ESTIMATION IN THE PRESENCE OF NONIGNORABLE MISSING DATA AND A MARKOV SUPERPOPULATION MODEL

Stephen M. Woodruff, Bureau of Labor Statistics
441 G St. N.W., Washington D.C. 20212

KEY WORDS: Survey Sampling, Imputation

1. INTRODUCTION

In sample surveys, auxiliary variables and covariates are often used to increase the precision of survey estimators. The value of these variables for variance reduction increases as their correlation with the variables of interest increases. It will normally be the case that similar correlations exist between the members of subsets of the survey variables themselves and that in cases of partial nonresponse this correlation structure may also be employed to increase the precision of the survey estimates. This paper deals with one such situation and derives a best linesr unbiased estimator when the probability of response to an item in a sample unit is related to the value of that item.

We explore survey estimation with partial nonresponse where these correlations are induced by a superpopulation model. The object is to estimate the finite population mean for each time j, in a sequence of k adjacent time periods from sample data collected at each time period from a single (fixed over time) sample of business establishments. A best linear unbiased estimator (BLUE) is derived under the combination of both the superpopulation model that describes the behavior over time of the data and the response model that captures the nonignorable response mechanism within the sampled units.

The superpopulation model is a Markov relationship under which the expected value of the item of interest at time j is proportional to it's value at time j-1 for each member of the population. The response mechanism relates the probability of response at a given time to the value for the item of interest at that time. Both this superpopulation model and this response mechanism can be combined to form a linear relation which defines a generalized least squares BLUE.

The Markov superpopulation model is similar to the linear regression model that can be used to predict finite population means. This predictive approach is given and analyzed in Cassel, Sarndal, and Wretman (1977), Royall and Cumberland (1981a,1981b), and Royall and Herson (1973), among many others. The methods suggested here are based on BLU estimation of superpopulation means rather than prediction of the corresponding finite population means. For moderate to large population sizes the estimation of a superpopulation mean is essentially the same as the prediction of the corresponding finite population mean. These models seem to capture data relationships in certain establishment surveys, they consist of simple relationships on the first two moments of a superpopulation distribution, and they directly imply a BLUE under squared error loss. For the problem of estimation in the presence of nonresponse the superpopulation model and its corresponding BLUE provide a useful alternative to maximum likelihood estimators based on the Normal distribution as discussed by Rubin and Little (1987).

The Markov multivariate superpopulation model describing the relationship between the item of interest for k time periods within each of the N units in the finite population is given next. For each member i, of the finite population let the item of interest for the k time periods be denoted as a k-dimensional column vector, $\{Z_i\}$, which is given as:

$$Z_i = \theta + \Delta_i \quad \text{for } i=1,2,3,\ldots\ldots,N. \tag{1.1}$$

Where Z_i, θ, and Δ_i are k-dimensional column vectors, the set of random vectors $\{\Delta_i\}$ are mutually independent each with expectation the zero vector and with covariance matrix, Σ_z. The stochastic relationship between the components of Z_i is as follows. The prime denotes matrix transpose.

Let $Z_i'=(z_{i1},z_{i2},\ldots\ldots\ldots,z_{ik})$ for $i=1,2,\ldots,N$ and let β_j for $j=1,2,3,\ldots\ldots,k$ be unknown positive constants then $z_{i1} = \beta_1 + \lambda_{i1}$ and $z_{ij}=\beta_j \cdot z_{ij-1} + \lambda_{ij}$ for $j=2,3,\ldots,k$.

Where letting $\Lambda_i' = (\lambda_{i1},\lambda_{i2},\ldots\ldots\ldots,\lambda_{ik})$ for $i=1,2,3,\ldots\ldots.N$, the $\{\Lambda_i\}$ are iid random vectors with diagonal covariance matrix and with the zero vector for mean, and $V(\lambda_{ij}|z_{ij-1})=r_j z_{ij-1}$ where r_j is an unknown constant. This variance condition has been shown to be both appropriate and robust for many of the data series at the Bureau of Labor Statistics (BLS), West (1981) and Royall (1981). The covariance matrix of a Z_i, Σ_z, is necessarily of the form (b_{ij}) where:

$$b_{ij} = \begin{cases} \sigma_i^2 \prod_{\ell=i+1}^{j} \beta_\ell & \text{for } j>i \\ \sigma_i^2 & \text{for } i=j \\ \sigma_j^2 \prod_{\ell=j+1}^{i} \beta_\ell & \text{for } j<i \end{cases} \tag{1.2}$$

and σ_i^2, $i=1,2,\ldots\ldots\ldots,k$ are positive real numbers.

A k-dimensional response indicator vector of zeroes and ones, $t_i=(t_{i1},t_{i2},t_{i3},\ldots\ldots,t_{ik})'$, given for each sample unit will be defined next. The j^{th} component of t_i, t_{ij}, is one if the data for time j in sample unit i gets a response and zero if not. Conditional on Z_i it is further assumed that these components of t_i are independent.

Thus we have:

$$P(t_i = u_\ell' \mid Z_i) = \prod_{j=1}^{k} P(t_{ij} = u_{\ell j} \mid Z_i)$$

where $(u_{\ell j})$ is the $2^k \times k$ matrix whos rows are all distinct k-tuples of zeroes and ones and where u_ℓ denotes the ℓ^{th} row of $(u_{\ell j})$.

In this paper we will consider BLU-estimation under (1.1) for the $\{Z_i\}$ and under nonignorable nonresponse given by the following:

$$P(t_{ij} = 1 \mid Z_i) = L_j(z_{ij})$$

for $1 \le j \le k$, $1 \le i \le N$, where $L_j(x)$ is a polynomial with its range in [0,1] for all x in the support of z_{ij}. Given Z_i, $L_j(z_{ij})$ is the probability of response to the j^{th} item. Note that in case L=constant then we have ignorable nonresponse. The above conditions on t_i will be referred to as model (1.3).

Under the stochastic structure given by (1.1) and (1.3) we will be interested in estimating the vector of finite population means for the item of interest at times 1 through k using a conditional BLUE for θ that is denoted by $\hat{\theta}$.

In the next section this BLUE under both (1.1) and (1.3) is derived together with its variance. Estimators for unknown model parameters are also suggested.

The empirical section will consider response mechanisms applied to real data where (1.1) and (1.3) are only a crude approximation to the processes which yield the usable survey data.

Model (1.1) will often provide an adequate description for data on total number of employees within a business establishment across time, West (1981). The model (1.1) may also be used to describe the wages paid at different levels of the same occupation and (1.3) may be used to model the reluctance to respond in cases where extreme differences exist between the amount paid and the average paid.

The sample must be chosen so that the sample units can themselves be reasonably described by model (1.1). Simple random sampling is one way to achieve this end. In more complex situations, the generally accepted principles of good survey design will usually yield such samples although (1.1) must be applied with more care in the case of highly stratified cluster sampling.

The mean square errors (MSEs) of four other estimators are compared to the MSE of $\hat{\theta}$. These four estimators are:

1) The BLUE under (1.1) from the responding data ignoring (1.3) and which is denoted $\hat{\theta}(1.1)$.

2) The sample means of the responders,

$$\bar{z} = (\bar{z}_1(s_1), \bar{z}_2(s_2), \ldots, \bar{z}_k(s_k)),$$

where s_j is the set of responders at time j.

3) The product estimators

$$LR = (\hat{\beta}_1, \hat{\beta}_1\hat{\beta}_2, \hat{\beta}_1\hat{\beta}_2\hat{\beta}_3, \ldots\ldots, \prod_{j=1}^{k} \hat{\beta}_j)$$

where $\hat{\beta}_j = \bar{z}_j(s_j \cap s_{j-1}) / \bar{z}_{j-1}(s_j \cap s_{j-1})$ for $j>1$ and $\hat{\beta}_1 = \bar{z}_1(s_1)$.

4) The composite estimator, $C = (C_1, C_2, \ldots C_k)$ where C_j is defined as: $C_j = \alpha\bar{z}_j(s_j) + (1-\alpha)\hat{\beta}_j C_{j-1}$ with α chosen to minimize the variance of C_j and with $C_1 = \bar{z}_1(s_1)$.

Some simulation results using a Bureau of Labor Statistics Consumer Price Index data base are summerized in section 4 where comparisons of the MSEs of the vector $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)$ to 1) through 4) above are tabulated. In the presence of nonignorable nonresponse given by (1.3) $\hat{\theta}$ seems to give the estimator with the minimum mean square error among those tested.

2. DERIVATION AND DETAILS

In this section the data vector, design matrix, and covariance matrix for the data vector are defined and the generalized least squares BLUE is derived. Estimates for unknown superpopulation parameters are suggested and variance estimators for the BLUE are also included.

$(u_{\ell j})$ is the $2^k \times k$ matrix whos rows are all distinct k-tuples of zeroes and ones and where u_ℓ denotes the ℓ^{th} row of $(u_{\ell j})$. Let m_ℓ be the sum over the components of u_ℓ, let u_{2^k} be the row vector of zeros and let u_1 be the row vector of ones.

For any matrix (d_{ij}), d with a single subscript (d_i) will be used to denote the i^{th} row of (d_{ij}).

For $1 \le i \le 2^k - 1$ and $1 \le j \le m_i$, let F_{ij}= the column of (u_{ij}) in which the j^{th} one in u_i occurs. Let $A_i = (a_{\ell j})$ where for $1 \le \ell \le m_i$ and $1 \le j \le k$, $a_{\ell j} = 1$ if $F_{i\ell} = j$ and $a_{\ell j} = 0$ otherwise. A_i is an $m_i \times k$ matrix for $i = 1, 2, 3, \ldots\ldots, 2^k - 1$.

For $2 \le i \le 2^k$ and $1 \le j \le k - m_i$, let F^c_{ij} be the column of (u_{ij}) in which the j^{th} zero in u_i occurs. Let $B_i = (c_{\ell j})$ where for $1 \le \ell \le k - m_i$ and $1 \le j \le k$, $c_{\ell j} = 1$ if $F^c_{i\ell} = j$ and $c_{\ell j} = 0$ otherwise. B_i is a $(k - m_i) \times k$ matrix for $i = 2, 3, \ldots\ldots, 2^k$. Note $\{B_i: 2 \le i \le 2^k\} = \{A_i: 1 \le i \le 2^k - 1\}$. Let $C_i = \{\ell: 1 \le \ell \le N \text{ and } t_\ell = u_i\}$. C_i is the set of rows in (z_{ij}) which have data in exactly those columns corresponding to the nonzero columns of u_i. Let n_i= the number of elements in C_i.

Let τ denote $E(t_i)$, the expected value over both (1.1) and (1.3) of the response vector for unit i and let:

$$\bar{z}_\alpha = (1/n_\alpha) \sum_{\ell \in s_\alpha} z_\ell$$

where s_α is a subset of the first N integers and n_α is the size of s_α. Define $\bar{t}_\alpha$ similarly. When the conditional response functions, $\{L_j\}$, defined in the previous sections are linear:

$$L_j(x) = f_j x + g_j \quad (= E(t_{ij} \mid z_{ij} = x))$$

for $1 \le j \le k$, and D is the diagonal matrix of $(f_1, f_2, f_3, \ldots, f_k)$ then we have:

$$E\begin{bmatrix}\bar z_\alpha\\ \bar t_\alpha\end{bmatrix} = \begin{bmatrix}\theta\\ \tau\end{bmatrix} \quad \text{and} \quad \mathrm{Cov}\begin{bmatrix}\bar z_\alpha\\ \bar t_\alpha\end{bmatrix} = (1/n_\alpha)\begin{bmatrix}\Sigma_z & \Sigma_z D\\ D\Sigma_z & \Sigma_t\end{bmatrix}$$

Let $w_\alpha = \bar z_\alpha - \Sigma_z D \Sigma_t^{-1}(\bar t_\alpha - \tau)$. Then $E(w_\alpha) = \theta$ and $\mathrm{Cov}(w_\alpha)=(1/n_\alpha)(\Sigma_z - \Sigma_z D\Sigma_t^{-1}D\Sigma_z)$, define Σ_w to be: $\Sigma_z - \Sigma_z D\Sigma_t^{-1}D\Sigma_z$. Since w_α and $\bar t_\alpha$ are uncorrelated the conditional expected value of w_α given $\bar t_\alpha$ is approximately θ and the conditional covariance matrix of w_α given $\bar t_\alpha$ is approximately $(1/n_\alpha)\Sigma_w$ (if n_α is large enough so that $(\bar z_\alpha, \bar t_\alpha)$ is approaching normality then these approximations become exact.) Now define w_i for all i such that $C_i \neq \emptyset$ as w_α with $s_\alpha = C_i$. Let $y_i = A_i w_i$. We can summarize the information given by models (1.1) and (1.3) together with all the available sample data consisting of both the response indicator vectors for each sample establishment and their responses in the expression:

$$Y = X\theta + \varepsilon \tag{2.1}$$

where $X' = (A_1', A_2', A_3', \ldots, A_{2^k-1}')$, $Y' = (y_1', y_2', y_3', \ldots, y_{2^k-1}')$, ε is the random vector with mean zero and covariance matrix, Σ, given as the block diagonal matrix of the $(1/n_i)A_i\Sigma_w A_i'$ for all i with $n_i>0$ from i=1 in the upper left to $i=2^k-1$ in the lower right.

The BLUE of θ is given as:

$$\hat\theta = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y \tag{2.2}$$

This expression reduces to:

$$\hat\theta = \left[\sum_{i\varepsilon S} A_i'\Sigma_i^{-1}A_i\right]^{-1}\sum_{i\varepsilon S} A_i'\Sigma_i^{-1}y_i \tag{2.3}$$

where $S = \{i: C_i \neq \emptyset\}$ and $\Sigma_i = (1/n_i)A_i\Sigma_w A_i'$.

$\hat\theta$ is a function of the matrices in $\{\Sigma_i^{-1}: i\varepsilon S\}$. Each of the $\{\Sigma_i\}$ matrices are functions of Σ_z, Σ_t, and D. Note that:

$$\Sigma_i^{-} = (A_i\Sigma_w^{-1}A_i' - A_i\Sigma_w^{-1}B_i'(B_i\Sigma_w^{-1}B_i')^{-1}B_i\Sigma_w^{-1}A_i')n_i \tag{2.4}$$

for all $i\varepsilon S$ such that $A_i \neq I_k$.

Also note: $\Sigma_w^{-1} = \Sigma_z^{-1} + D(\Sigma_t - D\Sigma_z D)^{-1}D$. $\Sigma_t - D\Sigma_z D$ is the diagonal matrix with $\{\xi_j(1-\xi_j)-\sigma_j^2 f_j^2\}_{j=1}^{k}$ on the diagonal where $\xi_j = f_j\theta_j + g_j$. As described in the following paragraphs it will be possible to estimate Σ_t and Σ_z (and hence Σ_w^{-1}) without directly estimating the $\{\xi_j\}$. The $\{\xi_j\}$ may also be estimated using $\hat\theta_j = (1/n(s_j))\sum_{\ell\varepsilon s_j} z_{\ell j}/(f_j z_{\ell j} + g_j)$, a Horwitz-Thompson estimator for θ_j using the probabilities of response to adjust for nonignorable nonresponse bias.

Since the response indicator vectors, $\{t_i\}$, suffer no nonresponse Σ_t is readily estimable through $(1/(n-1))(H'H - n\bar t\bar t')$ where H is the n×k matrix with rows, $\{t_i\}_{i\varepsilon s}$ and where $\bar t'$ is the vector of column means of H. Hence it suffices to know (or estimate) Σ_z^{-1} and this can be done as follows:

Let Σ_{zj} be the covariance matrix of the vector $\boldsymbol{z}_{ij} = (z_{i1}, z_{i2}, \ldots, z_{ij})$ for $1 \le j \le k$ and $1 \le i \le N$ (note: $\Sigma_{zk} = \Sigma_z$). Then $\Sigma_{z1} = \sigma_1^2$ and $\Sigma_{z1}^{-1} = 1/\sigma_1^2$.

Let $T_1 = \sigma_1^2$ and $T_j = \sigma_j^2 - \beta_j^2\sigma_{j-1}^2$ for $j=2,3,4,\ldots,k$. Then Σ_z^{-1} can be found from the following matrix difference equation.

$$\Sigma_{zj}^{-1} = (1/T_j)\begin{bmatrix} T_j\Sigma_{zj-1}^{-1} + \beta_j^2 M_{j-1}, & -\beta_j Q_j\\ -\beta_j Q_j', & 1\end{bmatrix}$$

for $j=2,3,4,\ldots k$ and where Q_j is the (j-1)-column vector of zeros except in the $(j-1)^{st}$ row where a one appears and M_{j-1} is the (j-1)×(j-1) matrix of zeros except for the bottom right entry which is a one.

This expression gives Σ_z^{-1} under (1.1) as a tridiagonal k×k matrix function of the $\{T_j\}$ and the $\{\beta_j\}$. Next note that T_j is the expected value of the conditional variance of z_{ij} given $\boldsymbol{z}_{ij-1}$ and that with this condition the BLUE of β_j is the ratio estimator, $\hat\beta_j$, given below and the expected value of this $\hat\beta_j$ is β_j independent of the response mechanism which generates the $\{t_{ij}\}$. Given this second moment condition, an estimator of T_j for j>1 is the product of $\hat\theta_{jHT} = (1/n)\sum z_{ij}/L_j(z_{ij})$ where this summation is over the units which have responses for time j and the estimator of r_j given by:

$$\hat r_j = (1/[n(s_{j,j-1})-1])\sum_{i\varepsilon s_{j,j-1}}(z_{ij} - \hat\beta_j z_{ij-1})^2/z_{ij-1}.$$

Where $s_{j,j-1}$ denotes the set of sample units which have observations for times j and j-1, $n(s_{j,j-1})$ denotes the size of $s_{j,j-1}$ and $\hat\beta_j = \bar z_j(s_{j,j-1})/\bar z_{j-1}(s_{j,j-1})$. $\bar z_j(s_{j,j-1})$ is the sample mean of the data for time j over $s_{j,j-1}$ and $\bar z_{j-1}(s_{j,j-1})$ is the sample mean for time j-1 over these units. As with $\hat{\beta}_j$, the expected value of $\hat{r}_j$ is independent of the response mechanism.

Also note that $E(\theta_{jHT}|\{z_{ij}\},s)=\bar{z}_j(s)$ where $\bar{z}_j(s)$ is $(1/n)\sum_{i\varepsilon s} z_{ij}$ and s=the full sample of n units.

Let $T_1 = \sigma_1^2$ be estimated with:

$$(1/(n-1))\sum_{i\varepsilon s}(z_{i1}-\hat{\theta}_{1HT})^2/L_1(z_{i1})$$

Let s_j denote the set of sample units which have observations for time j and $\bar{z}_j(s_j)$ denote the sample mean over these units of the data for time j.

If the sample data permits adequate estimators for the $\{T_i\}$ and $\{\beta_i\}$ (i.e. $\{s_{j,j-1}\}$ is sufficiently large) then (2.3) can be used to estimate $\theta' = (\theta_1, \theta_2, \theta_3, \ldots\ldots, \theta_k)$

$$= (\beta_1, \beta_1\beta_2, \beta_1\beta_2\beta_3, \ldots\ldots\ldots, \prod_{j=1}^{k} \beta_j) .$$

The covariance matrix of $\hat{\theta}$ is also estimated as $(X'\hat{\Sigma}^{-1}X)^{-1}$. If $\hat{\theta}$ varies "much" from the vector of sample means used to estimate the $\{T_i\}$ then this $\hat{\theta}$ should be used to reestimate the $\{T_i\}$ and in turn to reestimate θ. This may be continued until convergence.

3. A COMPARISON OF MEAN SQUARE ERRORS

The tables in this section compare the exact mean square error (MSE) under the full model given by (1.1) and (1.3) of:

1) The BLUE under (1.1) and (1.3) as given by (2.3) and denoted as Δ_1.

2) The BLUE under (1.1) alone as given by (2.3) with D=0 and denoted as Δ_2.

3) The vector of component means from responding units for each item that is denoted as Δ_3.

This comparison is made for the case k=2. The iid random vectors $Z_i = (z_{i1}, z_{i2})'$ have support in $[0,a_1]\times[0,a_2]$ and the conditional response functions for the two components are:

$L_1(x) = (f_1/a_1)x + K_1(1-f_1)$ for $0\le x\le a_1$

$L_2(x) = (f_2/a_2)x - f_2 + K_2(1+f_2)$ for $0\le x\le a_2$

where $0\le K_1\le 1$, $0\le K_2\le 1$, $0\le f_1\le 1$, $-1\le f_2\le 0$.

For these tables $\beta_1=.45a_1$, $\beta_2=1.05$, $\sigma_1^2=0.1$, and $\sigma_2^2=\beta_2^2\sigma_1^2 + (.02)\beta_1$.

Note that when f_1 and f_2 are zero we have ignorable nonresponse. In tables 1 and 2 we summarize a comparison of the MSEs of the above estimators for estimating (θ_1,θ_2). Let MSE(m,θ) denote the MSE under (1.1) and (1.3) of Δ_q for estimating θ_m. Then for each tabled pair of parameter pairs, $[(f_1,f_2),(K_1,K_2)]$, is given a 2×2 array consisting of:

MSE(1,1)/MSE(1,3)	MSE(2,1)/MSE(2,3)
MSE(1,2)/MSE(1,3)	MSE(2,2)/MSE(2,3)

Table 1. Ratios of Mean Square Errors for Steep Response Functions

(K_1,K_2)	(.8,.8)		(1.0,1.0)	
(f_1,f_2)				
(0,0)	.845	.845	1.00	1.00
	.845	.845	1.00	1.00
(0,-.6)	.875	.009	1.00	.010
	2.39	.051	1.00	.015
(.6,0)	.009	.883	.011	1.00
	.049	2.13	.016	1.00
(.6,-.6)	.008	.008	.009	.009
	.014	.029	.015	.026
(.9,-.9)	.002	.002	.002	.002
	.016	.039	.015	.036

Table 2. Ratios of Mean Square Errors for Flater Response Functions

(K_1,K_2)	(.8,.8)		(1.0,1.0)	
(f_1,f_2)				
(0,-.1)	.850	.318	.999	.413
	.883	.350	.999	.416
(0,-.2)	.855	.105	.999	.137
	.991	.149	.999	.142
(0,-.3)	.860	.046	.999	.059
	1.18	.092	.999	.064
(.2,-.2)	.118	.105	.151	.134
	.121	.116	.155	.141
(.3,-.3)	.051	.045	.063	.056
	.054	.060	.068	.066

All the entries in the upper right array in Table 1 are one since in the case of no nonresponse all three estimators are identical. The entries in the upper left array are all .845 because here we have ignorable nonresponse, Δ_1 is identical to Δ_2, and both are slightly more precise than Δ_3.

For $[(f_1,f_2),(K_1,K_2)]=[(0,-.6),(.8,.8)]$ we see Δ_1 has slightly less MSE than Δ_3 for θ_1 and vastly less for θ_2. Note that for this case Δ_3 is unbiased for θ_1 but negatively biased for θ_2. The entry, 2.39, results from the negative bias in Δ_2 for estimating θ_1. This negative bias is caused by two things; the response mechanism that gives Δ_2 its negative bias for θ_2 and the positive correlation between the components of Δ_2. The squared bias in Δ_2 for θ_1 overwhelms the variance of Δ_3.

For $[(f_1,f_2),(K_1,K_2)]=[(.6,0),(1,1)]$ we see that all three estimators have the same MSE for θ_2 but for θ_1 only Δ_1 is unbiased. Δ_2 and Δ_3 are biased but Δ_2 has considerably smaller variance than Δ_3.

Table 2 considers less extreme cases of nonignorable nonresponse and except for magnitudes of the entries is similar to table 1.

4. SIMULATION COMPARISONS

In this section computer simulation results are summarized. These simulations attempt to reflect the numerous practical difficulties of applying the forgoing survey methodology. In particular, the problems of estimating the superpopulation parameters in model (1.1) and the effects of misspecifying the nonresponse parameters in model (1.3).

The populations used for this simulation study are derived from a CPI data base with data for four time periods from 1110 establishments. Tables 3 contains simulation results using population I (ρ=.99) where ρ is the average correlation between data from adjacent time periods, i.e. the average correlations of (z_{ij}, z_{ij-1}) for j=2,3,4. Table 4 and Table 5 have the simulation MSEs for population II (ρ=.95). The sample size is 250 and for each of these 250 sample units the data for time j is a response if $f_j z_{ij} + g_j < \text{UNIF}(i,j)$ and is a nonresponse otherwise where the UNIF(i,j) are iid uniform (0,1) random variates. For all pairs (i,j), $0 \le z_{ij} \le 1$, $1 \le i \le 1110$ & $1 \le j \le 4$.

Each of the following tables contains estimated MSEs based on 400 replications of the sampling and estimation process. This number of replications sufficed to give the tabled entries a standard error of approximately 10^{-1} or less (Note that these tabled entries are the estimated MSE$\times 10^4$). Thus both practical differences and statistically significant differences are evident between most of the estimators under study.

Table 3 through Table 5 give estimated MSEs of $\hat{\theta}$, the BLUE from (2.3) and its four competitors which are defined at the end of section 1. The estimates of the superpopulation parameters in $\hat{\theta}$ are those given in section 2. For each of these tables the parameters $\{f_j\}$, $\{g_j\}$, $\{\hat{f}_j\}$, $\{\hat{g}_j\}$, and ρ are given below the table title. The hat indicates the value of f_j or g_j used to derive $\hat{\theta}$.

Table 3. Population I Simulation Results Estimated MSE$\times 10^4$

f_j=.8, $\hat{f}_j$=.75, g_j=.1, $\hat{g}_j$=.15, for $1 \le j \le 4$ & ρ=.99

Time	1	2	3	4
Estimator				
$\bar{z}$	2.63	2.83	2.71	2.50
LR	2.63	2.64	2.61	2.63
C	2.63	2.32	1.66	1.78
$\hat{\theta}(1.1)$	0.65	0.62	0.59	0.59
$\hat{\theta}$	0.43	0.41	0.39	0.39

Table 4. Population II Simulation Results Estimated MSE$\times 10^4$

f_j=.8, $\hat{f}_j$=.75, g_j=.1, $\hat{g}_j$=.15, for $1 \le j \le 4$ & ρ=.95

Time	1	2	3	4
Estimator				
$\bar{z}$	2.63	3.00	2.91	3.04
LR	2.63	2.98	3.20	3.59
C	2.63	2.30	1.02	1.24
$\hat{\theta}(1.1)$	0.84	0.77	0.72	0.85
$\hat{\theta}$	0.44	0.43	0.41	0.46

The actual values of f_j and g_j used to generate nonresponse in the simulation are given without a hat.

The best of the four biased competitors is $\hat{\theta}(1.1)$ and when the $\{f_j\}$ are large (tables 3 and 4) it has about a 50% larger MSE than does $\hat{\theta}$. This difference in their MSEs seems to be greater when ρ=.95 than when ρ=.99. When the $\{f_j\}$ are smaller in table 5 we see that the difference between $\hat{\theta}(1.1)$ and $\hat{\theta}$ is barely significant.

Both populations I and II are skewed to the left in the interval (0,1) at each time period. This means that in cases where the nonignorable nonresponse mechanism gives higher probabilities of response to smaller values of z_{ij}, $(f_j<0)$, we should see little difference between $\hat{\theta}(1.1)$ and $\hat{\theta}$. In the examples tested here the $\{f_j\}$ are all positive and there clearly is a pleasing reduction in MSE when this nonignorability is used to derive the BLUE $\hat{\theta}$.

The tables in this section indicate that $\hat{\theta}$, may offer useful reductions in MSE in the presence of moderate to extreme nonignorable nonresponse. When the nonignorable nonresponse mechanism is less extreme (the probability distribution of t_i approaches a distribution that occurs when the data is missing at random) $\hat{\theta}(1.1)$ is the estimator of choice.

Table 5. Population II Simulation Results Estimated MSE$\times 10^4$

f_j=.5, $\hat{f}_j$=.45, g_j=.3, $\hat{g}_j$=.35, for $1 \le j \le 4$ & ρ=.95

Time	1	2	3	4
Estimator				
$\bar{z}$	1.20	1.35	1.30	1.25
LR	1.20	1.31	1.38	1.52
C	1.20	0.88	0.73	0.68
$\hat{\theta}(1.1)$	0.48	0.42	0.43	0.46
$\hat{\theta}$	0.42	0.39	0.39	0.40

5. CONCLUSIONS

This article examines a way of reducing the MSE of sample survey estimates by using the relationships which often exist between subsets of survey variables. These relationships may be expressed in terms of superpopulation models from which a linear relationship (and its attendent BLUE) may be derived. The Markov superpopulation model discussed here has been found by the Bureau of Labor Statistics to accurately model the changes over time in establishment employment, West (1981) and Royall (1981).

Additional structure in the form of a nonignorable nonresponse mechanism was joined with this Markov model to modify the estimation process for cases where the probability of nonresponse is related to the would-be response for a particular time period in a given establishment. Both this nonresponse mechanism and this Markov Superpopulation model are combined in a single linear relationship where the response indicator vector may be treated as a covariate (it is completely known for each sample member). By conditioning on this known response indicator vector the response bias resulting from this nonresponse mechanism is removed from the observed data. A linear model using this conditioned data and involving the parameters we want to estimate, is used to construct a BLUE for the vector of mean values.

This BLUE was examined both theoretically and empirically in sections 3. and 4. where it performed well against its competitors. In section 3, the theoretical MSEs of three estimators were compared and some eyepoping results are seen in the tables of this section. In section 4 the more realistic simulation results bring us back to earth again. Although these simulations still show substantial reductions in MSE from the BLUE given by (2.3) as compared with the other biased estimators these differences in MSE are far less than those found in the theoretical MSEs of section 3. This is caused by the practical difficulties of estimating additional response mechanism parameters.

When nonignorability is most extreme (large $\{f_j\}$) $\hat{\theta}$ has the most to offer for reducing MSE. This appears to hold both when the correlation between adjacent time periods, ρ, is very close to one and when it is smaller.

Estimation of the response mechanism parameters $\{f_j\}$ and $\{g_j\}$ was not addressed here. These must be obtained from a source external to the available data and its stochastic structure as described in sections 1. and 2. The Bureau of Labor Statistics is currently establishing a survey cognition laboratory for measuring phenomena concerning the reaction of survey participants to questions being asked of them. In particular, data on a survey participants decision about whether or not to respond to a given question could become available for designing an appropriate response mechanism.

In cases where auxiliary data that is correlated with the survey variates is available, or weaker superpopulation structure must be assumed for the survey variates, the methods used here may still be applied whenever a good estimate of Σ, the covariance matrix of the vector consisting of the survey variates, the auxiliary variables, and any covariates, is available.

This paper was concerned with estimating a time series of population means from a sample that was fixed over time. The Markov superpopulation model, (1.1), may also capture the stochastic relationship between hierarchical variates like wages paid in different levels of the same occupation. In this case, the occupational levels which are missing in a given establishment may be treated as the item nonresponses and $\hat{\theta}$ or $\hat{\theta}(1.1)$ used as appropriate to estimate wages at all levels of the occupation.

In repeated sampling problems with sample overlap between adjacent time periods, (1.1), may describe the relation between survey variates from adjacent times. For such estimation problems, composite estimators are often used but the BLUE under (1.1) has been shown to be considerably more precise.

Superpopulation relationships like those used in this paper to derive an efficient estimator are routinely used in sample design (in stratified, systematic, and clustered designs). Apparently, there is still more to be gained by their use in the construction of survey estimators.

REFERENCES

1) Cassel,C. , Särndal, C. and Wretman, J.H. (1977), Foundations of Inference in Survey Sampling, John Wiley & Sons.

2) Little J. A. and Rubin B. R. (1987), "Statistical Analysis With Missing Data",Wiley.

3) Raj D. (1968), Sampling Theory, McGraw-Hill.

4) Royall, Richard M. (1981), "Study of the Role of Probability Models in 790 Survey Design and Estimation", Bureau of Labor Statistics contract Report 80-98.

5) Royall, R.M., and Cumberland. W.G. (1981a), "An Empirical Study of the Ratio Estimator and Estimators of its Variance," Journal of the American Statistical Association, 76, 66-77.

6) Royall, R.M., and Cumberland, W. G. (1981b), "The Finite-Population Linear Regression Estimator and Estimators of its Variance __ An Empirical Study," Journal of the American Statistical Association, 76, 924-930.

7) Royall, R. M., and Herson, J. H. (1973), "Robust Estimation in Finite Populations I," Journal of the American Statistical Association, 68, 880-889.

8) Scott A.J., and Smith T.M.F. (1974), "Analysis of Repeated Surveys Using Time Series Methods," Journal of the American Statistical Association, 69, 674-678.

9) Tenenbein, A. (1970), "A Double Sampling Scheme for Estimating from Misclassified Binomial Data," Journal of the American Statistical Association, 65, 1350-1361.

10) West, Sandra A. (1981), "Linear Models for Monthly All Employment Data", Bureau of Labor Statistics report.