The Annals of Statistics
1988, Vol. 16, No. 4, 1550–1566

ASYMPTOTIC RESULTS FOR MULTIPLE IMPUTATION[1]

By Nathaniel Schenker[2] and A. H. Welsh

U.S. Bureau of the Census and University of Chicago

Imputation and multiple-imputation procedures have been used in practice to handle the problem of ignorable nonresponse in sample surveys. We examine the large-sample properties of these procedures where covariates are available for the case when the complete-data analysis is based on least squares. The results provide a formal justification for the inference procedures discussed by Rubin and Schenker for the location problem and suggest new procedures for the regression problem.

1. Introduction. One of the main sources of error in surveys is missing data due to nonresponse [Cochran (1977), Chapter 13]. A review of the vast literature on nonresponse as well as several articles on how to deal with the problem appear in three volumes entitled *Incomplete Data in Sample Surveys* [Madow, Nisselson and Olkin (1983), Madow, Olkin and Rubin (1983) and Madow and Olkin (1983)], produced by the Panel on Incomplete Data of the Committee on National Statistics.

The discussions of ten case studies in Madow, Nisselson and Olkin (1983) reveal that a popular method of handling nonresponse in a survey is to impute (fill in) a value for each missing item in the survey. The survey data can then be analyzed using standard techniques for complete data. A major drawback of such single imputation followed by a standard analysis, as pointed out for example by Rubin (1978) and Ford (1983), is that the missing values are treated as if they were known. Thus the variability due to imputing the values is ignored and the resulting inferences are too sharp.

Rubin (1978) proposed multiple-imputation as a general Bayesian technique for handling nonresponse that allows assessment of uncertainty due to imputation. The practical idea is to replace each missing datum with two or more values representing a distribution of likely values. Each of the two or more resulting completed data sets is then analyzed using standard complete-data methods. These analyses are combined to reflect both within-imputation variability and between-imputation variability.

Some theoretical investigations of the properties of multiple-imputation interval estimates of the population mean can be found in Rubin (1979), Herzog and Rubin (1983) and Rubin and Schenker (1986). These studies all assume a simple-random-sample with no covariates, a scalar outcome variable and a nonresponse mechanism that is ignorable [Rubin (1976, 1983) and Little (1982)].

Received May 1986; revised June 1987.

[1]Manuscript prepared using computer facilities supported in part by National Science Foundation Grant DMS-84-04941 to the Department of Statistics at the University of Chicago.

[2]Research supported in part by National Science Foundation Grant SES-83-11428 to the National Opinion Research Center.

AMS 1980 *subject classifications*. Primary 62D05; secondary 62E20.

Key words and phrases. Bayesian inference, hot deck, missing data, nonresponse, sample surveys.

In this context, the assumption of ignorable nonresponse is equivalent to assuming that the respondents are just a simple random sample of the originally intended sample.

A more general situation is that of a survey with a scalar outcome variable Y and covariates X, in which there is ignorable nonresponse on Y. For many common sampling procedures (e.g., simple random sampling, stratified random sampling based on the covariates and sampling with probability proportional to a covariate value), ignorable nonresponse means that the probability of a unit responding, given X and Y, does not depend on Y.

The situation with covariates is often more relevant in practice than the simple-random-sample case examined in the preceding references for several reasons. First, many surveys have background information that is available for all units. Second, the assumption of ignorable nonresponse can approximate reality more closely when there are covariates available to "explain" the nonresponse. For example, many "hot-deck" imputation procedures assume that within adjustment cells defined by the covariates X, the distribution of Y for nonrespondents is similar to the distribution of Y for respondents [Ford (1983)]. This can be much more reasonable than assuming that the respondents and nonrespondents have similar Y distributions unconditionally. Finally, even if the nonresponse mechanism does not depend on X or Y, inferences from the data can be more precise if the information in X is used in creating imputations, as long as Y is related to X.

The purpose of this paper is to examine the large-sample properties of multiple-imputation procedures when the scalar outcome variable Y follows a linear model involving the covariates X and there is ignorable nonresponse on Y. The asymptotic sampling distributions of the multiple-imputation estimators are derived for the situation in which the complete-data analysis is based on standard least-squares procedures. The imputation procedures considered here have been used extensively in practice but few theoretical results are available. The present results may be thought of either as providing an alternative justification for multiple imputation or as establishing the calibration [see Dawid (1982)] properties of the Bayesian procedures. Finally, the results for the sample mean in the location problem referred to in Rubin and Schenker (1986) are contained in the present results.

Section 2 introduces notation and describes multiple-imputation procedures in the linear-model framework. Asymptotic distributions of multiple-imputation estimators are derived in Section 3. Some specific examples of imputation methods and their properties are given in Section 4. Section 5 contains a discussion of the application of the results and possible extensions to alternative estimation procedures.

2. Multiple imputation in the linear-model context. Suppose that Q is a (possibly vector-valued) quantity of interest and that when there is complete response, inference for Q may be based on

$$\hat{W}^{1/2}(Q - \hat{Q}) \sim N(0, I),$$

where $\hat{Q}$ and $\hat{W}$ estimate Q and the dispersion matrix of $Q - \hat{Q}$, respectively.

When there is nonresponse and several (say m) independent imputations of the missing values have been created under a single nonresponse model, there are m completed data sets and hence m values of $\hat{Q}$ and $\hat{W}$, say $\hat{Q}_{*l}$ and $\hat{W}_{*l}$, $l = 1, \ldots, m$. The multiple-imputation estimate of Q is

$$\hat{Q}_{*.} = m^{-1} \sum_{l=1}^{m} \hat{Q}_{*l}. \tag{2.1}$$

The estimated dispersion matrix of $Q - \hat{Q}_{*.}$ is

$$\hat{T} = \hat{W} + (1 + m^{-1})\hat{B}, \tag{2.2}$$

where

$$\hat{W} = m^{-1} \sum_{l=1}^{m} \hat{W}_{*l} \tag{2.3}$$

is the average within-imputation dispersion matrix of $Q - \hat{Q}_{*.}$ and

$$\hat{B} = (m-1)^{-1} \sum_{l=1}^{m} (\hat{Q}_{*l} - \hat{Q}_{*.})(\hat{Q}_{*l} - \hat{Q}_{*.})' \tag{2.4}$$

is the between-imputation dispersion matrix. [$\hat{W}$ in (2.3) equals the complete-data estimator of the dispersion matrix of $Q - \hat{Q}$ when $m = 1$.] When the observations are not identically distributed, formula (2.2) may need some modification; see Section 5.

A simple normal-based multiple-imputation inference for Q is based on

$$\hat{T}^{-1/2}(Q - \hat{Q}_{*.}) \sim N(0, I).$$

Li (1985) and Rubin and Schenker (1986) have proposed other approximations to correct for the fact that $\hat{B}$ may be based on only a few degrees of freedom. These approximations may be derived by assigning $Q - \hat{Q}_{*.}$ a normal distribution and $\hat{B}$ a Wishart distribution. These approximations will be discussed further in Section 5 after the derivation of the appropriate distributions in Sections 3 and 4. In the linear-model framework, suppose that when there is complete response, the data are of the form

$$\begin{pmatrix} Y_1 \\ X_1 \end{pmatrix}, \ldots, \begin{pmatrix} Y_n \\ X_n \end{pmatrix},$$

where the Y_i's are scalar outcome variables and the X_i's are p-dimensional covariate vectors. We assume throughout that the Y_i's follow the model

$$Y_i = X_i'\theta + \varepsilon_i, \tag{2.5}$$

where the ε_i's are independent and identically distributed random variables with mean 0 and variance σ^2 and θ is a p-dimensional vector of regression coefficients. For simplicity, we assume throughout that (2.5) contains an intercept term; if the model does not contain an intercept some minor modifications to what follows will be required. Inference for $Q = Q(\theta)$ may be based on results for θ so we first consider inference for $Q = \theta$.

With complete data and n not small, least-squares inferences for θ are based on

$$\hat{W}^{-1/2}(\theta - \hat{\theta}) \sim N(0, I),$$

where

$$\hat{\theta} = \Psi^{-1} \sum_{i=1}^{n} X_i Y_i$$

and

$$\hat{W} = s^2 \Psi^{-1},$$

with

$$\Psi = \sum_{i=1}^{n} X_i X_i' \quad \text{and} \quad s^2 = (n-p)^{-1} \sum_{i=1}^{n} \left(Y_i - X_i'\hat{\theta}\right)^2.$$

Suppose now that due to ignorable nonresponse, only $n_1 < n$ of the Y values are observed with $n_0 = n - n_1$ missing. The observed data can then be written in the form

$$\begin{pmatrix} Y_1 \\ X_1 \end{pmatrix}, \dots, \begin{pmatrix} Y_{n_1} \\ X_{n_1} \end{pmatrix}, \quad \begin{pmatrix} \cdot \\ X_{(1)} \end{pmatrix}, \dots, \begin{pmatrix} \cdot \\ X_{(n_0)} \end{pmatrix},$$

where the $X_{(i)}$'s are covariates associated with missing Y values. If the probability of nonresponse and/or $\mathrm{Var}(\varepsilon_i)$ depends on X_i, then $X_{(1)}, \dots, X_{(n_0)}$ may be extreme in the design space. In this case, imputation will amount to extrapolation and consequently will be extremely sensitive to the model used. It is therefore advisable in practice to impute with caution. Suppose multiple imputations for the missing Y values, say $(Y_{(1)l}, \dots, Y_{(n_0)l})$, $1 \le l \le m$, have been created. Then the complete-data statistics obtained from the multiply-imputed data are $(\hat{\theta}_{*l}, \hat{W}_{*l})$, $1 \le l \le m$, where

$$\hat{\theta}_{*l} = (\Psi_1 + \Psi_0)^{-1}\left(\Psi_1\hat{\theta}_1 + \Psi_0\hat{\theta}_{0l}\right)$$

and

$$\hat{W}_{*l} = s_{*l}^2(\Psi_1 + \Psi_0)^{-1},$$

with

$$\Psi_1 = \sum_{i=1}^{n_1} X_i X_i', \qquad \Psi_0 = \sum_{i=1}^{n_0} X_{(i)} X_{(i)}',$$

$$\hat{\theta}_1 = \Psi_1^{-1} \sum_{i=1}^{n_1} X_i Y_i, \qquad \hat{\theta}_{0l} = \Psi_0^{-1} \sum_{i=1}^{n_0} X_{(i)} Y_{(i)l}$$

and

$$s_{*l}^2 = (n-p)^{-1}\left[\sum_{i=1}^{n_1} \left(Y_i - X_i'\hat{\theta}_{*l}\right)^2 + \sum_{i=1}^{n_0} \left(Y_{(i)l} - X_{(i)}'\hat{\theta}_{*l}\right)^2\right].$$

It is convenient to note that we can represent s_{*l}^2 as a function of

$$s_1^2 = (n_1 - p)^{-1} \sum_{i=1}^{n_1} \left(Y_i - X_i'\hat{\theta}_1\right)^2$$

and

$$s_{0l}^2 = (n_0 - p)^{-1} \sum_{i=1}^{n_0} \left(Y_{(i)l} - X_{(i)}'\hat{\theta}_{0l}\right)^2$$

(see the proof of Theorem 1). By (2.1) (with $Q = \theta$), (2.3) and (2.4),

$$\hat{\theta}_{*\cdot} = m^{-1} \sum_{l=1}^{m} \hat{\theta}_{*l},$$

$$\hat{W} = m^{-1} \sum_{l=1}^{m} s_{*l}^2 (\Psi_1 + \Psi_0)^{-1}$$

and

$$\hat{B} = (m-1)^{-1} \sum_{l=1}^{m} (\hat{\theta}_{*l} - \hat{\theta}_{*\cdot})(\hat{\theta}_{*l} - \hat{\theta}_{*\cdot})'.$$

We will also require

$$\hat{W}_1 = m^{-1} \sum_{l=1}^{m} s_{*l}^2 \Psi_1^{-1}.$$

3. Asymptotic theory for multiple-imputation estimators. Our main result, Theorem 1, gives the asymptotic sampling behavior of the estimators $\hat{\theta}_{*\cdot}$, $\hat{B}$ and $\hat{W}$ when the imputed data satisfy certain conditions. These conditions are conveniently stated in terms of the conditional asymptotic behavior of the estimators $\hat{\theta}_{0l}$ and s_{0l}^2, $1 \le l \le m$, based on the imputed portion of the data. We discuss particular methods of imputation in Section 4 and show that these methods satisfy the conditions of the theorems.

We begin by proving a useful preliminary lemma. We suppose throughout that all random variables are defined on a common probability space.

LEMMA 1. *Let $\{V_i\}$ be a sequence of random variables such that, for some function h, as $n \to \infty$,*

$$h(V_1, \dots, V_n) \to_{\mathscr{D}} \Gamma,$$

where Γ has a distribution function G. If $\{U_i\}$ and $\{W_i\}$ are sequences of random variables such that

$$P\{W_n \le r, U_n - h(V_1, \dots, V_n) \le s \| V_1, \dots, V_n\} \to H(r)F(s) \tag{3.1}$$

almost surely for all $(r, s) \in \mathscr{R}^2$, where H and F are continuous distribution functions, then

$$P(W_n \le r, U_n \le t) \to H(r)(G * F)(t),$$

for all $(r, t) \in \mathscr{R}^2$, where "$*$" denotes convolution.

PROOF. Notice that

$$\begin{aligned} &|P(W_n \le r, U_n \le t) - H(r)(G*F)(t)| \\ &\quad \le E\Big[\sup_{-\infty < s < \infty} |P\{W_n \le r, U_n - h(V_1, \dots, V_n) \\ &\qquad\qquad \le s \| V_1, \dots, V_n\} - H(r)F(s)|\Big] \\ &\qquad + H(r)|E[F\{t - h(V_1, \dots, V_n)\} - (G*F)(t)]|. \end{aligned}$$

By the argument leading to Polya's theorem [Lemma 3.2 of Ranga Rao (1962)], the convergence in (3.1) holds uniformly in s so that the dominated convergence theorem ensures that the first term on the right-hand side converges to 0. Since F is bounded and continuous,

$$E\left[F\{t - h(V_1, \dots, V_n)\}\right] \to E\{F(t - \Gamma)\} = (G * F)(t)$$

and the result obtains. □

We are now able to prove Theorem 1.

THEOREM 1. *Let $m \geq 2$ be a fixed integer and let $n \to \infty$ such that $n_0/n_1 = \lambda$, $0 < \lambda < \infty$. Suppose that the model (2.5) holds and that $n_1^{-1}\Psi_1 \to \Delta_1$ and $n_0^{-1}\Psi_0 \to \Delta_0$, as $n \to \infty$, where Δ_1 and Δ_0 are positive definite matrices. If for almost all sample sequences, conditional on the observed data,*

$$n_0^{1/2}\left(\hat{\theta}_{0l} - \hat{\theta}_1\right)/\sigma \to_{\mathscr{D}} N(0, \Sigma), \qquad 1 \leq l \leq m, \tag{3.2}$$

for some dispersion matrix Σ, and

$$s_{0l}^2/\sigma^2 \to_P 1, \qquad 1 \leq l \leq m, \tag{3.3}$$

then it follows that

$$n^{1/2}\left(\hat{\theta}_{*} - \theta\right)/\sigma \to_{\mathscr{D}} Z,$$
$$n\hat{B}/\sigma^2 \to_{\mathscr{D}} \Xi,$$
$$n\hat{W}/\sigma^2 \to_P (1 + \lambda)(\Delta_1 + \lambda\,\Delta_0)^{-1}$$

and

$$n\hat{W}_1/\sigma^2 \to_P (1 + \lambda)\,\Delta_1^{-1},$$

where Z and Ξ are independent and have p-dimensional

$$N\left(0, (1 + \lambda)\left\{\Delta_1^{-1} + m^{-1}\lambda(\Delta_1 + \lambda\,\Delta_0)^{-1}\Delta_0\Sigma\,\Delta_0(\Delta_1 + \lambda\,\Delta_0)^{-1}\right\}\right)$$

and

$$\text{Wishart}\left(m - 1, (m - 1)^{-1}(1 + \lambda)\lambda(\Delta_1 + \lambda\,\Delta_0)^{-1}\Delta_0\Sigma\,\Delta_0(\Delta_1 + \lambda\,\Delta_0)^{-1}\right)$$

distributions, respectively.

PROOF. Write

$$\hat{\theta}_{*l} - \hat{\theta}_1 = (\Psi_1 + \Psi_0)^{-1}\Psi_0\left(\hat{\theta}_{0l} - \hat{\theta}_1\right), \tag{3.4}$$

so that, conditional on the observed data,

$$n_0^{1/2}\left(\hat{\theta}_{*l} - \hat{\theta}_1\right)/\sigma \to_{\mathscr{D}} N\left(0, \lambda^2(\Delta_1 + \lambda\,\Delta_0)^{-1}\Delta_0\Sigma\,\Delta_0(\Delta_1 + \lambda\,\Delta_0)^{-1}\right)$$

almost surely. Let $a \in \mathscr{R}^p$ be any fixed p-vector such that $a'a = 1$ and set

$$\tau' = (\tau_1, \dots, \tau_m),$$

where

$$\tau_l = n_0^{1/2}a'\left(\hat{\theta}_{*l} - \hat{\theta}_1\right)/\sigma, \qquad 1 \leq l \leq m.$$

Conditional on the observed data, the m imputations are independent so that

$$\tau \to_{\mathscr{D}} N\big(0, \lambda^2 a'(\Delta_1 + \lambda\,\Delta_0)^{-1}\Delta_0\Sigma\,\Delta_0(\Delta_1 + \lambda\,\Delta_0)^{-1}aI\big) \quad \text{almost surely.}$$

Let H be any $m \times m$ orthogonal matrix with all entries in the first row equal to $m^{-1/2}$. Applying the continuous mapping theorem [Billingsley (1979), page 330] to $H\tau$, we have that, conditional on the observed data,

$$\begin{aligned} n^{1/2}a'\big(\hat\theta_{*\cdot} - \hat\theta_1\big)/\sigma &= (n/n_0)^{1/2}m^{-1}\sum_{l=1}^{m}\tau_l \\ &= (n/n_0)^{1/2}m^{-1/2}(H\tau)_1 \\ &\to_{\mathscr{D}} Z_1 \end{aligned}$$

and

$$\begin{aligned} na'\hat B a/\sigma^2 &= (n/n_0)(m-1)^{-1}\sum_{l=1}^{m}(\tau_l - \bar\tau)^2 \\ &= (n/n_0)(m-1)^{-1}\sum_{l=1}^{m}(H\tau)_l^2 \\ &\to_{\mathscr{D}} \Xi_1 \end{aligned}$$

almost surely, where Z_1 and Ξ_1 are independently distributed as

$$N\big(0, m^{-1}(1+\lambda)\lambda a'(\Delta_1 + \lambda\,\Delta_0)^{-1}\Delta_0\Sigma\,\Delta_0(\Delta_1 + \lambda\,\Delta_0)^{-1}a\big)$$

and

$$(m-1)^{-1}(1+\lambda)\lambda a'(\Delta_1 + \lambda\,\Delta_0)^{-1}\Delta_0\Sigma\,\Delta_0(\Delta_1 + \lambda\,\Delta_0)^{-1}a\chi^2_{m-1},$$

respectively. But

$$n^{1/2}a'\big(\hat\theta_{*\cdot} - \theta\big)/\sigma = n^{1/2}a'\big(\hat\theta_{*\cdot} - \hat\theta_1\big)/\sigma + n^{1/2}a'\big(\hat\theta_1 - \theta\big)/\sigma,$$

so that by the central limit theorem for the least-squares estimator and Lemma 1,

$$n^{1/2}a'\big(\hat\theta_{*\cdot} - \theta\big)/\sigma \to_{\mathscr{D}} Z_2$$

and

$$na'\hat B a/\sigma^2 \to_{\mathscr{D}} \Xi_1,$$

where Z_2 is independent of Ξ_1 and has a

$$N\big(0, (1+\lambda)\{a'\Delta_1^{-1}a + m^{-1}\lambda a'(\Delta_1 + \lambda\,\Delta_0)^{-1}\Delta_0\Sigma\,\Delta_0(\Delta_1 + \lambda\,\Delta_0)^{-1}a\}$$

distribution. The first part of the theorem obtains.

To prove the last two parts of the theorem, it suffices to show that

$$s^2_{*l}/\sigma^2 \to_P 1, \qquad 1 \le l \le m.$$

Put

$$s_1^2 = (n_1 - p)^{-1}\sum_{i=1}^{n_1}\big(Y_i - X_i'\hat\theta_1\big)^2.$$

Then for $1 \le l \le m$,

$$\begin{aligned} s_{*l}^2 &= (n-p)^{-1}\left\{\sum_{i=1}^{n_1}\left(Y_i - X_i'\hat{\theta}_{*l}\right)^2 + \sum_{i=1}^{n_0}\left(Y_{(i)l} - X_{(i)}'\hat{\theta}_{*l}\right)^2\right\} \\ &= (n-p)^{-1}\{(n_1-p)s_1^2 + (\hat{\theta}_{*l}-\hat{\theta}_1)'\Psi_1(\hat{\theta}_{*l}-\hat{\theta}_1) \\ &\qquad + (n_0-p)s_{0l}^2 + (\hat{\theta}_{*l}-\hat{\theta}_{0l})'\Psi_0(\hat{\theta}_{*l}-\hat{\theta}_{0l})\} \\ &= (n-p)^{-1}\{(n_1-p)s_1^2 + (\hat{\theta}_{0l}-\hat{\theta}_1)'\Psi_0(\Psi_1+\Psi_0)^{-1} \\ &\qquad \times \Psi_1(\Psi_1+\Psi_0)^{-1}\Psi_0(\hat{\theta}_{0l}-\hat{\theta}_1) + (n_0-p)s_{0l}^2 \\ &\qquad + (\hat{\theta}_{0l}-\hat{\theta}_1)'\Psi_1(\Psi_1+\Psi_0)^{-1}\Psi_0(\Psi_1+\Psi_0)^{-1}\Psi_1(\hat{\theta}_{0l}-\hat{\theta}_1)\} \end{aligned}$$

by (3.4) and the fact that

$$\hat{\theta}_{*l} - \hat{\theta}_{0l} = (\Psi_1+\Psi_0)^{-1}\Psi_1(\hat{\theta}_1 - \hat{\theta}_{0l}).$$

Let $\varepsilon > 0$ be given. Then for $1 \le l \le m$,

$$\begin{aligned} P\{|s_{*l}^2 - \sigma^2| > 2\varepsilon\} &\le P\{|(n-p)^{-1}(n_1-p)s_1^2 - (1+\lambda)^{-1}\sigma^2| > \varepsilon\} \\ &\quad + E\big[P\{|(n-p)^{-1}(n_0-p)s_{0l}^2 + (n-p)^{-1}(\hat{\theta}_{0l}-\hat{\theta}_1)' \\ &\qquad \times \Psi_0(\Psi_1+\Psi_0)^{-1}\Psi_1(\Psi_1+\Psi_0)^{-1}\Psi_0(\hat{\theta}_{0l}-\hat{\theta}_1) \\ &\qquad + (n-p)^{-1}(\hat{\theta}_{0l}-\hat{\theta}_1)'\Psi_1(\Psi_1+\Psi_0)^{-1} \\ &\qquad \times \Psi_0(\Psi_1+\Psi_0)^{-1}\Psi_1(\hat{\theta}_{0l}-\hat{\theta}_1) \\ &\qquad - \lambda(1+\lambda)^{-1}\sigma^2| > \varepsilon \,\|\, \text{observed data}\}\big]. \end{aligned}$$

Now $s_1^2 \to \sigma^2$ almost surely [see the proof of Theorem 2.2 of Freedman (1981)] so that the first term on the right-hand side can be made arbitrarily small. Also, conditional on the observed data, $s_{0l}^2 \to_P \sigma^2$ and $\hat{\theta}_{0l} - \hat{\theta}_1 \to_P 0$ almost surely, by hypothesis, so that the second term on the right-hand side can be made arbitrarily small by the dominated convergence theorem and the result obtains. □

We will also require a variation on Theorem 1 that permits an alternative centering in (3.2). Specifically, if we replace (3.2) by the condition that for almost all sample sequences, conditional on the observed data,

$$(3.5) \qquad n_0^{1/2}(\hat{\theta}_{0l} - \theta)/\sigma \to_{\mathscr{D}} N(0, \Sigma), \qquad 1 \le l \le m,$$

then since

$$\hat{\theta}_{*l} - \theta = (\Psi_1+\Psi_0)^{-1}\{\Psi_0(\hat{\theta}_{0l} - \theta) + \Psi_1(\hat{\theta}_1 - \theta)\}$$

and

$$\hat{B} = (\Psi_1+\Psi_0)^{-1}\Psi_0\left\{(m-1)\sum_{l=1}^{m}(\hat{\theta}_{0l} - \hat{\theta}_{0\cdot})(\hat{\theta}_{0l} - \hat{\theta}_{0\cdot})'\right\}\Psi_0(\Psi_1+\Psi_0)^{-1},$$

where $\hat{\theta}_{0\cdot} = m^{-1}\Sigma_{l=1}^{m}\hat{\theta}_{0l}$, a similar argument to that used to prove Theorem 1 yields Theorem 2.

THEOREM 2. *Suppose that the conditions of Theorem* 1 *hold but with* (3.2) *replaced by* (3.5). *Then*

$$n^{1/2}(\hat{\theta}_{*\cdot} - \theta)/\sigma \to_{\mathscr{D}} Z,$$
$$n\hat{B}/\sigma^2 \to_{\mathscr{D}} \Xi,$$
$$n\hat{W}/\sigma^2 \to_P (1+\lambda)(\Delta_1 + \lambda\,\Delta_0)^{-1}$$

and

$$n\hat{W}_1/\sigma^2 \to_P (1+\lambda)\,\Delta_1^{-1},$$

where Z and Ξ *are independent and have p-dimensional*

$$N\big(0, (1+\lambda)(\Delta_1 + \lambda\,\Delta_0)^{-1}\{\Delta_1 + m^{-1}\lambda\,\Delta_0\Sigma\,\Delta_0\}(\Delta_1 + \lambda\,\Delta_0)^{-1}\big)$$

and

$$\text{Wishart}\big(m-1, (m-1)^{-1}(1+\lambda)\lambda(\Delta_1 + \lambda\,\Delta_0)^{-1}\Delta_0\Sigma\,\Delta_0(\Delta_1 + \lambda\,\Delta_0)^{-1}\big)$$

distributions, respectively.

Notice that the distribution of Z in Theorem 2 is different from that in Theorem 1.

4. Imputation methods. In this section, we discuss four methods of imputing a single set of n_0 missing Y values $Y_{(1)}, \dots, Y_{(n_0)}$ given the observed data. The multiple imputations are obtained by m independent applications of a method. We will assume throughout that the model (2.5) holds and that both $n_1^{-1}\Psi_1 \to \Delta_1$ and $n_0^{-1}\Psi_0 \to \Delta_0$ as $n \to \infty$, where Δ_1 and Δ_0 are positive definite matrices. We show that for each of the methods, condition (3.3) and either condition (3.2) of Theorem 1 or condition (3.5) of Theorem 2 are satisfied.

Hot-deck imputation. Suppose the covariates X take on b values and that there are several observations at each of these values. Rewrite the data in b blocks as

$$\begin{pmatrix} Y_{11} \\ X_1 \end{pmatrix}, \begin{pmatrix} Y_{12} \\ X_1 \end{pmatrix}, \dots, \begin{pmatrix} Y_{1n_{11}} \\ X_1 \end{pmatrix}, \underbrace{\begin{pmatrix} \cdot \\ X_1 \end{pmatrix}, \dots, \begin{pmatrix} \cdot \\ X_1 \end{pmatrix}}_{n_{01}}$$
$$\vdots$$
$$\begin{pmatrix} Y_{b1} \\ X_b \end{pmatrix}, \begin{pmatrix} Y_{b2} \\ X_b \end{pmatrix}, \dots, \begin{pmatrix} Y_{bn_{1b}} \\ X_b \end{pmatrix}, \underbrace{\begin{pmatrix} \cdot \\ X_b \end{pmatrix}, \dots, \begin{pmatrix} \cdot \\ X_b \end{pmatrix}}_{n_{0b}}$$

where $n_1 = \Sigma_{j=1}^{b} n_{1j}$ and $n_0 = \Sigma_{j=1}^{b} n_{0j}$. In each block, sample independently with replacement (i.e., with equal probability) the observed Y's to create imputations for the missing Y's. That is, in block j, independently sample n_{0j} observations with replacement from $\{Y_{j1}, \ldots, Y_{jn_{1j}}\}$, $1 \le j \le b$. This is a standard method of creating imputations; see Ford (1983) for references.

An alternative interpretation of hot-deck imputation is that observations $\varepsilon_{j(i)}$, $1 \le i \le n_{0j}$, are drawn independently with replacement from $\{\varepsilon_{j1}, \ldots, \varepsilon_{jn_{1j}}\}$ and used to construct

$$Y_{j(i)} = X_j'\theta + \varepsilon_{j(i)}, \qquad 1 \le i \le n_{0j}, 1 \le j \le b.$$

Now, with $\Psi_0 = \Sigma_{j=1}^b n_{0j} X_j X_j'$,

$$\begin{aligned} \hat{\theta}_0 - \theta &= \Psi_0^{-1} \sum_{j=1}^{b} \sum_{i=1}^{n_{0j}} X_j \varepsilon_{j(i)} \\ &= \sum_{j=1}^{b} A_j' E_{(j)}, \end{aligned}$$

where $E_{(j)} = (\varepsilon_{j(1)}, \ldots, \varepsilon_{j(n_{0j})})'$ and $A_j' = \Psi_0^{-1}(X_j, \ldots, X_j)_{p \times n_{0j}}$, $1 \le j \le b$. Let $E_j = (\varepsilon_1, \ldots, \varepsilon_{n_{1j}})'$ be independent n_{1j}-vectors, $1 \le j \le b$, where the ε's are independent random variables with distribution function F. It is straightforward to use Lemmas 8.4, 8.7 and 8.9 of Bickel and Freedman (1981) to show that conditional on the observed data, $n_0^{1/2}\Sigma_{j=1}^b A_j' E_{(j)}$ and $n_0^{1/2}\Sigma_{j=1}^b A_j' E_j$ have the same asymptotic distributions. That (3.5) holds with $\Sigma = \Delta_0^{-1}$ follows from the central limit theorem applied to $n_0^{1/2}\Sigma_{j=1}^b A_j' E_j$.

Next, notice that conditional on the observed data,

$$\begin{aligned} s_0^2 &= (n_0 - p)^{-1} \left\{ \sum_{j=1}^{b} \sum_{i=1}^{n_{0j}} \varepsilon_{j(i)}^2 + \left(\hat{\theta}_0 - \theta\right)' \Psi_0 \left(\hat{\theta}_0 - \theta\right) \right\} \\ &\to_P \sigma^2 \end{aligned}$$

almost surely by (3.5) and by Lemmas 8.5 and 8.6 of Bickel and Freedman (1981). [The argument is similar to the proof of part b of Theorem 2.1 of Bickel and Freedman (1981).]

Hot-deck imputation is feasible in large samples from models involving only a few covariates taking a small number of possible values. Consideration of alternative scenarios suggests that it is also useful to have available methods of imputation that utilize more of the structure of the underlying model.

Simple residual imputation. Put

$$r_i = Y_i - X_i'\hat{\theta}_1, \qquad 1 \le i \le n_1,$$

where $\hat{\theta}_1 = \Psi_1^{-1}\Sigma_{i=1}^{n_1} X_i Y_i$ is the least-squares estimate based on the complete data. Draw a sample $r_{(1)}, \ldots, r_{(n_0)}$ of size n_0 by sampling independently with replacement from $\{r_1, \ldots, r_{n_1}\}$ and construct

$$Y_{(i)} = X_{(i)}'\hat{\theta}_1 + r_{(i)}, \qquad 1 \le i \le n_0.$$

If the model does not contain an intercept, the residuals have to be centered about $\bar{r} = n_1^{-1}\Sigma_{i=1}^{n_1} r_i$ [see Freedman (1981)]; as a small sample adjustment, it is possible to rescale the residuals by $(n_1/(n_1 - p))^{1/2}$. Simple residual imputation has essentially been proposed and used by Kalton and Kish (1981) and David, Little, Samuhel and Triest (1986).

It follows from Theorem 2.2 of Freedman (1981) that the conditions (3.2) and (3.3) of Theorem 1 are satisfied with $\Sigma = \Delta_0^{-1}$.

Since the simple residual imputation method resamples from the observed residuals, no variability beyond that which is present in the complete-data portion of the data is introduced. However, by incorporating some additional structure, we can introduce more variability into the imputations. Suppose that in the model (2.5) the errors are normally distributed. Then if θ and σ^2 are treated as independent with prior density proportional to σ^{-2}, the marginal posterior distribution of σ^2 is $(n_1 - p)s_1^2/\chi^2_{n_1-p}$ and the conditional posterior distribution of θ given σ^2 is $N(\hat{\theta}_1, \sigma^2\Psi_1^{-1})$; see Box and Tiao [(1973), page 116]. The next imputation method uses the posterior distribution of (θ, σ^2) in generating imputations.

Normal imputation. Draw σ^{*2} from $(n_1 - p)s_1^2/\chi^2_{n_1-p}$ and then draw θ^* from $N(\hat{\theta}_1, \sigma^{*2}\Psi_1^{-1})$. Also draw n_0 independent observations $Z_{(1)}, \dots, Z_{(n_0)}$ from $N(0,1)$ and construct

$$Y_{(i)} = X'_{(i)}\theta^* + \sigma^* Z_{(i)}, \qquad 1 \le i \le n_0.$$

Herzog and Rubin (1983) applied the normal imputation method with $\sigma^{*2} = s_1^2$ to CPS income data; the present method reduces to the fully normal method of Rubin and Schenker (1986) in the location problem. It should be noted that the asymptotic results below do not require the errors in the model (2.5) to be normally distributed.

Let $a \in \mathscr{R}^p$ be any fixed vector such that $a'a = 1$. Then, with $\Sigma = \Delta_0^{-1} + \lambda\,\Delta_1^{-1}$,

$$\begin{aligned} &\left|P\{n_0^{1/2}a'(\hat{\theta}_0 - \hat{\theta}_1) \le x \| \text{observed data}\} - \Phi\{x/(\sigma(a'\Sigma a)^{1/2})\}\right| \\ &\quad \le E\Bigg[\sup_{-\infty \le s \le \infty}\Big|P\{n_0^{1/2}a'(\hat{\theta}_0 - \theta^*) \le s \| \text{observed data}, \theta^*, \sigma^*\} \\ &\qquad\qquad - \Phi\{s/(\sigma(a'\Delta_0^{-1}a)^{1/2})\}\Big| \,\Big\|\, \text{observed data}\Bigg] \qquad (4.1) \\ &\qquad + \Big|E\Big[\Phi\{(x - n_0^{1/2}a'(\theta^* - \hat{\theta}_1))/(\sigma(a'\Delta_0^{-1}a)^{1/2})\} \| \text{observed data}\Big] \\ &\qquad\qquad - \Phi\{(x/(\sigma(a'\Sigma a)^{1/2})\}\Big|. \end{aligned}$$

Conditional on the observed data, $\sigma^{*2}/s_1^2 \sim (n_1 - p)/\chi^2_{n_1-p}$ and $s_1^2 \to \sigma^2$ almost surely [see the proof of Theorem 2.2 of Freedman (1981)] so that $\sigma^{*2}/\sigma^2 \to_P 1$ almost surely. It follows that conditional on the observed data,

$$n_0^{1/2}a'(\theta^* - \hat{\theta}_1) \to_{\mathscr{D}} Z \sim N(0, \lambda\sigma^2 a'\Delta_1^{-1}a) \quad \text{almost surely}.$$

Since Φ is bounded and continuous,

$$E\left[\Phi\left\{\left(x - n_0^{1/2}a'\left(\theta^* - \hat{\theta}_1\right)\right)/\left(\sigma\left(a'\Delta_0^{-1}a\right)^{1/2}\right)\right\}\|\text{observed data}\right]$$
$$\to E\Phi\left\{(x - Z)/\left(\sigma\left(a'\Delta_0^{-1}a\right)^{1/2}\right)\right\}$$
$$= \Phi\left\{x/\left(\sigma\left(a'\Sigma a\right)^{1/2}\right)\right\}.$$

Now, conditional on the observed data, θ^*, and σ^*,

$$n_0^{1/2}a'\left(\hat{\theta}_0 - \theta^*\right) = n_0^{1/2}\sigma^* a'\Psi_0^{-1}\sum_{i=1}^{n_0} X_{(i)}Z_{(i)}$$
$$\to_{\mathscr{D}} N\left(0, \sigma^2 a'\Delta_0^{-1}a\right) \quad \text{almost surely.}$$

Thus the first term on the right-hand side of (4.1) converges to 0 almost surely by the conditional dominated convergence theorem and the argument leading to Polya's theorem. It follows that (3.2) holds with $\Sigma = \Delta_0^{-1} + \lambda\,\Delta_1^{-1}$.

With regard to (3.3), notice that almost surely, for any $\varepsilon > 0$,

$$P\left\{|s_0^2 - \sigma^{*2}| > \varepsilon\|\text{observed data}\right\}$$
$$\le \varepsilon^{-2}E\left\{\left(s_0^2 - \sigma^{*2}\right)^2\|\text{observed data}\right\}$$
$$= \varepsilon^{-2}E\left[E\left\{\left(s_0^2 - \sigma^{*2}\right)^2\|\theta^*, \sigma^{*2}, \text{observed data}\right\}\|\text{observed data}\right]$$
$$= 2\varepsilon^{-2}(n_0 - p)^{-1}$$
$$\to 0$$

so that conditional on the observed data, $s_0^2/\sigma^{*2} \to_P 1$ almost surely. Since $\sigma^{*2}/\sigma^2 \to_P 1$ almost surely, condition (3.3) holds.

The normal imputation method uses the complete data only through $\hat{\theta}_1$ and s_1^2. An interesting alternative procedure may be developed as a compromise between the simple residual and normal methods.

Adjusted normal imputation. As in the normal method, draw σ^{*2} from $(n_1 - p)s_1^2/\chi_{n_1-p}$ and then draw θ^* from $N(\hat{\theta}_1, \sigma^{*2}\Psi_1^{-1})$. Also, as in the simple residual method, independently draw n_0 observations $r_{(1)}, \dots, r_{(n_0)}$ with replacement from $\{r_1, \dots, r_{n_1}\}$. Then construct

$$Y_{(i)} = X_{(i)}'\theta^* + n_1^{1/2}\sigma^* r_{(i)}/\left\{(n_1 - p)s_1^2\right\}^{1/2}, \qquad 1 \le i \le n_0.$$

Notice that conditional on the observed data, $n_1^{1/2}r_{(i)}/\{(n_1 - p)s_1^2\}^{1/2}$ has mean 0 and variance 1 so that the imputed Y's have the right conditional moments but a distribution whose shape is adjusted to reflect that of $r_1, \dots, r_{n_1}$. This adjusted normal imputation method reduces to the method of imputation adjusted for uncertainty in the mean and variance given in Rubin and Schenker (1986) for the location problem. As with the normal imputation method, the results below do not assume that the normal model holds.

Conditional on the observed data and on σ^*,

$$n_0^{1/2}(\hat{\theta}_0 - \hat{\theta}_1)/\sigma^* = n_0^{1/2}(\hat{\theta}_0 - \theta^*)/\sigma^* + n_0^{1/2}(\theta^* - \hat{\theta}_1)/\sigma^*.$$

Conditional on the observed data,

$$n_0^{1/2}(\theta^* - \hat{\theta}_1) \to_{\mathscr{D}} N(0, \lambda\sigma^2\Delta_1^{-1})$$

almost surely and conditional on the observed data, θ^*, and σ^*,

$$n_0^{1/2}(\hat{\theta}_0 - \theta^*) = n_0^{1/2}\sigma^*\Psi_0^{-1}\sum_{i=1}^{n_0} X_{(i)}r_{(i)}/\{n_1^{-1}(n_1 - p)s_1^2\}^{1/2}$$
$$\to_{\mathscr{D}} N(0, \sigma^2\Delta_0^{-1})$$

almost surely by Theorem 2.2 of Freedman (1981). By the same argument as that applied to the normal imputation method, it follows that (3.2) holds with $\Sigma = \Delta_0^{-1} + \lambda\,\Delta_1^{-1}$. The proof that (3.3) holds is also similar to that for the normal imputation method.

The adjusted normal imputation method provides a neat compromise between imputing from the complete data alone and imputing from a model with a predetermined shape. An important point to note is that while the motivation for the normal and adjusted normal imputation methods is highly parametric, the results for these methods (and hence Theorem 1) do not assume that the normal model holds. Thus, their application is broader than their derivation might suggest.

5. Asymptotic inference. The results of the last two sections may be used to suggest asymptotic inference procedures. To simplify calculations, suppose that

$$n^{1/2}(\hat{\theta}_{*\cdot} - \theta) \sim \begin{cases} N(0, WW_1^{-1}W + B) & \text{for hot-deck imputation,} \\ N(0, W_1 + B) & \text{for simple residual, normal and} \\ & \text{adjusted normal imputation,} \end{cases}$$

$$m^{-1}(m-1)n\hat{B} \sim \text{Wishart}(m-1, B),$$

$$\hat{\theta}_{*\cdot}, \hat{B} \text{ are independent,}$$

$$n\hat{W}_1 = W_1$$

and

$$n\hat{W} = W,$$

where $W = \sigma^2(1+\lambda)(\Delta_1 + \lambda\,\Delta_0)^{-1}$, $W_1 = \sigma^2(1+\lambda)\,\Delta_1^{-1}$ and

$$B = \sigma^2 m^{-1}(1+\lambda)\lambda(\Delta_1 + \lambda\,\Delta_0)^{-1}\Delta_0\Sigma\,\Delta_0(\Delta_1 + \lambda\,\Delta_0)^{-1}.$$

Theorems 1 and 2 give conditions under which these statements hold asymptotically.

The general discussion in Section 2 and the preceding results suggest that the dispersion matrix $WW_1^{-1}W + B$ be estimated by

$$n\hat{W}\hat{W}_1^{-1}\hat{W} + nm^{-1}(m-1)\hat{B}$$

or, using the small m modification in Rubin and Schenker (1986), by

$$n\hat{U} = n\hat{W}\hat{W}_1^{-1}\hat{W} + n(1 + m^{-1})\hat{B},$$

and that $W_1 + B$ be estimated by

$$n\hat{W}_1 + nm^{-1}(m - 1)\hat{B},$$

or by

$$n\hat{V} = n\hat{W}_1 + n(1 + m^{-1})\hat{B}.$$

Let C be a $k \times p$ matrix of full rank and consider the quadratic forms

$$T_1 = (\hat{\theta}_{*\cdot} - \theta)'C'(C\hat{U}C')^{-1}C(\hat{\theta}_{*\cdot} - \theta)$$

and

$$T_2 = (\hat{\theta}_{*\cdot} - \theta)'C'(C\hat{V}C')^{-1}C(\hat{\theta}_{*\cdot} - \theta).$$

While the exact distributions of T_1 and T_2 are complicated, simple approximations that permit the practical use of T_1 and T_2 for constructing confidence contours and testing are available. Li (1985) developed several approximations that may be usefully applied in the present context. The multivariate approximate degrees of freedom method of Yao (1965) uses a multivariate t^2-distribution to approximate the distribution of T_1. This is equivalent to using

$$T_i \sim \frac{\nu_i k}{\nu_i - k + 1} F(k, \nu_i - k + 1), \qquad i = 1,2, \tag{5.1}$$

where

$$\nu_1 = \max\Big\{k, m^2(m-1)T_1^2/(m+1)^2 \times \Big\{(\hat{\theta}_{*\cdot} - \theta)'C'(C\hat{U}C')^{-1}C\hat{B}C'(C\hat{U}C')^{-1}C(\hat{\theta}_{*\cdot} - \theta)\Big\}^2\Big\}$$

and

$$\nu_2 = \max\Big\{k, m^2(m-1)T_2^2/(m+1)^2 \times \Big\{(\hat{\theta}_{*\cdot} - \theta)'C'(C\hat{V}C')^{-1}C\hat{B}C'(C\hat{V}C')^{-1}C(\hat{\theta}_{*\cdot} - \theta)\Big\}^2\Big\},$$

a result which is a direct multivariate analogue of the approximation suggested in Rubin and Schenker (1986).

In the special case that $\Delta_1 = \Delta_0 = \Delta$ say, the formulas can be simplified. We have that $\Sigma = \rho^2\Delta^{-1}$, where ρ depends on the imputation method, and moreover that

$$W = \sigma^2\Delta^{-1},$$
$$W_1 = (1 + \lambda)W$$

and

$$B = m^{-1}(1 + \lambda)^{-1}\lambda\rho^2 W,$$

so we can write the asymptotic dispersion matrix of $n^{1/2}(\hat{\theta}_{*\cdot} - \theta)$ as

$$\mathrm{Cov}\big(n^{1/2}(\hat{\theta}_{*\cdot} - \theta)\big) = \big(\delta + m^{-1}(1+\lambda)^{-1}\lambda\rho^2\big)W,$$

where $\delta = (1+\lambda)^{-1}$ for hot-deck imputation and $\delta = 1+\lambda$ for simple residual, normal and adjusted normal imputation. We can then incorporate the estimates $\hat{U}$ and $\hat{V}$ into a single estimator,

$$n\hat{U} = n\,\delta\hat{W} + n(1+m^{-1})\hat{B}.$$

Alternatively, it is clear that

$$n\hat{S} = \big(\delta + (1+m^{-1})\hat{\gamma}\big)n\hat{W},$$

where $\hat{\gamma} = (pm)^{-1}(m-1)\mathrm{trace}(\hat{W}^{-1}\hat{B})$, also estimates the asymptotic dispersion matrix of $n^{1/2}(\hat{\theta}_{*\cdot} - \theta)$. If C is a $k \times p$ matrix of full rank as before, we can consider the quadratic forms

$$T = \big(\hat{\theta}_{*\cdot} - \theta\big)'C'\big(C\hat{U}C'\big)^{-1}C\big(\hat{\theta}_{*\cdot} - \theta\big)$$

and

$$T_3 = \big(\hat{\theta}_{*\cdot} - \theta\big)'C'\big(C\hat{S}C'\big)^{-1}C\big(\hat{\theta}_{*\cdot} - \theta\big).$$

The approximation (5.1) applies to T. Li (1985) suggested the alternative approximation

$$T \sim kF(k, \nu/k),$$

where $\nu = k(m-1)\{1 + m\,\delta/\hat{\gamma}(m+1)\}^2$. This second approximation is motivated by the fact that applying the approximate degrees of freedom method to T_3 leads to the approximation

$$T_3 \sim kF(k, \nu).$$

Li's empirical investigations, while carried out in a different context, indicate that the quality of these approximations may be improved by increasing m, the number of imputations. Of course, as $m \to \infty$ the preceding F-distributions may be replaced by χ^2_k-distributions.

The preceding results for the case $k = 1$ lead immediately to t-based interval estimates and tests for individual linear contrasts in θ. Moreover, by the Cauchy–Schwarz inequality, for any $a \in \mathscr{R}^p$,

$$\big[a'\big(\hat{\theta}_{*\cdot} - \theta\big)\big]^2 \le a'Ma\big(\hat{\theta}_{*\cdot} - \theta\big)'M^{-1}\big(\hat{\theta}_{*\cdot} - \theta\big),$$

where M is any nonsingular $p \times p$ matrix, so that a $1-\alpha$ level simultaneous interval estimate for any contrast $a'\theta$ is given by

$$a'\hat{\theta}_{*\cdot} \pm \{a'MaF^*\}^{1/2},$$

where $M = \hat{U}$, $\hat{V}$ or $\hat{S}$ and F^* is the α percentage point of the F-distribution (with $k = p$) chosen to approximate T_1 if $M = \hat{U}$, T_2 if $M = \hat{V}$ and T_3 if $M = \hat{S}$. It is straightforward to construct prediction intervals. Inference for a nonlinear function of θ, provided the function is smooth enough, may be approached by means of the usual one-term Taylor series linearization and the preceding methods.

Finally, it is of interest to consider the use of multiple imputation with nonlinear alternatives to the least-squares procedure. It is reasonable to hope that linearization would lead to appropriate analogues of Theorems 1 and 2. A complete result for the imputation methods would depend on results for bootstrapping appropriate estimators of θ in the location problem (for hot-deck imputation) and in the regression problem (for the other imputation methods).

Acknowledgments. We are grateful to a referee and an Associate Editor for helpful comments that improved the paper.

Note added in proof. Since this paper was written, the following book on multiple imputation has been published: RUBIN, D. B. (1987), *Multiple Imputation for Nonresponse in Surveys*, Wiley, New York. Chapter 4 contains results related to those in this paper, developed within a different framework.

REFERENCES

BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217.

BILLINGSLEY, P. (1979). *Probability and Measure*. Wiley, New York.

BOX, G. E. P. and TIAO, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading, Mass.

COCHRAN, W. G. (1977). *Sampling Techniques*, 3rd ed. Wiley, New York.

DAVID, M., LITTLE, R. J. A., SAMUHEL, M. E. and TRIEST, R. K. (1986). Alternative models for CPS income imputation. *J. Amer. Statist. Assoc.* **81** 29–41.

DAWID, A. P. (1982). The well-calibrated Bayesian (with discussion). *J. Amer. Statist. Assoc.* **77** 605–613.

FORD, B. L. (1983). An overview of hot-deck procedures. In *Incomplete Data in Sample Surveys* **2** (W. G. Madow, I. Olkin and D. B. Rubin, eds.) 185–207. Academic, New York.

FREEDMAN, D. A. (1981). Bootstrapping regression models. *Ann. Statist.* **9** 1218–1228.

HERZOG, T. N. and RUBIN, D. B. (1983). Using multiple imputations to handle nonresponse in sample surveys. In *Incomplete Data in Sample Surveys* (W. G. Madow, I. Olkin and D. B. Rubin, eds.) **2** 209–245. Academic, New York.

KALTON, G. and KISH, L. (1981). Two efficient random imputation procedures. *Proc. Survey Research Methods Section Amer. Statist. Assoc.* 146–151.

LI, K. H. (1985). Hypothesis testing in multiple imputation—with emphasis on mixed-up frequencies in contingency tables. Ph.D. thesis, Univ. Chicago.

LITTLE, R. J. A. (1982). Models for nonresponse in sample surveys. *J. Amer. Statist. Assoc.* **77** 237–250.

MADOW, W. G. and OLKIN, I., EDS. (1983). *Incomplete Data in Sample Surveys* **3**. Academic, New York.

MADOW, W. G., NISSELSON, H. and OLKIN, I., EDS. (1983). *Incomplete Data in Sample Surveys* **1**. Academic, New York.

MADOW, W. G., OLKIN, I. and RUBIN, D. B., EDS. (1983). *Incomplete Data in Sample Surveys* **2**. Academic, New York.

RANGA RAO, R. (1962). Relations between weak and uniform convergence of measures with applications. *Ann. Math. Statist.* **33** 659–680.

RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63** 581–592.

RUBIN, D. B. (1978). Multiple imputations in sample surveys—a phenomenological Bayesian approach to nonresponse. *Proc. Survey Research Methods Section Amer. Statist. Assoc.* 20–34.

RUBIN, D. B. (1979). Illustrating the use of multiple imputations to handle nonresponse in sample surveys. *Proc. ISI*, Manila 517–532.

RUBIN, D. B. (1983). Conceptual issues in the presence of nonresponse. In *Incomplete Data in Sample Surveys* (W. G. Madow, I. Olkin and D. B. Rubin, eds.) 2 123–142. Academic, New York.

RUBIN, D. B. and SCHENKER, N. (1986). Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *J. Amer. Statist. Assoc.* **81** 366–374.

YAO, Y. (1965). An approximate degrees of freedom solution to the multivariate Behrens–Fisher problem. *Biometrika* **52** 139–147.

DIVISION OF BIOSTATISTICS
SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF CALIFORNIA
10833 LE CONTE AVENUE
LOS ANGELES, CALIFORNIA 90024

DEPARTMENT OF STATISTICS
AUSTRALIAN NATIONAL UNIVERSITY
G.P.O. BOX 4
CANBERRA ACT 2601
AUSTRALIA