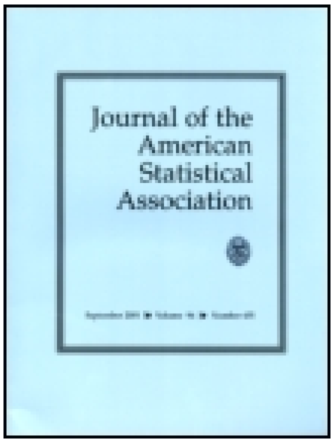

Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uasa20

A Bayesian Procedure for Imputing Missing Values in Sample Surveys

H. Y. Chiu [a] & J. Sedransk [b]
[a] Rolm Corporation , Santa Clara , CA , 95054 , USA
[b] Department of Statistics and Actuarial Science , University of Iowa , Iowa City , IA , 52242 , USA
Published online: 12 Mar 2012.

A Bayesian Procedure for Imputing Missing Values in Sample Surveys

H. Y. CHIU and J. SEDRANSK*

A new method is proposed for imputation of missing values in sample survey data. The procedure uses standard statistical methodology, permits a general specification of the nonresponse process, and does not impose specific model assumptions. Prior information from past similar surveys or from other sources may be incorporated in a routine manner.

KEY WORDS: Mean imputation; Random imputation; EM algorithm; Categorical data model; Dirichlet distribution; Order restrictions.

1. INTRODUCTION

Procedures for the treatment of missing values in large data sets have received considerable attention in the past several years. Emphasizing large-scale sample surveys, the Panel on Incomplete Data of the Committee on National Statistics, National Research Council, has investigated the field and has published their findings in three volumes (Madow, Nisselson, and Olkin 1983; Madow and Olkin 1983; Madow, Olkin, and Rubin 1983).

In many circumstances, it is desired to produce a *completed* (and consistent) data set to facilitate primary and secondary data analyses. To achieve this goal, an imputation procedure is commonly employed (Kalton and Kasprzyk 1982); that is, each missing value is replaced by a valid value of the random variable under study. Analysts of the data from a survey are frequently unfamiliar with details of the conduct of the survey. Such individuals will benefit from a more careful treatment of the missing values than they would be able to provide for themselves. This is particularly true if the missing data cannot be regarded as missing at random. Kalton and Kasprzyk (1982) provided a good summary of other reasons for the use of imputations and listed commonly employed imputation methods.

Many situations in which a completed data set is required also exhibit a sufficient lack of structure so that analysts are reluctant to postulate models such as multiple linear regression or multivariate normality to assist in making imputations. Current methodology for such situations is typified by the various "hot deck" procedures (Chapman 1976; Ford 1983; U.S. Bureau of the Census 1980a,b; Welniak and Coder 1980). Although there is no general agreement about the definition of a hot deck procedure, Ford (1983) defined it as "a duplication process . . . a reported value [from the sample] is duplicated to represent [the] missing value. The adjective 'hot' refers to imputing with values from the current sample" (p. 186). One type of hot deck method is exemplified by the sequential, previous observation, procedure where "the sample is put in some type of order . . . and for each missing value the previous reported value is duplicated" (Ford 1983, p. 187). Another prototypical hot deck method is the statistical matching procedure employed for the U.S. Census Bureau's March supplement to the Current Population Survey (CPS). Welniak and Coder (1980) described this statistical matching procedure as "a 'hot deck' system whereby nonrespondents and donors (i.e., respondents) are matched on detailed socio-economic characteristics using a sorting and merging procedure rather than a hot deck matrix. Once a matching donor is located, values reported by the donor are . . . imputed to the nonrespondent's record . . . all missing information is imputed from the same donor" (pp. 4–5).

It is difficult to assess the properties of such hot deck methods because they do not use standard statistical methodology. Analytical determination of properties of the sequential procedure are very difficult (Bailar and Bailar 1979, 1983; Ernst 1982), and an analytical determination of properties of the statistical matching procedure seems to be impossible. In addition, hot deck methods do not explicitly take into account the likely possibility that probability distributions for the respondent and nonrespondent subpopulations are different. One illustration of such differences is for the variable "income" in the March CPS supplement (see U.S. Bureau of the Census 1980b). Surveys of wholesale establishments provide a second illustration (see Secs. 2 and 3.2). With the hot deck methods, imputed values for nonrespondents are "donated" by respondents. The use of adjustment cells should mitigate the effect of using "respondent" donors but may not prevent substantial differences between respondents and nonrespondents within some of the cells.

Our objective is to suggest a method for imputation of missing values that uses standard statistical methodology, does not impose specific model assumptions, and permits a general specification of the nonresponse process. It would be desirable if such a procedure permitted prior information to be included in a routine manner.

Although our procedure is not intended for use in large-scale surveys, the methodology should be appropriate for many surveys in which (a) imputation is required, (b) there is insufficient structure to warrant the use for imputation

* H. Y. Chiu is Software Engineer, Rolm Corporation, Santa Clara, CA 95054. J. Sedransk is Professor and Chairman, Department of Statistics and Actuarial Science, University of Iowa, Iowa City, IA 52242. This work was partially supported by U.S. Bureau of the Census JSA 80-12. The authors thank J. H. Jinn for writing the computer programs needed for the analysis in Section 3, John Coder and Nash Monsour for providing the data, and the referee and an associate editor for constructive suggestions.

Journal of the American Statistical Association
September 1986, Vol. 81, No. 395, Theory and Methods

of specific statistical models (e.g., regression), and (c) there are expected to be substantial differences between respondents and nonrespondents.

In its simplest form, the method to be described assumes a categorized finite population, with each member having a probability of nonresponse. These nonresponse probabilities vary from category to category. Inferences about the number of nonrespondents who belong in each category are made using Bayesian predictive methodology. If desired, one may then apply directly the multiple imputation procedures described by Rubin (1978) and Herzog and Rubin (1983). (See Sec. 3.1.)

Although some aspects of the problem formulation and the use of the EM algorithm (Sec. 4.1) are similar to Little (1982), our objective is imputation rather than estimation of category proportions. In addition, we adopt a Bayesian approach. This permits us to incorporate formally in the inferential process the results of previous investigations such as surveys of the same population and special methodological studies of the relationships among respondents and nonrespondents. A special feature of the proposed prior distributions is the possibility of including known smoothness relationships among the category proportions.

The model and notation are introduced in Section 2. This section includes a specification of the nonresponse process and possible forms for the prior distribution. Data useful for evaluation of the prior distribution are described. The methodology for random imputation for a single random variable is presented in Section 3. Numerical illustrations are given, together with some results of a numerical investigation of the effect on imputations of incorrect specification of the parameters of the prior distribution. Alternative procedures are described in Section 4. These include (a) mean imputation, (b) model-based methods, and (c) use of an alternative sample design (i.e., the subsampling of nonrespondents). Finally, in Section 5 there are two extensions to the case of several random variables.

2. MODEL AND NOTATION

Consider a finite population of N distinguishable elements. Assume that the random variable of interest, Y, can take on the values $Y_{(1)} < Y_{(2)} < \cdots < Y_{(D)}$, which can be specified before sampling. The imputation methodology to be described is also appropriate when Y is a categorized continuous variable and $Y_{(i)}$ is a measure of central tendency for the ith category. Let P_i denote the unknown proportion of population elements with $Y = Y_{(i)}$. Assuming, for simplicity, a random sample of size n selected with replacement, r respondents are observed, with r_i having $Y = Y_{(i)} (\sum_{i=1}^D r_i = r)$. Among the $n - r$ nonrespondents, denote by t_i the unknown number of elements having $Y = Y_{(i)}$. The objective is to predict $\{t_i\}$. Denoting the predicted values by $\hat{t}_1, \ldots, \hat{t}_D (\sum_{i=1}^D \hat{t}_i = n - r)$, the imputed values for the $n - r$ nonrespondents are $\hat{t}_i$ repetitions of $Y_{(i)}$ for $i = 1, \ldots, D$. Thus the completed data set consists of $r_i + \hat{t}_i$ repetitions of $Y_{(i)}$ with sample mean $n^{-1} \sum_{i=1}^D Y_{(i)} (r_i + \hat{t}_i)$. When Y is a categorized continuous variable, it may be preferable to add random residuals to the $Y_{(i)}$ so that the imputed values are $\{Y_{(i)} + \varepsilon_{ij}: i = 1, \ldots, D; j = 1, \ldots, \hat{t}_i\}$. The imputation procedure just described can be applied to a poststratum (or "adjustment cell").

To describe our general nonresponse process, let θ_i denote the proportion of elements in the population with $Y = Y_{(i)}$ who would, if sampled, be nonrespondents; $1 - \theta_i$ denotes the corresponding proportion of respondents. Thus the nonresponse probabilities may vary with $Y_{(i)}$. Table 1 summarizes the notation; for example, $P_i(1 - \theta_i)$ is the proportion of elements with $Y = Y_{(i)}$ and who would respond if sampled.

Assuming that a random sample of size n has been selected, the likelihood of $\{P_i\}$ and $\{\theta_i\}$ is proportional to

$$\prod_{i=1}^{D} \{P_i(1 - \theta_i)\}^{r_i} \left(\sum_{i=1}^{D} P_i \theta_i \right)^{n-r}. \tag{2.1}$$

Although the parameters in (2.1) are not identifiable without further specification, they will be identifiable in two common situations: (a) there is a prior distribution for $\{P_i\}$ or for $\{\theta_i\}$; or (b) a subsample of the initial nonrespondents is selected, and an intensive follow-up yields responses for these subsampled nonrespondents. In addition, for the more typical situation of multivariate responses, a modest relaxation of assumptions will permit identifiability of all parameters (see Sec. 5).

A familiar prior distribution for $\{P_i\}$ and $\{\theta_i\}$ is given by

$$f'(\{P_i\}) \propto \prod_{i=1}^{D} P_i^{\beta_i - 1}$$

$$\text{for } 0 \le P_i \le 1 \ (i = 1, \ldots, D), \ \sum_{i=1}^{D} P_i = 1, \tag{2.2}$$

$$h_i'(\theta_i) \propto \theta_i^{a_i - 1}(1 - \theta_i)^{b_i - 1}$$

$$\text{for } 0 \le \theta_i \le 1 \ (i = 1, \ldots, D), \tag{2.3}$$

and $\theta_1, \ldots, \theta_D$ and $\{P_i\}$ are independent.

In surveys carried out periodically (or irregularly) over time it is important, and increasingly common, to have special methodological studies to assess the magnitudes of the $\{\theta_i\}$. This is done to evaluate the nature and extent of the differences between respondents and nonrespondents. The Bayesian approach permits one to incorporate the results from such investigations in data-based prior distributions for $\{P_i\}$ and $\{\theta_i\}$. One such study at the U.S. Census Bureau followed the completion of the 1982 Census of Wholesale Trade. For example, for each Standard Industrial Classification (SIC) and for each of several important

Table 1. Notation for Nonresponse Process

Category	1	i	D
Value of Y	$Y_{(1)}$	$Y_{(i)}$	$Y_{(D)}$
Proportion of respondents	$P_1(1 - \theta_1)$	$P_i(1 - \theta_i)$	$P_D(1 - \theta_D)$
Proportion of nonrespondents	$P_1\theta_1$	$P_i\theta_i$	$P_D\theta_D$

variables, the values for the nonrespondents were obtained from administrative records. Thus one has a frequency distribution of Y for each of the "respondent" and "nonrespondent" subpopulations and the overall response rate. These data can be employed to assign prior distributions to $\{P_i\}$ and $\{\theta_i\}$ for use in planning for and analyzing the results from future surveys of analogous types of business establishments. There also have been several matches of the CPS (March supplement) and Internal Revenue Service (IRS) files. One of these matches is described by Greenlees, Reece, and Zieschang (1982, sec. 4.1). A more recent investigation included matching the CPS and IRS files (1982 income) for individuals in sampled households in the March 1983 CPS. That is, the U.S. Census Bureau has IRS total wage and salary income for both the respondents and nonrespondents in the March 1983 CPS and the overall response rate. In Section 3.2 some of the data just described is used to illustrate the assignment of values to the parameters, $\{(\beta_i, a_i, b_i): i = 1, \ldots, D\}$, of the prior distribution.

In many repeated surveys there will be familiarity with either the nonresponse process or the relationship among the $\{P_i\}$. Then, (2.2) or (2.3) should be modified. For example, in some circumstances, one might wish to specify unimodal smoothness restrictions such as

$$P_1 \le \cdots \le P_k \ge P_{k+1} \ge \cdots \ge P_D \quad \text{or} \quad \theta_1 \ge \theta_2 \ge \cdots \ge \theta_t \le \theta_{t+1} \le \cdots \le \theta_D,$$

or bimodal restrictions such as

$$P_1 \le \cdots \le P_k \ge P_{k+1} \ge \cdots \ge P_r \le \cdots \le P_t \ge \cdots \ge P_D.$$

Such order restrictions can be accommodated in the imputation procedures to be described. It is our experience that when such smoothness restrictions are applicable, the location of the maximal and minimal values (e.g., k and t for the unimodal restrictions) exhibit little change from one survey to the next. Chiu (1982) showed that moderate misspecification of the location of a maximum or minimum value has little effect on subsequent inferences. Sedransk, Monahan, and Chiu (1985) presented Bayesian methodology that accommodates uncertainty about the location of such maxima or minima.

To illustrate the methodology, the prior distribution specified by (2.2) and (2.3) will be considered. The smoothness condition $P_1 \le \cdots \le P_k \ge P_{k+1} \ge \cdots \ge P_D$ is explicitly added in Section 4.1. Using (2.1), (2.2), and (2.3), the posterior distribution of $\{P_i\}$ and $\{\theta_i\}$ is given by

$$f''(\{P_i\}, \{\theta_i\}) \propto \prod_{i=1}^{D} \{P_i(1-\theta_i)\}^{r_i} \left(\sum_{i=1}^{D} P_i\theta_i\right)^{n-r} \left\{\prod_{i=1}^{D} P_i^{\beta_i - 1}\right\} \times \left\{\prod_{i=1}^{D} \theta_i^{a_i-1}(1-\theta_i)^{b_i-1}\right\} \quad (2.4)$$

for $0 \le P_i \le 1$, $P_1 \le \cdots \le P_k \ge \cdots \ge P_D$, $\Sigma_{i=1}^{D} P_i = 1$, and $0 \le \theta_i \le 1$.

3. METHODOLOGY FOR RANDOM IMPUTATION FOR A SINGLE RANDOM VARIABLE

3.1 Methodology

The most appropriate imputation for $\{t_i: i = 1, \ldots, D\}$ is the predictive distribution

$$p(\{t_i\} \mid \{r_i\}, n) = \int f(\{t_i\} \mid \{(P_i, \theta_i): i = 1, \ldots, D\}, n - r) \times f''(\{(P_i, \theta_i): i = 1, \ldots, D\})\, d\mathbf{P}\, d\boldsymbol{\theta} \quad (3.1)$$

for $t_i \ge 0$ and $\Sigma_{i=1}^{D} t_i = n - r$. In (3.1), $f''(\{(P_i, \theta_i): i = 1, \ldots, D\})$ is the posterior distribution defined by (2.4) and

$$f(\{t_i\} \mid \{(P_i, \theta_i): i = 1, \ldots, D\}, n - r) = (n-r)! \prod_{i=1}^{D} (P_i\theta_i)^{t_i} \Big/ \Pi t_i!(\Sigma P_j\theta_j)^{n-r} \quad (3.2)$$

is the conditional multinomial probability function.

In practice, one may use $p(\{t_i\} \mid \{r_i\}, n)$ by simulating values of $\{t_i\}$ from it. As suggested by Herzog and Rubin (1983), "the objective of multiple imputations is to simulate this posterior distribution" (p. 220). Specifically, the simulation proceeds by selecting $\{(P_i, \theta_i): i = 1, \ldots, D\}$ from f'' and then choosing $\{t_i\}$ from (3.2) using the values of $\{(P_i, \theta_i)\}$ selected from f''. The process may be repeated to obtain multiple imputations and then used, if desired, to form an interval estimate for the population mean (see Herzog and Rubin 1983, secs. 2.2 and 2.3).

First consider simulation from (2.4) without the order restrictions. Noting the dependence of $\{P_i\}$ and $\{\theta_i\}$ because of the term $(\Sigma_{j=1}^{D} P_j\theta_j)^{n-r}$, one may write

$$f''(\{P_i\}, \{\theta_i\}) = \sum_{A} f''(\{P_i\}, \{\theta_i\} \mid \{y_i\}) g(\{y_i\}),$$

where

$$A = \left\{\{y_i\}: 0 \le y_i, \sum_{i=1}^{D} y_i = n - r\right\},$$

$$f''(\{P_i\}, \{\theta_i\} \mid \{y_i\}) = \left[\prod_{i=1}^{D} \{\beta(a_i + y_i, b_i + r_i)\}^{-1} \times \theta_i^{a_i+y_i-1}(1-\theta_i)^{b_i+r_i-1}\right] \times \left[\Gamma\left\{\sum_{i=1}^{D} (r_i + \beta_i + y_i)\right\} \times \left\{\prod_{i=1}^{D} \Gamma(r_i + \beta_i + y_i)\right\}^{-1} \prod_{i=1}^{D} P_i^{r_i+\beta_i+y_i-1}\right],$$

$$g(\{y_i\}) = N(\{y_i\})/G(\{y_i\}),$$

$$N(\{y_i\}) = \prod_{i=1}^{D} \{\beta(r_i + y_i, b_i + r_i) \times \Gamma(r_i + \beta_i + y_i)/\Gamma(y_i + 1)\},$$

and

$$G(\{y_i\}) = \sum_A N(\{y_i\}), \tag{3.3}$$

Thus one selects $\{y_i\}$ from $g(\{y_i\})$ and then chooses $\{P_i\}$, $\{\theta_i\}$ from $f''(\{P_i\}, \{\theta_i\} \mid \{y_i\})$. In some situations (e.g., large values or $n - r$) one cannot calculate $G(\{y_i\})$. Then one may select $\{y_i\}$, using the following algorithm.

(a) Select $\{y_i\}$ with probability $h(\{y_i\})$, where $h(\{y_i\})$ is the "equal probability" multinomial

$$h(\{y_i\}) = (n - r)!\{D^{n-r} \prod_{i=1}^{D} \Gamma(y_i + 1)\}^{-1}.$$

(b) Then retain $\{y_i\}$ with probability $N(\{y_i\})/L \cdot h(\{y_i\})$, where L is an upper bound for $N(\{y_i\})/h(\{y_i\})$.

(c) If $\{y_i\}$ from (a) is *not* retained in (b), repeat steps (a)–(c) until a $\{y_i\}$ is retained in (b).

This algorithm ensures that $\{y_i\}$ is selected with probability $g(\{y_i\})$.

3.2 Properties

We have applied the methodology presented in Section 3.1, using as prior information each of the two data sets briefly described in Section 2. For each of several SIC's (e.g., wholesale distributors of beer and ale) the first data set consists of complete (categorized) frequency distributions of Y for the "respondent" and "nonrespondent" subpopulations of wholesale establishments. The variables are 1982 total sales and annual payroll. In the first part of the study we chose the values of the parameters of the prior distributions in (2.2) and (2.3) so that $\beta_i/\sum_{j=1}^{D} \beta_j$ equals the observed proportion of establishments with $Y = Y_{(i)}$ and $a_i/(a_i + b_i)$ equals the observed proportion of those establishments with $Y = Y_{(i)}$ who are nonrespondents. Envisioning a random sample of n establishments, the prior information is assigned a value equivalent to a sample of m establishments; that is $\sum_{j=1}^{D} \beta_j$ is taken to be m and $(a_i + b_i)$ is taken to be m times the proportion of establishments with $Y = Y_{(i)}$. Although selection of the value for m is judgmental, studies of the effect on subsequent inferences of the choice of m provide objective guidance. See, for example, the results of the second part of our numerical investigation.

The second data set consists of frequency distributions of (categorized) 1982 total wage and salary income as reported to the IRS for respondent and nonrespondent households in the March 1983 CPS. For illustration, we have considered only those households for which joint returns were filed. For the first part of the study parameters of the prior distribution were assigned as described previously.

In Tables 2 and 3 we present some of the frequency distributions that were used in the numerical studies, the results of which are reported forthwith. Note that in each table Panel A gives the frequency distributions corresponding to the observed data. Each frequency distribution in Panel A was derived from a more detailed one by aggregating some of the original (wage and salary or sales) classes.

A data set completed with imputed values typically will be analyzed using a wide variety of statistical techniques. For simplicity, however, we have considered only the completed sample mean,

$$\hat{\bar{Y}}_n = \sum_{i=1}^{D} Y_{(i)}(r_i + \hat{t}_i)/n. \tag{3.4}$$

A program has been written in FORTRAN by J. H. Jinn to make multiple selections of the $\{t_i\}$ from (3.1) using the methodology described in Section 3.1. Letting $\hat{V}_n$ denote the sample variance calculated from the completed data and $(\hat{\bar{Y}}_{nj}, \hat{V}_{nj})$ denote the value of $(\hat{\bar{Y}}_n, \hat{V}_n)$ from the jth

Table 2. Distributions of 1982 Sales, Y, for Wholesale Distributors of Beer and Ale for Respondents, Nonrespondents, and All Distributors; Probability of Nonresponse Given Y

Size class (in $1,000)	Pr(Y | Respondent)	Pr(Y | Nonrespondent)	Pr(Y)	Pr(Nonresponse | Y)
		A. Observed data		
0–249	.041	.255	.089	.642
250–999	.205	.276	.221	.281
1,000–2,999	.345	.254	.325	.176
3,000–9,999	.297	.147	.263	.126
10,000–50,000	.111	.070	.102	.155
		B. Alternative 1		
0–249			.060	.642
250–999			.240	.281
1,000–2,999			.330	.176
3,000–9,999			.270	.126
10,000–50,000			.100	.155
		C. Alternative 2		
0–249			.089	.600
250–999			.221	.250
1,000–2,999			.325	.150
3,000–9,999			.263	.100
10,000–50,000			.102	.100

Table 3. Distributions of Wage and Salary Income, Y, for 1982 Reported by Families in March 1983 CPS Who Filed Joint Returns: Respondents, Nonrespondents, and All Families; Probability of Nonresponse Given Y

Size class (in $1)	Pr(Y | Respondent)	Pr(Y | Nonrespondent)	Pr(Y)	Pr(Nonresponse | Y)
		A. Observed data		
0–9,999	.145	.139	.144	.117
10,000–19,999	.285	.245	.280	.106
20,000–29,999	.334	.314	.332	.115
30,000–39,999	.176	.216	.181	.145
40,000–49,999	.060	.087	.063	.169
		B. Alternative 1		
0–9,999			.110	.117
10,000–19,999			.290	.106
20,000–29,999			.342	.115
30,000–39,999			.191	.145
40,000–49,999			.067	.169
		C. Alternative 2		
0–9,999			.144	.117
10,000–19,999			.280	.106
20,000–29,999			.332	.115
30,000–39,999			.181	.100
40,000–49,999			.063	.080
		D. Alternative 3		
0–9,999			.144	.117
10,000–19,999			.280	.106
20,000–29,999			.332	.115
30,000–39,999			.181	.100
40,000–49,999			.063	.050

replication ($j = 1, \ldots, K$), we calculate

$$\hat{\hat{Y}}_{n\cdot} = \sum_{j=1}^{K} \hat{\hat{Y}}_{nj}/K,$$

$$\hat{V}_{n\cdot} = \sum_{j=1}^{K} \hat{V}_{nj}/K,$$

$$\hat{V}_B = \sum_{j=1}^{K} (\hat{\hat{Y}}_{nj} - \hat{\hat{Y}}_{n\cdot})^2/(K - 1),$$

and

$$\hat{V}_V = \sum_{j=1}^{K} (\hat{V}_{nj} - \hat{V}_{n\cdot})^2/(K - 1),$$

together with the sample mean, $\hat{\hat{Y}}_r$, and sample variance, $\hat{V}_r$, from the respondents' data. This set of quantities includes those needed to implement Herzog and Rubin's (1983) interval estimate for the population mean.

The objectives of the first part of the study are to investigate (a) the relationships of $\hat{\hat{Y}}_{n\cdot}$ to $\hat{\hat{Y}}_r$ and $\hat{V}_{n\cdot}$ to $\hat{V}_r$, and (b) the *variability* of $\hat{\hat{Y}}_{n\cdot}$, $\hat{V}_{n\cdot}$, and $\hat{V}_B$ over *different* selections of K replications. For this investigation we use the more extensive wage and salary distributions (with 10 wage and salary classes), take $K = 2$ (Herzog and Rubin 1983, p. 222), and choose r_i to be equal to its prior expected value, $n\beta_i b_i/(a_i + b_i) \sum \beta_j$. Thus the response rate in the sample is about .88. We consider $(n, m) = (200, 200)$, $(200, 100)$, $(400, 200)$, and $(400, 100)$, where n denotes the total sample size and m denotes the assigned value of the prior information.

Define $R_1 = \hat{\hat{Y}}_{n\cdot}/\hat{\hat{Y}}_r$ and $R_2 = \hat{V}_{n\cdot}/\hat{V}_r$. Then among 13 examples the maximal values of $|R_1 - 1|$ and $|R_2 - 1|$ are .02 and .08, respectively. The small sizes of these numbers are to be expected, because in the original distributions (having 10 wage and salary classes) the probability of nonresponse given Y varies only from .10 to .19 (also see Table 3, Panel A). For a given (n, m), we find that $\hat{\hat{Y}}_{n\cdot}$ and $\hat{V}_{n\cdot}$ vary little from one set of $K = 2$ replications to another. On the other hand, $\hat{V}_B$ varies substantially; the effect of this variation on $\{\hat{V}_B + (n^{-1})\hat{V}_{n\cdot}\}^{1/2}$, however, is relatively small because $\hat{V}_B$ is dominated by $n^{-1}\hat{V}_{n\cdot}$. These results also hold for two variants of the wage and salary distributions: (a) increasing Pr(nonresponse | Y) for each class by about .08, but retaining the overall distribution of Y; and (b) retaining the original values of Pr(nonresponse | Y), but adjusting the overall frequency distribution of Y so that the observed frequency for class i becomes that for class $(i + 1)$.

In the second part of the study, we investigate the effect on $\hat{\hat{Y}}_n$ [see (3.4)] of making poor choices of the prior parameters in (2.2) and (2.3) (e.g., basing the choice of $\{\beta_i, a_i, b_i\}$ on an outdated data set). We assume that the $\{r_i\}$ have been observed and that there are two choices for selection of the prior parameters: (a) the observed data (given in Panel A of Table 2 or Table 3), regarded here as being "correct," and (b) an alternative, "incorrect," data set. Using (a) or (b) will lead to different imputations. The expectations of $\{t_i: i = 1, \ldots, D\}$ (conditional on n and $\{r_i\}$) are denoted by $E_j''(t_i)$ ($i = 1, \ldots, D$) and where $j = 1$ and $j = 2$ correspond, respectively, to the "correct" and "incorrect" data. Thus the expected difference between the values of $\hat{\hat{Y}}_n$ obtained by using (a) and (b) is

$$n^{-1}\sum_{i=1}^{D} Y_{(i)}\{E_1''(t_i) - E_2''(t_i)\}$$

$$= n^{-1}(n - r)\sum_{i=1}^{D} Y_{(i)}\left\{E_1''\left(\frac{P_i\theta_i}{\sum P_k\theta_k}\right) - E_2''\left(\frac{P_i\theta_i}{\sum P_k\theta_k}\right)\right\}. \quad (3.5)$$

Since the magnitude of (3.5) will be greatly affected by the values of $\{Y_{(i)}: = i = 1, \ldots, D\}$, we consider instead $\{|\Delta_i|: i = 1, \ldots, D\}$, where

$$\Delta_i = E_1''\left(\frac{P_i\theta_i}{\sum P_k\theta_k}\right) - E_2''\left(\frac{P_i\theta_i}{\sum P_k\theta_k}\right). \quad (3.6)$$

For each of several alternative data sets, presented in Panels B and C of Table 2 and Panels B, C, and D of Table 3, we have estimated $\{\Delta_i: i = 1, \ldots, D\}$ by using 30 replications. Since with only 30 replications the standard errors of our estimates tend to be relatively large, our comparisons are made in terms of $\{|t(i)|: i = 1, \ldots, D\}$, where

$$t(i) = \{\overline{E}_1(i) - \overline{E}_2(i)\}/\{\hat{V}_{\overline{E}_1(i)} + \hat{V}_{\overline{E}_2(i)}\}^{1/2} \quad (3.7)$$

and $\overline{E}_j(i)$ and $\hat{V}_{\overline{E}_j(i)}$ estimate $E_j''(P_i\theta_i/\sum P_k\theta_k)$ and the sampling variance of $\overline{E}_j(i)$.

Table 4 presents the values of $\{|t(i)|: i = 1, \ldots, 5\}$ for comparisons of (a) the observed data with alternatives 1 and 2 for the sales distributions (Table 2) (here, $n = 40$ and $m = 10$ and 20), and (b) the observed data with alternatives 1–3 for the wages and salaries distributions (Table 3) (here, $n = 90$ and $m = 2$, 10, 20, and 30).

Although the quantities in Table 4 are subject to considerable sampling variability, several conclusions can be drawn. First, small to moderate changes in the prior specifications, together with at most a moderate weight given to the prior information, have a small effect on the imputations. For example, from Table 3 it is clear that alternatives 1 and 2 are each "close" to the observed data and from Table 4 the values of the $\{|t(i)|\}$ corresponding to these alternatives are all small, even for $m/n = 1/3$.

Larger changes in the prior specifications, together with a substantial weight given to the prior information, have a substantial effect on the imputations. In many applications, the latter problem can be avoided because there will be knowledge about changes between the initial study (whose data are used to choose the values of the prior parameters) and the current survey. For example, efforts to reduce the nonresponse rate for the larger units or changes in the distribution of Y due to general (economic) inflation will be known. Then, one may devalue the prior information (e.g., by decreasing m). The effect of decreasing m may be seen in Table 4 by looking at the results for alternatives 1 and 2 in Panel A and alternative 3 in Panel B. For the latter case, $|t(5)| = 2.85$, .68, and .41 for $m = 20$, 10, and 2, respectively. From Table 3 it can be seen easily that for the fifth class Pr(nonresponse | Y) is .169 for the observed, "correct" data and .050 for alternative 3. Similar results can be seen by looking at alternatives 1 and 2 in Panel A of Table 4 and comparing the distributions in Panels A–C of Table 2.

Finally, large changes in the prior specifications, together with a small weight given to the prior information, have only a modest effect on the imputations.

Table 4. Effect on Imputations of Changes in the Specification of Prior Parameters

Comparison of observed data with:	n	m	|t(1)|	|t(2)|	|t(3)|	|t(4)|	|t(5)|
		A. Sales distributions					
Alternative 1	40	20	1.70	2.36	.27	.06	1.29
		10	1.15	.75	.26	.50	.60
Alternative 2	40	20	.35	.38	.25	.46	1.94
		10	.41	.01	.10	.14	0.55
		B. Wage and salary distributions					
Alternative 1	90	30	1.04	.30	.04	1.33	.03
		20	.82	.03	.12	1.03	.39
		10	.24	1.13	1.04	.51	.41
Alternative 2	90	20	.46	.19	.16	.45	.52
		10	.54	1.00	1.26	.17	1.24
		2	.36	.50	.22	1.25	.62
Alternative 3	90	20	.59	.34	.54	.19	2.85
		10	.56	.57	.08	.64	.68
		2	.13	1.00	1.01	.74	.41

NOTE: The alternative distributions are given in Tables 2 and 3; $|t(i)|$ is defined in (3.7), and (n, m) are the sample size and value assigned to the prior information.

4. ALTERNATIVE METHODOLOGY

4.1 Mean Imputation

An alternative to the random imputation method described in Section 3.1 is to use mean imputation. One possibility is to take

$$\hat{t}_i = E''(t_i|\{r_i\}, n)$$

$$= (n - r)E''\left\{\left(P_i\theta_i \bigg/ \sum_{j=1}^{D} P_j\theta_j\right) \bigg| \{r_i\}, n\right\}.$$

In many circumstances an adequate approximation for $\hat{t}_i$ will be $(n - r)\hat{P}_i\hat{\theta}_i/\sum_{j=1}^{D} \hat{P}_j\hat{\theta}_j$, where $\hat{P}_i$ and $\hat{\theta}_i$ are the modes of the posterior distribution. A better approximation to $E''\{P_i\theta_i/\sum_{j=1}^{D} P_j\theta_j\}$ can be obtained by extending results presented in Lindley (1980).

To find the posterior mode of $\{P_i\}$ and $\{\theta_i\}$ one may use the EM algorithm (Dempster, Laird, and Rubin 1977). Letting $\mathbf{y}$ denote the observed data and $\mathbf{x}$ the complete data, $\mathbf{y} = (\{r_i\}, n)$ and $\mathbf{x} = (\{r_i\}, \{t_i\})$. The likelihood of $\{P_i\}$ and $\{\theta_i\}$, given the complete data $\mathbf{x}$, is

$$l(\{P_i\}, \{\theta_i\} \mid \mathbf{x}) \alpha \prod_{i=1}^{D} \{P_i(1 - \theta_i)\}^{r_i} \prod_{i=1}^{D} (P_i\theta_i)^{t_i}. \quad (4.1)$$

Using (2.2), (2.3), (4.1), and $P_1 \le \cdots \le P_k \ge \cdots \ge P_D$, the corresponding posterior distribution of $\{P_i\}$ and $\{\theta_i\}$ is

$$f''(\{P_i\}, \{\theta_i\} \mid \mathbf{x}) \alpha \prod_{i=1}^{D} P_i^{(r_i+t_i+\beta_i-1)} \prod_{i=1}^{D} \theta_i^{(t_i+a_i-1)}(1 - \theta_i)^{(r_i+b_i-1)}$$

for $0 \le P_i \le 1$, $P_1 \le \cdots \le P_k \ge \cdots \ge P_D$, $\sum_{i=1}^{D} P_i = 1$, and $0 \le \theta_i \le 1$.

The EM algorithm requires the following E steps and M steps.

E step. Calculate

$$\begin{aligned}
&E[\log f''(\{P_i\}, \{\theta_i\} \mid \mathbf{x}) \mid \mathbf{y}, \{P_i^{(v)}\}, \{\theta_i^{(v)}\}] \\
&= \text{constant} + \sum_{i=1}^{D} \left\{ (r_i + \beta_i - 1) + \left[(n - r) P_i^{(v)} \theta_i^{(v)} \right.\right. \\
&\quad \left.\left. \div \left(\sum_{j=1}^{D} P_j^{(v)} \theta_j^{(v)} \right) \right] \right\} \log P_i \\
&\quad + \sum_{i=1}^{D} \left\{ \left[\left\{ (n - r) P_i^{(v)} \theta_i^{(v)} \Big/ \sum_{j=1}^{D} P_j^{(v)} \theta_j^{(v)} \right\} \right.\right. \\
&\quad \left.\left. + a_i - 1 \right] \log \theta_i + (r_i + b_i - 1) \log(1 - \theta_i) \right\}, \quad (4.2)
\end{aligned}$$

where $P_i^{(v)}$ and $\theta_i^{(v)}$ denote the values of P_i and θ_i from the vth iteration. The prior expected values of P_i and θ_i (ignoring the order restrictions) may be used as the initial values, $P_i^{(0)}$ and $\theta_i^{(0)}$.

M step. Choose $\{P_i^{(v+1)}\}$ and $\{\theta_i^{(v+1)}\}$ to maximize (4.2) subject to $0 \le P_i \le 1$, $P_1 \le \cdots \le P_k \ge \cdots \ge P_D$, $\sum P_i = 1$, and $0 \le \theta_i \le 1$.

Using standard results in isotonic regression (Barlow, Bartholomew, Bremner, and Brunk 1972, sec. 1.2; Cryer and Robertson 1975),

$$\begin{aligned}
P_i^{(v+1)} &= \max_{1 \le s \le k} \max_{k \le t \le D} A(s, t) \equiv A(a, b) \\
&\qquad \text{for } i = a, \ldots, b \\
&= \max_{1 \le s \le i} \min_{i \le t \le a-1} A(s, t) \\
&\qquad \text{for } i = 1, \ldots, a - 1 \\
&= \min_{b+1 \le s \le i} \max_{i \le t \le D} A(s, t) \\
&\qquad \text{for } i = b + 1, \ldots, D, \quad (4.3)
\end{aligned}$$

where the first relation in (4.3) defines (a, b),

$$A(s, t) = \sum_{i=s}^{t} \bar{P}_i^{(v+1)} / (t - s + 1),$$

$$\begin{aligned}
\bar{P}_i^{(v+1)} = &\left\{ (n - r) \left[P_i^{(v)} \theta_i^{(v)} \Big/ \sum_{j=1}^{D} P_j^{(v)} \theta_j^{(v)} \right] \right. \\
&\left. + (r_i + \beta_i - 1) \right\} \\
&\times \left\{ n + \sum_{j=1}^{D} (\beta_j - 1) \right\}^{-1},
\end{aligned}$$

and for $i = 1, \ldots, D$,

$$\begin{aligned}
\theta_i^{(v+1)} = &\left\{ (n - r) \left[P_i^{(v)} \theta_i^{(v)} \Big/ \sum_{j=1}^{D} P_j^{(v)} \theta_j^{(v)} \right] + (a_i - 1) \right\} \\
&\times \left\{ (n - r) \left[P_i^{(v)} \theta_i^{(v)} \Big/ \sum_{j=1}^{D} P_j^{(v)} \theta_j^{(v)} \right] \right. \\
&\left. + (a_i - 1) + (r_i + b_i - 1) \right\}^{-1}.
\end{aligned}$$

The procedure is to repeat the EM algorithm until the change in $\{P_i^{(v)}\}$ and $\{\theta_i^{(v)}\}$ from one iteration to the next is sufficiently small. Note that algorithms to compute $\{P_i^{(v+1)}\}$ are given by Barlow et al. (1972, sec. 2) and Cryer and Robertson (1975).

Although both Herzog and Rubin (1983) and Kalton and Kasprzyk (1982) showed that mean imputation is sometimes unsatisfactory, the mean imputation method in this section is easier to implement than the random imputation method presented in Section 3.1. This is especially so when order restrictions are to be included.

4.2 Subsampling Nonrespondents

When there is substantial nonresponse in a sample survey, a common practice is to select a small subsample of the nonrespondents. More costly interviewing techniques are applied to this subsample, and it is expected that nearly complete responses will be obtained.

In the first phase of this two-phase sample design, a random sample of n observations is selected from the population with r_1 respondents. Among these r_1 respondents, r_{1i} have the value $Y_{(i)}(\sum_{i=1}^{D} r_{1i} = r_1)$. In the second phase, a random subsample of size $r_2 \le n - r_1$ is selected from the $(n - r_1)$ nonrespondents, and these elements are assumed to respond fully, with r_{2i} elements having the value $Y_{(i)}(\sum_{i=1}^{D} r_{2i} = r_2)$. Among the $n - r_1 - r_2$ remaining nonrespondents, let t_i denote the unknown number of elements having $Y = Y_{(i)}$.

With the prior distribution given by (2.2) and (2.3), the posterior distribution of $\{P_i\}$ and $\{\theta_i\}$ is given by

$$\begin{aligned}
f''(\{P_i\}, \{\theta_i\}) &\alpha \prod_{i=1}^{D} \{P_i(1 - \theta_i)\}^{r_{1i}} \prod_{i=1}^{D} (P_i\theta_i)^{r_{2i}} \\
&\times \left(\sum_{i=1}^{D} P_i\theta_i \right)^{n - r_1 - r_2} \prod_{i=1}^{D} P_i^{\beta_i - 1} \prod_{i=1}^{D} \theta_i^{a_i - 1}(1 - \theta_i)^{b_i - 1}, \quad (4.4)
\end{aligned}$$

where $0 \le P_i \le 1$, $P_1 \le \cdots \le P_k \ge \cdots \ge P_D$, $\sum_{i=1}^{D} P_i = 1$, and $0 \le \theta_i \le 1$.

Note from (4.4) that $\{P_i\}$ and $\{\theta_i\}$ are identifiable from the observed data alone. Thus if appropriate, noninformative prior distributions for $\{\theta_i\}$ or $\{P_i\}$ can be assumed. Another interesting possibility is to use noninformative prior distributions together with smoothness restrictions on $\{P_i\}$ or $\{\theta_i\}$, reflecting general familiarity with the distribution of Y or the nonresponse process.

Since (4.4) has the same form as (2.4), inference about $\{t_i\}$ may proceed exactly as described in Sections 3.1 and 4.1.

4.3 Model-Based Approaches

It was stated in Section 1 that our objective is to develop methodology for surveys in which there is insufficient structure to warrant the use for imputation of statistical models such as regression models. Model-based approaches have been proposed by Heckman (1976) and Greenlees et al. (1982) among others. Heckman's methodology is appropriate under restrictive assumptions (multiple linear regression, multivariate normality, and a rather specific nonresponse process). Although this methodology has been applied in missing data problems, we are unaware of any validation of the assumptions. The approach of Greenlees et al. (1982) is to postulate a linear regression between the variable of interest, Y, and covariates $\mathbf{X}$ while the probability of response is a logistic function of Y and other covariates, $\mathbf{Z}$. From their numerical evaluation, the Greenlees et al. (1982) method looks promising for applications for which their model assumptions are valid.

5. IMPUTATION METHODOLOGY FOR SEVERAL RANDOM VARIABLES

There are many possible generalizations to higher dimensions of the methodology described in Sections 3.1 and 4.1. In this section two possibilities are outlined. For simplicity, for the first procedure we consider only two categorical variables. For the second procedure we consider the general case of k variables but assume a monotone data pattern. In the multivariate case, in which the number of parameters is very large, it is anticipated that data-based or locally uniform prior distributions will be employed.

With the first method, it is assumed that each element in the finite population can be classified into a two-dimensional contingency table with regard to row variable X and column variable Y. In addition, the values of X and Y, which can be specified before sampling, are $\{X_{(i)}, Y_{(j)}: i = 1, \ldots, I, j = 1, \ldots, J\}$, where $X_{(1)} < \cdots < X_{(I)}$ and $Y_{(1)} < \cdots < Y_{(J)}$. Typically, survey data will be incomplete because of nonresponse in either the row or column variable. Often, census data is available. Then there is no missing data on X, say.

One nonresponse process is summarized in Table 5. Let P_{ij} denote the proportion of elements in the finite population with $X = X_{(i)}$ and $Y = Y_{(j)}$, $P_{i\cdot} = \sum_{j=1}^{J} P_{ij}$, and $P_{\cdot j} = \sum_{i=1}^{I} P_{ij}$. Denote by $\theta_{12(ij)}$ the proportion of elements with $(X = X_{(i)}, Y = Y_{(j)})$, who, if sampled, would respond to neither X nor Y; $\theta_{1(i)}$ the proportion of elements who would respond to X (i.e., $X = X_{(i)}$) but not Y; and $\theta_{2(j)}$ the proportion of elements who would respond to Y (i.e., $Y = Y_{(j)}$) but not X. Finally, define $\phi_{ij} = \{1 - \theta_{1(i)} - \theta_{2(j)} - \theta_{12(ij)}\}P_{ij}$. Then ϕ_{ij} is the proportion of elements in the finite population having $X = X_{(i)}$, $Y = Y_{(j)}$, and who would respond to both variables, if sampled. Similarly, $\theta_{2(j)}P_{\cdot j}$ is the proportion of elements having $Y = Y_{(j)}$, and who, if sampled, would respond $Y = Y_{(j)}$, but would not give any response about X. This nonresponse process is an extension of one first presented by Chen and Fienberg (1974, sec. 2).

Table 5. Response and Nonresponse Probabilities for a Finite Population Classified by Two Variables

	Fully Classified Table					
	$Y_{(1)}$		$Y_{(j)}$		$Y_{(J)}$	Row supplemental margins
$X_{(1)}$	ϕ_{11}	. . .	ϕ_{1j}	. . .	ϕ_{1J}	$\theta_{1(1)}P_{1\cdot}$
	$\vdots$		$\vdots$		$\vdots$	
$X_{(i)}$	ϕ_{i1}	. . .	ϕ_{ij}	. . .	ϕ_{iJ}	$\theta_{1(i)}P_{i\cdot}$
	$\vdots$		$\vdots$		$\vdots$	
$X_{(I)}$	ϕ_{I1}	. . .	ϕ_{Ij}	. . .	ϕ_{IJ}	$\theta_{1(I)}P_{I\cdot}$
			Column supplemental margins			Row × Column supplemental margin
	$\theta_{2(1)}P_{\cdot 1}$		$\theta_{2(j)}P_{\cdot j}$		$\theta_{2(J)}P_{\cdot J}$	$\sum_{i=1}^{I}\sum_{j=1}^{J} \theta_{12(ij)}P_{ij}$

NOTE: $\phi_{ij} = (1 - \theta_{1(i)} - \theta_{2(j)} - \theta_{12(ij)})P_{ij}$, $P_{i\cdot} = \sum_{j=1}^{J} P_{ij}$, and $P_{\cdot j} = \sum_{i=1}^{I} P_{ij}$ for $i = 1, \ldots, I$, $j = 1, \ldots, J$.

Assume as in Section 2 that a random sample of size n has been chosen with replacement. Let m_{ij} be the fully cross-classified count for the (i, j)th cell, r_i ($i = 1, \ldots, I$) the count of partially classified elements in the ith row without their column identity being known, c_j ($j = 1, \ldots, J$) the count of partially classified elements in the jth column without their row identity being known, and s the count of elements that respond to neither X nor Y. Finally, let r_{ij}, c_{ij}, and s_{ij} be the unknown counts for the (i, j)th cell corresponding to these r_i, c_j, and s elements, respectively.

For the observed data the likelihood is proportional to

$$\left[\prod_{i=1}^{I}\prod_{j=1}^{J} \{(1 - \theta_{1(i)} - \theta_{2(j)} - \theta_{12(ij)})P_{ij}\}^{m_{ij}}\right]\left\{\prod_{i=1}^{I} (\theta_{1(i)}P_{i\cdot})^{r_i}\right\} \times \left\{\prod_{j=1}^{J} (\theta_{2(j)}P_{\cdot j})^{c_j}\right\}\left\{\left(\sum_{i=1}^{I}\sum_{j=1}^{J} \theta_{12(ij)}P_{ij}\right)^{s}\right\}. \quad (5.1)$$

The parameters in (5.1) will be identifiable if any one of the following situations obtains: (a) $\theta_{12(ij)} = \theta_{1(i)}\theta_{2(j)}$ for all i, j; (b) $\theta_{12(ij)} = \theta$ for all i, j; (c) there is a prior distribution for $\{\theta_{12(ij)}\}$; or (d) a subsample of nonrespondents is selected and an intensive follow-up yields responses for these subsampled nonrespondents. Note that for the common situation in which X is known for all elements in the population condition (b) is satisfied.

The first step in the imputation procedure is to assign a prior distribution to $\{\theta_{1(i)}\}$, $\{\theta_{2(j)}\}$, $\{\theta_{12(ij)}\}$, and $\{P_{ij}\}$. For tractability, these may be of the beta or Dirichlet forms. Note that when X is known for all elements in the population $\theta_{12(ij)} = \theta_{2(j)} = 0$, thus simplifying the assignment of a prior probability distribution.

When $\theta_{12(ij)} = \theta_{2(j)} = 0$, the random imputation methodology presented in Section 3.1 can easily be applied to obtain (multiple) imputations for $\{r_{ij}: i = 1, \ldots, I, j = 1, \ldots, J\}$ conditional on $\{m_{ij}: i = 1, \ldots, I, j = 1, \ldots, J\}$ and $\{r_i: i = 1, \ldots, I\}$. For the general case, mean imputation is considerably simpler to use. As in Section 4.1, the EM algorithm may be employed to find the posterior modes of the parameters. These quantities are then used to predict the values of $\{r_{ij}\}$, $\{c_{ij}\}$, and $\{s_{ij}\}$. For further details, see Chiu (1982).

Finally, note that the nonresponse process just described can be generalized so that for elements with $X = X_{(i)}$ and $Y = Y_{(j)}$, the probabilities of nonresponse on X only, on Y only, and on both X and Y, are different for different i and j (Chiu 1982).

With the second method it is assumed that each element in the finite population can be classified by k categorical variables $Y_1, \ldots, Y_k$. For $j = 1, \ldots, k$, the values of Y_j are $\{y_i^{(j)}: i = 1, \ldots, c_j\}$. To illustrate, a monotone data pattern is postulated: Y_j is observed for all elements for which Y_{j+1} is observed. That is, of the k variables, variable 1 is the one that is the most observed. In sample surveys, reasonable approximations to monotone data patterns are often seen.

Assuming that a random sample of size n is selected with replacement, the response indicator function for a sampled element is specified by $\mathbf{d} = (d_1, d_2, \ldots, d_k)$, where d_j takes the values 1 or 0 according to whether Y_j is observed or not. For $i_1 = 1, \ldots, c_1$, let

$$P_{i_1} = \Pr(Y_1 = y_{i_1}^{(1)}),$$

$$\theta_{i_1} = \Pr(d_1 = 0 \mid Y_1 = y_{i_1}^{(1)}),$$

r_{i_1} = the number of elements in the sample having $Y_1 = y_{i_1}^{(1)}$ and $d_1 = 1$, and t_{i_1} = the unknown number of elements in the sample having $Y_1 = y_{i_1}^{(1)}$ and $d_1 = 0$. For $1 \le j \le k - 1$, $1 \le i_1 \le c_1, \ldots, 1 \le i_j \le c_j$, $1 \le i_{j+1} \le c_{j+1}$, let

$$P_{i_{j+1}|i_1,\ldots,i_j} = \Pr(Y_{j+1} = y_{i_{j+1}}^{(j+1)} \mid Y_l = y_{i_l}^{(l)} \quad \text{for } l = 1, \ldots, j),$$

$$\theta_{i_{j+1}|i_1,\ldots,i_j} = \Pr(d_{j+1} = 0 \mid Y_l = y_{i_l}^{(l)} \quad \text{for } l = 1, \ldots, j + 1),$$

$r_{i_{j+1}|i_1,\ldots,i_j}$ = the number of elements in the sample having $d_1 = d_2 = \cdots = d_{j+1} = 1$ and $Y_l = y_{i_l}^{(l)}$ for $l = 1, \ldots, j + 1$, and $t_{i_{j+1}|i_1,\ldots,i_j}$ = the unknown number of elements in the sample having $d_1 = \cdots = d_j = 1$, $d_{j+1} = 0$, and $Y_l = y_{i_l}^{(l)}$ for $l = 1, \ldots, j + 1$. Letting $\mathbf{Y}$ denote the observed data and $\mathbf{X}$ denote the complete data,

$$\mathbf{Y} = (\{r_{i_1}\}, \{r_{i_{j+1}|i_1,\ldots,i_j}\}, s, \{s_{i_1,\ldots,i_j}\})$$

and

$$\mathbf{X} = (\{r_{i_1}\}, \{r_{i_{j+1}|i_1,\ldots,i_j}\}, \{t_{i_1}\}, \{t_{i_{j+1}|i_1,\ldots,i_j}\}),$$

where

$$s = \sum_{i_1=1}^{c_1} t_{i_1}$$

and

$$s_{i_1,\ldots,i_j} = \sum_{i_{j+1}=1}^{c_{j+1}} t_{i_{j+1}|i_1,\ldots,i_j}.$$

First, we obtain the likelihood of $\boldsymbol{\phi} = (\{P_{i_1}\}, \{\theta_{i_1}\}, \{P_{i_{j+1}|i_1,\ldots,i_j}\}, \{\theta_{i_{j+1}|i_1,\ldots,i_j}\})$, given the complete data, $\mathbf{X}$. Then one may assign a prior distribution to $\boldsymbol{\phi}$ that is analogous to those postulated in Section 2 (i.e., independent beta and Dirichlet distributions). Finally, as in Section 4.1, one may apply the EM algorithm to the posterior distribution of $\boldsymbol{\phi}$ to obtain the modal values $\{\hat{P}_{i_1}\}$, $\{\hat{\theta}_{i_1}\}$, $\{\hat{P}_{i_{j+1}|i_1,\ldots,i_j}\}$, and $\{\hat{\theta}_{i_{j+1}|i_1,\ldots,i_j}\}$. These posterior modal values are then used to predict $u_{i_1,\ldots,i_j}$, the unknown number of elements in the sample having $Y_l = y_{i_l}^{(l)}$ for $l = 1, \ldots, j$ and $d_j = 0$ for $1 \le j \le k$, $1 \le i_1 \le c_1, \ldots, 1 \le i_j \le c_j$.

First, u_{i_1} is estimated by

$$\hat{u}_{i_1} = s\hat{P}_{i_1}\hat{\theta}_{i_1}\left(\sum_{l=1}^{c_1} \hat{P}_l\hat{\theta}_l\right)^{-1} \quad \text{for } i_1 = 1, \ldots, c_1.$$

Then, assuming that the $\{\hat{u}_{i_1}\}$ are integers, impute $\hat{u}_{i_1}$ repetitions of $y_{i_1}^{(1)}$. After this imputation, there are an estimated $(\hat{u}_{i_1} + s_{i_1})$ elements that are nonrespondents to Y_2 and have $Y_1 = y_{i_1}^{(1)}$. Impute $\hat{u}_{i_1,i_2}$ repetitions of $y_{i_2}^{(2)}$ to these $(\hat{u}_{i_1} + s_{i_1})$ elements, where

$$\hat{u}_{i_1,i_2} = (\hat{u}_{i_1} + s_{i_1})\hat{P}_{i_2|i_1}\hat{\theta}_{i_2|i_1}\left(\sum_{l=1}^{c_2} \hat{P}_{l|i_1}\hat{\theta}_{l|i_1}\right)^{-1} \quad \text{for } i_2 = 1, \ldots, c_2.$$

Continuing this process, impute $\hat{u}_{i_1,\ldots,i_{j+1}}$ repetitions of $y_{i_{j+1}}^{(j+1)}$ to the estimated $(\hat{u}_{i_1,\ldots,i_j} + s_{i_1,\ldots,i_j})$ elements that are nonrespondents to Y_{j+1} and have $Y_k = y_{i_k}^{(k)}$ $(k = 1, \ldots, j)$. Here

$$\hat{u}_{i_1,\ldots,i_{j+1}} = (\hat{u}_{i_1,\ldots,i_j} + s_{i_1,\ldots,i_j}) \times \hat{P}_{i_{j+1}|i_1,\ldots,i_j}\hat{\theta}_{i_{j+1}|i_1,\ldots,i_j}\left(\sum_{l=1}^{c_{j+1}} \hat{P}_{l|i_1,\ldots,i_j}\hat{\theta}_{l|i_1,\ldots,i_j}\right)^{-1}.$$

[*Received June 1984. Revised January 1986.*]

REFERENCES

Bailar, B. A., and Bailar, J. C. (1979), "Comparison of Two Procedures for Imputing Missing Survey Values," in *Proceedings of the Survey Research Methods Section, American Statistical Association*, pp. 462–467.

——— (1983), "Comparison of the Biases of the 'Hot-Deck' Imputation Procedure With an 'Equal Weights' Imputation Procedure," in *Incomplete Data in Sample Surveys: Proceedings of a Symposium*, New York: Academic Press, pp. 299–311.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D. (1972), *Statistical Inference Under Order Restrictions*, New York: John Wiley.

Chapman, D. W. (1976), "A Survey of Nonresponse Imputation Procedures," in *Proceedings of the Social Statistics Section, American Statistical Association*, pp. 245–251.

Chen, T., and Fienberg, S. E. (1974), "Two Dimensional Contingency Tables With Both Completely and Partially Cross-Classified Data," *Biometrics*, 30, 629–642.

Chiu, H. Y. (1982), "A New Bayesian Approach to Estimating Finite Population Parameters," unpublished Ph.D. thesis, State University of New York at Albany, Dept. of Mathematics and Statistics.

Cryer, J., and Robertson, T. (1975), "Isotonic Estimation of the Probability of Extinction of a Branching Process," *Journal of the American Statistical Association*, 70, 905–911.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood From Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society*, Ser. B, 39, 1–38.

Ernst, L. (1982), "Variance of the Estimated Mean for Several Imputation Procedures," technical report, U.S. Bureau of the Census.

Ford, B. (1983), "An Overview of Hot Deck Procedures," in *Incomplete Data in Sample Surveys: Theory and Bibliographies*, New York: Academic Press, pp. 185–207.

Greenlees, J., Reece, W., and Zieschang, K. (1982), "Imputation of Missing Values When the Probability of Response Depends on the Variable Being Imputed," *Journal of the American Statistical Association*, 77, 251–261.

Heckman, J. (1976), "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement*, 5, 475–492.

Herzog, T., and Rubin, D. (1983), "Using Multiple Imputations to Handle Nonresponse in Sample Surveys," in *Incomplete Data in Sample Surveys: Theory and Bibliographies*, New York: Academic Press, pp. 210–245.

Kalton, G., and Kasprzyk, D. (1982), "Imputing for Missing Survey Responses," in *Proceedings of the Survey Research Methods Section, American Statistical Association*, pp. 22–31.

Lindley, D. V. (1980), "Approximate Bayesian Methods," in *Bayesian Statistics*, eds. J. M. Bernardo, M. H. DeGroot, D.V. Lindley, and A. F. M. Smith, Valencia: University Press, 223–245.

Little, R. J. A. (1982), "Models for Nonresponse in Sample Surveys," *Journal of the American Statistical Association*, 77, 237–250.

Madow, W. G., Nisselson, H., and Olkin, I. (1983), *Incomplete Data in Sample Surveys: Report and Case Studies*, New York: Academic Press.

Madow, W. G., and Olkin, I. (1983), *Incomplete Data in Sample Surveys: Proceedings of the Symposium*, New York: Academic Press.

Madow, W. G., Olkin, I., and Rubin, D. (1983), *Incomplete Data in Sample Surveys: Theory and Bibliographies*, New York: Academic Press.

Rubin, D. (1978), "Multiple Imputations in Sample Surveys—A Phenomenological Bayesian Approach to Nonresponse," in *Proceedings of the Survey Research Methods Section, American Statistical Association*, pp. 20–28.

Sedransk, J., Monahan, J., and Chiu, H. Y. (1985), "Bayesian Estimation of Finite Population Parameters in Categorical Data Models Incorporating Order Restrictions," *Journal of the Royal Statistical Society*, Ser. B, 47, 519–527.

U.S. Bureau of the Census (1980a), "Revisions to the March CPS Processing System," technical report.

——— (1980b), "Income Nonresponse and Characteristics of Non-respondents on the March Current Population Survey (CPS) Supplement," technical report.

Welniak, E. J., and Coder, J. (1980), "A Measure of the Bias in the March CPS Earnings Imputation System and Results of a Simple Bias Adjustment Procedure," technical report, U.S. Census Bureau.