A MULTIVARIATE APPROACH TO ESTIMATION IN FINITE POPULATION SAMPLING WHEN NONIGNORABLE NONRESPONSE IS PRESENT

Stephen M. Woodruff, Bureau of Labor Statistics
Room 2128, 441 G St. N.W., Washington D.C. 20212

Key Words: Superpopulation Model,
Nonignorable Nonresponse.

1. INTRODUCTION

This paper describes a technique for estimation from *typical* sample survey data. Such data is riddled with nonresponse due to processes unknown and, in addition, may result from poorly implemented or documented sample designs. The multivariate structure usually found in sample survey data is used to compensate for these data deficiencies by building all available relevant information into the estimator. This estimator minimizes average error with respect to the stochastic structure defined by this available information.

The estimator derived here is a best linear unbiased estimator (BLUE). It is a generalization of the ordinary regression estimator, Cochran (1977). Regression transformations are used to condition on known outcomes of auxiliary variables and covariates. This is similar to superpopulation prediction theory since both produce estimators which are robust against extreme samples by conditioning on auxiliary data; see Cassel, Särndal, and Wretman (1977), Royall and Cumberland (1981a) (1981b), Royall and Herson (1973). The close tie between estimating first moments and estimating second moments parallels the similar tie in the EM–algorithm; see Little and Rubin (1987), Pfeffermann (1988), Srivastava and Carter (1986). The estimation problem that this paper solves is very similar to the one that Pfeffermann (1988) discusses. The solution is different from his in two major respects. The covariance matrix is estimated using an identity derived from a superpopulation model and conditioning on sample outcomes of covariate/auxiliary variables to remove bias is done by response group (to be defined).

Estimating finite population means is usually done by weighting sample responses inversely to their probability of selection. This is done item by item resulting in the Horvitz–Thompson estimator; see Cochran (1977), Raj (1968). In cases where auxiliary variables are available, these auxiliaries may be used to derive improved estimators for a single item mean. Ratio and regression estimators are examples of this, but these estimators remain essentially univariate.

Many different but related items of information are collected on sample survey questionnaires. These relationships suggest that survey estimators should be derived under a multivariate framework, which describes these item dependencies. This is particularly true when data items are missing due to either nonresponse or designed nonavailability, as in the case of rotation sampling.

The finite population to be sampled is described in terms of an N×k matrix, Z=(A,C,T), of random variables where the submatrix of auxiliary variables, A is N×m_a, the submatrix of covariates, C is N×m_c, and the submatrix of target variables, T is N×m_t. Let the realization of Z be denoted, $\mathcal{Z}$. Let, $\mathcal{Z}$=(a,c,t) where each row of $\mathcal{Z}$, $\mathcal{Z}_i$=(a_i,c_i,t_i) gives the outcomes (realizations) of the $k=m_a+m_c+m_t$ random variables attached to the i^{th} member of the finite population for $1\le i\le N$.

To illustrate the estimation methodology consider a simple example in which N=6 and k=4. The rows of this example of Z are normally and independently, identically distributed (iid), each with a mean vector, (μ_a,μ_c,μ_t).

Some of the columns of $\mathcal{Z}$ are known for each member of the finite population. These are referred to as auxiliary variables. A sampling distribution which is usually a function of these auxiliary variables generates a sample of n<N rows of $\mathcal{Z}$ (n=3 in the example). The statistician then attempts to observe the remaining components (non auxiliary variables) of $\mathcal{Z}_i$ for all i in the sample. Unfortunately, nonresponse may occur, and only some subset of the nonauxiliary components of $\mathcal{Z}_i$ may be observed. Usually, this subset will vary from sample member to sample member.

These nonauxiliary variables may be further divided into two groups — covariates and target variables. The goal of the sample survey is to measure the finite population means of each target variable. The target variables are observed only for sample members, and they suffer the nonresponse. The covariates are observed without nonresponse for all sample members.

For this example, μ_t is to be estimated from the following four pieces of information: 1) all the auxiliary variable outcomes in $\mathcal{Z}$, together with μ_a and μ_c, 2) all the covariate outcomes in $\mathcal{Z}$ for the sample members, 3) a subset of the target outcomes in $\mathcal{Z}$ for the sample members. This subset is the result of nonresponse and varies from sample member to sample member, and 4) the available information on the distribution of the random matrix Z.

Suppose that rows 2, 3, and 6 were sampled from Z. In the interests of simplicity the sampling distribution will be ignored in this example. The indicator functions are used in the derivation of the BLUE in subsequent sections. In row 2, T_{21} was a nonresponse, as was T_{32} in row 3, and in row 6 no nonresponse occurred. The available sample and auxiliary data for this example is summarized in the following matrix of random variables and realizations of random variables:

Auxilaries — Covariates — Target Variables

$$Z_{\text{post sampling}} = \begin{bmatrix} a_{11} & C_{11} & T_{11} & T_{12} \\ a_{21} & c_{21} & T_{21} & t_{22} \leftarrow \\ a_{31} & c_{31} & t_{31} & T_{32} \leftarrow \\ a_{41} & C_{41} & T_{41} & T_{42} \\ a_{51} & C_{51} & T_{51} & T_{52} \\ a_{61} & c_{61} & t_{61} & t_{62} \leftarrow \end{bmatrix} \text{ Sample}$$

The second source of information is the distribution of the random matrix Z. Since mean square error (MSE) is the estimation criterion, the covariance matrix of Z will suffice and since the rows of Z are iid the covariance matrix of a Z_i will suffice. Let Σ_z denote the covariance matrix of a Z_i. When Z contains all available correlated variables, its rows may be modeled as iid random vectors.

Z is initially a matrix of iid random vectors in which some random components are being replaced with their realizations. First, the auxiliary data arrives, then the sample data. When the auxiliary data is observed, Z becomes a mixture of this known auxiliary data, and as yet unobserved random variables. Conditional on the observed auxiliary data, the rows of Z are no longer iid since this auxiliary data will distinguish the distribution of different rows of Z. These distributional deviations between rows of

Z given the auxiliary data, are generally dealt with in the sampling literature by unequal probability sampling (stratification and clustering). Then the sample target variables are weighted inversely to selection probabilities, to get an estimator for the target variable means, which is unbiased with respect to the sampling distribution. This is the Horvitz–Thompson estimator.

This example shows a different approach to estimating target means which is based on transformations of the target components of Z. These transformations harness the knowledge contained in the distribution of Z to adjust the sample target variables for the auxiliary variable and covariate observations in the sample units.

These transformations are linear regression adjustments. They are applied to averages over each response group of sample row vectors of Z. A response group is a subset of sample units that responded to exactly the same target variables. In the example, there are three response groups, one for each sample member and the response group averages are the trivial averages (averages over a single item). To describe these response group transformations, partition Σ_z as follows:

$$\Sigma_z = \begin{bmatrix} H_g & H_{gt} \\ H_{tg} & \Sigma_t \end{bmatrix}$$ where H_g is the covariance matrix of (A_{i1},C_{i1}), Σ_t is the covariance matrix of (T_{i1},T_{i2}) and H_{gt} is the covariance matrix between these two subvectors. The transformation to recover iid subsequent to observing auxiliaries and covariates gives us a new vector, (W_{i1},W_{i2}), given by:

$$(W_{i1},W_{i2})=(T_{i1},T_{i2}) - (A_{i1}-\mu_a ,C_{i1}-\mu_c)H_g^{-1} H_{gt}$$

for i=2,3, and 6.

Note that conditional on the auxiliaries and covariates, $E(W_{i1},W_{i2})=(\mu_{t1},\mu_{t2})$ and $Cov(W_{i1},W_{i2})= \Sigma_t - H_{tg}H_g^{-1}H_{gt} = \Sigma_w$ for i=2,3, and 6. The last equality is the definition of Σ_w and it implies that $Var(W_{ij}|(A_i,C_i))=Var(T_{ij}|(A_i,C_i))\leq Var(T_{ij})$ for i in the sample and j=1 or 2. Nonresponse resulted in T_{21} and T_{32} being unobserved, and thus, W_{21} and W_{32} are unobservable. W_{22}, W_{31}, W_{61}, and W_{62} remain, and they are related to μ_t by the following expression:

$$Y = \begin{Bmatrix} W_{22} \\ W_{31} \\ W_{61} \\ W_{62} \end{Bmatrix} = \begin{Bmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{Bmatrix} \mu_t' + \varepsilon = X\mu_t' + \varepsilon, \quad (*)$$

where the first and last equalities in this line are definitions of Y and X. The covariance matrix of ε is the diagonal matrix of the appropriate submatrices of Σ_w. Let Σ denote the covariance matrix of ε. Then the least squares BLUE for μ_t' is $\hat{\mu}_t' = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y$ and its variance is $(X'\Sigma^{-1}X)^{-1}$.

The sampling distribution and the nonresponse mechanism may be included in this BLUE by including the sample indicator function in Z as an auxiliary variable and including the response indicator vector in Z as covariates. The response indicator vector is a vector of zeros and ones. It is the same dimension as the vector of target variables, and its l^{th} component is one if the l^{th} component of the target vector is observed (a response) and zero if not. The sample indicator function is defined for each member of the finite population as one if the member is selected and zero if not.

By properly designing Z the effects on the available data of stratification and clustering in the selection process are removable by conditioning the target variables on known auxiliary variables and covariates.

The reader by wonder why the regression adjustments by response group of target variables are necessary. In particular, why not delete the unobserved components of Z, line up the observed components of Z, as was done with the (W_{ij}), (see Pfeffermann 1988) to get a version of (*) with $\mu = (\mu_a,\mu_c,\mu_t)'$ on the right hand side? This version of (*) implies a BLUE for μ_t but this estimator for μ_t is inappropriate for two reasons.

First, the linear regression equations used in later sections to describe the relationship between components of a row of Z may induce linear dependencies between components of μ. Therefore, the desired solution solves a restricted minimization problem, but the normal equations referred to in the preceding paragraph with μ on the right hand side solve the unrestricted minimization.

The second reason may be more intuitive. It is usually desireable to condition decisions on all available relevant information (see conditionality in Cox and Hinkley, (1974)).

Note the following pounts about this estimation methodology:

1) there is extreme flexibility for including all available relevant data and data relationships

2) the BLUE is conditioned on all known and readily estimable quantities

3) when the response mechanism is known the response indicator vector may be included as a covariate

4) collapsing the data by response group and doing the regression adjustment for auxiliary and covariate means by response group may reduce the effect of nonignorable nonresponse in cases where the nonresponse mechanism is unknown

5) the sampling distribution may be integrated into the estimator by including the sample indicator function in Z as an auxiliary variable

6) a variance estimator for the BLUE is included in its derivation and this variance estimator, $(X'\hat{\Sigma}^{-1}X)^{-1}$, measures both variance with respect to the distribution the BLUE inherits from Z and, in some situations, the variance of the BLUE with respect to repeated sampling.

7) the BLUE derived here will usually outperform the Horvitz–Thompson estimator. A model based ratio estimator for the estimation problem considered in the rest of this paper is compared to the BLUE. This ratio estimator has been in use at BLS for several decades and in spite of considerable effort to improve it, nothing substantially better has been found until the BLUE derived here.

2. THE LINEAR REGRESSION EQUATION FOR AN EMPLOYMENT SURVEY

The Bureau of Labor Statistics' (BLS) Current Employment Statistics Survey (CES) is a monthly survey of over 275,000 nonagricultural business establishments. This survey produces estimates of total employment, women and production workers, and hours and earnings. These estimates are made for over 1500 industries. Monthly employment level and month to month change are of primary importance to the users of this survey. An estimator for monthly employment level is derived in section 3. This estimator is an upscale version of the estimator derived in section 1 and uses the data relationships to be defined next.

The linear regression equation states that expected employment in month j is proportional to actual employment during month j–1 for each member of the population. The constant of proportionality, β_j, is the same for each member of the finite population. A response mechanism is hypothesized, which relates the probability of response for a target variable to its realized value. This response mechanism was introduced as a way to test the robustness of the BLUE, derived under ignorable nonresponse. A second BLUE, which assumes an approximate knowledge of this response mechanism, is also derived. This second BLUE serves as a benchmark with which to compare the BLUE under ignorable nonresponse (or complete ignorance of the response mechanism).

The linear regression equation relating the components of Z_i is given next. There is a single auxiliary variable, benchmark employment, available for this employment survey. The benchmark employment is the total establishment employment for the initial month (j=0) and is known for every business establishment in the finite population. The benchmark employment for establishment i is denoted A_i for $1\le i\le N$. The total employment for establishment i during each month after the benchmark month are the target variables, denoted $T_{i1},T_{i2},\ldots T_{im_t}$. These are the employments in establishment i for j=1, 2,... ,m_t months after the benchmark month. The response indicator vector is the only set of potential covariates. Thus, this vector will be denoted, C_i, and not included in Z_i until later. Under ignorable nonresponse C_i is uncorrelated with (A_i,T_i) and the vectors, $Z_i=(A_i,T_{i1},T_{i2}\ldots T_{im_t})$, $1\le i\le N$, are iid. The components of Z_i are related as follows:

$$\text{Let } A_i=\beta_0+\lambda_{i0},\ T_{i1}=\beta_1 a_i+\lambda_{i1}, \text{ and}$$
$$T_{ij}=\beta_j t_{ij-1}+\lambda_{ij} \text{ for } j>1 \qquad (2.1)$$

where $\{(\lambda_{i0},\lambda_{i1},\ldots,\lambda_{im_t})': 1\le i\le N\}$ are iid random vectors with mean zero and diagonal covariance matrix. The $\{\beta_j\}$ are unknown constants, the variance of λ_{i0} exists, and the variance of λ_{ij} given t_{ij-1} is proportional to t_{ij-1} $(\mathrm{Var}(\lambda_{ij}|t_{ij-1})=r_j t_{ij-1})$ where the $\{r_j: 1\le j\le m_t\}$ are unknown constants $(\mathrm{Var}(\lambda_{i1}|a_i)=r_1 a_i)$.

This conditional variance property for the $\{\lambda_{ij}\}$ has been shown to be appropriate for total employment as measured in the CES (West (1981), Royall (1981)). The covariance matrix of any Z_i, Σ_z, is necessarily of the form (b_{lj}) where:

the diagonal entries are: $b_{ll}=\sigma_l^2$ for $l=0,1,2,\ldots m_t$ (2.2).

and the offdiagonal entries are: $b_{lj}=\sigma_l^2\prod_{k=l+1}^{j}\beta_k$ for $l<j\le m_t$

and where σ_l^2, for $l=0,1,2,\ldots m_t$ are unknown constants (functions of the $\{r_l\}$ and $\{\beta_l\}$).

$C_i=(C_{i1},C_{i2},\ldots,C_{im_t})$ where $C_{ij}=1$ if the data for target variable j in unit i is a response and $C_{ij}=0$ if not.

A nonignorable nonresponse mechanism is introduced next. The BLUE under this additional structure is used in the simulation study (section 5) as a benchmark with which to compare the estimators to be evaluated there. Conditional on t_i, it is assumed that the components of C_i are independent and that:

$$P(C_{ij}=1|T_{ij}=t_{ij})=f_j t_{ij}+g_j \text{ for } 1\le i\le N \text{ and } 1\le j\le m_t \quad (2.3).$$
$$\&\ P(C_i=u_l|T_i=t_i)=\prod_{j=1}^{m_t}P(C_{ij}=u_{lj}|T_i=t_i),$$

where the $\{(f_j,g_j)\}$ are defined so that $f_j t_{ij}+g_j$ is between zero and one for all t_{ij} and where (u_{lj}) is the $2^{m_t}\times m_t$ matrix with rows that consist of all possible distinct m_t–tuples of zeros and ones (its last row consists of all zeros). u_l denotes the l^{th} row of (u_{lj}). Note that when $f_j=0$ for all j, the nonresponse is ignorable.

μ_c is the expected value under both (2.1) and (2.3) of C_i.

Let $\bar{T}_\alpha=(1/n_\alpha)\sum_{l\in s_\alpha}T_l$ where s_α is an arbitrary subset of the first N integers and n_α is the size of s_α. Define $\bar{A}_\alpha$ and $\bar{C}_\alpha$ similarly.

Before considering either sampling or nonresponse, the stochastic structure given in (2.1) and (2.3) may be summarized for arbitrary subsets s_α as follows:

$$E(\bar{A}_\alpha,\bar{C}_\alpha,\bar{T}_\alpha)=\mu=(\mu_a,\mu_c,\mu_t) \quad \text{and}$$
$$\mathrm{Cov}(\bar{A}_\alpha,\bar{C}_\alpha,\bar{T}_\alpha)=(1/n_\alpha)\begin{bmatrix}\Sigma_a & \Sigma_{at}D & \Sigma_{at}\\ D\Sigma_{ta} & \Sigma_c & \Sigma_t D\\ \Sigma_{ta} & D\Sigma_t & \Sigma_t\end{bmatrix} \text{ where:}$$

1) D is the diagonal matrix of $(f_1,f_2,\ldots,f_{m_t})$

2) $\begin{bmatrix}\Sigma_a & \Sigma_{at}\\ \Sigma_{ta} & \Sigma_t\end{bmatrix}=(b_{lj})$ as given in (2.2). Σ_a is the variance of the auxiliary variable, Σ_t is the covariance matrix of the target variables, and Σ_{at} is the covariance matrix between the auxiliary variable and the target variables $(\Sigma'_{at}=\Sigma_{ta})$

3) Σ_c is the covariance matrix of the covariates with respect to both (2.1) and (2.3). It has off diagonal elements, which are the same as those of $D\Sigma_t D$, and diagonal elements, $\{\mu_{cj}(1-\mu_{cj})\}_{j=1}^{m_t}$ where μ_{cj} is the j^{th} component of μ_c. (2.4)

Following the methodology outlined in the introduction,

$E(\bar{T}_\alpha|\bar{A}_\alpha,\bar{C}_\alpha)=\mu_t+(\bar{A}_\alpha-\mu_a,\ \bar{C}_\alpha-\mu_c)H_g^{-1}H_{gt}$. If $\bar{W}_\alpha$ is defined as:

$$\bar{W}_\alpha=\bar{T}_\alpha-(\bar{A}_\alpha-\mu_a,\ \bar{C}_\alpha-\mu_c)H_g^{-1}H_{gt},$$

then $E(\bar{W}_\alpha|\bar{A}_\alpha,\bar{C}_\alpha)=\mu_t$, the vector to be estimated. The covariance matrix of $\bar{W}_\alpha$ is $(1/n_\alpha)\Sigma_w$ where $\Sigma_w=\Sigma_t-H_{tg}H_g^{-1}H_{gt}$, $H_{gt}=\begin{bmatrix}\Sigma_{at}\\ \Sigma_t D\end{bmatrix}$ and $H_g=\begin{bmatrix}\Sigma_a & \Sigma_{at}D\\ D\Sigma_{ta} & \Sigma_c\end{bmatrix}$.

$\bar{W}_\alpha$ and $(\bar{A}_\alpha,\bar{C}_\alpha)$ are uncorrelated, therefore, the conditional expected value of $\bar{W}_\alpha$ given $(\bar{A}_\alpha,\bar{C}_\alpha)=(\bar{a}_\alpha,\bar{c}_\alpha)$ is approximately μ_t and the conditional covariance matrix of $\bar{W}_\alpha$ given $(\bar{A}_\alpha,\bar{C}_\alpha)=(\bar{a}_\alpha,\bar{c}_\alpha)$ is approximately $(1/n_\alpha)\Sigma_w$ (if n_α is large enough so that $(\bar{A}_\alpha,\bar{C}_\alpha,\bar{T}_\alpha)$ is approaching

normality, then these approximations approach exact equality).

Now suppose a probability sample is selected according to a sampling scheme, which is a function of the auxiliary variable. Let π_i denote the probability of selection for the i^{th} member of the finite population (i^{th} row of Z).

After the sample has been selected and the data (minus the nonresponse) has been observed for the m_t target variables, the sample units are grouped according to their response patterns. Let $\{\mathscr{G}_l\}$ be the set of these groups, and $\mathscr{G}_l=\{i:1\le i\le N \text{ and } C_i=u_l\}$. $\mathscr{G}_l$ is the set of sample units which have responses for exactly those target variables corresponding to the nonzero components of u_l. For all l such that $\mathscr{G}_l\neq\phi$, let $\overline{W}_l$ be the $\overline{W}_\alpha$ defined above with $s_\alpha=\mathscr{G}_l$ and let n_l be the number of sample units in $\mathscr{G}_l$.

A modest generalization of the methodology outlined in the introduction will produce a BLUE for the target finite population means. This BLUE is derived in section 3 together with estimates of the unknown regression equation parameters.

3. THE BEST LINEAR UNBIASED ESTIMATOR AND ESTIMATORS OF THE UNKNOWNREGRESSION PARAMETERS

Let G_l be the response indicator matrix for the l^{th} response group, $\mathscr{G}_l$, formed by deleting from the $m_t\times m_t$ identity matrix all columns, j, such that $u_{lj}=0$. Then $y_l=\overline{W}_l G_l$ consists of only those components of $\overline{W}_l$, which are observable (responses). The observed data and data relationships are summarized:

$$Y=X\mu_t' + \varepsilon \qquad (3.1)$$

where $X'=(G_1,G_2,\dots,G_{2^{m_t}-1})$,

$Y'=(y_1,y_2,\dots,y_{2^{m_t}-1})$.

and ε is the random vector with mean, zero and covariance matrix, Σ given as the block diagonal matrix of the $\Sigma_l=(1/n_l)G_l'\Sigma_w G_l$ for all l with $n_l>0$ from $l=1$ in the upper left to $l=2^{m_t}-1$ in the lower right. X, Y, and ε contain only those arrays $\{G_l, y_l, \text{ and } (\varepsilon_l,\Sigma_l)\}$ such that $n_l>0$.

The BLUE, $\hat{\mu}_t'$ is: $(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y$

$$=\left[\sum_{l\in\mathscr{S}} G_l\Sigma_l^{-1}G_l'\right]^{-1}\sum_{l\in\mathscr{S}} G_l\Sigma_l^{-1}y_l \qquad (3.2)$$

where $\mathscr{S}=\{l:n_l>0\}$.

Σ_z, D, μ_a, and μ_c are necessary to derive the $\{y_l\}$, Σ_w, and $\hat{\mu}_t'$. D is known. μ_c, the vector of population response indicator means, contains the means of the only covariates for this problem and is estimable with the Horvitz–Thompson estimator (recall the covariates suffer no nonresponse). Denote this estimator, $\hat{\mu}_c$. Σ_z is estimated by inverting $\hat{\Sigma}_z^{-1}$, where Σ_z^{-1} is estimated by inserting estimates of β_j and τ_j for $0\le j\le m_t$ into Σ_z^{-1}. Recall the $\{\beta_j\}$ are defined in (2.1). $\tau_0=\sigma_0^2$, and $\tau_j=\sigma_j^2-\beta_j^2\sigma_{j-1}^2$ for $1\le j\le m_t$, where the $\{\sigma_j^2\}$ are defined in (2.2). (2.2) also implies that the inverse of Σ_z is the tridiagonal matrix with diagonal elements given by:

$(1/\tau_j) + (\beta_{j+1}^2/\tau_{j+1})$ for $0\le j<m_t$ and with the last diagonal entry, $1/\tau_{m_t}$.

The $(j,j+1)^{th}$ and $(j+1,j)^{th}$ off–diagonal elements of Σ_z^{-1} for $0\le j<m_t$ are $-(\beta_{j+1}/\tau_{j+1})$. All other entries of Σ_z^{-1} are zero.

β_o is estimated with the finite population mean, $\bar{a}=\hat{\beta}_0$, of the auxiliary variables $\{a_i\}_{i=1}^N$. $\beta_1,\beta_2,\dots\ \beta_{m_t}$ are estimated with the ratio of matched sums of target variables. For example, $\hat{\beta}_j=(\sum_{i\in s_j} t_{ij} / \sum_{i\in s_j} t_{ij-1})$, where s_j is the set of sample units with responses for both target variable j and target variable j–1. Let $\hat{\theta}_j=\prod_{l=0}^{j}\hat{\beta}_l$ for $j=0,1,\dots,m_t$. Initially, let $\hat{\mu}_a=\hat{\theta}_0$ and $\hat{\mu}_t^o=(\hat{\theta}_1,\hat{\theta}_2,\dots,\hat{\theta}_{m_t})$. (3.3)

τ_j is the expected value of the conditional variance of T_{ij} given $(a_i,t_{i1},\dots,t_{ij-1})$ for $j\ge 1$. $\tau_j=r_j\mu_{tj-1}$ for $j=1,2,\dots,m_t$. Let $\hat{\tau}_j=\hat{r}_j\hat{\theta}_{j-1}$ for $j=1,2,\dots,m_t$.

where: $\hat{r}_j=(1/[n(s_j)-1])\sum_{i\in s_j}(t_{ij}-\hat{\beta}_j t_{ij-1})^2/t_{ij-1}$ and $n(s_j)$ is the size of s_j.

$$\hat{\tau}_0=\hat{\sigma}_0^2=(1/N-1)\sum_{i=1}^{N}(a_i-\bar{a})^2.$$

This completes the estimation of Σ_z, and with this $\hat{\Sigma}_z$, $\hat{\mu}_c$ and $\hat{\mu}_a$ the $\{\overline{W}_l\}$ and $\hat{\Sigma}_w$ may be calculated and in turn, $\hat{\mu}_t$ as given in (3.2). With the exception of one additional covariate, this is how the BLUE, $\hat{\mu}_t$, was derived in the simulation study to be described in section 5. This BLUE was calculated in two versions. The first version assumes D=0 and would be used when the nonresponse mechanism is either ignorable or unknown. The second version assumes an approximate knowledge of D to derive the BLUE. This second BLUE provides a benchmark with which to compare the BLUE under the weakened model (assuming D=0).

If $\hat{\mu}_t$ varies much from the initial estimate $\hat{\mu}_t^o=(\hat{\theta}_1,\hat{\theta}_2,\dots\ \hat{\theta}_{m_t})$, which is defined above, then this $\hat{\mu}_t$ may be used in place of $\hat{\mu}_t^o$ to reestimate Σ_z (and the $\{\overline{W}_l\}$) and thus to reestimate $\hat{\mu}_t$. This may be continued until convergence. The simulation results described in section 5 do not use this iterative reestimation process, which characterizes the EM–algorithm.

4. THE EFFECT OF IGNORING PERTINENT DATA

The effect of disregarding either useful data or known data relationships is quantified next. This is done by computing the MSE under the full (or strong) model of the BLUE derived under a model (weak) that assumes only some proper subset of the full models data and data relationships. This MSE is always larger than the variance of the BLUE derived under the full model. The bias and variance components of this additional MSE are derived below.

Suppose the same subset of the auxiliary variables and covariates is deleted from each Z_i. Let these random

variables be denoted by, Z_{i1}. Let the remaining auxiliary variables and covariates be Z_{i2}. Then the two random vectors, (Z_{i1},Z_{i2}) and (A_i,C_i), contain exactly the same random variables but they may be ordered differently.

Let: $E(Z_{i1},Z_{i2},T_i)=(\mu_1,\mu_2,\mu_t)$ and $Cov(Z_{i1},Z_{i2},T_i)=\begin{bmatrix} \Sigma_1 & \Sigma_{12} & \Sigma_{1t} \\ \Sigma_{21} & \Sigma_2 & \Sigma_{2t} \\ \Sigma_{t1} & \Sigma_{t2} & \Sigma_t \end{bmatrix}$.

Let $\hat{\mu}_t^c$ denote the BLUE under the model where the rows of the matrix, Z, are (Z_{i2},T_i). In this case, $\overline{W}_l^c = \overline{T}_l - (Z_{l2}-\mu_2)\Sigma_2^{-1}\Sigma_{2t} = \overline{W}_l + b_l$, where $\overline{W}_l$ is as defined in section 2 and $\overline{W}_l^c$ is its analogue for this weak model. The difference between $\overline{W}_l^c$ and $\overline{W}_l$ is b_l and given by the expression:

$b_l = O_{l1}L\Sigma_{1t} - O_{l1}LM\Sigma_{2t} - O_{l2}M'L\Sigma_{1t} + O_{l2}M'LM\Sigma_{2t}$, where:

$O_{l1} = (\overline{Z}_{l1}-\mu_1)$, $O_{l2} = (\overline{Z}_{l2}-\mu_2)$,

$L = (\Sigma_1 - \Sigma_{12}\Sigma_2^{-1}\Sigma_{21})^{-1}$, and $M = \Sigma_{12}\Sigma_2^{-1}$.

If y_l^c denotes the weak model version of y_l, then $y_l^c = \overline{W}_l^c G_l = \overline{W}_l G_l + b_l G_l = y_l + b_l G_l$. Recall that since all auxiliary variables and covariates are being conditioned upon, the conditional variance of b_l is zero. $Cov(y_l) = Cov(y_l^c) = (1/n_l)G_l'\Sigma_w G_l$.

The BLUE under the weak model is: $\hat{\mu}_t^{c'} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y^c$, where X and Σ are exactly as given in section 3 and $Y^c=(y_1^c,y_2^c,\ldots\ y_{2m_t-1}^c)'$. This weak model BLUE is related to the BLUE under the full model by: $\hat{\mu}_t^{c'} = \hat{\mu}_t' + HB$ where

$H = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}$ and
$B=(b_1G_1,b_2G_2,\ldots\ ,b_{2m_t-1}G_{2m_t-1})'$. Since HB is constant under the full model, $\hat{\mu}_t^c$ and $\hat{\mu}_t$ differ, only by this constant bias, HB.

$MSE(\hat{\mu}_t^{c'}) = Cov(\hat{\mu}_t') + HBB'H'$ and this implies that the MSE of each component of $\hat{\mu}_t^c$ is larger than the MSE (variance) of the corresponding component of $\hat{\mu}_t$.

Just as ignoring pertinent covariates or auxiliary variables will add only bias to the BLUE, it can be shown that ignoring pertinent target variables adds only variance to the BLUE.

5. SIMULATION STUDY

The empirical MSEs of four estimators are compared here. The four estimators are:

1) $\hat{\mu}_t$ under (2.1) and (2.3), D≠0

2) the BLUE, $\hat{\mu}_t^c$, under (2.1) alone (= $\hat{\mu}_t$ with D=0)

3) HT, the Horvitz–Thompson estimator for the target variables adjusted for nonresponse. The nonresponse adjustment factor is the ratio of total sample weights to total responding unit weights for each target variable

4) LR, the link relative estimator given by (3.3), $(\hat{\theta}_2,\hat{\theta}_3,\hat{\theta}_4,\hat{\theta}_5)$. This is the estimator currently used by the BLS for the employment survey referred to in the introduction.

The populations used for this study were derived from a BLS data base with data for six consecutive months from 1110 establishments. $Z_i=(A_i,T_{i1},T_{i2},\ldots\ T_{i5})$, where the components of Z_i are related by (2.1) and lie in the interval (0,1000). For this simulation T_{i1} is treated as a covariate (no nonresponse) and conditioned upon by using the Horvitz–Thompson estimator to estimate μ_{t1}. The vector, $(\bar{t}_2,\bar{t}_3,\bar{t}_4,\bar{t}_5)$, is to be estimated where, $\bar{t}_j$ is the finite population mean of the j^{th} target variable, j=2,3,4,5. A sample of size n=200 is selected for each of the 500 replications. Sampling is by probability proportionate to size as measured by the $\{a_i\}$.

The target variables are subjected to three different response mechanisms. A separate table is given for each of these response mechanisms. For each target variable and response mechanism an interval is tabulated. The lower limit is the probability of response when the target variable is 0, and the upper limit is the probability of response when the target variable is 1000. For a target value between 0 and 1000 the probability of response is the linear interpolation between these two endpoints.

Response Mechanisms

RM\Targets	2	3	4	5
1	(.5,1.0)	(.4,1.0)	(.2,.9)	(.2,1.0)
2	(.9,.9)	(.8,.8)	(.65,.65)	(.5,.5)
3	(.8,1.0)	(.75,1.0)	(.6,1.0)	(.5,.7)

For each target variable and estimator in the tables, the estimated MSE is the average squared difference between the target variable estimator and the finite population mean for that target variable. This average is

Table 1. Estimated MSEs, Biases, and Variances.
MSE /(Bias) for Response Mechanism 1.

Target Variable / Estimator	2	3	4	5
HT	66.9	83.0	213.1	228.7
	(6.6)	(7.3)	(12.6)	(13.3)
LR	4.1	11.5	15.1	22.3
	(.55)	(1.17)	(1.5)	(1.89)
$\hat{\mu}_t^c$	3.4	7.6	7.8	10.9
	(.38)	(.71)	(.80)	(1.07)
$\hat{\mu}_t$	3.2	7.1	7.1	9.7
	(.18)	(.39)	(.09)	(–.06)

The Variance Estimator of $\hat{\mu}_t^c$ from $(X'\hat{\Sigma}^{-1}X)^{-1}$; Its Estimated Expected Value and Its Estimated Standard Error.

$\hat{E}(\hat{V}(\hat{\mu}_t^c))$	3.8	5.8	9.4	12.6
$\hat{\sigma}$	.5	.7	1.2	1.5

Table 2. Estimated MSEs, Biases, and Variances.
MSE /(Bias) for Response Mechanism 2.

Target Variable / Estimator	2	3	4	5
HT	9.4	17.2	34.2	55.8
	(.13)	(–.12)	(.59)	(–.31)
LR	3.0	6.9	9.4	15.3
	(.12)	(.10)	(0.0)	(–.14)
$\hat{\mu}_t^c$	2.8	6.1	7.0	10.7
	(.15)	(.15)	(.09)	(–.04)

The Variance Estimator of $\hat{\mu}_t^c$ from $(X'\hat{\Sigma}^{-1}X)^{-1}$; Its Estimated Expected Value and Its Estimated Standard Error.

$\hat{E}(\hat{V}(\hat{\mu}_t^c))$	3.3	5.2	8.4	12.5
$\hat{\sigma}$	.2	.5	0.9	1.4

Table 3. Estimated MSEs, Biases, and Variances.

MSE /(Bias) for Response Mechanism 3.

Target Variable / Estimator	2	3	4	5
HT	16.1 (2.3)	21.3 (2.69)	48.0 (4.81)	49.9 (2.75)
LR	3.1 (.25)	7.3 (.44)	8.0 (.55)	11.7 (.51)
$\hat{\mu}_t^c$	2.9 (.19)	6.3 (.27)	6.4 (.24)	9.3 (.11)
$\hat{\mu}_t$	2.8 (.14)	6.3 (.20)	6.3 (.07)	9.3 (−.09)

The Variance Estimator of $\hat{\mu}_t^c$ from $(X'\hat{\Sigma}^{-1}X)^{-1}$; Its Estimated Expected Value and Its Estimated Standard Error.

$\hat{E}(\hat{V}(\hat{\mu}_t^c))$	3.3	5.1	7.9	11.5
$\hat{\sigma}$	.4	.5	.3	.5

over the 500 replications of the sampling and estimation process. This number of replications gives the MSE estimates (upper entry – unbracketed), relative errors of around 5% and nearly always less than 10%.

The lower portion of each table evaluates $(X'\hat{\Sigma}^{-1}X)^{-1}$ as a variance estimator. The results of section 4 say that this estimated covariance matrix may be used to estimate the variance of both $\hat{\mu}_t^c$ and $\hat{\mu}_t$ (recall that $\hat{\mu}_t^c$ and $\hat{\mu}_t$ differ only by the covariates in C_i). The average of the diagonals of the $\{(X'\hat{\Sigma}^{-1}X)^{-1}\}$ from the 500 replications is displayed in the first row. The estimators of standard error for the diagonal of a single $(X'\hat{\Sigma}^{-1}X)^{-1}$ is given below each of these estimated means.

In Table 2, the estimated MSE of $\hat{\mu}_t^c$ for target variable 5 is 10.7, and the variance estimator for the estimator of target variable 5 is in the 2σ–interval, $12.5 \pm 2(1.4)$, with high probability, (if normality holds, $\cong .95$). For Table 2, MSE and variance of $\hat{\mu}_t^c$ are the same because nonresponse is ignorable and $\hat{\mu}_t^c = \hat{\mu}_t$.

In Table 1, where the nonignorability is most extreme, $\hat{\mu}_t^c$ still does relatively well. As predicted in section 4, the differences in MSE between $\hat{\mu}_t$ and $\hat{\mu}_t^c$ appear to be mostly bias. Squaring the biases (bracketed entries) for $\hat{\mu}_t^c$ and subtracting these from the corresponding MSEs gives results that are very close to the MSEs of $\hat{\mu}_t$. For example, the MSE of $\hat{\mu}_t^c$ for target variable 5 is 10.9, its bias is 1.07, and $10.9 - (1.07)^2 = 9.8 \cong$ MSE for target variable 5 of $\hat{\mu}_t$ (9.7).

The performance of $\hat{\mu}_t^c$ in all of the tables is only slightly inferior to $\hat{\mu}_t$. Other simulations were done on test populations with different amounts of correlation between adjacent months. The results on these populations were similar to those tabled here.

6. CONCLUSIONS

The BLUE, $\hat{\mu}_t^c$, is a generalization of the ordinary regression estimator (Cochran (1977)). The close tie in this methodology between estimating first and second moments parallels the similar tie in the EM–algorithm. This dependency between estimates of first and second moments implies that estimates of variance for the BLUE are produced as a fallout of these procedures.

The potential computional complexity of this BLUE presents few problems. Computer languages like SASPROC MATRIX, SAS IML, GAUSS, and APL can make quick work of the matrix arithmetic. The simulation results in section five cost about $50.00 per table for 500 replications of the sampling and estimation on an IBM main frame.

For the example presented here the sampling was pps. This allowed the use of the Horvitz–Thompson estimator as a second benchmark for comparison with the BLUE. Although its relative error was small, it still provided a fairly gross upper bound of MSE compared to the other estimators. The precision of $\hat{\mu}_t^c$ may be further improved by selecting a best sample with respect to the distribution of Z. This type of sampling may be dangerous in cases where the distribution of Z is erroneously specified, but the procedures suggested here are to be used when a long history of sampling experience confirms the correct form of this distribution. For reasons beyond statistics, probability sampling will always be a necessary part of applied sampling theory. Therefore, it may be useful to think of the sampling distribution as the unique case of a ***known*** nonignorable nonresponse mechanism acting on the sampling universe.

REFERENCES

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

Cochran W.C. (1977), *Sampling Techniques*, John Wiley & Sons.

Cox D.R. and Hinkley D.V. (1974), *Theoretical Statistics*, Chapman and Hall.

Little J.A. and Rubin D.B. (1987), *Statistical Analysis with Missing Data*, John Wiley & Sons.

Pfeffermann D. (1988), "The Effect of Sampling Design and Response Mechanism on Multivariate Regression Based Predictors", *Journal of the American Statistical Association*, 83, 824–833.

Raj D. (1968), *Sampling Theory*, McGraw–Hill.

Rao C.R. (1973), *Linear Statistical Inference and Its Applications*, John Wiley & Sons.

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

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.

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.

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

Rubin D.B. (1987), *Multiple Imputation for Nonresponse in Surveys*, John Wiley & Sons.

Srivastava M.S. and Carter E.M. (1986), "The Maximum Likelihood Method for Nonresponse in Sample Surveys", *Survey Methodology*, Vol 12, Number 1, 61–72.

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