Journal of Official Statistics
Vol. 10. No. 2, 1994, pp. 167–179

Bayesian Predictive Inference for Multivariate Sample Surveys

Balgobin Nandram[1]

Abstract: Multivariate observations are available from units in a longitudinal two-stage cluster sample design in which the same units or different units can be observed within the same clusters over occasions. The data from all variables are analyzed simultaneously and using the hierarchical Bayesian multivariate normal linear model, an estimator of a general finite population quantity, linear in the population values (e.g., change in finite population mean from one occasion to another), is constructed. Some properties of the point estimator are obtained when the variance components are assumed known. Numerical methods are used when the variance components are unknown. We analyze data on the Patterns of Care Studies, two-stage cluster samples of cancer patients each having two scores (bivariate) on two occasions. In particular, we describe the numerical computation of the finite population means (and changes in these means over the two occasions) of the two scores simultaneously.

Key words: Deleted residuals; Gibbs sampler; longitudinal; mean squared error; re-transformation; two-stage sampling.

1. Introduction

Many sample surveys are multivariate and there is a recognition that the multivariate nature of these surveys should be taken into account explicitly. Fuller and Harter (1987) used a non-Bayesian method to estimate the finite population mean for a small area assuming a multivariate regression model with components of variance error structure. It is becoming popular to use Bayesian predictive inference for finite population parameters, especially for problems such as estimation for small areas (Dempster and Raghunathan 1987). Sedransk and Malec (1985), Calvin and Sedransk (1991) and Nandram and Sedransk (1993) used hierarchical Bayesian normal linear models and showed, with examples, that such assumptions are consistent with data from some typical multi-stage cluster sample designs. Nandram and Sedransk (1993) extended the work of Calvin and Sedransk (1991) to cover longitudinal sample surveys. They emphasized a two-stage cluster sample design with univariate data and two occasions.

We emphasize multivariate surveys in which several characteristics are measured on the same unit. In addition, we can accommodate many longitudinal surveys by treating the responses on the same unit over several occasions as one multivariate

[1] Department of Mathematical Sciences, Worcester Polytechnic Institute, Room 103, Stratton Hall, Worcester, MA 01609, U.S.A.
Acknowledgement: The author thanks Joseph Sedransk for his assistance on Section 2 and the three referees for their contributions to the presentation.

observation. While our models are very flexible, and any number of occasions can be accommodated, in practice for repeated surveys conducted on many occasions, one might still consider a time series approach. Autoregressive models (e.g., Markovian) or the Kalman filter are usually the choice. However, such models are inappropriate for a small number of occasions (e.g., two or three). Moreover, for two-stage cluster samples within occasions our models apply when all clusters remain the same over all occasions. In general, from one occasion to another, the units in a particular cluster can be different, they can remain the same, or a rotation scheme can be followed. Also some clusters can be rotated out of the sample. Cochran (1977, ch. 12) described various patterns and methods for repeated sampling of the same population.

The Patterns of Care Studies (PCS) are a set of investigations studying the quality of treatment received by cancer patients whose primary treatment modality is radiation therapy. The principal statistical objectives of the PCS are assessment of the current status of radiation therapy care in the United States and assessment of changes over time. Data were collected in three different years (1973, 1978, 1983) and for several different sites (e.g., cervix, larynx, prostate), but only data from 1978 and 1983 are available. The variables under study relate to the processes involved in radiation therapy practice, and include indicators of the quality of both the pretreatment evaluation and the planning and actual delivery of therapy. To measure the completeness of the pretreatment evaluation and the therapy planning and monitoring, the "workup" and "treatment" scores were set up. The larger the score, the closer the patient's care conforms to acceptable standards of care. Each score lies in the interval (0,1) and the bivariate data on each patient are complete.

Let $\underset{\sim}{Y}$ denote the vector of population values and suppose interest is on $\omega(\underset{\sim}{Y})$ (e.g., a finite population mean). To perform a Bayesian predictive inference, the values $\underset{\sim}{Y}$ (or after a transformation the transformed values $\underset{\sim}{Z}$) follow a standard hierarchical normal linear model. However, to make inference about $\omega(\underset{\sim}{Y})$, one must re-transform $\underset{\sim}{Z}$ to the original scale, a problem of current interest in Bayesian predictive inference. Calvin and Sedransk (1991) used a first order Taylor series approximation and also suggested an alternative approximation to make inference on the original scale. Nandram and Sedransk (1993) used a similar approximation. One feature of the present research is to show that inference can be made on the original scale without using such uncertain approximations.

While the hierarchical Bayesian normal linear model (Lindley and Smith 1972) is appropriate for modeling multivariate sample surveys, the problem of unknown variance components is intractable. It is easy to incorporate the multivariate nature into the model, but the analysis with unknown covariance matrices, having conditional distributions such as the Wishart, is difficult. We provide a simple solution to this problem by using the Gibbs sampler algorithm (Gelfand and Smith 1990).

The objectives of this paper are both to describe and to apply methodology for Bayesian predictive inference that is appropriate for many multivariate and longitudinal sample surveys. In Section 2, after describing a model that might be appropriate for such surveys, we obtain the posterior distribution of a general linear function of the finite population values under the assumption that variance components are known. We study the properties of the posterior mean and show that it

has an optimality property. These results provide insight about the behavior of the models for the more complicated unknown variance case. In Section 3, using the Patterns of Care Studies (PCS) data we describe the computations involved to obtain the posterior distributions of the finite population quantities when there are unknown variance components. Section 4 has concluding remarks.

2. Bayesian Predictive Inference

In this section we consider inference about $\underset{\sim}{\omega}$, a set of general linear functions of the values of units in a finite population. (Inference for nonlinear functions can be obtained using the sampling based methods in Section 3.) We present a probabilistic specification appropriate for many longitudinal, multi-stage cluster sample designs where there are multivariate observations. In the PCS many characteristics are measured on each patient, and there are essentially two occasions. With known variance components we obtain the Bayes estimator of $\underset{\sim}{\omega}$ under a quadratic loss function and study its properties in Theorem 1.

2.1. *Modeling finite population values*

Assume that the survey is carried out on q occasions, and that on the jth occasion there are N_j clusters in the population, $j = 1, 2, \ldots, q$. There are M_{jk} units in the kth cluster on the jth occasion, $k = 1, 2, \ldots, N_j$, and the number of variables measured for each unit is p. We accommodate changes in the population of clusters by taking $M_{jk} = 0$ if cluster k is not a member of the population on occasion j. (Note that clusters are not nested within occasions.) Let $\underset{\sim}{Y}$ denote the $M \times 1$ vector of values for all units in the population where $M = p\Sigma_{j=1}^{q}\Sigma_{k=1}^{N}M_{jk}$ and N is the total number of distinct clusters in the population over all occasions.

Let $Y_{ijk\ell}$ denote the value of Y for variable i for the ℓth unit in the kth cluster on the jth occasion, and $\underset{\sim}{Y}_{jk\ell} = (Y_{1jk\ell}, \ldots, Y_{pjk\ell})'$, the vector of values for unit $(jk\ell)$ where $\ell = 1, \ldots, M_{jk}$, $k = 1, \ldots, N$ and $j = 1, \ldots, q$. For example, for the PCS we have $p = 2$, corresponding to the workup and treatment scores. Also, let

$$\underset{\sim}{Y}_{jk} = (\underset{\sim}{Y}'_{jk1}, \ldots, \underset{\sim}{Y}'_{jkM_{jk}})',$$

$$\underset{\sim}{Y}_{k} = (\underset{\sim}{Y}'_{1k}, \ldots, \underset{\sim}{Y}'_{qk})'$$

and

$$\underset{\sim}{Y} = (\underset{\sim}{Y}'_{1}, \ldots, \underset{\sim}{Y}'_{N})'.$$

Denoting the mean vector of a unit in cluster k on the jth occasion by $\underset{\sim}{\mu}_{jk}$,

$$\underset{\sim}{\mu}_{jk} = (\mu_{1jk}, \ldots, \mu_{pjk})',$$

$$\underset{\sim}{\mu}_{k} = (\underset{\sim}{\mu}'_{1k}, \ldots, \underset{\sim}{\mu}'_{qk})'$$

and

$$\underset{\sim}{\mu} = (\underset{\sim}{\mu}'_{1}, \ldots, \underset{\sim}{\mu}'_{N})'.$$

Note that $\underset{\sim}{\mu}_{jk}$ is the same for all units in cluster k on the jth occasion. Also let Σ_{jk} $(p \times p)$, not necessarily diagonal, be the covariance matrix of a unit in cluster k on the jth occasion. Then, conditional on $\underset{\sim}{\mu}_{jk}$ and Σ_{jk}, we assume that

$$\underset{\sim}{Y}_{jk1}, \underset{\sim}{Y}_{jk2}, \ldots, \underset{\sim}{Y}_{jkM_{jk}} | \underset{\sim}{\mu}_{jk}, \Sigma_{jk} \text{ i.i.d. } N(\underset{\sim}{\mu}_{jk}, \Sigma_{jk})$$

$j = 1, 2, \ldots, q$; $k = 1, 2, \ldots, N$. Letting $\text{cov}(\underset{\sim}{Y} | \underset{\sim}{\mu}, \Sigma) = \Sigma$, Σ can be chosen to represent the desired covariance structure, for example, between $\underset{\sim}{Y}_{jk\ell}$ and $\underset{\sim}{Y}_{j'k\ell}$ and between $\underset{\sim}{Y}_{jk\ell}$ and $\underset{\sim}{Y}_{j'k\ell'}$. In matrix notation, the preceding specification of the distribution of $\underset{\sim}{Y}$ conditional on $\underset{\sim}{\mu}$ and Σ is given by

$$\underset{\sim}{Y} | \underset{\sim}{\mu}, \Sigma \sim N(A_1\underset{\sim}{\mu}, \Sigma) \tag{2.1}$$

where A_1 is the matrix expressing the relationship between $\underset{\sim}{Y}$ and $\underset{\sim}{\mu}$. (Note that A_1, a matrix of zeros and ones, picks out the appropriate elements of $\underset{\sim}{\mu}$.)

For the second stage of the hierarchical model it is assumed that

$$\underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2, \ldots, \underset{\sim}{\mu}_N | \underset{\sim}{\theta}, \Delta \text{ i.i.d } N(\underset{\sim}{\theta}, \Delta)$$

where Δ will typically include correlation over both occasions and variables. If cluster k is in the population on only q_k occasions then, independently for $k = 1, \ldots, N$

$$\underset{\sim}{\mu}_k | \underset{\sim}{\theta}_k, \Delta_k \sim N(\underset{\sim}{\theta}_k, \Delta_k)$$

where $\underset{\sim}{\theta}_k$ and Δ_k are the corresponding components of $\underset{\sim}{\theta}$ and Δ. Summarizing, the second stage of the hierarchical model is

$$\underset{\sim}{\mu} | \underset{\sim}{\theta}, K \sim N(A\underset{\sim}{\theta}, K) \tag{2.2}$$

where A is the matrix expressing the relationship between $\underset{\sim}{\mu}$ and $\underset{\sim}{\theta}$; and $K = \text{diag}(\Delta_1, \Delta_2, \ldots, \Delta_N)$. (Note that, like A_1, A picks out the appropriate elements of $\underset{\sim}{\theta}$.)

At the last stage a noninformative reference prior on $\underset{\sim}{\theta}$ is specified; i.e.,

$$p(\underset{\sim}{\theta}) = \text{constant.} \tag{2.3}$$

From (1.1) and (1.2), letting $\Omega = (\Sigma, K)$,

$$E(\underset{\sim}{Y} | \underset{\sim}{\theta}, \Omega) = A^* \underset{\sim}{\theta}$$

where $A^* = A_1 A$, and

$$\text{var}(\underset{\sim}{Y} | \underset{\sim}{\theta}, \Omega) = A_1 K A_1' + \Sigma = V.$$

Since Σ is positive definite, V is positive definite.

Let n be the total number of distinct clusters sampled over the q occasions, and m_{jk} the number of units sampled from cluster k on occasion j. If cluster k is not sampled on occasion j or cluster k is not in the population on occasion j, $m_{jk} = 0$. It is assumed that the multivariate data for each sampled unit are complete. Moreover, it is assumed that the sample design is not informative. Specifically, letting S be the set of all possible samples, the probability of selecting sample $s \in S, p_s$ does not depend on $\underset{\sim}{Y}$, and if p_s depends on design variables they are assumed to be known for all units in the population. For weaker conditions and further discussion see Sugden and Smith (1984). Finally, it is assumed that any lack of knowledge about the labels associated with the units can be ignored; see Scott and Smith (1973).

We wish to make inferences about $\underset{\sim}{\omega}$, a $(t \times 1)$ vector of finite population parameters, where $\underset{\sim}{\omega} = L' \underset{\sim}{Y}$ with L an $(M \times t)$ matrix. (Note that $t = \Sigma_{i=1}^{p} t_i$ where t_i is the number of different finite population quantities considered for variable i.) For the PCS example one might take $t = 6$, $t_1 = 3$ and $t_2 = 3$ corresponding to the population means on occasions 1 and 2 and the difference between the two population means for the two scores. Partitioning $\underset{\sim}{Y}$ and L into sampled (s) and nonsampled (ns) parts we have

$$\underset{\sim}{Y} = \begin{bmatrix} \underset{\sim}{Y}_s \\ \cdots \\ \underset{\sim}{Y}_{ns} \end{bmatrix}, \quad L = \begin{bmatrix} L_s \\ \cdots \\ L_{ns} \end{bmatrix}$$

and

$$\underset{\sim}{\omega} = L_s' \underset{\sim}{Y}_s + L_{ns}' \underset{\sim}{Y}_{ns}. \tag{2.4}$$

That is, $\underset{\sim}{\omega}$ consists of two quantities containing (a) sampled values and (b) nonsampled values. Thus, the quantity in (a) is known and the quantity in (b) is to be obtained by Bayesian predictive inference. (Note that $\underset{\sim}{Y}_s$ is a $m \times 1$ vector of sampled values where

$$m = p \sum_{j=1}^{q} \sum_{k \epsilon s} m_{jk}.)$$

Like $\underset{\sim}{Y}$ and L for posterior inference about $\underset{\sim}{\omega}$ in (2.4) we partition all vectors and matrices into sampled (s) and nonsampled (ns) parts

$$A_1 = \begin{bmatrix} A_s \\ \cdots \\ A_{ns} \end{bmatrix}; \quad \Sigma = \begin{bmatrix} \Sigma_s & \vdots & \Sigma_{s,ns} \\ \cdots & \vdots & \cdots \\ \Sigma_{s',ns} & \vdots & \Sigma_{ns} \end{bmatrix}$$

$$A^* = \begin{bmatrix} A_s^* \\ \cdots\cdots \\ A_{ns}^* \end{bmatrix} = \begin{bmatrix} A_s A \\ \cdots\cdots \\ A_{ns} A \end{bmatrix}$$

and

$$V = \begin{bmatrix} V_S & \vdots & V_{s,ns} \\ \cdots\cdots & \vdots & \cdots\cdots \\ V_{s,ns}' & \vdots & V_{ns} \end{bmatrix}.$$

Letting $q_T = p\Sigma_{k=1}^N q_k$, A_1 is an $M \times q_T$ matrix and A is a $q_T \times qp$ matrix. Since at least one cluster is sampled on each occasion A_s^* has full column rank. We present $E(\underset{\sim}{\omega}|\underset{\sim}{Y}_s, \Omega)$ and $\text{var}(\underset{\sim}{\omega}|\underset{\sim}{Y}_s, \Omega)$ in Appendix A.

2.2. *Properties of the Bayes estimator*

In this section we consider the general linear estimator of $\underset{\sim}{\omega}, \hat{\underset{\sim}{\omega}}_s$, where

$$\hat{\underset{\sim}{\omega}}_s = G_s' \underset{\sim}{Y}_s + \underset{\sim}{C}_s. \tag{2.5}$$

In (2.5) G_s is an $m \times t$ matrix of constants and $\underset{\sim}{C}_s$ is a $t \times 1$ vector of constants; G_s and $\underset{\sim}{C}_s$ change from sample to sample.

Assuming the quadratic loss function

$$\text{loss}(\hat{\underset{\sim}{\omega}}_s, \underset{\sim}{\omega}) = (\hat{\underset{\sim}{\omega}}_s - \underset{\sim}{\omega})' H (\hat{\underset{\sim}{\omega}}_s - \underset{\sim}{\omega}) \tag{2.6}$$

with H a $t \times t$ positive definite matrix, the Bayes estimator of $\underset{\sim}{\omega}$, denoted by $\hat{\underset{\sim}{\omega}}_s^*$, is $E(\underset{\sim}{\omega}|\underset{\sim}{Y}_s, \Omega)$. In this section we evaluate the Bayes estimator, $\hat{\underset{\sim}{\omega}}_s^*$, assuming (2.1) and (2.2) but without the normality assumptions. We assume $\underset{\sim}{\theta}$ and Ω are fixed throughout.

Using the form of $E(\underset{\sim}{\omega}|\underset{\sim}{Y}_s, \Omega)$ in Appendix A, after considerable algebraic manipulation, one may write

$$\hat{\underset{\sim}{\omega}}_s^* = B_s' \underset{\sim}{Y}_s \tag{2.7}$$

where

$$\begin{aligned} B_s' = {} & L_s' + L_{ns}' V_{s,ns}' V_s^{-1} \\ & + L_{ns}'(A_{ns}^* - V_{s,ns}' V_s^{-1} A_s^*) \\ & \times (A_s^{*\prime} V_s^{-1} A_s^*)^{-1} A_s^{*\prime} V_s^{-1}. \end{aligned}$$

Observe that $\hat{\underset{\sim}{\omega}}_s^*$ is a member of the general class of linear estimators given by (2.5) with $G_s = B_s$ and $\underset{\sim}{C}_s = \underset{\sim}{0}$.

The risk (in θ) under the quadratic loss function (2.6) is given by $E_{\theta, s\in S}\{(\hat{\underset{\sim}{\omega}}_s - \underset{\sim}{\omega})' H(\hat{\underset{\sim}{\omega}}_s - \underset{\sim}{\omega})\}$ where the expectation is taken over both the randomization distribution and the distribution specified by (2.1) and (2.2), but without the normality assumptions. Since the sample design is assumed to be not informative, it is easy to show that

$$\begin{aligned} & E_{\underset{\sim}{\theta}, s\in S}\{(\hat{\underset{\sim}{\omega}}_s - \underset{\sim}{\omega})' H(\hat{\underset{\sim}{\omega}}_s - \underset{\sim}{\omega})\} \\ & \quad = \sum_{s\in S} p_s\{\text{trace}(D_s H) + \underset{\sim}{g}_s' H \underset{\sim}{g}_s\} \end{aligned} \tag{2.8}$$

where

$$\underset{\sim}{g}_s = E_{\underset{\sim}{\theta}}(\hat{\underset{\sim}{\omega}}_s - \underset{\sim}{\omega}) = (G_s' A_s^* - L' A^*)\underset{\sim}{\theta} + \underset{\sim}{C}_s$$

and

$$\begin{aligned} D_s = {} & (G_s - L_s)' V_s (G_s - L_s) \\ & + L_{ns}' V_{ns} L_{ns} - (G_s - L_s)' V_{s,ns} L_{ns} \\ & - L_{ns}' V_{ns,s}(G_s - L_s). \end{aligned}$$

(Note that the expectation in (2.8) depends on $\underset{\sim}{\theta}$ only through $\underset{\sim}{g}_s$.)

Under the model given by (2.1) and (2.2), with loss function (2.6) but without the normality assumptions, it can be shown that $\hat{\underset{\sim}{\omega}}_s$ in (2.5) has bounded risk (in $\underset{\sim}{\theta}$) if, and only if

$$G_s' A_s^* = L' A^* \tag{2.9}$$

for all $s \in S$ with $p_s > 0$. Moreover, the bounded risk of $\hat{\underset{\sim}{\omega}}_s$ is

$$\begin{aligned} & E_{\underset{\sim}{\theta}, s\in S}\{(\hat{\underset{\sim}{\omega}}_s - \underset{\sim}{\omega})' H(\hat{\underset{\sim}{\omega}}_s - \underset{\sim}{\omega})\} \\ & \quad = \sum_{s\in S} p_s\{\text{trace}(HD_s) + \underset{\sim}{C}_s' H \underset{\sim}{C}_s\}. \end{aligned} \tag{2.10}$$

In addition, it is not difficult to show that the Bayes estimator, $\hat{\underset{\sim}{\omega}}_s^*$ in (2.7), is also design unbiased; see Cassel, Särndal and Wretman (1977, ch. 4).

Next we state Theorem 1.

Theorem 1

Assume for fixed $\underset{\sim}{\theta}$ the specifications in (2.1) and (2.2) without normality. Then, for the quadratic loss function (2.6), in the class of linear estimators of $\underset{\sim}{\omega}$, (2.5), with

bounded risk (in $\underset{\sim}{\theta}$), the Bayes estimator in (2.7) has minimal risk $\Sigma_{s\in S} p_s \text{trace}(HD_s)$, where $D_s = \text{var}_{\underset{\sim}{\theta}}(B_s'\underset{\sim}{Y}_s - L'\underset{\sim}{Y})$ and $p_s > 0$ for all $s \in S$.

Proof: See Appendix B.

The practical import of this theorem is twofold. First, a practitioner should be more confident about using $\hat{\underset{\sim}{\omega}}_s^*$ when, as will often be the case, there is the belief that the general structure in (2.1) and (2.2) is appropriate but the normality assumptions are more tenuous. Examples of this general structure are in Malec and Sedransk (1985) and Nandram and Sedransk (1993). Second, it has been shown that the Bayes estimator, $\hat{\underset{\sim}{\omega}}_s^*$, has an optimal *frequentist* property (i.e., minimal bounded mean squared error). This should make the use of $\hat{\underset{\sim}{\omega}}_s^*$ more attractive to statisticians preferring a frequentist approach to inference.

We note that Theorem 1 generalizes similar theorems proved by Scott and Smith (1969), Royall (1976), and Malec and Sedransk (1985). In addition, using a mixed linear model with many stages for univariate data, Datta and Ghosh (1991) assumed that the error variance is unknown but all variance ratios are *known*, and proved that the Bayes estimator is the best linear unbiased predictor of a finite population quantity linear in the population values. However, realistic analytical results for the unknown variance case are very difficult to obtain for multivariate data models.

3. Methodology for Bivariate Data with Two Occasions

In this section we describe the methodology necessary to analyze the data from the PCS. We present the main features of the PCS in Section 3.1. Since there are no analytical results when the variance components are unknown, the decision-theoretic results in Section 2 give insight into the performance of the estimators obtained in this section.

3.1. *Main features of PCS*

An important feature of the PCS is that on each occasion the set of clusters (radiation therapy facilities) is the same, but the set of units (patients) is completely different. (For details about the PCS see Calvin and Sedransk 1991). This trait is shared by many large surveys such as the Hospital Discharge Survey (conducted annually by the National Center for Health Statistics). Here we consider the PCS on two occasions, 1978 and 1983, and patients with cervix cancer. The populations of radiation therapy facilities are essentially the same in the two years.

A description of the design is in order. A sample of n_1 of the N facilities was taken on the first occasion. Then a sample of m_{1k} of the M_{1k} patients was taken from the kth sampled facility. On the second occasion a subsample of n_{12} facilities was selected from the facilities sampled on the first occasion together with a sample of n_2 of the facilities not selected on the first occasion. Then a sample of size m_{2k} of the M_{2k} patients was taken from the kth sampled facility. Our population is the set of $N = 895$ facilities in existence in both the 1978 and 1983 surveys. (There are few births and deaths.) The number of facilities sampled only in 1978 is 47, the number sampled in both 1978 and 1983 is 24; and the number sampled only in 1983 is 21. Thus, a sample of size $n = 92$ facilities was taken. The first, second and third quartile of the distribution of facility sizes in 1978 are 3, 8, and 11, respectively and in 1983 they are 4, 7, and 14. The corresponding quartiles for the sample sizes are 2, 6, and 8 in 1978 and in 1983 they are 3, 5, and 5.

To satisfy the principal objectives of providing estimates of the current status of

quality of care and changes over time, point estimates and measures of variability are required for the associated finite population quantities. Nandram and Sedransk (1993) used models for the workup and treatment scores *separately*. We extend their models to accommodate the workup and treatment scores simultaneously. (That is, we treat the PCS data as arising from a two-stage survey with bivariate responses on two occasions.)

3.2. *Modeling the scores*

We note that the bivariate scores are not normally distributed. Taking each score separately but the data on both occasions simultaneously, Nandram and Sedransk (1993) found in their models that after a transformation the scores are approximately normally distributed. They used a squared transformation for the workup scores and a cubed transformation for the treatment scores. They also found that the transformed scores are approximately homogeneous across facilities.

To maintain the spirit of the general discussion in Section 2, we entertain the following specifications: Let $\underset{\sim}{Y}_{jk\ell} = (Y_{1jk\ell}, Y_{2jk\ell})'$ be the vector of scores of patient ℓ in facility k on occasion j (1: workup; 2: treatment; $j = 1,2$, $k = 1,2,\ldots,N$ and $\ell = 1, 2,\ldots,M_{jk}$). Let $\underset{\sim}{T}_{jk\ell} = (T_{1jk\ell}, T_{2jk\ell})'$ where $T_{1jk\ell} = g_1^{-1}(Y_{1jk\ell})$ and $T_{2jk\ell} = g_2^{-1}(Y_{2jk\ell})$. (We start with $g_1^{-1}(x) = x^2$ and $g_2^{-1}(x) = x^3$.) Then

$$\underset{\sim}{T}_{jk1},\ldots,\underset{\sim}{T}_{jkM_{jk}}|\underset{\sim}{\mu}_{jk},\Sigma_j \text{ i.i.d } N(\underset{\sim}{\mu}_{jk},\Sigma_j). \tag{3.1}$$

Note that each component of $\underset{\sim}{T}_{jk\ell}$ is restricted to be in (0,1). (That is, the random variables in (3.1) have a common truncated multivariate normal distribution.) Now letting $\underset{\sim}{\mu}_k = (\underset{\sim}{\mu}'_{1k}, \underset{\sim}{\mu}'_{2k})'$ and $\underset{\sim}{\mu} = (\underset{\sim}{\mu}'_1, \underset{\sim}{\mu}'_2,\ldots,\underset{\sim}{\mu}'_N)'$ then

$$\underset{\sim}{\mu}_1,\underset{\sim}{\mu}_2,\ldots,\underset{\sim}{\mu}_N|\theta,\Gamma \text{ i.i.d } N(\theta,\Delta). \tag{3.2}$$

Finally, letting $\Omega = (\Sigma_1,\Sigma_2,\Delta)$ for the parameters of $\Psi = (\theta,\Omega)$ we assume

$$p(\Psi) \propto |\Sigma_1\Sigma_2|^{-3/2}|\Delta|^{-5/2}. \tag{3.3}$$

We make inferences about the change in the finite population mean from 1978 to 1983 for each of the two variables, workup and treatment score, but we consider these two variables simultaneously. The two populations are those patients having cancer of the cervix first diagnosed in 1978 and 1983. Thus the quantities of interest are the components of the vector $\underset{\sim}{\omega}$ where

$$\underset{\sim}{\omega} = \{\omega_{ij},\ i = 1,2,\ j = 1,2;$$
$$\omega_{33} = \omega_{12} - \omega_{11},\ \omega_{44} = \omega_{22} - \omega_{21}\}$$

and

$$\omega_{ij} = \sum_{k=1}^{N}\sum_{\ell=1}^{M_{jk}} Y_{ijk\ell}\left\{\sum_{k=1}^{N} M_{jk}\right\}^{-1} \quad i,j = 1,2. \tag{3.4}$$

Using a first order Taylor's series expansion Nandram and Sedransk (1993) approximated the posterior distribution of $\underset{\sim}{\omega}$ given the variance components by a normal distribution. Instead we use the Gibbs sampler and a bootstrap method to fill in the nonsampled value of $\underset{\sim}{Y}$ after re-transforming $\underset{\sim}{T}$ to $\underset{\sim}{Y}$.

We assess the *entire* model by using a complete Bayesian cross validation method to study "deleted" residuals and predictors; see Gelfand, Dey and Chang (1992). Let $\underset{\sim}{T}_{jk\ell}$ be the vector of all transformed values for patient ℓ in facility k on occasion j and let $\underset{\sim}{T}_{(jk\ell)}$ be the vector of *all* transformed values *excluding* those for patient ℓ in facility k on occasion j. Now letting

$$E[\underset{\sim}{T}_{jk\ell}|\underset{\sim}{T}_{(jk\ell)}] = \underset{\sim}{e}_{jk\ell}$$
$$\text{var}[\underset{\sim}{T}_{jk\ell}|\underset{\sim}{T}_{(jk\ell)}] = V_{jk\ell}^{-1}$$

and

$$V_{jk\ell}^{-1} = S'_{jk\ell}S_{jk\ell}$$

the Cholesky's decomposition of the 2×2 matrix $V_{jk\ell}^{-1}$, we define the standardized deleted residuals for patient ℓ in facility k on occasion j as

$$\underset{\sim}{R}_{jk\ell} = S_{jk\ell}(\underset{\sim}{T}_{jk\ell} - \underset{\sim}{e}_{jk\ell}). \quad (3.5)$$

We use the Gibbs sampler to compute $\underset{\sim}{R}_{jk\ell}$ in (3.5) in a manner similar to that outlined by Gelfand et al. (1992). For example, we compute

$$E(\underset{\sim}{T}_{jk\ell} | \underset{\sim}{T}_{(jk\ell)}) = E_{\Psi | \underset{\sim}{T}_{jk(\ell)}}\{E(\underset{\sim}{T}_{jk\ell} | \underset{\sim}{T}_{jk(\ell)}, \Psi)\} \quad (3.6)$$

where $\underset{\sim}{T}_{jk(\ell)}$ is the vector of values for all patients in facility k on occasion j deleting the values of patient ℓ. Given $\underset{\sim}{T}_{jk(\ell)}$ and Ψ the components of $\underset{\sim}{T}_{jk\ell}$ have a multivariate normal distribution which is omitted in the interest of space. The outer expectation in (3.6) is computed using weights as in Gelfand et al. (1992).

If Σ_1, Σ_2 and Δ do not vary too much, then the components of $\underset{\sim}{R}_{jk\ell}$ in (3.5) are approximately independent standard normal random variables. We plotted each component of $\underset{\sim}{R}_{jk\ell}$ versus the corresponding component of $\underset{\sim}{e}_{jk\ell}$ to obtain a residual plot analogous to those used by Waternaux, Laird and Ware (1989). We also used a normal probability plot of $\underset{\sim}{R}_{jk\ell}$ versus normal scores. Figure 1(a) shows the residual plot and Figure 1(b) the normal probability plot. The plots show eleven patients with absolute residual values larger than 3.0. Omitting these patients shows only a minor improvement in these plots. Nine of the patients had at least one of the workup and treatment scores equal to one. As is expected, the large absolute residuals are associated with scores much different from others within a facility. (There are 1274 points in each of these plots.) Other transformations (including power) did not fit as well as the square for the workup and cube for the treatment scores. This is consistent with Sedransk and Malec (1985) who fitted

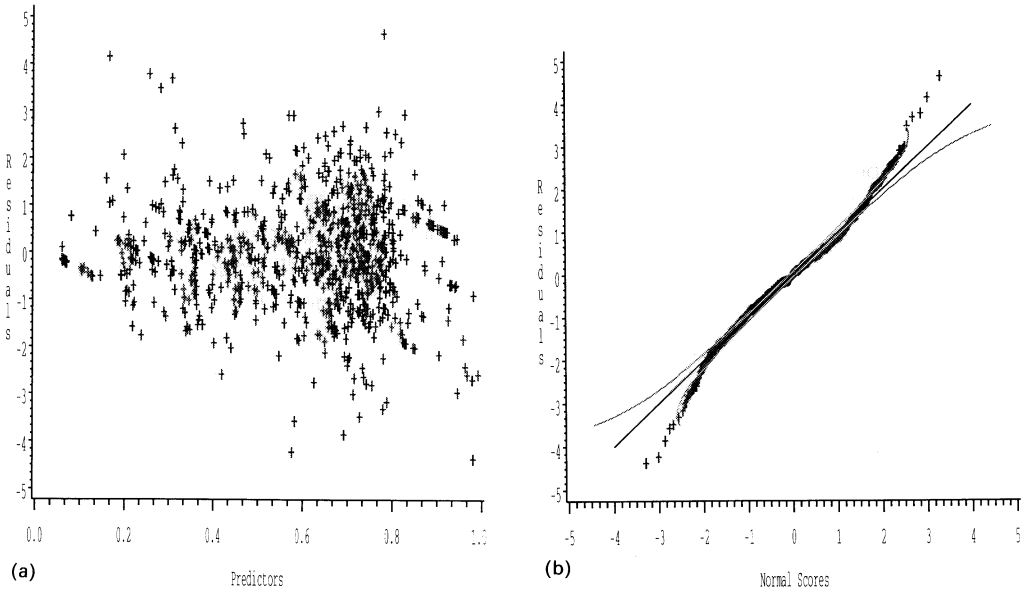

Fig. 1. (a) Plot of deleted residuals versus deleted predictors. (b) Normal probability plot of deleted residuals: + + + observed values; – – – expected 45 degrees line; – – – 95% pointwise critical bands

many different transformations to the PCS data for the simpler models discussed by Malec and Sedransk (1985).

3.3. *Inference for finite population quantities*

We proceed in two stages. First we run the Gibbs sampler (Gelfand and Smith 1990) on the n = 92 sampled facilities since there are 895 facilities in the survey on both occasions. Then using the iterates from the Gibbs sampler, we fill in the nonsampled values of $\underset{\sim}{Y}$.

We used the Gibbs sampler to obtain iterates from the unconditional posterior distribution of Ψ. Thus we next describe the conditional posterior distributions needed for the Gibbs sampler.

Let s_j denote the set of facilities sampled on occasion j, j = 1,2. Also let $\bar{\underset{\sim}{T}}_k' = (\bar{\underset{\sim}{T}}_{1k}', \bar{\underset{\sim}{T}}_{2k}')$ and $\bar{\underset{\sim}{T}}_{jk} = (\bar{T}_{1jk}, \bar{T}_{2jk})'$ where

$$\bar{T}_{ijk} = \begin{cases} m_{jk}^{-1} \sum_{\ell=1}^{m_{jk}} T_{ijk\ell}; & k \in s_j \\ 0; & \text{else} \end{cases} \tag{3.7}$$

i, j = 1,2; k = 1,2,..., N. Let $\underset{\sim}{T}_s$ denote the vector of transformed sampled values and

$$\Lambda_k = \begin{cases} \left\{ \begin{bmatrix} m_{1k}\Sigma_1^{-1} & 0 \\ 0 & m_{2k}\Sigma_2^{-1} \end{bmatrix} + \Delta^{-1} \right\}^{-1} \begin{bmatrix} m_{1k}\Sigma_1^{-1} & 0 \\ 0 & m_{2k}\Sigma_2^{-1} \end{bmatrix}, \\ \qquad\qquad k \in s_1 \quad \text{or} \quad k \in s_2 \\ 0, \qquad\qquad \text{else.} \end{cases}$$

First, $\underset{\sim}{\mu}_k | \underset{\sim}{T}_s, \Psi$ are independent 4-variate normal vectors with

$$E[\underset{\sim}{\mu}_k | \underset{\sim}{T}_s, \Psi] = \Lambda_k \bar{\underset{\sim}{T}}_k + (I - \Lambda_k)\underset{\sim}{\theta}$$

and

$$\text{cov}[\underset{\sim}{\mu}_k | \underset{\sim}{T}_s, \Psi] = (I - \Lambda_k)\Delta \tag{3.8}$$

where I is the 4 × 4 identity matrix. Second, $\underset{\sim}{\theta} | \underset{\sim}{T}_s, \underset{\sim}{\mu}, \Omega$ is a 4-variate normal vector with

$$E[\underset{\sim}{\theta} | \underset{\sim}{T}_s, \underset{\sim}{\mu}, \Omega] = n^{-1} \sum_{k=1}^{n} \underset{\sim}{\mu}_k = \bar{\underset{\sim}{\mu}}$$

and

$$\text{cov}[\underset{\sim}{\theta} | \underset{\sim}{T}_s, \underset{\sim}{\mu}, \Omega] = n^{-1}\Delta. \tag{3.9}$$

Third,

$$\Sigma_j^{-1} | \underset{\sim}{T}_s, \underset{\sim}{\theta}, \underset{\sim}{\mu}, \Delta \sim W_2 \left\{ \sum_{k \in s_j} m_{jk}, \left\{ \sum_{k \in s_j} \sum_{\ell=1}^{m_{jk}} (\underset{\sim}{Y}_{jk\ell} - \underset{\sim}{\mu}_{jk})(\underset{\sim}{Y}_{jk\ell} - \underset{\sim}{\mu}_{jk})' \right\}^{-1} \right\} \tag{3.10}$$

and

$$\Delta^{-1} | \underset{\sim}{T}_s, \underset{\sim}{\theta}, \Sigma_1, \Sigma_2 \sim W_4 \left\{ n, \left\{ \sum_{k=1}^{n} (\underset{\sim}{\mu}_k - \bar{\underset{\sim}{\mu}})(\underset{\sim}{\mu}_k - \bar{\underset{\sim}{\mu}})' \right\}^{-1} \right\} \tag{3.11}$$

where $W_p(\nu, A)$ denotes a p × p matrix Wishart distribution with degrees of freedom ν and parameter A.

We ran the Gibbs sampler to obtain "good" iterates of Ψ. Let $\underset{\sim}{\mu} = (\underset{\sim}{\mu}_s', \underset{\sim}{\mu}_{ns}')'$ with $\underset{\sim}{\mu}_s = (\underset{\sim}{\mu}_1', \underset{\sim}{\mu}_2', \ldots, \underset{\sim}{\mu}_n')'$ and $\underset{\sim}{\mu}_{ns} = (\underset{\sim}{\mu}_{n+1}', \underset{\sim}{\mu}_{n+2}', \ldots, \underset{\sim}{\mu}_N')'$ where $\underset{\sim}{\mu}_s$ denotes the vector of population means of all clusters sampled on at least one occasion. Starting with method of moments estimates of $\underset{\sim}{\theta}, \Sigma_1, \Sigma_2$ and Δ the Gibbs sampler proceeds by drawing $\underset{\sim}{\mu}_s$ from (3.8), $\underset{\sim}{\theta}$ from (3.9), Σ_1 and Σ_2 from (3.10), and Δ from (3.11). Using the most recent values for the requisite parameters at each iteration, the current parameters are drawn from the respective conditional posterior distributions, and the whole procedure is repeated until convergence. (Note that there are 388 parameters in the Gibbs sampler; one

iterate consists of $\underset{\sim}{\mu}_s$ and the components of Ψ.)

We experienced difficulties in the convergence of the Gibbs sampler with long-term dependence among iterates especially for components of $\underset{\sim}{\theta}$ and Δ. We therefore, followed the advice of Zeger and Karim (1991, sec. 5.4) on the optimization of the algorithm. To obtain a sample of 1000 values of Ψ we chose, after initial convergence ($\approx$ 1000 iterates), every fiftieth observation from continued runs of the Gibbs sampler. (The estimated autocorrelation function indicated that taking every fiftieth observation was sufficient to remove the dependence between successive iterates.) The total of about 51000 iterate values yielded 1000 values of Ψ. We assessed convergence graphically. We plotted the estimated univariate densities for all components of Ω using the first 200, 400, 600, 800, 1000 iterates. For the off-diagonal elements of the covariance matrices, Σ_1, Σ_2 and Δ, we used nonparametric density estimation (Silverman 1986). The last three sets of iterates exhibit no significant change. Thus, we use the 1000 virtually independent iterates for future computations (e.g., diagnostics and finite population posterior distributions).

At the second stage we observe

$$[\underset{\sim}{T}_{ns}, \underset{\sim}{\mu}, \underset{\sim}{\theta}, \Sigma_1\Sigma_2, \Delta \mid \underset{\sim}{T}_s] = [\underset{\sim}{T}_{ns} \mid \underset{\sim}{\mu}, \Sigma_1, \Sigma_2] \times [\underset{\sim}{\mu} \mid \underset{\sim}{T}_s, \Psi][\Psi \mid \underset{\sim}{T}_s] \quad (3.12)$$

where $[\underset{\sim}{U} \mid \underset{\sim}{V}]$ denotes the conditional distribution of $\underset{\sim}{U}$ given $\underset{\sim}{V}$. (Note that the breakdown in (3.12) is essential because $N \gg n$ and the Gibbs sampler is run only with $\underset{\sim}{\mu}_s$). We use (3.12) to obtain realizations of the vector $\underset{\sim}{Y} = (\underset{\sim}{Y}_s', \underset{\sim}{Y}_{ns}')'$ of population values. First, draw one iterate from the empirical distribution of $[\Psi \mid \underset{\sim}{T}_s]$ obtained from the Gibbs sampler. (That is, one iterate is drawn at random from the 1000 "good" iterates and the components of Ψ are stripped off.) Second, draw a vector $\underset{\sim}{\mu}$ from (3.8). Third, draw a vector $\underset{\sim}{T}_{ns}$ from (3.1). Finally, using $Y_{ijk\ell} = g_i(T_{ijk\ell})$ i = 1,2 for workup and treatment scores we obtain a realization of $\underset{\sim}{Y}^{(q)} = (\underset{\sim}{Y}_s', \underset{\sim}{Y}_{ns}')'$. The entire procedure is repeated Q times to obtain a random sample $\underset{\sim}{Y}^{(1)}, \ldots, \underset{\sim}{Y}^{(Q)}$ which in turn provides a random sample $\underset{\sim}{\omega}^{(1)}, \underset{\sim}{\omega}^{(2)}, \ldots, \underset{\sim}{\omega}^{(Q)}$ of Q values of $\underset{\sim}{\omega}$ in (3.4). Then as estimates of the posterior mean and variance of $\underset{\sim}{\omega}$ we simply use

$$E(\underset{\sim}{\omega} \mid \underset{\sim}{T}_s) \approx Q^{-1} \sum_{q=1}^{Q} \underset{\sim}{\omega}^{(q)} = \bar{\underset{\sim}{\omega}}$$

and

$$\text{var}(\underset{\sim}{\omega} \mid \underset{\sim}{T}_s) \approx Q^{-1} \sum_{q=1}^{Q} (\underset{\sim}{\omega}^{(q)} - \bar{\underset{\sim}{\omega}})(\underset{\sim}{\omega}^{(q)} - \bar{\underset{\sim}{\omega}})'. \quad (3.13)$$

We also evaluated (3.13) for the first t iterates (after initial convergence), t = 200, 400, 600, 800, 1000, for the posterior means and standard deviations of the elements of $\underset{\sim}{\omega}$. As expected, convergence for $E(\omega_{ij} \mid \underset{\sim}{T}_s)$ and $\{\text{var}(\omega_{ij} \mid \underset{\sim}{T}_s)\}^{1/2}$ i, j = 1,2 was much more rapid than for the components of Ψ.

We present in Table 1(a) for each variable the values of the estimated posterior means and standard deviations obtained from (3.13), for the finite population means for

Table 1a. Estimates of finite population means on two occasions for workup and treatment scores

Quantity	1978	1983	Change
i. Workup			
Mean	0.7598	0.8201	0.0623
Standard Error	0.0115	0.0098	0.0151
ii. Treatment			
Mean	0.7480	0.7721	0.0241
Standard Error	0.0175	0.0152	0.0213

Note: The means and standard errors are obtained from (3.13).

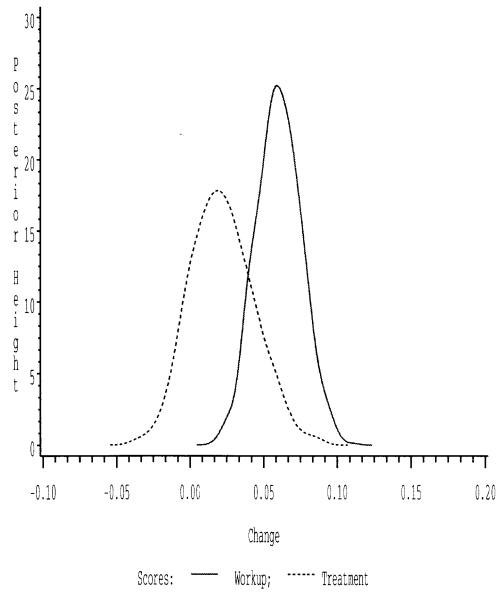

Fig. 2. Posterior densities of the change in finite population mean

1978 and 1983 and for the difference in finite population means. In Table 1(a), we see a substantial improvement in the quality of care as measured by the quality of the workup performed, but no improvement in the quality of treatment. This result is much tighter than that given by Nandram and Sedransk (1993). Based on the 1000 "good" iterates of the Gibbs sampler we also obtain a density estimate for the change in finite population mean. A density estimator with a normal kernel was used with window width suggested by Silverman (1986, p. 48). Figure 2 confirms that the increase in quality of care is substantial with respect to workup but not treatment.

Finally, we observe that the effort on the multivariate methodology is worthwhile as there are correlations among the components of $\omega_{ij}, i, j = 1, 2$; see Table 1(b). For example, the estimated posterior correlation between ω_{11} and ω_{12} is 0.4091. This is expected as there was some similarity between the quality of care for workup and treatment in 1978.

4. Concluding Remarks

It is possible to develop reasonable models for complex sample surveys which are both multivariate and longitudinal. The use of resampling methods such as the Gibbs sampler makes it possible to avoid many approximations and to provide inferences that include all known sources of variation.

We have extended the work of Calvin and Sedransk (1991) and Nandram and Sedransk (1993) in several ways. First, our method is multivariate not univariate. Second, even for these complex models we avoid the use of Taylor's series approximation. Third, our diagnostic procedure is completely Bayesian and checks the entire model simultaneously, not parts of it. Fourth, our method is sampling based throughout and thus it avoids the use of complicated formulas.

In addition to showing that the model provides a reasonable fit, Theorem 1 also adds credence to the hierarchical Bayesian linear model for applications to many survey problems. Moreover, Theorem 1 extends results in Malec and Sedransk (1985) in two directions, multivariate and longitudinal. One may easily extend Theorem 1 to accommodate models appropriate for many situations where multi-stage cluster sampling is used; see, for example, Malec and Sedransk (1985).

Table 1b. Estimates of correlation of ω_{ij}, $i, j = 1, 2$

Pair	11 : 12	11 : 21	11 : 22	12 : 21	12 : 22	21 : 22
Correlation	0.4091	0.0133	0.0092	0.0255	0.1507	−0.1465

Note: ω_{ij}, i, j = 1, 2, are given by (3.4); correlations are obtained from (3.13).

One area for further research is with respect to the transformation chosen. It is desirable to incorporate the uncertainty in the choice of the transformation. When uncertainty in a transformation is built into a model, the variability of another estimator (e.g., an estimator of a finite population mean) tends to be inflated; see, for example, Carroll and Ruppert (1981).

Appendix

Appendix A: Posterior Moments of $\underset{\sim}{\omega}$

First,

$$E(\underset{\sim}{\omega}|\underset{\sim}{Y}_s, \Omega) = L_s'\underset{\sim}{Y}_s + L_{ns}'E(\underset{\sim}{Y}_{ns}|\underset{\sim}{Y}_s, \Omega)$$

and

$$\text{var}(\underset{\sim}{\omega}|\underset{\sim}{Y}_s, \Omega) = L_{ns}'\text{var}(\underset{\sim}{Y}_{ns}|\underset{\sim}{Y}_s, \Omega)L_{ns}. \quad (A.1)$$

Under the model specified by (2.1), (2.2) and (2.3), and using results in Lindley and Smith (1972)

$$E(\underset{\sim}{\mu}|\underset{\sim}{Y}_s, \Omega) = (\text{var}(\underset{\sim}{\mu}|\underset{\sim}{Y}_s, \Omega))A_s'\Sigma_s^{-1}\underset{\sim}{Y}_s \quad (A.2)$$

and

$$\begin{aligned}\text{var}(\underset{\sim}{\mu}|\underset{\sim}{Y}_s, \Omega) &= \{A_s'\Sigma_s^{-1}A_s + K^{-1}\}^{-1} \\ &\times \{I + K^{-1}A[A'K^{-1}A - A'K^{-1} \\ &\times [A_s'\Sigma_s^{-1}A_s + K^{-1}]^{-1}K^{-1}A]^{-1}A'K^{-1} \\ &\times [A_s'\Sigma_s^{-1}A_s + K^{-1}]^{-1}\}. \quad (A.3)\end{aligned}$$

Thus, using (A.2) and (A.3),

$$\begin{aligned}E(\underset{\sim}{Y}_{ns}|\underset{\sim}{Y}_s, \Omega) &= E_{\underset{\sim}{\mu}|\underset{\sim}{Y}_s}(E(\underset{\sim}{Y}_{ns}|\underset{\sim}{Y}_s, \underset{\sim}{\mu}, \Omega)) \\ &= [A_{ns} - \Sigma_{ns,s}\Sigma_s^{-1}A_s] \times E(\underset{\sim}{\mu}|\underset{\sim}{Y}_s, \Omega) \\ &+ \Sigma_{ns,s}\Sigma_s^{-1}\underset{\sim}{Y}_s\end{aligned}$$

and

$$\begin{aligned}\text{var}(\underset{\sim}{Y}_{ns}|\underset{\sim}{Y}_s, \Omega) &= (\Sigma_{ns} - \Sigma_{s,ns}'\Sigma_S^{-1}\Sigma_{s,ns}) \\ &+ (A_{ns} - \Sigma_{s,ns}'\Sigma_s^{-1}A_s)\text{var}(\underset{\sim}{\mu}|\underset{\sim}{Y}_s, \Omega) \\ &\times (A_{ns} - \Sigma_{s,ns}'\Sigma_s^{-1}A_s)'.\end{aligned}$$

Appendix B: Sketch of Proof of Theorem 1

By (2.10), for any linear estimator $\hat{\underset{\sim}{\omega}}_s = G_s'\underset{\sim}{Y}_s + \underset{\sim}{C}_s$ with bounded risk (in $\underset{\sim}{\theta}$),

$$\begin{aligned}&E_{\underset{\sim}{\theta}, s\in S}\{(\hat{\underset{\sim}{\omega}}_s - \underset{\sim}{\omega})'H(\hat{\underset{\sim}{\omega}}_s - \underset{\sim}{\omega})\} \\ &= \sum_{s\in S} p_s\{\text{trace}(HD_s) + \underset{\sim}{C}_s'H\underset{\sim}{C}_s\} \quad (B.1)\end{aligned}$$

where $D_s = \text{var}_{\underset{\sim}{\theta}}(G_s'\underset{\sim}{Y}_s = L'\underset{\sim}{Y})$.

Now D_s may be rewritten as

$$\begin{aligned}D_s &= \text{var}_{\underset{\sim}{\theta}}(B_s'\underset{\sim}{Y}_s - L'\underset{\sim}{Y}) \\ &+ \text{var}_{\underset{\sim}{\theta}}((G_s - B_s)'\underset{\sim}{Y}_s) + U + U' \quad (B.2)\end{aligned}$$

where

$$U = \text{cov}_{\underset{\sim}{\theta}}[B_s'\underset{\sim}{Y}_s - L'\underset{\sim}{Y}, (G_s - B_s)'\underset{\sim}{Y}_s].$$

It is easy to show that $U = [(B_s - L_s)'V_s - L_{ns}'V_{s,ns}'](G_s - B_s)$. Using the definition of B_s in (2.7), and applying (2.9) to both $\hat{\underset{\sim}{\omega}}_s$ and $\hat{\underset{\sim}{\omega}}_s^*$, it follows that $U = 0$. Then, from (B.1) and (B.2)

$$\begin{aligned}&E_{\underset{\sim}{\theta}, s\in S}\{(\hat{\underset{\sim}{\omega}}_s - \underset{\sim}{\omega})'H(\hat{\underset{\sim}{\omega}}_s - \underset{\sim}{\omega})\} \\ &= E_{\underset{\sim}{\theta}, s\in S}\{(\hat{\underset{\sim}{\omega}}_s^* - \underset{\sim}{\omega})'H(\hat{\underset{\sim}{\omega}}_s^* - \underset{\sim}{\omega})\} \\ &+ \sum_{s\in S} p_s[\text{trace}\{H\,\text{var}_{\underset{\sim}{\theta}}(G_s - B_s)'\underset{\sim}{Y}_s\} \\ &+ \underset{\sim}{C}_s'H\underset{\sim}{C}_s]. \quad (B.3)\end{aligned}$$

The remainder of the proof consists of showing that since $p_s > 0$ for all $s \in S$ the second term on the right side of (B.3) is (a) nonnegative and (b) equal to zero if, and only if, $G_s = B_s$ and $\underset{\sim}{C}_s = \underset{\sim}{0}$. Thus, the only linear estimator with minimal bounded risk (in $\underset{\sim}{\theta}$) is the Bayes estimator, $\hat{\underset{\sim}{\omega}}_s^* = B_s'\underset{\sim}{Y}_s$, in (2.7).

5. References

Calvin, J. and Sedransk, J. (1991). Bayesian and Frequentist Predictive Inference for the Patterns of Care Studies. Journal of the American Statistical Association, 86, 36–48.

Carroll, R.J. and Ruppert, D. (1981). On Prediction and the Power Transformation Family. Biometrika, 68, 609–615.

Cassel, C.M., Särndal, C.E., and Wretman, J.H. (1977). Foundations of Inference in Survey Sampling. New York: John Wiley.

Cochran, W.G. (1977). Sampling Techniques. New York: John Wiley.

Datta, G. and Ghosh, M. (1991). Bayesian Prediction in Linear Models: Applications to Small Area Estimation. Annals of Statistics, 19, 1748–1770.

Dempster, A.P. and Raghunathan, T.E. (1987). Using a Covariate for Small Area Estimation: A Common Sense Bayesian Approach. In Small Area Statistics, ed. R. Platek, J.N.K. Rao, C.E. Särndal and M.P. Singh. New York: John Wiley.

Fuller, W.A. and Harter, R.M. (1987). The Multivariate Components of Variance Model for Small Area Estimation. In Small Area Statistics, ed. R. Platek, J.N.K. Rao, C.E. Särndal and M.P. Singh. New York: John Wiley.

Gelfand, A.E., Dey, D.K., and Chang, H. (1992). Model Determination using Predictive Distributions with Implementation via Sampling-Based Methods. Bayesian Statistics 4, ed. J. Bernardo. New York: Oxford University Press.

Gelfand, A.E. and Smith, A.F.M. (1990). Sampling-Based Approaches to Calculating Marginal Densities. Journal of the American Statistical Association, 85, 398–409.

Lindley, D.V. and Smith, A.F.M. (1972). Bayes Estimates for the Linear Model (with Discussion). Journal of the Royal Statistical Society, Ser. B, 34, 1–41.

Malec, D.J. and Sedransk, J. (1985). Bayesian Methodology for Predictive Inference for Finite Population Parameters in Multistage Cluster Sampling. Journal of the American Statistical Association, 80, 891–902.

Nandram, B. and Sedransk, J. (1993). Bayesian Predictive Inference for Longitudinal Sample Surveys. Biometrics, 49, 1045–1055.

Royall, R.M. (1976). The Linear Least-Squares Prediction Approach to Two-Stage Sampling. Journal of the American Statistical Association, 71, 657–664.

Scott, A. and Smith, T.M.F. (1969). Estimation in Multi-Stage Surveys. Journal of the American Statistical Association, 64, 830–840.

Scott, A. and Smith, T.M.F. (1973). Survey Design, Symmetry and Posterior Distributions. Journal of the Royal Statistical Society, Ser. B, 35, 57–60.

Sedransk, J. and Malec, D.J. (1985). Bayesian Predictive Inference for Surveys to Access the Quality of Care of Cancer Patients. Bulletin of the International Statistical Institute, 50, 12.1.1–12.1.15.

Silverman, B.W. (1986). Density Estimation for Statistics and Data Analysis. London: Chapman and Hall.

Sugden, R.A. and Smith, T.M.F. (1984). Ignorable and Informative Designs. Biometrika, 71, 495–506.

Waternaux, C., Laird, N., and Ware, J. (1989). Methods for the Analysis of Longitudinal Data Blood Lead Concentrations and Cognitive Development. Journal of the American Statistical Association, 84, 33–41.

Zeger, S.L. and Karim, M.R. (1991). Generalized Linear Models with Random Effects; A Gibbs Sampling Approach. Journal of the American Statistical Association, 86, 79–86.

Received September 1991
Revised January 1994