T}_{n_1}$ can be expressed as the total bias in the estimator minus the bias that would occur if the sample were complete. The incomplete data bias for an estimator $\hat{T}_{n_1}$ may be defined as

$$E_\xi (\hat{T}_{n_1} - T) - E_\xi(\hat{T}_n - T) = E_\xi (\hat{T}_{n_1} - \hat{T}_n)$$

where $\hat{T}_n$ ($\hat{T}_{n_1}$) is an estimator of T based on values of Y from $n(n_1)$ units, and $n_1 \leq n$. The expectation is taken with respect to the superpopulation model of y.

We now show that the estimated total monthly sales by the Horvitz-Thompson estimator is $p\xi$ - unbiased under models A or B and there is no incomplete data bias when the current imputation procedure, or the alternative ratio imputation procedures are used. The four imputation procedures for imputing the missing items y are $\hat{y}_i = \hat{R}^{(j)} x_i$, j=1,...,4, where $\hat{R}^{(j)}$ is as follows:

1. $\hat{R}^{(1)} = (\sum_i y_i/\pi_i) \;/\; (\sum_i x_i/\pi_i)$ (3.1)

2. $\hat{R}^{(2)} = \sum_i y_i / \sum_i x_i$ (3.2)

3. $$\hat{R}^{(3)} = (\sum_i \frac{1}{\pi_i} \frac{y_i}{x_i}) \,/\, (\sum_i \frac{1}{\pi_i}) \qquad (3.3)$$

4. $$\hat{R}^{(4)} = \frac{1}{n_r} \sum_{i=1}^{n_r} (y_i/x_i) \qquad (3.4)$$

All summations are over the number of reported units n_r.

A. pξ - Unbiasedness Criterion of the Estimated Total

When nonresponse occurs, under the current stratified sample design and the current imputation procedure, the estimated total $\hat{Y}$ of monthly sales from all sample units (reported or imputed) by the Horvitz-Thompson estimator is

$$\hat{Y} = \sum_k \hat{Y}_k$$

$$= \sum_k \sum_h \frac{1}{\pi_h} [\sum_{i=1}^{n_{khr}} y_{khi} + \sum_{i=n_{khr+1}}^{n_{kh}} \hat{y}_{khi}]$$

where

$\hat{Y}_k$ is the estimated total from each imputation cell k,

$\hat{y}_{khi} = \hat{R}_k^{(1)} x_{khi}$, is the imputed value for missing item y_{khi}, where

$$\hat{R}_k^{(1)} = (\sum_h \sum_{i=1}^{n_{khr}} y_{khi}/\pi_h) \,/\, (\sum_h \sum_{i=1}^{n_{khr}} x_{khi}/\pi_h) , \qquad (3.5)$$

n_{khr} is the number of response sampling units in stratum h and imputation cell k, and

π_h is the sampling inclusion probability in stratum h.

We'll show $\hat{Y}$ is $p\xi$ - unbiased under models A or B. For each imputation cell k, and stratum h, model B is defined as follows:

$$y_{khi} = R_k x_{khi} + e_{khi} , \quad e_{khi} \sim (0, x_{khi}\sigma^2) \qquad (3.6)$$

$k = 1,\ldots,K$,

$h = 1,\ldots,H$,

$i = 1,\ldots,N_{kh}$.

where

K is the number of imputation cells,

H is the number of strata ,

N_{kh} is the number of units in cell k and stratum h .

The finite population total of y is Y, and $Y = \sum_{k}^{K} \sum_{h}^{H} \sum_{i}^{N_{kh}} y_{khi}$.

Note that $E_{\xi_B} (\hat{Y} - Y) = \sum_{k}^{K} R_k (\hat{X}_k - X_k) \neq 0$. i.e., $\hat{Y}$ is not ξ_B unbiased.

Under model B, $E_{\xi_B} (\hat{R}_k^{(1)}) = R_k$, and

$$E_p E_{\xi_B} (\hat{Y}_k) = R_k E_p \sum_{h} \sum_{i=1}^{n_{kh}} \frac{x_{khi}}{\pi_h} = R_k X_k .$$

Hence

$$E_p E_{\xi_B}(\hat{Y}) = E_p E_{\xi_B}\left(\sum_k^K \hat{Y}_k\right) = \sum_k^K R_k X_k$$

Now

$$E_{\xi_B}(Y) = E_{\xi_B} \sum_k^K \sum_h^H \sum_i^{N_{kh}} y_{khi}$$

$$= \sum_k^K R_k \sum_h^H \sum_i^{N_{kh}} x_{khi}$$

$$= \sum_k^K R_k X_k$$

Hence we have

$$E_p E_{\xi_B}(\hat{Y} - Y) = 0 .$$

We can also show that $\hat{Y}$ is $p\xi$ - unbiased under model A, where model A is the same as model B except the error variance is $x^2\sigma^2$ instead of $x\sigma^2$. This is because $E_{\xi_A}(\hat{R}_k^{(1)}) = R_k$, and $E_p E_{\xi_A}(\hat{Y}_k) = R_k X_k$.

If alternative imputation ratios are used, for example:

$$R_k^{(2)} = \sum_h \sum_{i=1}^{n_{khr}} y_{khi} \Big/ \sum_h \sum_{i=1}^{n_{khr}} x_{khi} , \tag{3.7}$$

$$R_k^{(3)} = \left(\sum_h \sum_i \frac{1}{\pi_k}\right)^{-1} \sum_h \sum_{i=1}^{n_{khr}} \frac{1}{\pi_k} \frac{y_{khi}}{x_{khi}} , \tag{3.8}$$

$$R_k^{(4)} = \left(\sum_h n_{khr}\right)^{-1} \sum_h \sum_{i=1}^{n_{khr}} \frac{y_{khi}}{x_{khi}} , \tag{3.9}$$

the estimated total $\hat{Y}$ using $\hat{R}_k^{(i)} x$, $i=2,3,4$, to impute missing y, will still be pξ - unbiasedness under models A or B, because $\hat{R}_k^{(i)}$, i=2,3,4 are model unbiased under models A or B.

If the imputation model is (2.4) with any model error defined in (2.5) - (2.8), the estimated total $\hat{Y}$ by using $\hat{R}_k^{(i)}x$, i=1,...,4, is also pξ - unbiased.

B. Incomplete Data Bias Criterion of the Estimated Total

In the following, Schaible's (1979) incomplete data bias is used to assess the current imputation procedure. Recall, the incomplete data bias in an estimator $\hat{T}_{n_1}$ is the bias in the estimator $\hat{T}_{n_1}$ minus the bias in the estimator $\hat{T}_n$ that would occur if the sample were complete, i.e.,

$$E_\xi (\hat{T}_{n_1} - \hat{T}_n) = E_\xi(\hat{T}_{n_1} - T) - E_\xi(\hat{T}_n - T) .$$

In the monthly retail trade survey, $\hat{T}_{n_1}$ is the estimated total using all the sample unit values (reported or imputed),

$$\hat{T}_{n_1} = \sum_k \sum_h \frac{1}{\pi_h} [\sum_{i=1}^{n_{khr}} y_{khi} + \sum_{i=n_{khr+1}}^{n_{kh}} \hat{y}_{khi}] ,$$

where

$$\hat{y}_{khi} = \hat{R}_k^{(1)} x_{khi} ,$$

$\hat{R}_k^{(1)}$ is the ratio estimator defined in (3.5) using reported data,

$\hat{T}_n$ is the estimated totals using all sample units as if they were all reported data.

$$\hat{T}_n = \sum_k \sum_h \frac{1}{\pi_h} \sum_{i=1}^{n_{kh}} y_{khi} \quad .$$

Since $E_\xi(\hat{R}_K^{(1)}) = R_k$, under models A or B, hence

$$E_\xi(\hat{T}_{n_1}) = \sum_k \sum_h \frac{1}{\pi_h} R_k \sum_{i=1}^{n_{kh}} x_{khi} = E_\xi\ (\hat{T}_n) \ ,$$

where E_ξ is the expectation under models A or B. Hence $E_\xi(\hat{T}_{n_1} - \hat{T}_n) = 0$ under models A or B using current imputation ratio . This is also true for the other 3 ratio imputation procedures, since $E_\xi(\hat{R}_k^{(j)}) = R_k$, j=2, 3, 4.

C. Incomplete Data Bias of Trend Under Models A or B

The monthly trend is defined to be the ratio of the total estimates of the current month sales by the previous month sales

$$\hat{T} = \left(\sum_k \sum_h \sum_{i\varepsilon r_1} \frac{y_{khi}}{\pi_h} + \sum_k \sum_h \sum_{i\varepsilon \bar{r}_1} \frac{\hat{y}_{khi}}{\pi_h}\right) \Big/ \left(\sum_k \sum_h \sum_{i\varepsilon r_2} \frac{x_{khi}}{\pi_h} + \sum_k \sum_h \sum_{i\varepsilon \bar{r}_2} \frac{\hat{x}_{khi}}{\pi_h}\right)$$

where

$$\hat{y}_{khi} = \hat{R}_k\, x_{khi} \ , \quad \hat{R}_k = \left(\sum_k \sum_h \sum_{i\varepsilon r_1} \frac{y_{khi}}{\pi_h}\right) \Big/ \left(\sum_k \sum_h \sum_{i\varepsilon r_1} \frac{x_{khi}}{\pi_h}\right)$$

$$\hat{x}_{khi} = \hat{\beta}_k\, z_{khi} \ , \quad \hat{\beta}_k = \left(\sum_k \sum_h \sum_{i\varepsilon r_2} \frac{x_{khi}}{\pi_h}\right) \Big/ \left(\sum_k \sum_h \sum_{i\varepsilon r_2} \frac{z_{khi}}{\pi_h}\right)$$

r_1 is the sample units of the reported y (current month sales) data, $\tilde{r}_1$, is the sample units of the nonreported y,

r_2 is the sample units of the reported x (previous month sales) data, $\tilde{r}_2$ is the sample units of the nonreported x,

z is the previous month sales of 3 month's ago data for the rotating panel.

Since x's are always imputed before imputing y's, we can assume all x's are fixed and are available for all sample units when we impute y, hence $\hat{T}$ can be written as

$$\hat{T} = \left(\sum_k \sum_h \sum_{i \varepsilon r_1} \frac{y_{khi}}{\pi_h} + \sum_k \sum_h \sum_{i \varepsilon \tilde{r}_1} \frac{\hat{y}_{khi}}{\pi_h}\right) / \left(\sum_k \sum_h \sum_i^{n_{kh}} \frac{x_{khi}}{\pi_h}\right) .$$

The trend estimate using a complete data set is

$$\hat{T}_c = \left(\sum_k \sum_h \sum_i^{n_{kh}} \frac{y_{khi}}{\pi_h}\right) / \left(\sum_k \sum_h \sum_i^{n_{kh}} \frac{x_{khi}}{\pi_h}\right) .$$

The incomplete data bias of trend is zero under models B or A, because

$$E_\xi \left(\hat{T} - \hat{T}_c\right) = \left(\sum_k R_k \hat{X}_k / \sum_k \hat{X}_k\right) - \left(\sum_k R_k \hat{X}_k / \sum_k \hat{X}_k\right) = 0 .$$

D. **Variance of the Estimated Total Under Models A or B**

So far we have considered the model unbiasedness of the estimated totals or trend when we have nonresponse under models A or B. We found that all four ratio type imputation procedures give $p\xi$ - unbiased estimated totals under models A or B, and zero incomplete data bias.

We now present the variance of the estimated total using four imputation procedures under models A or B.

The monthly estimated total $\hat{Y}$ using any of the four ratio type imputation procedures is

$$\hat{Y}^{(j)} = \sum_k \hat{Y}_k^{(j)}$$

$$= \sum_k \left(\sum_h \sum_{i=1}^{n_{khr}} \frac{y_{khi}}{\pi_h} + \sum_h \sum_{i=n_{khr+1}}^{n_{kh}} \frac{\hat{y}_{khi}^{(j)}}{\pi_h} \right)$$

where

$$\hat{y}_{khi}^{(j)} = \hat{R}_k^{(j)} \, x_{khi} , \qquad j=1,2,3,4 ,$$

$\hat{R}_k^{(j)}$ is defined in (3.5), (3.7) - (3.9).

$\hat{Y}^{(j)}$ is a $p\xi$ - unbiased estimate of total Y under models A or B, where $Y = \sum_k Y_k = \sum_k \sum_h \sum_i^{N_{kh}} Y_{khi}$.

The variance of $\hat{Y}^{(j)}$ under models A or B is defined to be $V_\xi\,(\hat{Y}^{(j)} - Y)$, and

$$V_\xi\,(\hat{Y}^{(j)} - Y) = V_\xi\,(\sum_k \hat{Y}_k^{(j)} - \sum_k Y_k) = \sum_k V_\xi\,(\hat{Y}_k^{(j)} - Y_k) \quad .$$

For each imputation cell k, we have

$$V_\xi\,(\hat{Y}_k^{(j)} - Y_k)$$

$$= V_\xi \Big(\sum_h \sum_{i=1}^{n_{khr}} \frac{y_{khi}}{\pi_h} + \sum_h \sum_{i=n_{khr}+1}^{n_{kh}} \frac{\hat{y}_{khi}^{(j)}}{\pi_h} - \sum_h \sum_{i=1}^{N_{kh}} y_{khi}\Big)$$

$$= \sum_h \sum_{i=1}^{n_{khr}} \frac{\nu_{khi}}{\pi_h^2} + \sum_h \sum_{i=1}^{N_{kh}} \nu_{khi} + \sum_h \sum_{i=n_{khr}+1}^{n_{kh}} \frac{V_\xi(\hat{R}_k^{(j)})\, x_{khi}^2}{\pi_h^2}$$

$$+ 2 \sum_h \sum_{i=1}^{n_{khr}} Cov_\xi\Big(\frac{e_{khi}}{\pi_h}, \hat{R}_k^{(j)} \sum_h \sum_{i=n_{khr}+1}^{n_{kh}} \frac{x_{khi}}{\pi_h}\Big)$$

$$- 2 \sum_h \sum_{i=1}^{n_{khr}} \sum_{j=1}^{N_{kh}} Cov_\xi\Big(\frac{e_{khi}}{\pi_h}, e_{khj}\Big) - 2 \sum_h \sum_{i=1}^{N_{kh}} Cov_\xi\Big(e_{khi}, \hat{R}_k^{(j)} \sum_h \sum_{i=n_{khr}+1}^{n_{kh}} \frac{x_{khi}}{\pi_h}\Big)$$

$$= \sum_h \sum_{i=1}^{n_{khr}} \frac{\nu_{khi}}{\pi_h^2} + \sum_h \sum_{i=1}^{N_{kh}} \nu_{khi} + V_\xi(\hat{R}_k^{(j)}) \sum_h \sum_{i=n_{khr}+1}^{n_{kh}} \frac{x_{khi}^2}{\pi_h^2} - 2 \sum_h \sum_{i=1}^{n_{khr}} \frac{\nu_{khi}}{\pi_h}$$

$$+ 2 \Big(\sum_h \sum_{i=n_{khr}+1}^{n_{kh}} \frac{x_{khi}}{\pi_h}\Big) \Big[\sum_h \sum_{i=1}^{n_{khr}} Cov_\xi\Big(\frac{e_{khi}}{\pi_h}, \hat{R}_k^{(j)}\Big) - \sum_h \sum_{i=1}^{N_{kh}} Cov_\xi(e_{khi}, \hat{R}_k^{(j)})\Big]$$

where

$$\nu_{khi} = x_{khi}^2\, \sigma^2 \qquad \text{for model A}$$

$$= x_{khi}\, \sigma^2 \qquad \text{for model B .}$$

To compare variances of any two estimators $\hat{Y}_k^{(l)}$, $\hat{Y}_k^{(m)}$ in each imputation cell k, we have

$$V_\xi\, (\hat{Y}_k^{(l)} - Y_k) - V_\xi\, (\hat{Y}_k^{(m)} - Y_k)$$

$$= \sum_h \sum_{i=n_{khr}+1}^{n_{kh}} \frac{x_{khi}^2}{\pi_h^2} \big(V_\xi\, (\hat{R}_k^{(l)}) - V_\xi\, (\hat{R}_k^{(m)})\big)$$

$$+ 2 \left(\sum_h \sum_{i=n_{khr}+1}^{n_{kh}} \frac{x_{khi}}{\pi_h}\right) \sum_h \sum_{i=1}^{n_{khr}} [Cov_\xi(\frac{e_{khi}}{\pi_h}, \hat{R}_k^{(1)}) - Cov_\xi(\frac{e_{khi}}{\pi_h}, \hat{R}_k^{(m)})]$$

$$- 2 \left(\sum_h \sum_{i=n_{khr}+1}^{n_{kh}} \frac{x_{khi}}{\pi_h}\right) \sum_h \sum_{i=1}^{N_{kh}} [Cov_\xi(e_{khi}, \hat{R}_k^{(1)}) - Cov_\xi(e_{khi}, \hat{R}_k^{(m)})] ,$$

$$1 \neq m = 1,\ldots,4 ,$$

where

$$V_\xi (\hat{R}_k^{(1)}) = \left(\sum_h \sum_{i=1}^{n_{khr}} \frac{\nu_{khi}}{\pi_h^2}\right) / \left(\sum_h \sum_{i=1}^{n_{khr}} \frac{x_{khi}}{\pi_h}\right)^2 ,$$

$$V_\xi (\hat{R}_k^{(2)}) = \left(\sum_h \sum_{i=1}^{n_{khr}} \nu_{khi}\right) / \left(\sum_h \sum_{i=1}^{n_{khr}} x_{khi}\right)^2 ,$$

$$V_\xi (\hat{R}_k^{(3)}) = \left(\sum_h \sum_{i=1}^{n_{khr}} \frac{\nu_{khi}}{\pi_h^2\, x_{khi}^2}\right) / \left(\sum_h \sum_{i=1}^{n_{khr}} \frac{1}{\pi_h}\right)^2 ,$$

$$V_\xi (\hat{R}_k^{(4)}) = \frac{1}{n_{khr}^2} \sum_h \sum_{i=1}^{n_{khr}} \frac{\nu_{khi}}{x_{khi}^2} ,$$

$$Cov_\xi(\frac{e_{khi}}{\pi_h}, \hat{R}_k^{(1)}) = (\nu_{khi}/\pi_h^2) / \left(\sum_h \sum_{i=1}^{n_{khr}} x_{khi}/\pi_h\right)$$

$$Cov_\xi(\frac{e_{khi}}{\pi_h}, \hat{R}_k^{(2)}) = (\nu_{khi}/\pi_h) / \left(\sum_h \sum_i x_{khi}\right)$$

$$Cov_\xi(\frac{e_{khi}}{\pi_h}, \hat{R}_k^{(3)}) = (\nu_{khi}/\pi_h^2\, x_{khi}) / \left(\sum_h \sum_i \frac{1}{\pi_h}\right)$$

$$Cov_\xi(\frac{e_{khi}}{\pi_h}, \hat{R}_k^{(4)}) = (\nu_{khi}/\pi_h\, x_{khi}) / \left(\sum_h n_{khr}\right)$$

V. Monte Carlo Study

In Huang (1984), a Monte Carlo study was carried out to evaluate different imputation procedures based on a given set of complete data (reported list sample from SIC 562 in the December 1982 Retail Trade Survey). Five sets of incomplete data were generated from the complete data. Missing items of incomplete data were randomly suppressed from each imputation cell of complete data according to its current imputation rate for each of the five sets. The reader is cautioned that since only five incomplete data sets were used, the results of the comparisons may not give an accurate picture. In the following, the bias and MSE of the estimated total for a given complete data set were derived under the assumption that the missing data are a random sample of the complete data set.

Without loss of generality, only one imputation cell is assumed. A sample of size n is assumed to be drawn from a population of size N. The sampling unit i has inclusion probability π_i . In a sample of size n, there are n_r units reported, and $n_{\bar{r}}$ units not reported. In the following we treat these reported n_r units as our complete data set. Assuming the nonresponse mechanism is ignorable, i.e., the data are missing at random, the incomplete data sets are generated in which $n'_{\bar{r}}$ units are suppressed randomly from the complete data set of n_r units, and different ratio type imputation procedures are used to impute $n'_{\bar{r}}$ missing units of y values using the auxiliary variable x, which is available for all n_r units.

Let Y be the estimated total using the complete data of n_r units

$$Y = \sum_{i=1}^{n_r} y_i/\pi_i .$$

Let $\hat{Y}$ be the estimated total using incomplete data of n_r units, of which n'_r units are reported, and $n'_{\bar{r}}$ units are imputed, i.e.,

$$\hat{Y} = \sum_{i=1}^{n'_r} y_i/\pi_i + \sum_{i=1}^{n'_{\bar{r}}} \hat{y}_i/\pi_i$$

where

$$\hat{y}_i = \hat{R}_{n'_r} x_i \ , \quad i=1,\ldots,n'_{\bar{r}} \ ,$$

$$n_r = n'_r + n'_{\bar{r}} \ ,$$

$\hat{R}_{n'_r}$ is one of the four ratio type estimators in (3.1)-(3.4) using all n'_r units.

We also assume that the nonresponse rate is such that

$$\lim_{n'_{\bar{r}} \to \infty} f_{n'_{\bar{r}}} = n'_{\bar{r}}/n_r = f < 1 \ .$$

Lemma 1. Under above notations and assumptions, for large n'_r ,

$$E((\hat{Y} - Y) \mid n_r) \doteq -(n'_{\bar{r}}/n_r) \sum_{i=1}^{n_r} (\hat{e}_i/\pi_i) \ ,$$

$$= 0 \ , \text{ if } \quad \hat{R}_{n'_r} = \hat{R}^{(1)}$$

where

$$\hat{e}_i = y_i - \hat{R}_{n_r} x_i \ ,$$

$\hat{R}_{n_r}$ is any of the four ratio type estimators (3.1)-(3.4) using the complete data of size n_r ,

$E(\cdot \mid n_r)$ is the expectation over all possible samples of size $n'_{\bar{r}}$ drawn from n_r.

Proof:

$E((\hat{Y} - Y) \mid n_r)$

$$= E\left(\left(\sum_{i=1}^{n'_r} y_i/\pi_i + \sum_{i=1}^{n'_{\bar{r}}} \hat{y}_i/\pi_i - \sum_{i=1}^{n_r} y_i/\pi_i\right) \mid n_r\right)$$

$$= n'_{\bar{r}}\, E\left(\left(\frac{1}{n'_{\bar{r}}} \sum_{i=1}^{n'_{\bar{r}}} (\hat{y}_i - y_i) / \pi_i\right) \mid n_r\right)$$

Since $\hat{y}_i = \hat{R}_{n'_r}\, x_i$, we'll prove in the following that $\hat{R}_{n'_r} = \hat{R}_{n_r} + O_p(n'_r{}^{-1/2})$ for $\hat{R}_{n'_r}$ being any form defined in (3.1) - (3.4), we then have

$$E\left(\frac{1}{n'_{\bar{r}}} \sum_{i=1}^{n'_{\bar{r}}} (\hat{y}_i/\pi_i) \mid n_r\right)$$

$$\doteq E\left(\frac{1}{n'_{\bar{r}}} \sum_{i=1}^{n'_{\bar{r}}} (x_i/\pi_i)\, \hat{R}_{n_r} \mid n_r\right)$$

To prove $\hat{R}_{n'_r} = \hat{R}_{n_r} + O_p(n'_r{}^{-1/2})$, e.g.,

if $\hat{R}_{n'_r}$ is $\hat{R}^{(2)}$ (ratio of means) defined in (3.2) using n'_r units,

$$E\,(\hat{R}_{n'_r} \mid n_r) = \hat{R}_{n_r} + \frac{1-f}{n'_r}\,(C_{xx}(n_r) - C_{yx}(n_r))\,\hat{R}_{n_r}$$

$$\lim_{n'_r \to \infty} E\,(\hat{R}_{n'_r} \mid n_r) = \hat{R}_{n_r}$$

and $V(\hat{R}_{n'_r} | n_r) = O(n'^{-1}_r)$

where $\hat{R}_{n_r}$ is the same form of $\hat{R}^{(2)}$, but using the complete reported data set of size n_r,

$$C_{xx}(n_r) = \frac{s_x^2(n_r)}{\bar{x}^2_{n_r}} ,$$

$$C_{yx}(n_r) = \frac{s_y(n_r)\ s_x(n_r)}{\bar{x}_{n_r}\ \bar{y}_{n_r}} ,$$

$f = n'_r/n_r$.

By corollary 5.1.3.2 in Fuller (1976), we have $\hat{R}_{n'_r} = \hat{R}_{n_r} + O_p(n'^{-1/2}_r)$

If $\hat{R}_{n'_r}$ is $\hat{R}^{(1)}$ or $\hat{R}^{(3)}$ (defined in (3.1) and (3.3)), it can also be treated as a ratio of means and has a similar bias term as above with slightly different target and auxiliary variables. For example, for $\hat{R}^{(1)}$, the target variable is y_i/π_i, the auxiliary variable is x_i/π_i ; for $\hat{R}^{(3)}$, the target variable is $y_i/x_i\pi_i$, and the auxiliary variable is $1/\pi_i$.

If $\hat{R}_{n'_r}$ is of form $\hat{R}^{(4)}$ (mean of ratios) defined in (3.4) but using n'_r units, we have

$$E\ (\hat{R}_{n'_r} \mid n_r) = (n_r)^{-1} \sum_{i=1}^{n_r} (y_i/x_i) = \hat{R}_{n_r} .$$

Thus we have

$E\ ((\hat{Y} - Y) \mid n_r)$

$$\doteq -\frac{n'_{\bar r}}{n_r}\sum_{i=1}^{n_r}\frac{1}{\pi_i}(y_i - \hat R_{n_r} x_i)$$

where $\hat R_{n_r}$ is any of four ratio estimators (3.1) - (3.4) using a complete data set of size n_r.

When $\hat R_{n_r}$ is a form of $\hat R^{(1)}$, $\sum_{i=1}^{n_r}\frac{1}{\pi_i}(y_i - \hat R_{n_r} x_i) = 0$ and hence $E((\hat Y - Y) \mid n_r) = 0$

Lemma 2. Under the notation and assumptions defined in this section, for large n'_r,

$$E((\hat Y - Y)^2 \mid n_r)$$

$$\doteq \frac{n'_{\bar r}}{n_r}\left(\sum_{i=1}^{n_r}\frac{\hat e_i^2}{\pi_i^2} + \frac{(n'_{\bar r} - 1)}{(n_r - 1)}\sum_{i\neq j}^{n_r}\frac{\hat e_i}{\pi_i}\frac{\hat e_j}{\pi_j}\right)$$

where

$$\hat e_i = y_i - \hat R_{n_r} x_i$$

$\hat R_{n_r}$ is an estimator defined in (3.1) - (3.4) using the complete data set of size n_r .

Proof:

$$E((\hat Y - Y)^2 \mid n_r)$$

$$= E((\sum_{i=1}^{n'_{\bar r}}(\hat y_i - y_i)/\pi_i)^2 \mid n_r)$$

$$= E(\sum_{i=1}^{n'_{\bar r}}((y_i - \hat y_i)/\pi_i)^2 \mid n_r)$$

$$+ E\left(\sum_{i \neq j}^{n'_{\bar{r}}} ((y_i - \hat{y}_i)/\pi_i)((y_j - \hat{y}_j)/\pi_j) \mid n_r\right)$$

$$= E\left(n'_{\bar{r}} \frac{1}{n'_{\bar{r}}} \sum_{i=1}^{n'_{\bar{r}}} \left(\frac{y_i^2}{\pi_i^2} - 2\frac{y_i \hat{y}_i}{\pi_i^2} + \frac{\hat{y}_i^2}{\pi_i^2}\right) \mid n_r\right)$$

$$+ E\left(n'_{\bar{r}}(n'_{\bar{r}} - 1) \frac{1}{n'_{\bar{r}}(n'_{\bar{r}} - 1)} \sum_{i \neq j}^{n'_{\bar{r}}} \left(\frac{y_i y_j}{\pi_i \pi_j} - \frac{\hat{y}_i y_j}{\pi_i \pi_j} - \frac{y_i \hat{y}_j}{\pi_i \pi_j} + \frac{\hat{y}_i \hat{y}_j}{\pi_i \pi_j}\right) \mid n_r\right)$$

$$\doteq \frac{n'_{\bar{r}}}{n_r} \sum_{i=1}^{n_r} \left(\frac{y_i^2}{\pi_i^2} - 2\frac{y_i x_i}{\pi_i^2} \hat{R}_{n_r} + \frac{x_i^2}{\pi_i^2} \hat{R}_{n_r}^2\right)$$

$$+ \frac{n'_{\bar{r}}(n'_{\bar{r}} - 1)}{n_r(n_r - 1)} \sum_{i \neq j}^{n_r} \left\{ \frac{y_i y_j}{\pi_i \pi_j} - \left(\frac{y_j x_i}{\pi_j \pi_i} + \frac{y_i x_j}{\pi_i \pi_j}\right) \hat{R}_{n_r}\right.$$

$$\left. + \frac{x_i x_j}{\pi_i \pi_j} \hat{R}_{n_r}^2 \right\}$$

by the fact that $\hat{y}_i = \hat{R}_{n'_r} x_i$, and $\hat{R}_{n'_r} = \hat{R}_{n_r} + O_p(n'^{-1/2}_r)$ for $\hat{R}_{n'_r}$ being any forms defined in (3.1)-(3.4).

Let $\hat{y}_{ci} = \hat{R}_{n_r} x_i$, $i=1,\ldots,n_r$, we have

$$E((\hat{Y} - Y)^2 \mid n_r)$$

$$\doteq \frac{n'_{\bar{r}}}{n_r} \sum_{i=1}^{n_r} \left[\frac{y_i^2}{\pi_i^2} - 2\frac{y_i \hat{y}_{ci}}{\pi_i^2} + \frac{\hat{y}_{ci}^2}{\pi_i^2}\right]$$

$$+ \frac{n'_{\bar{r}}(n'_{\bar{r}} - 1)}{n_r(n_r - 1)} \sum_{i \neq j}^{n_r} \left(\frac{y_i y_j}{\pi_i \pi_j} - \frac{\hat{y}_{ci} y_j}{\pi_i \pi_j} - \frac{y_i \hat{y}_{cj}}{\pi_i \pi_j} + \frac{\hat{y}_{ci} \hat{y}_{cj}}{\pi_i \pi_j}\right)$$

$$= \frac{n'_{\bar{r}}}{n_r}\left[\sum_{i=1}^{n_r} \left(\frac{y_i - \hat{y}_{ci}}{\pi_i}\right)^2 + \frac{(n'_{\bar{r}} - 1)}{(n_r - 1)} \sum_{i \neq j}^{n_r} \frac{(y_i - \hat{y}_{ci})}{\pi_i} \frac{(y_j - \hat{y}_{cj})}{\pi_j}\right]$$

$$= \frac{n'_{\bar{r}}}{n_r}\left(\sum_{i=1}^{n_r} \frac{\hat{e}_i^2}{\pi_i^2} + \frac{(n'_{\bar{r}} - 1)}{(n_r - 1)} \sum_{i \neq j}^{n_r} \frac{\hat{e}_i}{\pi_i} \frac{\hat{e}_j}{\pi_j}\right)$$

Note that if $\hat{R}_{n_r}$ is of form of $R^{(1)}$, then

$$\sum_{i=1}^{n_r} \frac{\hat{e}_i^2}{\pi_i^2} = - \sum_{i \neq j}^{n_r} \frac{\hat{e}_i}{\pi_i} \frac{\hat{e}_j}{\pi_j}, \quad \text{because} \quad \sum_{i=1}^{n_r} \frac{e_i}{\pi_i} = 0 .$$

Lemma 3. Under the notation and assumptions defined in this section, redefine

$\hat{y}_i = Rx_i$ for $i=1,\ldots, n'_{\bar{r}}$, where R is a preassigned value, then

(1) $$E((\hat{Y} - Y) \mid n_r) = - \frac{n'_{\bar{r}}}{n_r} \sum_{i=1}^{n_r} \frac{1}{\pi_i} (y_i - Rx_i) ,$$

(2) $$E((\hat{Y} - Y) \mid n_r)) = 0 \quad , \quad \text{iff} \quad R = \left(\sum_{i=1}^{n_r} \frac{y_i}{\pi_i}\right) / \left(\sum_{i=1}^{n_r} \frac{x_i}{\pi_i}\right) ,$$

(3) $E((\hat{Y} - Y)^2 \mid n_r))$

$$= \frac{n'_{\bar{r}}}{n_r} \sum_{i=1}^{n_r} \frac{1}{\pi_i^2} (y_i - Rx_i)^2$$

$$+ \frac{n'_{\bar{r}} (n'_{\bar{r}} - 1)}{n_r (n_r - 1)} \sum_{i \neq j}^{n_r} \frac{1}{\pi_i \pi_j} (y_i - Rx_i)(y_j - Rx_j) ,$$

(4) The value of R that minimizes (3) is

$$\hat{R}_{opt} = \frac{\left(\sum_{i=1}^{n_r} \frac{x_i y_i}{\pi_i^2} + \frac{(n'_{\bar{r}} - 1)}{(n_r - 1)} \sum_{i \neq j}^{n_r} \frac{x_i y_j}{\pi_i \pi_j}\right)}{\left(\sum_{i=1}^{n_r} \frac{x_i^2}{\pi_i^2} + \frac{(n'_{\bar{r}} - 1)}{(n_r - 1)} \sum_{i \neq j}^{n_r} \frac{x_i x_j}{\pi_i \pi_j} \right)} . \tag{5.1}$$

Proof:

(1) $E((\hat{Y} - Y) \mid n_r)$

$$= E\left(\left(\sum_{i=1}^{n'_r} \frac{y_i}{\pi_i} + \sum_{i=1}^{n'_{\bar{r}}} \frac{Rx_i}{\pi_i} - \sum_{i=1}^{n_r} \frac{y_i}{\pi_i}\right) \mid n_r\right) \quad , \quad \text{and } n_r = n'_r + n'_{\bar{r}}$$

$$= - E\left(\sum_{i=1}^{n'_{\bar{r}}} \frac{1}{\pi_i} (y_i - Rx_i) \mid n_r\right)$$

$$= - \frac{n'_{\bar{r}}}{n_r} \sum_{i=1}^{n_r} \frac{1}{\pi_i} (y_i - Rx_i)$$

(2) It is clear that

$$E((\hat{Y} - Y) \mid n_r) = 0 \text{ if and only if } R = \left(\sum_{i=1}^{n_r} \frac{y_i}{\pi_i}\right) / \left(\sum_{i=1}^{n_r} \frac{x_i}{\pi_i}\right) .$$

(3) $E((\hat{Y} - Y)^2 \mid n_r)$

$$= E\left(\left(\sum_{i=1}^{n'_{\bar{r}}} - (y_i - Rx_i)/\pi_i\right)^2 \mid n_r\right)$$

$$= E\left(\sum_{i=1}^{n'_{\bar{r}}} (y_i - Rx_i)^2/\pi_i^2 + \sum_{i\neq j} (y_i - Rx_i)(y_j - Rx_j)/\pi_i\pi_j \mid n_r\right)$$

$$= \frac{n'_{\bar{r}}}{n_r} \sum_{i=1}^{n_r} \frac{1}{\pi_i^2} (y_i - Rx_i)^2$$

$$+ \frac{n'_{\bar{r}} (n'_{\bar{r}} - 1)}{n_r (n_r - 1)} \sum_{i\neq j}^{n_r} \frac{1}{\pi_i\pi_j} (y_i - Rx_i) (y_j - Rx_j)$$

(4) Minimize (3) with respect to R, we have

$$2 \frac{n'_{\bar{r}}}{n_r} \sum_{i=1}^{n_r} \frac{1}{\pi_i^2} (y_i - Rx_i) (- x_i)$$

$$+ \frac{n'_{\bar{r}} (n'_{\bar{r}} - 1)}{n_r (n_r - 1)} \sum_{i\neq j}^{n_r} \frac{1}{\pi_i\pi_j} [(- x_i)(y_j - Rx_j) + (y_i - Rx_i)(- x_j)] = 0 \quad ,$$

and

$$\hat{R}_{opt} = \frac{\sum_{i=1}^{n_r} \frac{x_i y_i}{\pi_i^2} + \frac{(n'_{\bar{r}} - 1)}{(n_r - 1)} \sum_{i \neq j}^{n_r} \frac{x_i y_j}{\pi_i \pi_j}}{\sum_{i=1}^{n_r} \frac{x_i^2}{\pi_i^2} + \frac{(n'_{\bar{r}} - 1)}{(n_r - 1)} \sum_{i \neq j}^{n_r} \frac{x_i x_j}{\pi_i \pi_j}} .$$

Lemma 4. The bias and MSE of the estimated total $\hat{Y}$ given n_r, by using $\hat{R}_{opt}$ to impute missing y_i, is

$$E((\hat{Y} - Y) | n_r) = - \frac{n'_{\bar{r}}}{n_r} \sum_{i=1}^{n_r} \frac{1}{\pi_i} (y_i - \hat{R}_{opt} x_i) ,$$

$$E((\hat{Y} - Y)^2 | n_r) = \frac{n'_{\bar{r}}}{n_r} \sum_{i=1}^{n_r} \frac{1}{\pi_i^2} (y_i - \hat{R}_{opt} x_i)^2$$

$$+ \frac{n'_{\bar{r}} (n'_{\bar{r}} - 1)}{n_r (n_r - 1)} \sum_{i \neq j}^{n_r} \frac{1}{\pi_i \pi_j} (y_i - \hat{R}_{opt} x_i)(y_j - \hat{R}_{opt} x_j) .$$

Proof:

$\hat{R}_{opt}$ is a function of n_r units of a complete data set which is the population that the incomplete data samples are randomly generated in the Monte Carlo study. For a given complete data set of size n_r, $\hat{R}_{opt}$ is a fixed value. Following the same proof as in Lemma 3, we have the results.

An estimator of $\hat{R}_{opt}$ by using n'_r reported units in the Monte Carlo study is

$$\tilde{R}_{opt} = \frac{\frac{1}{n'_r}\left(\sum_{i=1}^{n'_r}\frac{x_iy_i}{\pi_i^2} + \frac{n'_{\bar{r}}-1}{n'_r-1}\sum_{i\neq j}^{n'_r}\frac{x_iy_j}{\pi_i\pi_j}\right)}{\frac{1}{n'_r}\left(\sum_{i=1}^{n'_r}\frac{x_i^2}{\pi_i^2} + \frac{n'_{\bar{r}}-1}{n'_r-1}\sum_{i\neq j}^{n'_r}\frac{x_ix_j}{\pi_i\pi_j}\right)} \quad \text{for } n'_r > 2 . \tag{5.2}$$

Lemma 5. The bias and MSE of the estimated total $\hat{Y}$ given n_r, by using $\tilde{R}_{opt}$ to impute missing y_i is

$$E((\hat{Y} - Y)|n_r) \doteq -\frac{n'_{\bar{r}}}{n_r}\sum_{i=1}^{n_r}\frac{1}{\pi_i}(y_i - \hat{R}_{opt}\,x_i)$$

$$E((\hat{Y} - Y)^2|n_r) \doteq \frac{n'_{\bar{r}}}{n_r}\sum_{i=1}^{n_r}\frac{1}{\pi_i^2}(y_i - \hat{R}_{opt}\,x_i)^2 + \frac{n'_{\bar{r}}\,(n'_{\bar{r}}-1)}{n_r\,(n_r-1)}\sum_{i\neq j}^{n_r}\frac{1}{\pi_i\pi_j}(y_i - \hat{R}_{opt}\,x_i)(y_j - \hat{R}_{opt}\,x_j) .$$

Proof:

Following the similar proofs in lemmas 1 and 2, we only need to prove for large n'_r ,

$$\tilde{R}_{opt} = \hat{R}_{opt} + O_p({n'_r}^{-1/2}) ,$$

Let $\tilde{R}_{opt} = \frac{N_{n'_r}}{D_{n'_r}}$, and $\hat{R}_{opt} = \frac{N_{n_r}}{D_{n_r}}$

The numerator of $\tilde{R}_{opt}$ is

$$N_{n'_r} = \frac{1}{n'_r} \sum_{i=1}^{n'_r} \frac{x_i y_i}{\pi_i^2} + \frac{(n'_r - 1)}{n'_r (n'_r - 1)} \sum_{i \neq j}^{n'_r} \frac{x_i y_j}{\pi_i \pi_j} ,$$

$$E(N_{n'_r} \mid n_r) = \frac{1}{n_r} \sum_{i=1}^{n_r} \frac{x_i y_i}{\pi_i^2} + \frac{(n'_r - 1)}{n_r (n_r - 1)} \sum_{i \neq j}^{n_r} \frac{x_i y_j}{\pi_i \pi_j} = N_{n_r}$$

and N_{n_r} denotes the numerator of $\hat{R}_{opt}$.

The denominator of $\tilde{R}_{opt}$ is

$$D_{n'_r} = \frac{1}{n'_r} \sum_{i=1}^{n'_r} \frac{x_i^2}{\pi_i^2} + \frac{(n'_r - 1)}{n'_r (n'_r - 1)} \sum_{i \neq j}^{n'_r} \frac{x_i x_j}{\pi_i \pi_j}$$

$$E(D_{n'_r} \mid n_r) = \frac{1}{n_r} \sum_{i=1}^{n_r} \frac{x_i^2}{\pi_i^2} + \frac{(n'_r - 1)}{n_r (n_r - 1)} \sum_{i \neq j}^{n_r} \frac{x_i x_j}{\pi_i \pi_j} = D_{n_r}$$

and D_{n_r} denotes the denominator of $\hat{R}_{opt}$.

Now by the Taylor series expansion

$$D_{n'_r}^{-1} = D_{n_r}^{-1} - D_{n_r}^{-2} (D_{n'_r} - D_{n_r}) + O_p\left(\frac{1}{n'_r}\right)$$

$$V(D_{n'_r} \mid n_r) = \frac{S_{n_r}^2 \left(\frac{x^2}{\pi^2}\right)}{n'_r} + \frac{(n'_r - 1)^2}{n'_r (n'_r - 1)} S_{n_r}^2 \left(\frac{x_i x_j}{\pi_i \pi_j}\right) = O\left(\frac{1}{n'_r}\right)$$

$$D_{n'_r}^{-1} = D_{n_r}^{-1} + O_p\left(\frac{1}{\sqrt{n'_r}}\right)$$

$$\tilde{R}_{opt} \doteq \frac{N_{n'_r}}{D_{n_r}}$$

We have

$$\lim_{n'_r \to \infty} E(\tilde{R}_{opt} \mid n_r) = \frac{N_{n_r}}{D_{n_r}} = \hat{R}_{opt},$$

$$V(\tilde{R}_{opt} \mid n_r) \doteq \frac{V(N_{n'_r})}{D^2_{n_r}} = O\left(\frac{1}{n'_r}\right), \text{ and}$$

$$\tilde{R}_{opt} = \hat{R}_{opt} + O_p(n'^{-1/2}_r) .$$

To use $\tilde{R}_{opt}$ we need to know the number of nonresponse items $n'_{\bar{r}}$, and the number of response items n'_r in the sample. If the factor $(n'_{\bar{r}} - 1)(n'_r - 1)^{-1}$ is not used in $\tilde{R}_{opt}$, then

$$\tilde{R}_{opt} = \frac{\sum_{i=1}^{n'_r} \frac{x_i y_i}{\pi_i^2} + \sum_{i \neq j}^{n'_r} \frac{x_i y_j}{\pi_i \pi_j}}{\sum_{i=1}^{n'_r} \frac{x_i^2}{\pi_i^2} + \sum_{i \neq j}^{n'_r} \frac{x_i x_j}{\pi_i \pi_j}} = \frac{[\sum_{i=1}^{n'_r} \frac{x_i}{\pi_i}][\sum_{i=1}^{n'_r} \frac{y_i}{\pi_i}]}{[\sum_{i=1}^{n'_r} \frac{x_i}{\pi_i}][\sum_{i=1}^{n'_r} \frac{x_i}{\pi_i}]} = \frac{\sum_{i=1}^{n'_r} \frac{y_i}{\pi_i}}{\sum_{i=1}^{n'_r} \frac{x_i}{\pi_i}} .$$

$\tilde{R}_{opt}$ is reduced to the current imputation ratio $R^{(1)}$.

Another estimator of $\hat{R}_{opt}$ is

$$R^{(5)} = \left(\sum_{i=1}^{n'_r} \frac{x_i y_i}{\pi_i^2}\right) / \left(\sum_{i=1}^{n'_r} \frac{x_i^2}{\pi_i^2}\right) . \qquad (5.3)$$

It can be shown that when $R^{(5)} x_i$ is used to impute missing y_i, for large n'_r, the bias and MSE of $\hat{Y}$ for a given complete data set n_r are given in lemma 1 and 2, where

$$\hat{R}_{n_r} = \left(\sum_{i=1}^{n_r} \frac{x_i y_i}{\pi_i^2} \right) / \left(\sum_{i=1}^{n_r} \frac{x_i^2}{\pi_i^2} \right) .$$

If the inclusion probability π is not used in (5.3), we have

$$R^{(6)} = \left(\sum_{i=1}^{n'_r} x_i y_i \right) / \left(\sum_{i=1}^{n'_r} x_i^2 \right) \tag{5.4}$$

which is the least squares estimate of R of the ratio model with constant error variance (i.e., $y = R\, x + e$, e is independently identically distributed with mean zero and variance σ^2).

It can be shown that the bias and MSE of $\hat{Y}$ for a given complete data set are given in lemma 1 and 2 with

$$\hat{R}_{n_r} = \left(\sum_{i=1}^{n_r} x_i y_i \right) / \left(\sum_{i=1}^{n_r} x_i^2 \right) .$$

If the ordinary regression estimator is used to impute missing item y_i , $i = 1, \ldots, n'_{\bar{r}}$,

$$\hat{y}_i = \hat{\alpha}_{n'_r} + \hat{\beta}_{n'_r} x_i ,$$

where

$$\hat{\alpha}_{n'_r} = \bar{y}_{n'_r} - \hat{\beta}_{n'_r} \bar{x}_{n'_r} , \tag{5.5}$$

$$\bar{y}_{n'_r} = \frac{1}{n'_r} \sum_{i=1}^{n'_r} y_i ,$$

$$\bar{x}_{n'_r} = \frac{1}{n'_r} \sum_{i=1}^{n'_r} x_i ,$$

$$\hat{\beta}_{n'_r} = \sum_{i=1}^{n'_r} (x_i - \bar{x}_{n'_r})(y_i - \bar{y}_{n'_r}) \Big/ \sum_{i=1}^{n'_r} (x_i - \bar{x}_{n'_r})^2 \quad , \qquad (5.6)$$

then the bias and MSE of $\hat{Y}$ for a given complete data set n_r are given in lemma 1 and 2, with

$$\hat{e}_i = y_i - \hat{\alpha}_{n_r} - \hat{\beta}_{n_r} x_i \quad , \quad \text{and} \qquad (5.7)$$

$$\hat{\alpha}_{n_r} = \bar{y}_{n_r} - \hat{\beta}_{n_r} \bar{x}_{n_r} \quad ,$$

$$\bar{y}_{n_r} = \frac{1}{n_r} \sum_{i=1}^{n_r} y_i \quad ,$$

$$\bar{x}_{n_r} = \frac{1}{n_r} \sum_{i=1}^{n_r} x_i \quad ,$$

$$\hat{\beta}_{n_r} = \sum_{i=1}^{n_r} (x_i - \bar{x}_{n_r})(y_i - \bar{y}_{n_r}) \Big/ \sum_{i=1}^{n_r} (x_i - \bar{x}_{n_r})^2 \quad .$$

The above results can be extended to more than one imputation cell if we generate the incomplete data set from the complete data set independently for each imputation cell.

Let $\hat{Y}_k$, Y_k be the estimated totals of the incomplete data and the complete data respectively from imputation cell k, k=1,...,K. Then

$$\hat{Y} = \sum_{k=1}^{K} \hat{Y}_k \quad , \quad Y = \sum_{k=1}^{K} Y_K \ .$$

Let n_{r_k} be the sample size of the reported data of imputation cell k, and we randomly suppress n'_{r_k} units from this complete data set.
Let n_r be the sample size of the complete data set from all K imputation cells, and $n_r = \sum_{k=1}^{K} n_{r_k}$.

The bias of the estimated total $\hat{Y}$ given the complete data set is

$$E((\hat{Y} - Y) \mid n_r)$$

$$= \sum_{k=1}^{K} E((\hat{Y}_k - Y_k) \mid n_{r_k})$$

$$\doteq - \sum_{k=1}^{K} \frac{n'_{-r_k}}{n_{r_k}} \sum_{i=1}^{n_{r_k}} \frac{1}{\pi_{ki}} \hat{e}_{ki} \quad ,$$

where

$$\hat{e}_{ki} = y_{ki} - \hat{R}_{n_{r_k}} x_{ki} \;, \; k=1,\ldots,K, \; i=1,\ldots,n_{r_k} .$$

$\hat{R}_{n_{r_k}}$ is a ratio estimator of (3.1) to (3.4) and (5.2)-(5.3) using the complete data set of size n_{r_k} from each imputation cell k. For the regression estimator defined in (5.5), (5.6), $\hat{e}_i$ is defined in (5.7) for each imputation cell.

Similarly, the mean square errors of the estimated total $\hat{Y}$ given the complete data set can be written as

$$E((\hat{Y} - Y)^2 \mid n_r)$$

$$= \sum_{k=1}^{K} E((\hat{Y}_k - Y_k)^2 \mid n_{r_k})$$

$$\doteq \sum_{k=1}^{K} \frac{n'_{-r_k}}{n_{r_k}} \left(\sum_{i=1}^{n_{r_k}} \frac{\hat{e}^2_{ki}}{\pi^2_{ki}} + \frac{(n'_{-r_k} - 1)}{(n_{r_k} - 1)} \sum_{i \neq j}^{n_{r_k}} \frac{\hat{e}_{ki}}{\pi_{ki}} \frac{\hat{e}_{kj}}{\pi_{kj}} \right) \quad .$$

Under the assumption that the data are missing at random, we have already shown that the bias and MSE of the estimated total given the complete data set using various ratio and regression imputation procedures are functions of residuals of the complete data, and the nonresponse rate of each imputation cell. To compare different ratio and regression imputation procedures defined in (3.1) - (3.4) and (5.2) - (5.6) empirically, we can thus compute these biasses and MSE's using Monthly Retail Trade Survey reported data and current nonresponse rates without randomly generating all possible incomplete samples.

The Monthly Retail Trade Survey reported data of December 1982 for nine SIC's were used to compare the bias and MSE of the estimated totals of the different ratio and regression type imputation procedures. The trends ($\hat{R}_{n_{r_k}}$) calculated from the reported data of each imputation cell by these different estimators are tabulated in Table 4.1. The trends calculated by the optimum ratio procedure ($\hat{R}_{opt}$) and the current imputation procedure ($\hat{R}^{(1)}$) are fairly close for most SIC's. The bias and MSE of the estimated totals by using these different imputation procedures are tabulated in Tables 4.2 and 4.3, respectively. Algebraically, we have already shown that given the complete data set, the current imputation procedure is unbiased with respect to the estimated reported total for each imputation cell, and so are the empirical results. The relative biases of the other ratio imputation procedures are relatively small, less than 3% for most data.

The optimum ratio imputation procedure, $\tilde{R}_{opt}$, gave the minimum mean square error among all the ratio type imputation procedures. However, the gain in mean square error of $\tilde{R}_{opt}$ in comparing with the current imputation procedure is at most 0.002. The current imputation procedure is fairly competitive with the optimum ratio imputation procedure and is easier to compute.

Note that all the inferences of the Monte Carlo study are restricted to the data we used. The derivations of the bias and MSE are based on the assumption that the data are missing at random. The data used for the Monte Carlo study were examined to investigate the validity of this assumption. The imputation rates by sales classes of each imputation cell were tabulated in Table 4.4. There is no apparent relationship between item nonresponse rates and sales classes. The imputation rates by regions of each imputation cell were also tabulated in Table 4.5. The imputation rates are different for different regions but there is no specific pattern.

Based on the current imputation procedure, we also used mean square error (MSE) criterion to evaluate different imputation cell definitions, e.g., to answer the question of what quantiles (median, 1/4 or 1/8 or 1/16 quantiles) should be used for the cutoff of sales size if sales sizes are used within each group (group I and II) for imputation cell definition as opposed to the current fixed cutoff. The reported data of 9 SIC's of December 1982 were used. The empirical results showed that for SIC 562 the smaller the imputation cell is, the better the MSE. However, the most drastic reduction in MSE is the cell definition using 1/4 quantiles as sales cutoffs. There was an approximate 44% reduction in MSE as compared to the MSE under the current imputation cell definition. Using 1/8 quantiles as sales cutoffs a further 6% reduction over 1/4 quantiles was observed; and using 1/16 quantiles a further 3% reduction over 1/8 quantiles was observed. (See Table 4.6) Overall, the empirical results varied by SIC's. In 6 of 9 SIC's, the reductions in MSE ranged from 12% (-3%) to 59% (44%) by using 1/8 (1/4) quantiles instead of the current fixed cutoff. Most of these reductions in MSE came from group II. For SIC's 541, 551, and 5813, there was little gain in using any of the quantiles considered. (See Table 4.7)

VI. Summary

We have evaluated the bias of the estimated totals using four ratio type imputation procedures (including the currently used imputation procedure) under models and a Monte Carlo study for a given data set.

Under models A or B in (3.6) and the current sample design, the estimators of total using four ratio type imputation procedures defined in (3.5), (3.7) - (3.9) when nonresponse occurs are $p\xi$ - unbiased. The difference of the variances of estimated totals for each imputation cell using any of two ratio type imputation procedures under the model was derived.

The incomplete data bias of the estimated total using any of four imputation procedures defined in (3.5), (3.7) - (3.9) is zero under models A or B.

Under the assumption that the data are missing at random, the bias and MSE of the estimated total using different ratio type imputation procedures with respect to the estimated reported total were derived for the given reported data. An optimum ratio imputation estimator was also derived along with several variants. The bias and MSE were calculated for each of nine SIC's using December 1982 retail sales data. For the given data set, the empirical results showed that the estimated total using the current imputation procedure is unbiased and has the second smallest MSE among all ratio type imputation procedures in the study.

Since the gain of the MSE by using the optimum imputation procedure is trivial, and extra computation and information are needed to implement this optimum imputation procedure, we do not recommend any changes of the current ratio type imputation procedure in the Monthly Retail Trade Survey.

For the given data set, there is no apparent relationship of nonresponse rate with sales within each imputation cell for all nine SIC's.

By using different imputation cell definitions, the data of Retail Trade Survey of December 1982 of SIC 562 were used to examine the validity of two models A or B. That is, whether the error variance of the ratio model of current month sales y is proportional to previous month sales x or x^2. In Huang (1984), with the current imputation cell definition, it is x^a, where $1.25 \leq a \leq 2.21$. In this study, different imputation cell definitions were used, and the error variance of the ratio model of each imputation cell is proportional to x^b where b is as follows:

	Imputation Cells	No. of Cells	
1.	GP	2	$1.79 \leq b \leq 1.90$
2.	GP x sales (use median as cutoff)	4	$1.43 \leq b \leq 2.15$
3.	GP II x sales (use 1/4 quantiles)	4	$1.46 \leq b \leq 2.21$
4.	GP II x sales (use 1/8 quantiles)	8	$0.34 \leq b \leq 2.72$
5.	SMSA	2	$1.81 \leq b \leq 1.94$
6.	SMSA x GP	4	$1.88 \leq b \leq 2.23$
7.	Geographic division	9	$1.83 \leq b \leq 2.45$
8.	Firm	4	$1.79 \leq b \leq 2.32$
9.	Firm x sales (use median as cutoff)	8	$1.43 \leq b \leq 2.49$
10.	Firm 2 x sales (use 1/4 quantiles)	16	$0.47 \leq b \leq 2.82$
11.	Region	4	$1.83 \leq b \leq 2.05$
12.	Region x sales (use median as cutoff)	8	$1.27 \leq b \leq 2.52$

In the current imputation cells, for some SIC's, the number of establishments of Group II dominates Group I; for other SIC's, the number of establishments of Group I dominates Group II. However, the number of establishments in each region seems more evenly distributed for the data we examined. Using region by sales size (with the median as cutoff) as an alternative imputation cell definition will double the current imputation cells (8 instead of 4), and give much more even numbers of establishments within the cell. However, many big chain stores are spread over all regions in the country and reports often cover more than 1 region. Thus, using regional breakdowns for imputation cell definitions would cause problems to implement in practice. The empirical results suggested that for some SIC's of December 1982's data, we can do better by using alternative imputation cells, i.e., use sales quantiles as cutoffs within groups as opposed to the current fixed sales cutoffs. The decrease in MSE in 6 of 9 SIC's ranges from 12% to 59% by using 1/8 quantiles. We recommend that changes in the current imputation cells definition be considered, especially where empirical studies show that a significant reduction in the MSE can be achieved by increasing the number of imputation cells. We also suggest further similar empirical study be carried out on recent monthly data to provide further bases for changes in cell definitions. This will tell us whether there is a gain in using alternative imputation cells and what quantiles to use for a given SIC in a given month.

VII. Acknowledgement

The author gratefully acknowledges the helpful comments and criticisms of Cary Isaki, Nash Monsour and Carl Konshnick. Appreciation is also given to Linda Moore for computing Table 4.7.

VIII. References

1. Cassel, C.M., Särndal, C.E., and Wretman, J.H. (1979). "Some Uses of Statistical Models in Connection with the Nonresponse Problem," Symposium on Incomplete Data: Preliminary Proceedings, pp. 188-215.

2. Cochran, W.G. (1977). Sampling Techniques. John Wiley and Sons, New York.

3. Fuller, W.A. (1976). Introduction to Statistical Time Series. John Wiley and Sons, New York.

4. Huang, E.T. (1984). "An Imputation Study for the Monthly Retail Trade Survey," Proceedings of the American Statistical Association, Section on Survey Research Methods, pp. 610-615.

5. Schaible, W.L. (1979). "Estimation of Finite Population Totals from Incomplete Sample Data: Prediction Approach," Symposium on Incomplete Data: Preliminary Proceedings, pp. 170-187.

TABLE II.1 The Estimated Trends by Four Ratio Procedures, Correlation Coefficients, and Error Variance Parameters (λ, ρ) for Each Imputation Cell (Group x Sales - $50,000 as Cutoff) December 1982 - SIC 562

Current Imputation Cell Group	Sales	n	$R^{(1)}$	$R^{(2)}$	$R^{(4)}$	$R^{(3)}$	r	r^2	λ	ρ
Group 1	< $50,000	430 (20)	1.420	1.445	1.502	1.380	0.93294	0.87037	1.808576	1.77976
Group 1	> $50,000	355 (20)	1.396	1.334	1.497	1.477	0.76495	0.58514	0.039531	2.12919
		354 (20)	1.427	1.505	1.500	1.478	0.95845	0.91862	0.014775	2.21910
Group 2	< $50,000	249 (20)	1.666	1.722	1.731	1.627	0.65477	0.42873	8.48059	1.65939
		247 (20)	1.634	1.662	1.682	1.605	0.85129	0.72470	358.03	1.25194
Group 2	> $50,000	411 (20)	1.521	1.541	1.617	1.613	0.98741	0.97498	0.05136	2.07087

In cell (Group 1, Sales > $50,000), the outlier is (x,y) = ($8,168,230, $3,120,007).

In cell (Group 2, Sales < $50,000), the outliers are (x,y) = ($40,868, $370,307) ($25,043, $162,257), where y is the current month sales and x is the previous month sales.

TABLE II.2 The Estimated Trends by Four Ratio Procedures, Correlation Coefficients, and Error Variance Parameters (λ, ρ) for Each Imputation Cell (Group x Sales (Use Median as Cutoff)) December 1982 - SIC 562

Imputation Cell Group	Sales	n	$R^{(1)}$	$R^{(2)}$	$R^{(4)}$	$R^{(3)}$	r	r^2	λ	ρ
Group 1	< Median ($41,590)	392 (20)	1.410	1.429	1.498	1.375	0.94061	0.88475	1.0275	1.84316
Group 1	> Median ($41,590)	393 (20) 392 (20)	1.408 1.438	1.341 1.507	1.502 1.505	1.491 1.491	0.76558 0.95912	0.58611 0.91991	0.0878 0.0343	2.05 2.14324
Group 2	< Median ($80,100)	330 (20) 327 (20)	1.666 1.620	1.683 1.640	1.708 1.666	1.633 1.603	0.7869 0.8954	0.61918 0.8017	32.16 61.70	1.5262 1.4304
Group 2	> Median ($80,100)	330 (20)	1.496	1.538	1.612	1.581	0.9868	0.9737	0.0178	2.1494

In Cell 3 (Group 2, Sales < Median), the outliers are (x,y) = ($40,868, $370,307), ($65,906, $208,776), ($25,043, $162,257).

The low ρ in this cell was caused by 12 same reported values which lowered the group variance.

In Cell 2 (Group 1, Sales > Median) the outliers is (x,y) = ($8,168,230, $3,120,007).

TABLE II.3 The Estimated Trends by Four Ratio Procedures, Correlation Coefficients, and Error Variance Parameters (λ , ρ) for Imputation Cells--Group II x Sales Size (Use Quantiles as Cutoff) December 1982 - SIC 562

Imputation Cell Group	Sales	n	$R^{(1)}$	$R^{(2)}$	$R^{(4)}$	$R^{(3)}$	r	r^2	λ	ρ
A. Use 1/4 Quantiles as sales size cutoff										
(1) Group 2	(3/4 Quantile, ∞)	165 (20)	1.429	1.521	1.577	1.428	0.98393	0.96812	0.0065	2.2119
(2) Group 2	(median, 3/4 Quantile)	165 (20)	1.627	1.639	1.648	1.635	0.76799	0.58981	3.6163	1.7033
(3) Group 2	(1/4 Quantile, median)	165 (20)	1.687	1.669	1.682	1.693	0.51390	0.26409	1690542	0.5330
		163 (20)	1.623	1.622	1.627	1.629	0.72415	0.52439	10.4596	1.5903
(4) Group 2	(0, 1/4 Quantile)	165 (20)	1.616	1.720	1.734	1.583	0.67547	0.45626	10.2717	1.6302
		164 (20)	1.614	1.685	1.705	1.581	0.77251	0.59676	48.50	1.4576
B. Use 1/8 Quantiles as sales size cutoff										
(1) Group 2	(7/8 Quantile, ∞)	83 (7)	1.405	1.507	1.572	1.347	0.97948	0.95939	17.1295	1.6219
(2) Group 2	(3/4 Quantile, 7/8 Quantile)	82 (10)	1.493	1.578	1.581	1.484	0.73324	0.53764	0.0002	2.4984
(3) Group 2	(5/8 Quantile, 3/4 Quantile)	83 (10)	1.559	1.596	1.594	1.559	0.62553	0.39131	43605332	0.3368
(4) Group 2	(Median, 5/8 Quantile)	82 (10)	1.659	1.706	1.703	1.657	0.57586	0.33161	0.00027	2.7172
(5) Group 2	(3/8 Quantile, median)	83 (10)	1.679	1.633	1.640	1.679	0.40704	0.16083	9830.16	0.9810
		82 (7)	1.615	1.614	1.622	1.617	0.43485	0.18909	661.78	1.2095
(6) Group 2	(1/4 Quantile, 3/8 Quantile)	82 (10)	1.697	1.723	1.723	1.704	0.20679	0.04276	9.6849	-0.5334
		81 (7)	1.633	1.635	1.633	1.640	0.49465	0.24468	72.3634	1.3849

II.3 - continued

Imputation Cell Group	Sales	n	$R^{(1)}$	$R^{(2)}$	$R^{(4)}$	$R^{(3)}$	r	r^2	λ	ρ
(7) Group 2	(1/8 Quantile, 1/4 Quantile)	83 (10)	1.716	1.722	1.669	1.654	0.25765	0.06638	508291	0.5661
		82 (7)	1.665	1.664	1.664	1.649	0.19085	0.43687	1.1226-8	3.4585
		82 (10)	1.665	1.664	1.664	1.649	0.43687	0.19085	0.00017	2.4496
		70 (10)	1.663	1.624	1.624	1.646	0.46121	0.21272	18260667	0.2028
(8) Group 2	(0, 1/8 Quantile)	82 (10)	1.542	1.727	1.746	1.534	0.64509	0.41614	0.1198	2.0988

TABLE II.4 The Estimated Trends by Four Ratio Procedures, Correlation Coefficients and Error Variance Parameters (λ , ρ) for Each Imputation Cell (SMSA x GP)
December 1982 - SIC 562

Imputation Cell	n	$R^{(1)}$	$R^{(2)}$	$R^{(4)}$	$R^{(3)}$	r	r^2	λ	ρ
Collapsed Cell									
I. Non-SMSA	730 (20)	1.519	1.550	1.611	1.479	0.99091	0.98190	0.38991	1.939
II. SMSA	715 (20)	1.414	1.447	1.534	1.370	0.87927	0.77312	1.444	1.815
<u>Imputation Cell</u>									
1. SMSA GP2	307 (20)	1.532	1.571	1.675	1.627	0.98353	0.96734	0.3055	1.952
2. SMSA GP1	408 (20)	1.300	1.169	1.428	1.285	0.82694	0.68384	1.3467	1.8216
	407 (20)	1.330	1.375	1.431	1.285	0.97397	0.94862	0.7385	1.8820
3. Non-SMSA GP2	353 (20)	1.567	1.522	1.647	1.617	0.99409	0.98821	0.4054	1.94186
4. Non-SMSA GP1	377 (20)	1.486	1.628	1.577	1.449	0.96552	0.93222	0.0172	2.2338

In Cell (SMSA, GP1), the outlier is (x, y) = ($8,168,230, $3,120,007), where y is current month sales, x is previous month sales.

TABLE II.5 The Estimated Trends by Four Ratio Procedures, Correlation Coefficients, and Error Variance Parameters (λ , ρ) for Each Imputation Cell (Geographic Division)
December 1982 - SIC 562

Imputation Cell	n	$R^{(1)}$	$R^{(2)}$	$R^{(4)}$	$R^{(3)}$	r	r^2	λ	ρ
1. New England Div.	83 (20)	1.410	1.629	1.619	1.270	0.93874	0.88123	0.021116	2.23351
2. Middle Atlantic	238 (20)	1.366	1.249	1.488	1.342	0.75515	0.57026	0.249571	2.02213
	237 (20)	1.407	1.487	1.492	1.343	0.94483	0.89271	0.108064	2.10309
3. East North Central	220 (20)	1.515	1.536	1.535	1.436	0.99317	0.98640	0.276366	1.96379
4. West North Central	89 (20)	1.379	1.333	1.390	1.388	0.97822	0.95691	0.159922	2.06205
5. South Atlantic	230 (20)	1.461	1.542	1.585	1.383	0.97333	0.94737	0.045076	2.15520
6. East South Central	71 (20)	1.487	1.570	1.601	1.279	0.98170	0.96374	0.155529	2.06613
7. West South Central	178 (20)	1.549	1.732	1.625	1.591	0.98966	0.97943	1.906025	1.82640
8. Mountain	55 (20)	1.348	1.472	1.394	1.331	0.9927	0.98546	0.004209	2.44826
9. Pacific	281 (20)	1.564	1.485	1.705	1.696	0.99521	0.99044	2.5607+11	-0.63281
	280 (20)	1.568	1.492	1.706	1.696	0.99397	0.98797	0.128111	2.06785

Note:
In Cell 2, Middle Atlantic Division, the outlier is (x, y) = ($8,168,230, $3,120,007).
In Cell 9, Pacific Division, the low ρ , -0.63281 was caused by zero variance in the last group with 1 unit in it.
The resulting ρ , 2.06985, was computed by deleting the last group.

TABLE II.6 The Estimated Trends by Four Ratio Procedures, Correlation Coefficients, and Error Variance Parameters (λ, ρ) for Each Imputation Cell (Group or Firm)
December 1982 - SIC 562

Imputation Cell	n	$R^{(1)}$	$R^{(2)}$	$R^{(4)}$	$R^{(3)}$	r	r^2	λ	ρ
I.									
Full reported data	1445 (20)	1.472	1.493	1.573	1.440	0.92234	0.85071	1.3394	1.8169
	1444 (20)	1.480	1.534	1.574	1.440	0.98703	0.97424	1.3663	1.8151
II.									
1. GP 1	785 (20)	1.409	1.356	1.500	1.394	0.76996	0.59284	1.1298	1.8271
	784 (20)	1.423	1.492	1.501	1.395	0.96282	0.92702	0.5363	1.8995
2. GP 2	660 (20)	1.549	1.549	1.660	1.621	0.9884	0.9769	1.8858	1.7873
III.									
1. Firm 2	306 (20)	1.353	1.060	1.399	1.352	0.90949	0.82717	0.7530	1.8921
	305 (20)	1.372	1.372	1.403	1.352	0.96269	0.92677	1.3562	1.8353
2. Firm 3	267 (20)	1.451	1.479	1.535	1.514	0.95710	0.91604	0.0065	2.32177
3. Firm 4	212 (20)	1.623	1.641	1.600	1.631	0.98766	0.97547	0.0213	2.21442
4. Firm 6	660 (20)	1.549	1.549	1.660	1.621	0.9884	0.9769	1.8858	1.7873

The outlier for the full reported data is (x, y) = ($8,168,230, $3,120,007), which is also in Group 1 and Firm 2.

TABLE II.7 The Estimated Trends by Four Ratio Procedures, Correlation Coefficients and Error Variance Parameters (λ , ρ) for Each Imputation Cell (Firm size x Sales size (Use Median as Cutoff))
December 1982 - SIC 562

Imputation Cell											
	Firm	Sales	n	$R^{(1)}$	$R^{(2)}$	$R^{(4)}$	$R^{(3)}$	r	r^2	λ	ρ
1.	Firm 2	(0, Median)	153 (20)	1.326	1.292	1.381	1.321	0.98167	0.96368	0.200835	2.04874
		(0, $33,800)									
2.	Firm 2	(Median, ∞)	153 (20)	1.376	1.021	1.417	1.484	0.9225	0.85100	0.011321	2.27264
		($33,800, ∞)	152 (20)	1.413	1.393	1.424	1.484	0.9550	0.91203	1.339371	1.81692
3.	Firm 3	(0, Median)	133 (20)	1.451	1.427	1.540	1.526	0.90496	0.81895	0.66633	1.86902
		(0, $43,450)									
4.	Firm 3	(Median, ∞)	134 (20)	1.451	1.491	1.530	1.495	0.95264	0.90753	0.001327	2.48916
		($43,450, ∞)									
5.	Firm 4	(0, Median)	106 (10)	1.673	1.616	1.618	1.652	0.90450	0.81812	21.72	1.5390
		(0, $58,400)									
6.	Firm 4	(Median ∞)	106 (10)	1.572	1.647	1.583	1.578	0.98822	0.97658	0.000934	2.42714
		($58,400, ∞)									
7.	Firm 6	(0, Median)	330 (20)	1.666	1.683	1.708	1.633	0.7869	0.6192	32.16	1.5262
		(0, $80,100)	327 (20)	1.620	1.640	1.666	1.603	0.8954	0.8017	61.70	1.4304
8.	Firm 6	(Median, ∞)	330 (20)	1.496	1.538	1.612	1.581	0.9868	0.9737	0.0178	2.1494
		($80,100, ∞)									

In Cell (Firm 6, Sales < Median), after 3 outliers were deleted, there are 12 units with some reported values (x, y) = ($26,667, $50,710), this would lower the group variance and hence the ρ .

TABLE II.8 The Estimated Trends by Four Ratio Procedures, Correlation Coefficients and Error Variance Parameters (λ , ρ) For Each Imputation Cell (Firm x Sales (Use 1/4 Quantiles as cutoff)) December 1982 - SIC 562

Imputation Cell			n	$R^{(1)}$	$R^{(2)}$	$R^{(4)}$	$R^{(3)}$	r	r^2	λ	ρ
	Firm	Sales									
1.	Firm 2	(0, 1/4 Q) (0, $18,500)	76 (10)	1.239	1.208	1.313	1.246	0.99169	0.98346	0.041588	2.20597
2.	Firm 2	(1/4 Q, Median) ($18,500, $33,800)	77 (10)	1.428	1.373	1.449	1.518	0.87163	0.75974	0.333467	1.96197
3.	Firm 2	(Median, 3/4 Q) ($33,800, $80,300)	76 (10)	1.406	1.381	1.426	1.464	0.78062	0.60937	0.654947	1.80639
4.	Firm 2	(3/4 Q, ∞) ($80,300, ∞)	77 (10) 76 (10)	1.351 1.419	0.955 1.397	1.409 1.422	1.521 1.522	0.93301 0.94736	0.87050 0.89749	0.092224 0.175554	2.07687 2.02440
5.	Firm 3	(0, 1/4 Q) (0, $22,600)	66 (10)	1.454	1.321	1.538	1.575	0.96904	0.93904	2.2736	1.72413
6.	Firm 3	(1/4 Q, Median) ($22,600, $43,450)	67 (10)	1.449	1.522	1.542	1.478	0.80645	0.65037	0.000035	2.81923
7.	Firm 3	(Median, 3/4 Q) ($43,450, $92,050)	67 (10)	1.483	1.542	1.561	1.539	0.90233	0.8142	0.38222	1.91607
8.	Firm 3	(3/4 Q, ∞) ($92,050, ∞)	67 (10)	1.429	1.478	1.499	1.408	0.94524	0.89348	0.002945	2.3891
9.	Firm 4	(0, 1/4 Q) (0, $30,400)	53 (5)	1.693	1.706	1.641	1.627	0.93034	0.86554	0.265782	1.98090

II.8 - continued

Firm	Sales	n	$R^{(1)}$	$R^{(2)}$	$R^{(4)}$	$R^{(3)}$	r	r^2	λ	ρ
10. Firm 4	(1/4 Q, Median) ($30,400, $58,400)	53 (5)	1.659	1.580	1.594	1.701	0.84126	0.70771	2162401	0.47233
11. Firm 4	(Median, 3/4 Q) ($58,400, $100,400)	53 (5)	1.524	1.450	1.526	1.567	0.85201	0.72593	16.90	1.52738
12. Firm 4	(3/4 Q, ∞) ($100,400, ∞)	53 (5)	1.650	1.715	1.639	1.610	0.98982	0.97974	0.000021	2.67882
13. Firm 6	(0, 1/4 Q) (0, $34,250)	165 (20) 164 (20)	1.616 1.614	1.720 1.685	1.734 1.705	1.583 1.581	0.6755 0.7725	0.4563 0.5968	10.27 48.50	1.6302 1.4576
14. Firm 6	(1/4 Q, Median) ($34,250, $80,100)	165 (20) 163 (20)	1.687 1.623	1.669 1.622	1.682 1.627	1.693 1.629	0.5139 0.7242	0.2641 0.5244	1690542 10.46	0.5330 1.5903
15. Firm 6	(Median, 3/4 Q) ($80,100, $211,700)	165 (20)	1.627	1.639	1.648	1.635	0.7680	0.5898	3.62	1.7033
16. Firm 6	(3/4 Q, ∞)	165 (20)	1.429	1.521	1.577	1.428	0.9839	0.9681	0.0065	2.2119

TABLE II.9 The Estimated Trends by Four Ratio Procedures, Correlation Coefficients and Error Variance Parameters (λ , ρ) For Each Imputation Cell (Region)
December 1982 - SIC 562

Imputation Cell Region	n	$R^{(1)}$	$R^{(2)}$	$R^{(4)}$	$R^{(3)}$	r	r^2	λ	ρ
1. Northeast	321 (20)	1.378	1.330	1.522	1.323	0.75383	0.56826	1.4169+15	-1.3364
	320 (20)	1.408	1.523	1.525	1.324	0.94396	0.89106	0.1548	2.0427
2. North Central	309 (20)	1.493	1.510	1.493	1.421	0.99288	0.98582	1.128659	1.82576
3. South	479 (20)	1.499	1.632	1.602	1.424	0.97665	0.95385	0.559551	1.90275
4. West	336 (20)	1.518	1.484	1.654	1.623	0.99495	0.98993	0.1519	2.05163

TABLE II.10 The Estimated Trends by Four Ratio Procedures, Correlation Coefficients and Error Variance Parameters (λ, ρ) For Each Imputation Cell (Region x Sales Size (Use Median as Cutoff))

Imputation Cell Region	Sales	n	$R^{(1)}$	$R^{(2)}$	$R^{(4)}$	$R^{(3)}$	r	r^2	λ	ρ
1. Northeast	< Median $78,510	160 (20)	1.321	1.454	1.454	1.279	0.92524	0.85607	1.805161	1.83342
2. Northeast	> Median $78,510	161*(20)	1.424	1.311	1.589	1.577	0.74202	0.55059	4.1262+28	-3.69819
		160 (20)	1.481	1.536	1.597	1.578	0.93327	0.87001	0.000575	2.51685
3. North Central	< Median $48,020	154 (20)	1.468	1.424	1.509	1.405	0.93788	0.87962	0.188048	2.01626
4. North Central	> Median $48,020	155 (20)	1.506	1.519	1.477	1.483	0.99196	0.98398	0.068077	2.07287
5. South	< Median $49,780	239 (20)	1.453	1.589	1.610	1.401	0.83725	0.70098	138.10	1.36723
6. South	> Median $49,780	240 (20)	1.554	1.638	1.594	1.567	0.97106	0.94296	0.067503	2.08208
7. West	< Median $50,000	168 (20)	1.689	1.673	1.728	1.643	0.76389	0.58353	51.39	1.45591
		167 (20)	1.654	1.611	1.684	1.626	0.90678	0.82225	257.59	1.26726
8. West	> Median $50,000	168 (20)	1.442	1.470	1.580	1.563	0.99483	0.98969	0.145494	2.04225

In Cell (Northeast, Sales > Median), the outlier is (x, y) = (8,168,230, 3,120,007).

In Cell (West, Sales < Median), the outlier is (x, y) = (40,868, 370,307).

TABLE II.11 The Number of Units in Each Cell
(Region x Firm)
December 1982 - SIC 562

Region	Firm 2	3	4	6	Total
Northeast	87	46	34	154	321
North Central	74	66	22	147	309
South	90	96	83	210	479
West	55	59	73	149	336
Total	306	267	212	660	1445

The data are reported list sample with current month sales and previous month sales greater than zero.

Table II.12 The Estimated λ , ρ for the Weighted Observation in Each Imputation Cell

			December 1982			February 1983		
	Group	Sales	n	λ	ρ	n	λ	ρ
SIC 562 Imputation Cells								
1	2	> $50,000	411	0.026606	2.11329	402	0.032087	1.99166
2	2	< $50,000	249	0.260983	2.02471	265	0.143026	1.93844
3	1	> $50,000	354	0.195763	1.98888	276	8.474825	1.66948
4	1	< $50,000	431	0.437924	1.93984	612	0.381239	1.85591

TABLE 4.1 The Trend Estimates from the Reported Data
By Using Different Imputation Procedures
December 1982 & February 1983

Imputation Cell	n	Current Imputation Rate	$R^{(1)}$	$R^{(2)}$	$R^{(3)}$	$R^{(4)}$	$\tilde{R}_{opt}$	$R^{(5)}$	$R^{(6)}$	Regression Intercept	Regression Slope
December 1982											
SIC 562 (Women's Ready-to-Wear Stores)											
1. (GP2, Sales > $50,000)	411	0.5036	1.52106	1.54094	1.61267	1.61725	1.52057	1.47189	1.44666	(48,453)	1.41406
2. (GP2, Sales < $50,000)	249	0.3789	1.66597	1.72212	1.62664	1.73094	1.66547	1.64721	1.71827	(695)	1.69748
3. (GP1, Sales > $50,000)	354	0.1418	1.39459	1.33449	1.47357	1.49737	1.38769	1.25059	0.59034	(125,271)	0.50748
4. (GP1, Sales < $50,000)	431	0.1265	1.42068	1.44370	1.38074	1.50176	1.41811	1.36481	1.29816	(8,009)	1.17653
SIC 521 (Building Materials Stores)											
1. (GP2, Sales > $183,333)	137	0.6437	0.92677	0.97203	0.93626	1.00237	0.92657	0.91061	0.94173	(54,547)	0.82222
2. (GP2, Sales < $183,333)	84	0.4852	0.89277	0.94650	0.87658	0.93483	0.89124	0.85254	0.95098	(-2,959)	0.97571
3. (GP1, Sales > $183,333)	192	0.1351	0.97218	0.93648	0.95190	0.93007	0.97351	0.99685	0.93147	(4,940)	0.92513
4. (GP1, Sales < $183,333)	222	0.1395	1.00451	1.04942	0.89526	0.91191	0.91057	0.88611	0.88595	(1,770)	0.87266
SIC 531 (Department Stores)											
1. (GP2, Sales > $501,667)	5319	0.1519	1.56952	1.56952	1.62702	1.62702	1.56941	1.50744	1.50744	(224,943)	1.42432
2. (GP2, Sales < $501,667)	1757	0.1706	1.68143	1.68143	1.66428	1.66428	1.68146	1.69219	1.69219	(-36,876)	1.79356
3. (GP1, Sales > $501,667)	148	0.0570	1.62703	1.60580	1.57068	1.56788	1.55795	1.51036	1.47331	(271,120)	1.38477
4. (GP1, Sales < $501,667)	333	0.0641	1.65118	1.66224	1.71847	1.71609	1.64963	1.63046	1.64721	(14,772)	1.61042
SIC 541 (Grocery Stores)											
1. (GP2, Sales > $146,667)	1462	0.4161	1.18624	1.18228	1.17634	1.17354	1.18626	1.19425	1.18533	(-6,236)	1.19131
2. (GP2, Sales < $146,667)	149	0.5729	1.16267	1.18361	1.20410	1.22414	1.16226	1.13410	1.16892	(2,996)	1.12967
3. (GP1, Sales > $146,667)	506	0.1481	1.13239	1.17594	1.13218	1.16311	1.13212	1.12197	1.18322	(-14,199)	1.19580
4. (GP1, Sales < $146,667)	311	0.1541	1.07197	1.16040	1.06719	1.13326	1.07133	1.05306	1.16198	(-151)	1.16322
SIC 551 (Motor Vehicle Dealer)											
1. (Sales > $375,000)	613	0.0982	0.89171	0.90502	0.92589	0.92641	0.89091	0.86217	0.89504	(50,409)	0.88255
2. (Sales < $375,000)	161	0.0640	0.90397	0.89842	0.97571	0.98100	0.89546	0.83573	0.86909	(22,887)	0.85596

TABLE 4.1 - continued

Imputation Cell	n	Current Imputation Rate	$R^{(1)}$	$R^{(2)}$	$R^{(3)}$	$R^{(4)}$	R_{opt}	$R^{(5)}$	$R^{(6)}$	Regression Intercept	Regression Slope
December 1982											
SIC 572 (Household Appliance Stores, Radio and TV Stores)											
1. (GP2, Sales > $58,333)	130	0.2460	1.32286	1.31982	1.35369	1.35555	1.32056	1.28299	1.24406	(54,777)	1.17382
2. (GP2, Sales < $58,333)	13	0.3043	1.08801	1.36970	1.33216	1.35183	1.05899	1.00803	1.42811	(-17,377)	1.80425
3. (GP1, Sales > $58,333)	225	0.1660	1.17643	1.28487	1.15201	1.27356	1.17250	1.11195	1.32427	(-14,837)	1.33793
4. (GP1, Sales < $58,333)	132	0.1953	1.33613	1.40222	1.50160	1.36319	1.33486	1.32143	1.64107	(-11,433)	1.76060
SIC 5812											
1. (GP2, Sales > $34,167)	474	0.6611	1.13026	1.12915	1.08311	1.09259	1.09258	1.08842	1.06906	(2,352)	1.06447
2. (GP2, Sales < $34,167)	200	0.3833	1.05649	1.11535	1.06011	1.10525	1.05544	1.02338	1.11639	(-187)	1.12386
3. (GP1, Sales > $34,167)	539	0.1195	1.11626	1.10026	1.07663	1.08092	1.07947	1.04750	1.02372	(8,251)	0.98091
4. (GP1, Sales < $34,167)	318	0.1011	1.09324	1.04328	1.06914	1.05289	1.05189	1.03715	0.99321	(2,357)	0.95030
SIC 5813											
1. (Sales > $7,500)	350	0.0690	1.04785	1.03866	1.06325	1.05779	1.06044	1.09098	1.07959	(-1,479)	1.08996
2. (Sales < $7,500)	70	0.0278	1.07722	1.05088	1.04612	1.05132	1.04872	1.04632	1.03314	(203)	1.02679
SIC 592											
1. (GP2, Sales > $32,500)	127	0.4622	1.64730	1.59919	1.70852	1.66140	1.64680	1.62668	1.50877	(24,464)	1.42892
2. (GP2, Sales < $32,500)	31	0.0882	1.92127	2.10487	1.99401	2.19026	1.75293	1.56630	2.12616	(-1,993)	2.21153
3. (GP1, Sales > $32,500)	253	0.0761	1.40029	1.36244	1.45297	1.43284	1.43518	1.46576	1.37062	(14,031)	1.31861
4. (GP1, Sales < $32,500)	131	0.1149	1.30098	1.25924	1.31623	1.27545	1.27305	1.24715	1.32227	(-551)	1.33933
February 1983											
SIC 562 (Women's Ready-to-Wear Stores)											
1. (GP2, Sales > $50,000)	402	0.4858	0.99044	0.98497	0.99421	1.01155	0.99038	0.98350	0.96377	(9,363)	0.95606
2. (GP2, Sales < $50,000)	265	0.3073	0.99974	1.04596	0.98244	1.02930	0.99927	0.98393	1.06110	(-3,101)	1.15453
3. (GP1, Sales > $50,000)	276	0.1083	0.95900	0.96977	1.01274	1.02742	0.95794	0.94658	0.89824	(20,688)	0.85777
4. (GP1, Sales < $50,000)	612	0.0751	1.01611	1.00743	1.08986	1.03371	1.01606	1.01537	0.98805	(1,936)	0.94450

TABLE 4.2 The Bias (Relative Bias (%)) of the Estimated Total
By Using Different Imputation Procedures
December 1982 & February 1983

Unit: U.S. Dollars

SIC	n	Estimated Reported Total	$R^{(1)}$	$R^{(2)}$	$R^{(3)}$	$R^{(4)}$	R_{opt}	$R^{(5)}$	$R^{(6)}$	Regression Estimator
December 1982										
562 (Women's Ready-to-Wear Stores)	1445	1,636,658,834	0	4,745,104	18,303,757	29,244,857	-495,262	-18,584,772	-49,527,438	73,566,518
(%)			(0)	(0.290)	(1.118)	(1.787)	(-0.030)	(-1.136)	(-3.026)	(4.495)
521 (Building Materials Stores)	635	1,933,849,833	0	7,282,673	14,289,364	24,869,789	-133,689	-5,946,523	-850,347	8,085,767
(%)			(0)	(0.377)	(0.739)	(1.286)	(-0.007)	(-0.307)	(-0.044)	(0.418)
531 (Department Stores)	7557	14,758,285,090	0	202,901	71,895,263	72,247,748	-340,476	-77,969,143	-78,411,093	425,515
(%)			(0)	(0.001)	(0.487)	(0.490)	(-0.002)	(-0.528)	(-0.531)	(0.003)
541 (Grocery Stores)	2428	12,374,995,572	0	51,782,545	-13,933,939	24,927,603	-409,814	-1,945,308	59,579,151	39,159,736
(%)			(0)	(0.418)	(-0.113)	(0.201)	(-0.003)	(-0.016)	(0.481)	(0.316)
551 (Motor Vehicle Dealers)	753	14,565,413,603	0	14,544,341	59,652,581	61,644,105	-3,169,391	-53,164,307	-5,047,197	57,052,376
(%)			(0)	(0.100)	(0.410)	(0.423)	(-0.022)	(-0.365)	(-0.035)	(0.392)
572 (Household Appliances, Radio,TV Stores)	500	571,806,693	0	6,876,815	2,920,688	5,941,472	-266,227	-4,071,594	13,020,508	-2,565,749
(%)			(0)	(1.203)	(0.511)	(1.039)	(-0.047)	(-0.712)	(2.277)	(-0.449)
5812 (Eating Places)	1531	6,055,819,018	0	-2,022,721	31,094,652	47,959,481	-720,701	-18,773,433	-41,890,604	24,760,883
(%)			(0)	(-0.033)	(0.513)	(0.792)	(-0.012)	(-0.310)	(-0.692)	(0.409)
5813 (Drinking Places)	420	642,146,909	0	-163,904	-647,688	-151,990	59,308	1,057,819	634,619	-2,087,831
(%)			(0)	(-0.026)	(-0.101)	(-0.024)	(0.009)	(0.165)	(0.099)	(-0.325)
592 (Liquor Stores)	542	1,740,095,873	0	-3,303,540	3,987,841	1,672,854	-345,361	-2,725,450	-18,100,886	2,683,752
(%)			(0)	(-0.190)	(0.229)	(0.096)	(-0.020)	(-0.157)	(-1.040)	(0.154)
February 1983										
562 (Women's Ready-to-Wear Stores)	1555	1,203,030,196	0	707,857	4,269,095	7,423,561	-61,613	-2,156,555	-5,351,261	10,393,002
(%)			(0)	(0.059)	(0.355)	(0.617)	(-0.005)	(-0.1792)	(-0.445)	(0.864)

TABLE 4.3 The MSE of the Estimated Total By Using Different Imputation Procedures
(And the Ratio to its Current Imputation Procedure)
December 1982 & February 1983

Unit: $\$10^6$

SIC	n	$R^{(1)}$	$R^{(2)}$	$R^{(3)}$	$R^{(4)}$	R_{opt}	$R^{(5)}$	$R^{(6)}$	Regression Estimator
December 1982									
562 (Women's Ready-to-Wear Stores)	1445	122,250,188	149,423,484	485,083,727	542,620,951	122,151,733	254,714,015	1,384,741,428	2,236,292,027
		(1)	(1.2223)	(3.9680)	(4.4386)	(0.9992)	(2.0835)	(11.327)	(18.293)
521 (Building Materials Stores)	635	373,293,260	508,178,456	634,679,773	810,125,338	373,238,170	405,332,341	425,909,403	482,528,485
		(1)	(1.3613)	(1.7002)	(2.1702)	(0.9999)	(1.0858)	(1.1410)	(1.2926)
531 (Department Stores)	7557	148,129,951	148,187,203	5,362,890,768	5,363,246,499	148,069,922	6,176,598,292	6,178,473,892	121,809,138
		(1)	(1.0004)	(36.204)	(36.206)	(0.9996)	(41.697)	(41.710)	(0.8223)
541 (Grocery Stores)	2428	975,155,627	2,732,057,367	1,383,526,763	2,492,746,633	975,070,021	1,306,876,310	2,902,580,498	2,341,096,627
		(1)	(2.8017)	(1.4188)	(2.5563)	(0.9999)	(1.3402)	(2.9765)	(2.4007)
551 (Motor Vehicle Dealers)	753	2,617,510,636	2,906,824,784	4,919,457,310	5,040,731,723	2,610,996,151	4,108,583,221	2,689,858,236	5,023,064,502
		(1)	(1.1105)	(1.8794)	(1.9258)	(0.9975)	(1.5697)	(1.0276)	(1.9190)
572 (Household Appliances, Radio/TV Stores)	500	25,604,966	60,624,441	41,244,917	52,676,300	25,562,855	35,419,411	134,778,334	83,056,483
		(1)	(2.3677)	(1.6108)	(2.0573)	(0.9984)	(1.3833)	(5.2638)	(3.2438)
5812 (Eating Places)	1531	410,794,971	477,812,049	1,134,373,917	1,339,936,313	410,512,304	551,848,720	1,168,011,026	669,950,846
		(1)	(1.1631)	(2.7614)	(3.2618)	(0.9993)	(1.3434)	(2.8433)	(1.6309)
5813 (Drinking Places)	420	4,150,594	4,205,883	4,691,465	4,309,152	4,145,352	5,229,200	4,577,926	8,792,564
		(1)	(1.0133)	(1.1303)	(1.0382)	(0.9987)	(1.2599)	(1.1030)	(2.1184)
592 (Liquor Stores)	542	110,351,071	150,489,266	183,571,784	118,555,857	110,152,107	119,873,690	428,523,608	144,471,410
		(1)	(1.3637)	(1.6635)	(1.0744)	(0.9982)	(1.0863)	(3.8833)	(1.3092)
February 1983									
562 (Women's Ready-to-Wear Stores)	1555	10,231,391	14,006,547	21,046,490	30,399,492	10,229,762	12,064,582	40,815,179	54,881,884
		(1)	(1.3690)	(2.0571)	(2.9712)	(0.9998)	(1.1792)	(3.9892)	(5.3641)

TABLE 4.4 The Imputation Rate by Sales Classes of Each Imputation Cell
(SIC 562 Women's Ready-to-Wear Stores)
December 1982

Sales Classes	Total Number of Establishments[1] Cell 1 (GP 2, Sales > $50,000)	Cell 2 (GP 2, Sales < $50,000)	Cell 3 (GP 1, Sales > $50,000)	Cell 4 (GP 1, Sales < $50,000)	Imputation Rate[2] (%) Cell 1 (GP 2, Sales > $50,000)	Cell 2 (GP 2, Sales < $50,000)	Cell 3 (GP 1, Sales > $50,000)	Cell 4 (GP 1, Sales < $50,000)
1 (< 9 x $\$10^3$)		24		71		58.33		15.5
2 (9 - 15)		67		59		41.79		18.6
3 (15 - 20)		55		76		49.09		10.5
4 (20 - 25)		51		62		41.18		12.9
5 (25 - 30)		64		70		34.38		10.0
6 (30 - 35)		37		52		24.32		9.6
7 (35 - 45)		76		78		36.84		11.54
8 (45 - 50)		42		29		21.43		13.80
9 (50 - 60)	43		49		41.86		12.24	
10 (60 - 75)	63		78		28.57		11.54	
11 (75 - 85)	31		36		35.48		8.33	
12 (85 - 95)	318		35		94.65		5.71	
13 (95 - 110)	52		24		46.15		8.33	
14 (110 - 130)	48		33		27.08		3.03	
15 (130 - 150)	34		30		26.47		10.00	
16 (150 - 170)	31		21		22.58		4.76	
17 (170 - 180)	11		3		9.09		0.	
18 (180 - 200)	16		14		25.00		14.29	
19 (200 - 220)	16		8		18.75		25.00	
20 (220 - 240)	16		14		0		7.14	
21 (240 - 280)	15		11		20.00		0.	
22 (280 - 300)	9		5		11.11		20.0	
23 (300 - 500)	55		24		7.27		20.83	
24 (500 - 1000)	45		22		4.44		63.64	
25 (1000+)	34		10		5.88		70.00	
Total	837	416	417	497	50.30	37.98	14.15	12.68

1 The data used are monthly list sample with current month and previous month sales greater than 0. The establishment totals do not include RICM code = 5.

2 The imputation rate as calculated by dividing the number of establishments with RICM code = 2 or 3 by the total number of establishments (not including RICM code = 5).

TABLE 4.4 The Imputation Rate by Sales Classes of Each Imputation Cell
(SIC 521 - Building Materials Stores)
December 1982

	Total Number of Establishments[1]				Imputation Rate[2] (%)			
Sales Classes	Cell 1 (GP 2, Sales > $183,333)	Cell 2 (GP 2, Sales < $183,333)	Cell 3 (GP 1, Sales > $183,333)	Cell 4 (GP 1, Sales < $183,333)	Cell 1 (GP 2, Sales > $183,333)	Cell 2 (GP 2, Sales < $183,333)	Cell 3 (GP 1, Sales > $183,333)	Cell 4 (GP 1, Sales < $183,333)
1 (< 9 $\$10^3$)		1		12		100.00		41.67
2 (9 - 15)		3		9		66.67		11.11
3 (15 - 20)		2		5		50.00		20.00
4 (20 - 25)		3		8		100.00		25.00
5 (25 - 30)		1		8		0		37.50
6 (30 - 35)		4		8		25.00		25.00
7 (35 - 45)		8		19		25.00		15.79
8 (45 - 50)		3		7		33.33		0
9 (50 - 60)		8		24		37.50		8.33
10 (60 - 75)		17		33		17.65		6.06
11 (75 - 85)		6		19		16.67		5.26
12 (85 - 95)		6		13		16.67		0
13 (95 - 110)		6		19		33.33		31.58
14 (110 - 130)		18		21		38.89		14.29
15 (130 - 150)		17		24		11.76		8.33
16 (150 - 170)		40		21		75.00		9.52
17 (170 - 180)		17		6		82.35		0.
18 (180 - 200)	40	9	12	2	67.50	88.89	16.67	50.00
19 (200 - 220)	33		17		63.64		23.53	
20 (220 - 240)	39		15		58.97		26.67	
21 (240 - 280)	67		20		62.69		20.00	
22 (280 - 300)	36		13		77.78		15.38	
23 (300 - 500)	133		63		64.66		6.35	
24 (500 - 1000)	53		57		58.49		8.77	
25 (1000+)	6		25		66.67		20.00	
Total	407	169	222	258	64.37	48.52	13.51	13.95

TABLE 4.4 The Imputation Rate by Sales Classes of Each Imputation Cell
(SIC 531 - Department Stores)
December 1982

Sales Classes	Total Number of Establishments[1]				Imputation Rate[2] (%)			
	Cell 1 (GP 2, Sales > $501,667)	Cell 2 (GP 2, Sales < $501,667)	Cell 3 (GP 1, Sales > $501,667)	Cell 4 (GP 1, Sales < $501,667)	Cell 1 (GP 2, Sales > $501,667)	Cell 2 (GP 2, Sales < $501,667)	Cell 3 (GP 1, Sales > $501,667)	Cell 4 (GP 1, Sales < $501,667)
1 (< 9 x $$10^3$)		2		3		100.00		33.33
2 (9 - 15)		0		5		0		0
3 (15 - 20)		0		1		0		0
4 (20 - 25)		2		4		0		0
5 (25 - 30)		3		3		66.67		0
6 (30 - 35)		3		0		33.33		0
7 (35 - 45)		7		5		57.14		0
8 (45 - 50)		1		2		100.00		0
9 (50 - 60)		9		3		55.56		0
10 (60 - 75)		15		12		33.33		8.33
11 (75 - 85)		6		9		33.33		0
12 (85 - 95)		8		10		0		0
13 (95 - 110)		5		12		0		8.33
14 (110 - 130)		20		14		0		7.14
15 (130 - 150)		41		30		7.32		13.33
16 (150 - 170)		64		23		6.25		8.70
17 (170 - 180)		36		9		8.33		0.
18 (180 - 200)		67		24		5.97		4.17
19 (200 - 220)		90		17		5.56		0
20 (220 - 240)		75		12		12.00		8.33
21 (240 - 280)		184		33		10.33		3.03
22 (280 - 300)		121		14		14.05		0
23 (300 - 500)		1362		114		20.12		8.77
24 (500 - 1000)	2836	7	85		20.49	42.86	3.53	
25 (1000+)	3477		73		10.87		8.22	
Total	6313	2128	158	359	15.19	17.06	5.70	6.41

TABLE 4.4 The Imputation Rate by Sales Classes of Each Imputation Cell
(SIC 541 - Grocery Stores)
December 1982

Sales Classes	Total Number of Establishments[1]				Imputation Rate[2] (%)			
	Cell 1 (GP 2, Sales > \$146,667)	Cell 2 (GP 2, Sales < \$146,667)	Cell 3 (GP 1, Sales > \$146,667)	Cell 4 (GP 1, Sales < \$146,667)	Cell 1 (GP 2, Sales > \$146,667)	Cell 2 (GP 2, Sales < \$146,667)	Cell 3 (GP 1, Sales > \$146,667)	Cell 4 (GP 1, Sales < \$146,667)
1 (< 9 x $\$10^3$)		2		29		50.00		34.48
2 (9 - 15)		6		31		0		19.35
3 (15 - 20)		11		26		0		19.23
4 (20 - 25)		17		24		41.18		29.17
5 (25 - 30)		22		28		18.18		10.71
6 (30 - 35)		27		27		14.81		11.11
7 (35 - 45)		33		36		12.12		2.78
8 (45 - 50)		174		19		92.53		10.53
9 (50 - 60)		28		29		25.00		27.59
10 (60 - 75)		25		30		16.00		6.67
11 (75 - 85)		10		35		0		2.86
12 (85 - 95)		14		14		50.00		14.29
13 (95 - 110)		10		15		20.00		20.00
14 (110 - 130)		17		29		11.76		13.79
15 (130 - 150)	4	12	4	14	0	25.00	0	0
16 (150 - 170)	94		18		88.30		5.56	
17 (170 - 180)	12		5		25.00		0	
18 (180 - 200)	35		28		25.71		7.14	
19 (200 - 220)	32		17		28.13		17.65	
20 (220 - 240)	41		14		12.20		21.43	
21 (240 - 280)	87		33		26.44		15.15	
22 (280 - 300)	52		11		30.77		9.09	
23 (300 - 500)	619		145		30.21		24.14	
24 (500 - 1000)	1014		190		43.10		14.21	
25 (1000+)	562		156		51.25		8.97	
Total	2552	408	621	370	41.61	50.49	14.65	15.41

TABLE 4.4 The Imputation Rate by Sales Classes of Each Imputation Cell
(SIC 551 Motor Vehicle Dealers (Franchised))
December 1982

	Total Number of Establishments[1]				Imputation Rate[2] (%)			
Sales Classes	Cell 1 (GP 2, Sales > $375,000)	Cell 2 (GP 2, Sales < $375,000)	Cell 3 (GP 1, Sales > $375,000)	Cell 4 (GP 1, Sales < $375,000)	Cell 1 (GP 2, Sales > $375,000)	Cell 2 (GP 2, Sales < $375,000)	Cell 3 (GP 1, Sales > $375,000)	Cell 4 (GP 1, Sales < $375,000)
1 (< 9 x $$10^3$)		42		42		4.76		4.76
2 (9 - 15)		1		1		0		0
3 (15 - 20)		4		4		0		0
4 (20 - 25)		0		0		0		0
5 (25 - 30)		0		0		0		0
6 (30 - 35)		1		1		0		0
7 (35 - 45)		6		6		16.67		16.67
8 (45 - 50)		3		3		0		0
9 (50 - 60)		1		1		0		0
10 (60 - 75)		3		4		0		0
11 (75 - 85)		0		0		0		0
12 (85 - 95)		4		4		25.00		25.00
13 (95 - 110)		5		5		0		0
14 (110 - 130)		6		6		16.67		16.67
15 (130 - 150)		6		6		0		0
16 (150 - 170)		3		5		0		0
17 (170 - 180)		3		3		0		0
18 (180 - 200)		5		5		0		0
19 (200 - 220)		11		11		9.09		9.09
20 (220 - 240)		7		7		14.29		14.29
21 (240 - 280)		18		19		5.56		5.26
22 (280 - 300)		7		7		14.29		14.29
23 (300 - 500)	45	32	47	32	11.11	6.25	10.64	6.25
24 (500 - 1000)	144		156		12.50		14.10	
25 (1000+)	464		479		6.90		8.35	
Total	653	168	682	172	8.42	6.55	9.82	6.40

TABLE 4.4 The Imputation Rate by Sales Classes of Each Imputation Cell
(SIC 572 - Household Appliance Stores, Radio and TV Stores)
December 1982

Sales Classes	Total Number of Establishments[1] Cell 1 (GP 2, Sales > \$58,333)	Cell 2 (GP 2, Sales < \$58,333)	Cell 3 (GP 1, Sales > \$58,333)	Cell 4 (GP 1, Sales < \$58,333)	Imputation Rate[2] (%) Cell 1 (GP 2, Sales > \$58,333)	Cell 2 (GP 2, Sales < \$58,333)	Cell 3 (GP 1, Sales > \$58,333)	Cell 4 (GP 1, Sales < \$58,333)
1 (< 9 x $\$10^3$)		1		36		0		33.33
2 (9 - 15)		3		18		66.67		27.78
3 (15 - 20)		2		16		50.00		18.75
4 (20 - 25)		1		29		100.00		10.34
5 (25 - 30)		1		17		100.00		5.88
6 (30 - 35)		7		9		0		0
7 (35 - 45)		2		21		0		14.29
8 (45 - 50)		6		5		33.33		20.00
9 (50 - 60)	2		4	18	50.00		0	27.78
10 (60 - 75)	9		29		22.22		31.03	
11 (75 - 85)	5		14		0		21.43	
12 (85 - 95)	5		13		20.00		15.38	
13 (95 - 110)	11		18		63.64		16.67	
14 (110 - 130)	15		13		26.67		15.38	
15 (130 - 150)	14		18		21.43		0	
16 (150 - 170)	11		15		9.09		6.67	
17 (170 - 180)	6		9		33.33		0	
18 (180 - 200)	12		15		41.67		33.33	
19 (200 - 220)	6		13		33.33		7.69	
20 (220 - 240)	6		13		33.33		0	
21 (240 - 280)	14		22		14.29		0	
22 (280 - 300)	6		7		33.33		0	
23 (300 - 500)	33		30		18.18		16.67	
24 (500 - 1000)	21		30		23.81		43.33	
25 (1000+)	11		8		9.09		0	
Total	187	23	271	169	24.60	30.43	16.61	19.53

TABLE 4.4 The Imputation Rate by Sales Classes of Each Imputation Cell
(SIC 5812 Eating Places)
December 1982

Sales Classes	Total Number of Establishments[1]				Imputation Rate[2] (%)			
	Cell 1 (GP 2, Sales > $34,167	Cell 2 (GP 2, Sales < $34,167)	Cell 3 (GP 1, Sales > $34,167)	Cell 4 (GP 1, Sales < $34,167)	Cell 1 (GP 2, Sales > $34,167)	Cell 2 (GP 2, Sales < $34,167)	Cell 3 (GP 1, Sales > $34,167)	Cell 4 (GP 1, Sales < $34,167)
1 (< 9 x 10^3)		53		97		60.38		5.15
2 (9 - 15)		49		93		32.65		5.48
3 (15 - 20)		47		65		40.43		12.31
4 (20 - 25)		65		38		32.31		26.32
5 (25 - 30)		63		47		36.51		8.51
6 (30 - 35)	11	70	9	36	45.45	31.43	22.22	13.89
7 (35 - 45)	131		83		38.93		3.61	
8 (45 - 50)	69		43		52.17		4.66	
9 (50 - 60)	136		72		52.21		8.33	
10 (60 - 75)	170		96		54.71		9.38	
11 (75 - 85)	86		41		60.47		9.76	
12 (85 - 95)	488		33		94.26		3.03	
13 (95 - 110)	92		52		67.39		9.62	
14 (110 - 130)	79		46		58.23		17.39	
15 (130 - 150)	48		32		50.00		40.63	
16 (150 - 170)	27		22		44.44		4.55	
17 (170 - 180)	10		6		40.00		16.67	
18 (180 - 200)	15		8		46.67		25.00	
19 (200 - 220)	10		10		30.00		10.00	
20 (220 - 240)	10		11		40.00		18.18	
21 (240 - 280)	14		13		50.00		15.38	
22 (280 - 300)	2		4		100.00		0	
23 (300 - 500)	25		25		32.00		40.00	
24 (500 - 1000)	12		7		41.67		14.29	
25 (1000+)	8		6		25.00		16.67	
Total	1443	347	619	356	66.11	38.33	11.95	10.11

TABLE 4.4 The Imputation Rate by Sales Classes of Each Imputation Cell
(SIC 5813 Drinking Places)
December 1982

Sales Classes	Total Number of Establishments[1]				Imputation Rate[2] (%)			
	Cell 1 (GP 1, Sales > $7,500)	Cell 2 (GP 1, Sales < $7,500)	Collapsed Cell 1, (Sales > $7,500)	Collapsed Cell 2, (Sales < $7,500)	Cell 1 (GP 1, Sales > $7,500)	Cell 2 (GP 1, Sales < $7,500)	Collapsed Cell 1 (Sales > $7,500)	Collapsed Cell 1 (Sales < $7,500)
1 (< 9 x 10^3)	24	71	24	72	12.50	2.82	12.50	2.78
2 (9 - 15)	59		62		8.47		8.06	
3 (15 - 20)	42		47		4.76		8.51	
4 (20 - 25)	32		35		0		0	
5 (25 - 30)	20		20		5.00		5.00	
6 (30 - 35)	17		20		5.88		5.00	
7 (35 - 45)	24		30		4.17		3.33	
8 (45 - 50)	12		14		0.		0.	
9 (50 - 60)	12		19		8.33		5.26	
10 (60 - 75)	19		26		5.26		7.69	
11 (75 - 85)	4		11		0.		0.	
12 (85 - 95)	12		15		0.		0.	
13 (95 - 110)	6		9		0.		0.	
14 (110 - 130)	7		9		14.29		11.11	
15 (130 - 150)	9		9		33.33		33.33	
16 (150 - 170)	4		5		25.00		20.00	
17 (170 - 180)	5		5		0.		0.	
18 (180 - 200)	4		4		0.		0.	
19 (200 - 220)	1		1		100.		100.	
20 (220 - 240)	3		3		66.67		66.67	
21 (240 - 280)	2		2		0.		0.	
22 (280 - 300)	1		1		0.		0.	
23 (300 - 500)	4		4		0.		0.	
24 (500 - 1000)	3		3		0.		0.	
25 (1000+)	0				0.			
Total	326	71	378	72	7.06	2.82	6.88	2.78

TABLE 4.4 The Imputation Rate by Sales Classes of Each Imputation Cell
(SIC 592 - Liquor Stores)
December 1982

Sales Classes	Total Number of Establishments[1]				Imputation Rate[2] (%)			
	Cell 1 (GP 2, Sales > \$32,500)	Cell 2 (GP 2, Sales < \$32,500)	Cell 3 (GP 1, Sales > \$32,500)	Cell 4 (GP 1, Sales < \$32,500)	Cell 1 (GP 2, Sales > \$32,500)	Cell 2 (GP 2, Sales < \$32,500)	Cell 3 (GP 1, Sales > \$32,500)	Cell 4 (GP 1, Sales < \$32,500)
1 (< 9 x $\$10^3$)		9		27		22.22		7.41
2 (9 - 15)		5		19		20.00		15.79
3 (15 - 20)		6		32		0.		3.13
4 (20 - 25)		4		28		0.		14.29
5 (25 - 30)		5		26		0.		19.23
6 (30 - 35)	4	5	18	16	0.	0.	11.11	12.50
7 (35 - 45)	12		39		8.33		2.56	
8 (45 - 50)	7		15		28.57		0.	
9 (50 - 60)	4		32		50.00		6.25	
10 (60 - 75)	10		29		30.00		17.24	
11 (75 - 85)	7		19		0.		5.26	
12 (85 - 95)	10		11		10.00		9.09	
13 (95 - 110)	46		17		71.74		0.	
14 (110 - 130)	68		23		75.00		4.35	
15 (130 - 150)	19		15		21.05		13.33	
16 (150 - 170)	17		14		23.53		7.14	
17 (170 - 180)	4		5		0.		0.	
18 (180 - 200)	8		3		25.00		33.33	
19 (200 - 220)	3		2		33.33		0.	
20 (220 - 240)	2		7		50.00		14.29	
21 (240 - 280)	6		5		16.67		0.	
22 (280 - 300)	3		3		66.67		0.	
23 (300 - 500)	3		10		66.67		10.00	
24 (500 - 1000)	4		7		0.		28.57	
25 (1000+)	1		2		0.		0.	
Total	238	34	276	148	46.22	8.82	7.61	11.49

TABLE 4.5 The Imputation Rate By Region of Each Imputation Cell
December 1982

	Total Number of establishments[1]					Imputation Rate[2] %				
	NE	NC	South	West	Total	NE	NC	South	West	Overall
SIC 562 (Women's Ready-to-Wear Stores)										
Cell 1 (GP2, Sales > $50,000)	219	243	236	139	1370	44.75	62.14	51.27	36.69	50.30
2 (GP2, Sales < $50,000)	45	109	151	111	416	22.22	46.79	32.45	43.24	37.98
3 (GP1, Sales > $50,000)	116	83	133	85	417	30.34	26.51	5.26	3.53	14.15
4 (GP1, Sales < $50,000)	92	130	159	116	497	13.04	20.77	8.81	8.62	12.68
SIC 521 (Building Materials Stores)										
Cell 1 (GP2, Sales > $183,333)	58	137	142	70	407	50.00	75.91	74.65	32.86	64.37
2 (GP2, Sales < $183,333)	18	37	73	41	169	50.00	62.16	57.53	19.51	48.52
3 (GP1, Sales > $183,333)	33	58	59	72	222	15.15	22.41	13.56	5.56	13.51
4 (GP1, Sales < $183,333)	48	78	94	38	258	14.58	21.79	8.51	10.53	13.95
SIC 531 (Department Stores)										
Cell 1 (GP2, Sales > $501,667)	1164	1790	2080	1279	6313	14.95	13.69	19.95	9.77	15.19
2 (GP2, Sales < $501,667)	352	553	1019	204	2128	7.67	11.39	25.02	8.82	17.06
3 (GP1, Sales > $501,667)	63	35	37	23	158	7.94	5.71	5.41	0	5.70
4 (GP1, Sales < $501,667)	98	107	127	27	359	3.06	8.41	7.87	3.70	6.41
SIC 541 (Grocery Stores)										
Cell 1 (GP2, Sales > $146,667)	481	659	630	782	2552	30.77	52.35	23.33	53.96	41.61
2 (GP2, Sales < $146,667)	35	186	141	46	408	14.29	83.87	24.11	23.91	50.49
3 (GP1, Sales > $146,667)	102	266	141	112	621	24.51	11.28	9.22	21.43	14.81
4 (GP1, Sales < $146,667)	90	90	147	43	370	12.22	13.33	17.01	20.93	15.41

1 The data used are monthly list sample with current month and previous month sales greater than 0. The establishment totals do not include RICM code = 5.

2 The imputation rate is calculated by dividing the number of establishments with RICM code = 2 or 3 by the total number of establishments (not including RICM code = 5).

TABLE 4.5 The Imputation Rate By Region of Each Imputation Cell
December 1982

	Total Number of Establishments					Imputation Rate %				
	NE	NC	South	West	Total	NE	NC	South	West	Overall
SIC 551 (Motor Vehicle Dealers)										
Cell 1 (GP1, Sales > $375,000)	87	139	275	152	653	10.34	15.11	6.91	3.95	8.42
2 (GP1, Sales < $375,000)	36	52	48	32	168	0	5.77	12.50	6.25	6.55
Collapsed Cell 1 (Sales > $375,000)	87	139	287	670	1183	10.34	15.11	6.97	2.54	5.66
2 (Sales < $375,000)	37	52	48	35	172	0	5.77	12.50	5.71	6.83
SIC 572 (Household Appliance Stores)										
Cell 1 (GP2, Sales > $58,333)	46	59	48	34	187	45.65	25.42	20.83	0	24.60
2 (GP2, Sales < $58,333)	4	8	10	1	23	50.00	25.00	30.00	0	30.43
3 (GP1, Sales > $58,333)	72	57	93	49	271	33.33	14.04	11.83	4.08	16.61
4 (GP1, Sales < $58,333)	46	43	54	26	169	30.43	13.95	20.37	7.69	19.53
SIC 5812										
Cell 1 (GP2, Sales > $34,167)	312	184	812	135	1443	78.21	45.65	73.52	21.48	66.11
2 (GP2, Sales < $34,167)	79	94	117	57	347	56.96	50.00	21.37	28.07	38.33
3 (GP1, Sales > $34,167)	130	165	191	133	619	18.46	12.12	12.57	4.51	11.95
4 (GP1, Sales < $34,167)	77	107	124	48	356	1.30	13.08	14.52	6.25	10.11
SIC 5813										
Cell 1 (GP1, Sales > $7,500)	79	109	66	72	326	1.27	5.50	13.64	9.72	7.06
2 (GP1, Sales < $7,500)	23	24	16	8	71	4.35	0	6.25	0	2.82
Collapsed Cell 1 (Sales > $7,500)	83	112	89	93	377	3.61	5.36	11.24	7.53	6.90
2 (Sales < $7,500)	83	113	89	93	378	3.61	5.31	11.24	7.53	6.88
SIC 592										
Cell 1 (GP2, Sales > $32,500)	33	80	57	68	238	15.15	88.75	3.50	47.06	46.22
2 (GP2, Sales < $32,500)	3	7	6	18	34	0	42.86	0	0	8.82
3 (GP1, Sales > $32,500)	56	63	95	62	276	7.14	3.17	7.37	12.90	7.61
4 (GP1, Sales < $32,500)	28	37	49	34	148	25.00	5.41	4.08	14.71	10.81

TABLE 4.6 The MSE of Estimated Total by Using Current Imputation Procedure for Selected Imputation Cells Definition - December 1982

Imputation Cell	No. of Cells	No. of units in the cells	Estimated Total ($)	MSE ($$10^6$)	Ratio of MSE
SIC 562					
1. Current cells GP x Sales (Use $50,000 as cutoff)	4	(411, 249) (354, 431)	1,636,658,834	122,250,188	1
2. Alternative cells					
a. GP x Sales (Use median)	4	(330, 330) (393, 392)		119,566,417	0.97805
b. GP x Sales (Use 1/4 Quantiles)	8	(165, 165) (196, 197)		68,342,418	0.55904
c. GP x Sales (Use 1/8 Quantiles)	16	(82, 83) (98, 99)		61,021,393	0.49915
d. GP x Sales (Use 1/16 Quantiles)	32	(41, 42) (49, 50)		57,783,350	0.47267

TABLE 4.7 The MSE of the Estimated Totals by Using Current Imputation Procedure for Selected Imputation Cells Definitions - December 1982 Unit=U.S. $\$10^6$

		Current Procedure		GP x Sales (Median)		GP x Sales (1/4 quantiles)		GP x Sales (1/8 quantiles)	
SIC	n	(1)	Ratio of (1) to (1)	(2)	Ratio of (2) to (1)	(3)	Ratio of (3) to (1)	(4)	Ratio of (4) to (1)
562 (Women's Ready-to-Wear Stores)									
G1	785	36,952,577	1	37,019,470	1.0018	35,471,786	0.9599	35,027,136	0.9479
G2	660	85,297,591	1	82,546,947	0.9678	32,870,632	0.3854	25,994,256	0.3047
Total	1445	122,250,168	1	119,566,917	0.9780	68,342,418	0.5590	61,021,392	0.4992
521 (Building Materials Stores)									
G1		345,448,465	1	338,747,820	0.9806	356,841,719	1.0330	270,743,235	0.7837
G2		27,821,592	1	27,772,071	0.9982	27,470,249	0.9874	22,038,788	0.7921
Total	653	373,293,260	1	366,519,898	0.9819	384,311,968	1.0295	292,782,023	0.7843
531 (Department Stores)									
G1		4,693,806	1	5,716,520	1.2179	5,522,438	1.1765	3,088,053	0.6579
G2		143,436,144	1	106,422,269	0.7419	79,967,690	0.5575	57,825,719	0.4031
Total	7557	148,129,951	1	112,138,789	0.7570	85,490,128	0.5771	60,913,771	0.4112
541 (Grocery Stores)									
G1		239,747,880	1	240,844,904	1.0045	230,482,546	0.9613	226,599,170	0.9452
G2		735,407,747	1	741,121,153	1.0078	730,461,660	0.9933	734,406,538	0.9986
Total	2428	975,155,627	1	981,966,057	1.0070	960,944,205	0.9854	961,005,708	0.9855
551 (Motor Vehicle Dealers)									
Total	774	2,617,510,636	1	2,891,157,789	1.1045	2,833,811,341	1.0826	2,978,240,874	1.1378
572 (Household Appliance Stores, Radio and TV)									
G1		25,248,467	1	26,710,885	1.0519	24,194,844	0.9582	22,552,323	0.8932
G2		355,999	1	328,867	0.9238	256,819	0.7214	112,882	0.3171
Total	500	25,604,966	1	37,039,753	1.0560	24,451,663	0.9549	22,665,205	0.8852

5812 (Eating Places)										
G1		203,465,246	1	294,987,982	1.4498	276,687,952	1.3599	235,945,661	1.1596
G2		207,329,724	1	183,271,236	0.8839	56,074,361	0.2705	42,320,846	0.2041
Total	1531	410,794,971	1	478,259,218	1.1642	332,762,313	0.8100	278,266,507	0.6774
5813 (Drinking Places)									
Total	420	4,150,594	1	4,101,101	0.9881	4,218,577	1.0164	4,817,800	1.1607
592 (Liquor Stores)									
G1		29,888,405	1	30,699,921	1.0271	30,219,902	1.0111	26,297,719	0.8798
G2		80,462,665	1	83,199,618	1.0340	76,610,800	0.9521	48,256,563	0.5997
Total	542	119,351,071	1	113,890,539	1.0321	106,830,702	0.9681	74,554,781	0.6756