Journal of the Korean
Statistical Society
Vol. 26, No. 2, 1997

On Frequentist Properties of Some Hierarchical Bayes Predictors for Small Domain Data in Repeated Surveys

Narinder K. Nangia [1] and Dal Ho Kim[2]

ABSTRACT

The paper shows that certain hierarchical Bayes (HB) predictors for small domain data in repeated surveys "universally" or "stochastically" dominate all linear unbiased predictors. Also, the HB predictors are "best" within the class of all equivariant predictors under a certain group of transformations.

Key Words : Hierarchical Bayes; Small areas; Repeated survey; Universal domination; Best equivariant prediction.

1. INTRODUCTION

Bayesian methods have been used quite extensively in recent years for solving small area (or domain) estimation problem. Particularly effective in this regard has been the hierarchical or empirical Bayes (HB or EB) approach which is especially suitable for a systematic connection of local areas through

[1]Trilogy Consulting Corporation, Waukegan, Illinois 60085, USA.
[2]Department of Statistics, Kyungpook National University, Taegu, 702-701, Korea.

models. We may refer to Fay and Herriot (1979), Ghosh and Meeden (1986), Ghosh and Lahiri (1987), Battese, Harter and Fuller (1988), Prasad and Rao (1990), Datta and Ghosh (1991a), among others. But most model-based inferences for small area estimation considered "borrowing strength" only from similar other areas to improve small area estimation.

Often small area populations are subject to change in time. Usually the time series information is available through the repeated surveys, carried out at regular time intervals by industries and government agencies. In such surveys, one has at one's disposal not only the current data, but also data from similar past experiments. Typically, only a few samples are available from an individual area. Recently, Ghosh, Nangia and Kim (1996) considered the HB model for borrowing strength over time in addition to borrowing strength from other local areas in small area estimation based on data from repeated surveys.

Datta and Ghosh (1991a) considered a HB procedure for prediction in mixed linear models and utilized the results for small area estimation. Datta and Ghosh (1991b) showed the frequentist properties of some HB predictors in finite population sampling when ratios of variance components are known. In this paper we extend the results of Datta and Ghosh (1991b) in the context of repeated surveys.

The outline of the remaining sections is as follows. In Section 2 we consider a HB time series model as in Ghosh, Nangia and Kim (1996) and provide the predictive distribution of unobserved population units at all the time points given the time series data on the sampled units. In Section 3 and Section 4, it is shown that proposed HB predictors have similar frequentist properties as in Datta and Ghosh (1991b) even in the context of repeated surveys. In specific, these HB predictors dominate "universally" or "stochastically" the linear unbiased predictors in the sense of Hwang (1985). Also, under a suitable group of transformations, the HB predictors are "best" within the class of all equivariant predictors for elliptically symmetric distributions.

2. THE MODEL AND THE HB PREDICTORS

In this section, we consider a HB model for prediction of the characteristic of interest (e.g. the population mean for each small area or domain) in the context of repeated surveys. Let Y_{ijk} denote some characteristic of interest associated with the k^{th} unit at time j in the i^{th} small area $(k = 1, \cdots, N_{ij}; j = 1, \cdots, t; i = 1, \cdots, m)$. Let $\mathbf{Y}_{ij}^{(1)} = (Y_{ij1}, \cdots, Y_{ij,n_{ij}})^T$

and $\mathbf{Y}_{ij}^{(2)} = (Y_{ij,n_{ij}+1}, \cdots, Y_{ij,N_{ij}})^T$. The superscripts (1) and (2) are used to denote the sampled and non-sampled units in the i^{th} small area at time j, respectively. Here N_{ij} is the population size of the i^{th} small area at time j from which a sample of size n_{ij} is taken ($1 \leq i \leq m; 1 \leq j \leq t$). Known design matrices $\mathbf{X}_{ij}^{(1)}$ $(n_{ij} \times p)$ and $\mathbf{Z}_{ij}^{(1)}$ $(n_{ij} \times q)$ correspond to the sampled units and $\mathbf{X}_{ij}^{(2)}$ $((N_{ij} - n_{ij}) \times p)$ and $\mathbf{Z}_{ij}^{(2)}$ $((N_{ij} - n_{ij}) \times q)$ correspond to the unsampled units. Assume $\text{rank}(\mathbf{X}_{ij}^{(1)}) = \text{rank}(\mathbf{X}_{ij}^{(2)}) = p$.

We consider the following Bayesian model:

$$\text{I.} \quad \begin{pmatrix} \mathbf{Y}_{ij}^{(1)} \\ \mathbf{Y}_{ij}^{(2)} \end{pmatrix} \Bigg| \begin{pmatrix} \theta_{ij}^{(1)} \\ \theta_{ij}^{(2)} \end{pmatrix}, r \stackrel{ind}{\sim} N\left(\begin{pmatrix} \theta_{ij}^{(1)} \\ \theta_{ij}^{(2)} \end{pmatrix}, r^{-1}\mathbf{V}_{ij} \right);$$

$$(i = 1, \cdots, m; j = 1, \cdots, t)$$

$$\text{II.} \quad \begin{pmatrix} \theta_i j^{(1)} \\ \theta_i j^{(2)} \end{pmatrix} \Bigg| \alpha, \mathbf{b}_j, r \stackrel{ind}{\sim} N\left(\begin{pmatrix} \mathbf{X}_{ij}^{(1)} \\ \mathbf{X}_{ij}^{(2)} \end{pmatrix} \alpha + \begin{pmatrix} \mathbf{Z}_{ij}^{(1)} \\ \mathbf{Z}_{ij}^{(2)} \end{pmatrix} \mathbf{b}_j, r^{-1}\Psi_j \right);$$

$$(i = 1, \cdots, m; j = 1, \cdots, t)$$

$$\text{III.} \quad \mathbf{b}_j | \mathbf{b}_{j-1}, r \stackrel{ind}{\sim} N(\mathbf{H}_j \mathbf{b}_{j-1}, \ r^{-1}\mathbf{W}); (j = 1, \cdots, t)$$

We assume that $\mathbf{V}_{ij}$, $\Psi_j, (i = 1, \cdots, m; j = 1, \cdots, t)$, and $\mathbf{W}$ are known positive definite (p.d.) matrices. Note that the proposed model assumes known ratios of variance components for simplicity as in Datta and Ghosh (1991b). $\mathbf{H}_j (q \times q), (j = 1, \cdots, t)$ are known matrices. $\mathbf{b}_0$ is assumed to be known. Without any loss of generality, we assume $\mathbf{b}_0 = \mathbf{0}$.

Before writing the stages I–III of the above model in a compact form, we need to introduce a few notations. In what follows, we shall use the notation $\mathbf{I}_u$ for an identity matrix of order u and $\mathbf{1}_u$ for a u-component column vector with each element equal to 1. $\emptyset$ denotes a null matrix of suitable order. Also, let $\mathbf{A}_{p\times q} \otimes \mathbf{B}_{m\times n}$ denote the matrix $\begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1q}\mathbf{B} \\ \vdots & & \vdots \\ a_{p1}\mathbf{B} & \cdots & a_{pq}\mathbf{B} \end{bmatrix}$ and let $\oplus_{i=1}^m \mathbf{A}_i$ denote the matrix $\begin{bmatrix} \mathbf{A}_1 & \cdots & \emptyset \\ \vdots & & \vdots \\ \emptyset & \cdots & \mathbf{A}_m \end{bmatrix}$

We define

$$\mathbf{C}_t^{(1)} = \left(\mathbf{C}_{1t}^{(1)^T}, \mathbf{C}_{2t}^{(1)^T}, \cdots, \mathbf{C}_{mt}^{(1)^T}\right)^T$$

$$\mathbf{C}_{it}^{(1)} = \left(\mathbf{Y}_{i1}^{(1)^T}, \mathbf{Y}_{i2}^{(1)^T}, \cdots, \mathbf{Y}_{it}^{(1)^T}\right)^T ; \ (i = 1, \cdots, m)$$

$$\mathbf{C}_t^{(2)} = \left(\mathbf{C}_{1t}^{(2)^T}, \mathbf{C}_{2t}^{(2)^T}, \cdots, \mathbf{C}_{mt}^{(2)^T}\right)^T$$

$$\mathbf{C}_{it}^{(2)} = \left(\mathbf{Y}_{i1}^{(2)^T}, \mathbf{Y}_{i2}^{(2)^T}, \cdots, \mathbf{Y}_{it}^{(2)^T}\right)^T ; \ (i = 1, \cdots, m)$$

$$\mathbf{A}_t^{(1)} = \left(\mathbf{X}_1^{(1)^T}, \mathbf{X}_2^{(1)^T}, \cdots, \mathbf{X}_m^{(1)^T}\right)^T$$

$$\mathbf{X}_i^{(1)} = \left(\mathbf{X}_{i1}^{(1)^T}, \mathbf{X}_{i2}^{(1)^T}, \cdots, \mathbf{X}_{it}^{(1)^T}\right)^T ; \ (i = 1, \cdots, m)$$

$$\mathbf{A}_t^{(2)} = \left(\mathbf{X}_1^{(2)^T}, \mathbf{X}_2^{(2)^T}, \cdots, \mathbf{X}_m^{(2)^T}\right)^T$$

$$\mathbf{X}_i^{(2)} = \left(\mathbf{X}_{i1}^{(2)^T}, \mathbf{X}_{i2}^{(2)^T}, \cdots, \mathbf{X}_{it}^{(2)^T}\right)^T ; \ (i = 1, \cdots, m)$$

$$\mathbf{U}_{jl} = \begin{cases} \prod_{k=l}^{j} \mathbf{H}_k \text{ for } l = 1, 2, \cdots, j; \ j = 1, \cdots, t \\ \mathbf{I}_q \text{ for } l = j+1; \ j = 1, \cdots, t \end{cases}$$

$$\mathbf{B}_{it}^{(1)} = \begin{bmatrix} \mathbf{Z}_{i1}^{(1)} & \emptyset & \cdots & \emptyset \\ \mathbf{Z}_{i2}^{(1)}\mathbf{U}_{22} & \mathbf{Z}_{i2}^{(1)} & \cdots & \emptyset \\ \vdots & \vdots & & \vdots \\ \mathbf{Z}_{it}^{(1)}\mathbf{U}_{t2} & \mathbf{Z}_{it}^{(1)}\mathbf{U}_{t3} & \cdots & \mathbf{Z}_{it}^{(1)} \end{bmatrix}$$

$$\mathbf{B}_{it}^{(2)} = \begin{bmatrix} \mathbf{Z}_{i1}^{(2)} & \emptyset & \cdots & \emptyset \\ \mathbf{Z}_{i2}^{(2)}\mathbf{U}_{22} & \mathbf{Z}_{i2}^{(2)} & \cdots & \emptyset \\ \vdots & \vdots & & \vdots \\ \mathbf{Z}_{it}^{(2)}\mathbf{U}_{t2} & \mathbf{Z}_{it}^{(2)}\mathbf{U}_{t3} & \cdots & \mathbf{Z}_{it}^{(2)} \end{bmatrix}$$

$$\mathbf{B}_t^{(1)} = \begin{bmatrix} \mathbf{B}_{1t}^{(1)} \\ \mathbf{B}_{2t}^{(1)} \\ \vdots \\ \mathbf{B}_{mt}^{(1)} \end{bmatrix}, \quad \mathbf{B}_t^{(2)} = \begin{bmatrix} \mathbf{B}_{1t}^{(2)} \\ \mathbf{B}_{2t}^{(2)} \\ \vdots \\ \mathbf{B}_{mt}^{(2)} \end{bmatrix}$$

$$\mathbf{V}_t = \text{ Block Diag } (\mathbf{V}_{11}, \cdots, \mathbf{V}_{1t}, \mathbf{V}_{21}, \cdots, \mathbf{V}_{2t}, \cdots, \mathbf{V}_{m1}, \cdots, \mathbf{V}_{mt})$$

$$\Psi_t = \text{ Block Diag } (\Psi_1, \cdots, \Psi_t).$$

We now write the model as a compact linear mixed model which incorporates all the information up to and including time t for all the small areas. We have, without any loss of generality, listed the sampled units first. To this end, write

$$\mathbf{C}_t = \mathbf{A}_t\alpha + \mathbf{B}_t\omega + \eta \tag{2.1}$$

where

$$\omega \sim N\left(\mathbf{0},\ r^{-1}\left(\mathbf{I}_t \otimes \mathbf{W}\right)\right),$$

$$\eta \sim N\left(\mathbf{0},\ r^{-1}\left(\left(\mathbf{I}_m \otimes \Psi_t\right) + \mathbf{V}_t\right)\right),$$

$$\mathbf{C}_t = \begin{pmatrix} \mathbf{C}_t^{(1)} \\ \mathbf{C}_t^{(2)} \end{pmatrix}, \ \mathbf{A}_t = \begin{pmatrix} \mathbf{A}_t^{(1)} \\ \mathbf{A}_t^{(2)} \end{pmatrix}, \ \mathbf{B}_t = \begin{pmatrix} \mathbf{B}_t^{(1)} \\ \mathbf{B}_t^{(2)} \end{pmatrix}, \ \eta = \begin{pmatrix} \eta^{(1)} \\ \eta^{(2)} \end{pmatrix}$$

Given α and r,

$$\mathbf{C}_t \sim N\left(\mathbf{A}_t\,\alpha,\ r^{-1}\,\mathbf{\Sigma}_t\right) \tag{2.2}$$

where

$$\mathbf{\Sigma}_t = \mathbf{B}_t\left(\mathbf{I}_t \otimes \mathbf{W}\right)\mathbf{B}_t^T + \left(\mathbf{I}_m \otimes \Psi_t\right) + \mathbf{V}_t.$$

Furthermore, we partition $\mathbf{\Sigma}_t$ as

$$\mathbf{\Sigma}_t = \begin{bmatrix} \mathbf{\Sigma}_{t11} & \mathbf{\Sigma}_{t12} \\ \mathbf{\Sigma}_{t21} & \mathbf{\Sigma}_{t22} \end{bmatrix}$$

where $\mathbf{\Sigma}_{t11}$ is $n_t \times n_t$, $\mathbf{\Sigma}_{t12}$ $\left(= \mathbf{\Sigma}_{t21}^T\right)$ is $n_t \times (N_t - n_t)$, and $\mathbf{\Sigma}_{t22}$ is $(N_t - n_t) \times (N_t - n_t)$. Here $n_t = \sum_{i=1}^m \sum_{j=1}^t n_{ij}$ and $N_t = \sum_{i=1}^m \sum_{j=1}^t N_{ij}$. Define

$$\mathbf{\Sigma}_{t22.1} = \mathbf{\Sigma}_{t22} - \mathbf{\Sigma}_{t21}\mathbf{\Sigma}_{t11}^{-1}\mathbf{\Sigma}_{t12}.$$

In model-based approach in survey sampling, the primary objective is to find the conditional (predictive) distribution of $\mathbf{C}_t^{(2)}$ given $\mathbf{C}_t^{(1)} = \mathbf{c}_t^{(1)}$. Before stating Theorem 2.1, we need to define certain known distributions. A random variable Z is said to have a Gamma(α, β) distribution if it has pdf $f(z) = [\exp(-\alpha z)\alpha^\beta z^{\beta-1}/\Gamma(\beta)]I_{[z>0]}$. A random vector $\mathbf{T} = (T_1, \cdots, T_p)^T$ is said to have a multivariate t-distribution with location parameter μ, scale parameter Φ and degrees of freedom ν if it has pdf

$$g(\mathbf{t}) \propto |\Phi|^{-1/2}[\nu + (\mathbf{t}-\mu)^T \Phi^{-1}(\mathbf{t}-\mu)]^{-(\nu+p)/2}.$$

For details regarding multivariate t-distribution see Zellner (1971) or Press (1972).

Theorem 2.1. Consider the model given in (2.1) or (2.2). Assume that α and r are independently distributed with

$$\begin{aligned} \alpha &\sim \text{Uniform}\,(R^p) \\ r &\sim \text{Gamma}\,(\tfrac{1}{2}a_0, \tfrac{1}{2}b_0);\ a_0 > 0, b_0 > 0. \end{aligned}$$

Also, assume that $n_t + b_0 - p > 2$. Then conditional on $\mathbf{C}_t^{(1)} = \mathbf{c}_t^{(1)}$, $\mathbf{C}_t^{(2)}$ is distributed as multivariate t with degrees of freedom $n_t + b_0 - p$, location parameter $M_t \mathbf{c}_t^{(1)}$ and scale parameter

$$\Omega_t = (n_t + b_0 - p)^{-1}\left(a_0 + \mathbf{c}_t^{(1)^T}\mathbf{K}_t \mathbf{c}_t^{(1)}\right)\mathbf{G}_t, \tag{2.3}$$

where

$$\begin{aligned} \mathbf{K}_t &= \mathbf{\Sigma}_{t11}^{-1} - \mathbf{\Sigma}_{t11}^{-1}\mathbf{A}_t^{(1)}\left(\mathbf{A}_t^{(1)^T}\mathbf{\Sigma}_{t11}^{-1}\mathbf{A}_t^{(1)}\right)^{-1}\mathbf{A}_t^{(1)^T}\mathbf{\Sigma}_{t11}^{-1} \\ \mathbf{M}_t &= \mathbf{\Sigma}_{t21}\mathbf{K}_t + \mathbf{A}_t^{(2)}\left(\mathbf{A}_t^{(1)^T}\mathbf{\Sigma}_{t11}^{-1}\mathbf{A}_t^{(1)}\right)^{-1}\mathbf{A}_t^{(1)^T}\mathbf{\Sigma}_{t11}^{-1} \\ \mathbf{G}_t &= \mathbf{\Sigma}_{t22.1} + \left(\mathbf{A}_t^{(2)} - \mathbf{\Sigma}_{t21}\mathbf{\Sigma}_{t11}^{-1}\mathbf{A}_t^{(1)}\right) \\ &\quad \times \left(\mathbf{A}_t^{(1)^T}\mathbf{\Sigma}_{t11}^{-1}\mathbf{A}_t^{(1)}\right)^{-1}\left(\mathbf{A}_t^{(2)} - \mathbf{\Sigma}_{t21}\mathbf{\Sigma}_{t11}^{-1}\mathbf{A}_t^{(1)}\right)^T \end{aligned} \tag{2.4}$$

The proof of Theorem 2.1 is technical, and is deferred to the Appendix. We are interested in predicting linear functions of the form

$$\xi_t \equiv \xi_t(\mathbf{C}_t^{(1)}, \mathbf{C}_t^{(2)}) = \mathbf{P}_1\mathbf{C}_t^{(1)} + \mathbf{P}_2\mathbf{C}_t^{(2)}$$

where $\mathbf{P}_1$ and $\mathbf{P}_2$ are known matrices, on the basis of the observed $\mathbf{c}_t^{(1)}$. For example, when $\mathbf{P}_1 = \oplus_{i=1}^m \mathbf{P}_{1i}$ and $\mathbf{P}_2 = \oplus_{i=1}^m \mathbf{P}_{2i}$ where $\mathbf{P}_{1i} = [\emptyset,\ \emptyset, \cdots, \mathbf{1}_{n_{ij}}^T, \cdots, \emptyset]$ and $\mathbf{P}_{2i} = [\emptyset,\ \emptyset, \cdots, \mathbf{1}_{N_{ij}-n_{ij}}^T, \cdots, \emptyset]$, $\xi_t(\mathbf{C}_t^{(1)}, \mathbf{C}_t^{(2)})$ reduces to the vector of finite population totals for the m small areas at time j, $(j = 1, 2, \cdots, t)$. In particular, we may be interested in predicting retrospectively the population totals for the small areas at the previous time points based on all the available sampled data at a given time.

By Theorem 2.1, the HB predictor of $\xi_t(\mathbf{C}_t^{(1)}, \mathbf{C}_t^{(2)})$ conditional on $\mathbf{C}_t^{(1)} = \mathbf{c}_t^{(1)}$ is given by

$$\mathbf{e}_t^B(\mathbf{c}_t^{(1)}) = (\mathbf{P}_1 + \mathbf{P}_2\mathbf{M}_t)\,\mathbf{c}_t^{(1)}. \tag{2.5}$$

Note that the predictor $\mathbf{e}_t^B(\mathbf{c}_t^{(1)})$ does not depend on the choice of the prior (proper) distribution of r. In this sense, the predictor $\mathbf{e}_t^B(\mathbf{c}_t^{(1)})$ is robust against the choice of priors for r.

3. UNIVERSAL AND STOCHASTIC DOMINATION

To investigate the frequentist properties of the proposed HB predictors given in Section 2, we consider the frequentist framework, where α and r are unknown. We do not assign any prior distributions for α and r and treat them as unknown parameters. Inference is done conditionally on these parameters. Write $\phi = (\alpha^T,\ r)^T$. Also, we write $\mathbf{e}^* = (\omega^T,\ \eta^T)^T$ where ω and η are as defined in Section 2. We dispense with the normality assumptions on ω and η.

A predictor $\delta(\mathbf{C}_t^{(1)})$ is said to be linear if $\delta(\mathbf{C}_t^{(1)})$ has the form $\mathbf{L}_t\mathbf{C}_t^{(1)}$ for some known $u \times n_t$ matrix $\mathbf{L}_t$. If in addition $E_\Phi\left[\delta(\mathbf{C}_t^{(1)}) - \xi_t(\mathbf{C}_t^{(1)}, \mathbf{C}_t^{(2)})\right] = \mathbf{0}$ for all Φ, we say that $\delta(\mathbf{C}_t^{(1)})$ is a linear unbiased predictor (LUP) of $\xi_t(\mathbf{C}_t^{(1)}, \mathbf{C}_t^{(2)})$. We consider the following quadratic loss

$$\begin{aligned} L_1(\xi_t, \delta) &= \|\delta - \xi_t\|_\omega^2 \\ &= (\delta - \xi_t)^T\omega(\delta - \xi_t) \\ &= tr\left[\omega L_0(\xi_t,\ \delta)\right], \end{aligned}$$

where ω is a nonnegative definite (n.n.d.) matrix and L_0 is the matrix loss such that

$$L_0(\xi_t, \delta) = (\delta - \xi_t)(\delta - \xi_t)^T.$$

We shall refer to such a loss as generalized Euclidean error w.r.t. ω. Our question to ask now is whether the risk optimality of $\mathbf{e}_t^B(\mathbf{C}_t^{(1)})$ holds within the class of all unbiased predictors, or at least within the class of all LUP's under certain criterion for a broader family of distributions of $\mathbf{e}^*$. To investigate this question, we need the notion of "universal" or "stochastic" domination as given in Hwang (1985). Let $R_L(\Phi, \xi_t; \delta) = E_\Phi \left[L\left(\left\|\delta(\mathbf{C}_t^{(1)}) - \xi_t(\mathbf{C}_t^{(1)},\ \mathbf{C}_t^{(2)})\right\|_\Omega^2\right)\right]$ be the risk function of the predictor δ for predicting ξ_t under a loss function which is a function of generalized Euclidean error w.r.t. ω for some function L. The following definition is adapted from Hwang (1985).

Definition 3.1. An estimator $\delta_1(\mathbf{C}_t^{(1)})$ universally dominates $\delta_2(\mathbf{C}_t^{(1)})$ (under the generalized Euclidean error loss w.r.t. Ω) if for every Φ, and every nondecreasing loss function L, $R_L(\Phi,\ \xi_t;\ \delta_1) \leq R_L(\Phi,\ \xi_t;\ \delta_2)$ holds and for a particular loss, the risk functions are not identical.

In Hwang (1985), it has been shown that δ_1 universally dominates δ_2 under the generalized Euclidean error w.r.t Ω if and only if $\left\|\delta_1(\mathbf{C}_t^{(1)}) - \xi_t(\mathbf{C}_t^{(1)},\ \mathbf{C}_t^{(2)})\right\|_\Omega^2$ is stochastically smaller than $\left\|\delta_2(\mathbf{C}_t^{(1)}) - \xi_t(\mathbf{C}_t^{(1)},\ \mathbf{C}_t^{(2)})\right\|_\Omega^2$. We say that a random variable Z_1 is stochastically smaller than Z_2 if $P_\theta(Z_1 > x) \leq P_\theta(Z_2 > x)$ for all x and all θ, and for some θ, Z_1, and Z_2 have distinct distributions.

We now assume that $\mathbf{e}^*$ has an elliptically symmetric distribution with pdf given by

$$h(\mathbf{e}^* \mid \Delta, r) \ \propto \ \left|r^{-1}\Delta\right|^{-\frac{1}{2}} f(r\mathbf{e}^{*T}\Delta^{-1}\mathbf{e}^*) \tag{3.1}$$

where

$$\delta \ = \ \text{Block Diag } \left(\mathbf{B}_t(\mathbf{I}_t \otimes \mathbf{W})\mathbf{B}_t^T,\ (\mathbf{I}_m \otimes \Psi_t) + \mathbf{V}_t\right).$$

Note that the normality of $\mathbf{e}^*$ with mean $\mathbf{0}$ and variance covariance matrix $r^{-1}\Delta$ is sufficient but not necessary for (3.1) to hold.

Now write $\mathbf{D}_{tj}^* = \mathbf{B}_t{}^{(j)}\mathbf{w} + \mathbf{e}^{(j)}$ $(j = 1, 2)$. Then $\mathbf{D}_t^* = (\mathbf{D}_{t1}^{*\,T},\ \mathbf{D}_{t2}^{*\,T})^T$ has also an elliptically symmetric pdf given by

$$h(\mathbf{d}_t^* | \mathbf{\Sigma}_t, r) \propto \left|r^{-1}\mathbf{\Sigma}_t\right|^{-\frac{1}{2}} f(r\mathbf{d}_t^{*T}\mathbf{\Sigma}_t^{-1}\mathbf{d}_t^*).$$

The next theorem shows that for a general class of elliptically symmetric distributions of $\mathbf{e}^*$, $\mathbf{e}_t^B(\mathbf{C}_t^{(1)})$ universally dominates every LUP of $\xi_t(\mathbf{C}_t^{(1)},\ \mathbf{C}_t^{(2)})$, $\mathbf{S}_t\mathbf{C}_t^{(1)}$, under every generalized Euclidean error loss w.r.t a n.n.d. Ω.

Theorem 3.1. Under the model (2.1) and (3.1), $\mathbf{e}_t^{B*}(\mathbf{C}_t^{(1)})$ universally dominates every LUP $\delta(\mathbf{C}_t^{(1)}) = \mathbf{S}_t\mathbf{C}_t^{(1)}$ of $\xi_t(\mathbf{C}_t^{(1)},\ \mathbf{C}_t^{(2)})$ for every p.d. Ω.

Theorem 3.1 is based on the following lemma.

Lemma 3.1. If $\mathbf{D}(N_t \times 1)$ has pdf $h(\mathbf{d}_t|\mathbf{I}_{N_t}, r)$, then for every $\mathbf{L}(u \times N_t)$, $u \leq N_t$, $\mathbf{LD} \stackrel{\mathrm{d}}{=} (\mathbf{LL}^T)^{\frac{1}{2}}\mathbf{D}_u$, where $\mathbf{D}_u = (\mathbf{I}_u, \emptyset)\mathbf{D}$, $\emptyset$ $(u \times (N_t - u))$ is a null matrix and $\stackrel{\mathrm{d}}{=}$ means equal in distribution.

The proofs of Lemma 3.1 and Theorem 3.1 are similar as Lemma 1 and Theorem 2 in Datta and Ghosh (1991b), and are omitted.

4. BEST EQUIVARIANT PREDICTION

In this section we concentrate on the equivariant prediction of $\xi_t(\mathbf{C}_t^{(1)}, \mathbf{C}_t^{(2)})$ on the basis of $\mathbf{C}_t^{(1)}$ under the suitable group of transformations. We assume the elliptic symmetry of $\mathbf{e}^*$. Consider the group of transformations

$$\mathcal{G} = \{\mathbf{g}_{\beta,d}, \beta \in R^p, d > 0 : \mathbf{g}_{\beta,d}(\mathbf{C}_t) = d\ \mathbf{C}_t + \mathbf{A}_t\beta\}\,. \tag{4.1}$$

Also assume that we partition $\mathbf{C}_t$ and $\mathbf{A}_t$ as in (2.1) and only $\mathbf{C}_t^{(1)} = \mathbf{c}_t^{(1)}$ is observed. If $\delta(\mathbf{C}_t^{(1)})$ estimates $(\mathbf{P}_1\mathbf{A}_t^{(1)} + \mathbf{P}_2\mathbf{A}_t^{(2)})\alpha$, then $d\delta(\mathbf{C}_t^{(1)}) + \mathbf{P}_1\mathbf{A}_t^{(1)}\beta + \mathbf{P}_2\mathbf{A}_t^{(2)}\beta$ should estimate $(\mathbf{P}_1\mathbf{A}_t^{(1)} + \mathbf{P}_2\mathbf{A}_t^{(2)})(d\alpha + \beta) = (\mathbf{P}_1\mathbf{P}_2)\mathbf{A}_t(d\alpha + \beta)$. Treating $\mathbf{A}_t(d\alpha + \beta)$ as the new location parameter, one may expect that $\delta(d\mathbf{C}_t^{(1)} + \mathbf{A}_t^{(1)}\beta)$ will estimate $(\mathbf{P}_1\mathbf{A}_t^{(1)} + \mathbf{P}_2\mathbf{A}_t^{(2)})(d\alpha + \beta)$. So we should have

$$\delta(d\mathbf{C}_t^{(1)} + \mathbf{A}_t^{(1)}\beta) = d\delta(\mathbf{C}_t^{(1)}) + \mathbf{P}_1\mathbf{A}_t^{(1)}\beta + \mathbf{P}_2\mathbf{A}_t^{(2)}\beta, \tag{4.2}$$

for all β and all $d > 0$. Now if we are interested in predicting $\xi_t(\mathbf{C}_t^{(1)}, \mathbf{C}_t^{(2)})$ instead of $E_\Phi(\xi_t(\mathbf{C}_t^{(1)}, \mathbf{C}_t^{(2)})) = \mathbf{P}_1\mathbf{A}_t^{(1)}\alpha + \mathbf{P}_2\mathbf{A}_t^{(2)}\alpha$, we can still use $\delta(\mathbf{C}_t^{(1)})$ and again we will impose (4.2) on δ.

Note that the induced group of transformations on the parameter space is given by

$$\overline{\mathcal{G}} = \left\{\overline{\mathbf{g}}_{\beta,d}, \beta \in R^p, d > 0 : \overline{\mathbf{g}}_{\beta,d}(\Phi) = (d^{-2}r, d\alpha^T + \beta^T)^T\right\}. \tag{4.3}$$

A loss function $L(\xi_t, \Phi; \delta)$ for predicting $\xi_t(\mathbf{C}_t^{(1)}, \mathbf{C}_t^{(2)})$ by $\delta(\mathbf{C}_t^{(1)})$ is invariant under the group of transformations $\mathcal{G}$ if

$$\begin{aligned} &L(d\xi_t(\mathbf{C}_t^{(1)}, \mathbf{C}_t^{(2)}) + (\mathbf{P}_1\mathbf{A}_t^{(1)} + \mathbf{P}_2\mathbf{A}_t^{(2)})\beta, \overline{\mathbf{g}}_{\beta,d}(\Phi);\\ &\quad d\delta(\mathbf{C}_t^{(1)}) + (\mathbf{P}_1\mathbf{A}_t^{(1)} + \mathbf{P}_2\mathbf{A}_t^{(2)})\beta)\\ &\quad = L(\xi_t(\mathbf{C}_t^{(1)}, \mathbf{C}_t^{(2)}), \Phi; \delta(\mathbf{C}_t^{(1)})) \end{aligned} \tag{4.4}$$

for all $\mathbf{C}_t^{(1)}, \mathbf{C}_t^{(2)}, \beta, d(>0)$ and Φ. The two losses $L_2 = rL_0$ and $L_3 = rL_1$ both satisfy (4.4).

We will now be interested in the best equivariant predictor of $\xi_t(\mathbf{C}_t^{(1)}, \mathbf{C}_t^{(2)})$. The following two lemmas provide a useful characterization of the class of equivariant predictors of $\xi_t(\mathbf{C}_t^{(1)}, \mathbf{C}_t^{(2)})$.

Lemma 4.1. Let $\delta_0(\mathbf{C}_t^{(1)})$ be an equivariant predictor for $\xi_t(\mathbf{C}_t^{(1)}, \mathbf{C}_t^{(2)})$. Then a necessary and sufficient condition for a predictor $\delta(\mathbf{C}_t^{(1)})$ of $\xi_t(\mathbf{C}_t^{(1)}, \mathbf{C}_t^{(2)})$ to be equivariant is that

$$\delta(\mathbf{C}_t^{(1)}) = \delta_0(\mathbf{C}_t^{(1)}) + \mathbf{h}(\mathbf{C}_t^{(1)}) \tag{4.5}$$

where $\mathbf{h}$ satisfies

$$\mathbf{h}(d\mathbf{C}_t^{(1)} + \mathbf{A}_t^{(1)}\beta) = d\mathbf{h}(\mathbf{C}_t^{(1)}) \tag{4.6}$$

for all $\mathbf{C}_t^{(1)}, \beta$ and $d > 0$.

The proof of the Lemma 4.1 is straightforward, and is ommitted. To find a representation of the function $\mathbf{h}$ in (4.6), we define

$$\mathbf{p}(\mathbf{K}_t\mathbf{C}_t^{(1)}) = \{\mathbf{K}_t\mathbf{C}_t^{(1)}/(\mathbf{C}_t^{(1)^T}\mathbf{K}_t\mathbf{C}_t^{(1)})^{\frac{1}{2}}\}I_{[\mathbf{C}_t^{(1)T}\mathbf{K}_t\mathbf{C}_t^{(1)}>0]}. \tag{4.7}$$

Note that since $\mathbf{K}_t\mathbf{\Sigma}_{t11}\mathbf{K}_t = \mathbf{K}_t$, so $\mathbf{C}_t^{(1)^T}\mathbf{K}_t\mathbf{C}_t^{(1)} = (\mathbf{K}_t\mathbf{C}_t^{(1)})^T\mathbf{\Sigma}_{t11}(\mathbf{K}_t\mathbf{C}_t^{(1)})$ and $\mathbf{p}$ is indeed a function of $\mathbf{K}_t\mathbf{C}_t^{(1)}$. It can be shown that $\mathbf{p}(\mathbf{K}_t\mathbf{C}_t^{(1)})$ is a maximal invariant under the group of transformations

$$\mathcal{G}' = \left\{\mathbf{g}'_{\beta,d}, \beta \in R^p, d > 0 : \mathbf{g}'_{\beta,d}(\mathbf{C}_t^{(1)}) = d\mathbf{C}_t^{(1)} + \mathbf{A}_t^{(1)}\beta\right\} \tag{4.8}$$

induced by $\mathcal{G}$ in the $\mathbf{C}_t^{(1)}$-space.

The following lemma characterizes the class of functions $\mathbf{h}(\mathbf{C}_t^{(1)})$ satisfying (4.6). We will use this lemma to characterize the class of equivariant predictors.

Lemma 4.2. A function $\mathbf{h}(\mathbf{C}_t^{(1)})(u \times 1)$ satisfies (4.6) if and only if $\mathbf{h}$ has the representation

$$\mathbf{h}(\mathbf{C}_t^{(1)}) = (\mathbf{C}_t^{(1)^T}\mathbf{K}_t\mathbf{C}_t^{(1)})^{\frac{1}{2}}\mathbf{s}(\mathbf{p}(\mathbf{K}_t\mathbf{C}_t^{(1)})) \tag{4.9}$$

where $\mathbf{s}(u \times 1)$ is an arbitrary function of $\mathbf{p}(\mathbf{K}_t\mathbf{C}_t^{(1)})$.

Proof. If. Assume $\mathbf{h}$ has the representation given in (4.9). Now, since $(d\mathbf{C}_t^{(1)} + \mathbf{A}_t^{(1)}\beta)^T\mathbf{K}_t(d\mathbf{C}_t^{(1)} + \mathbf{A}_t^{(1)}\beta) = d^2\mathbf{C}_t^{(1)^T}\mathbf{K}_t\mathbf{C}_t^{(1)}$ and since $\mathbf{p}(\mathbf{K}_t\mathbf{C}_t^{(1)})$

is a maximal invariant under $\mathcal{G}'$, so $\mathbf{p}(\mathbf{K}_t(d\mathbf{C}_t^{(1)} + \mathbf{A}_t^{(1)}\beta)) = \mathbf{p}(\mathbf{K}_t\mathbf{C}_t^{(1)})$ and consequently

$$\begin{aligned}\mathbf{h}(d\mathbf{C}_t^{(1)} + \mathbf{A}_t^{(1)}\beta) &= (d^2\mathbf{C}_t^{(1)^T}\mathbf{K}_t\mathbf{C}_t^{(1)})^{\frac{1}{2}}\mathbf{s}(\mathbf{p}(\mathbf{K}_t\mathbf{C}_t^{(1)})) \\ &= d\mathbf{h}(\mathbf{C}_t^{(1)}).\end{aligned}$$

Hence (4.6) is satisfied.

Only if. Since $\mathbf{h}$ satisfies (4.6) for all $\mathbf{C}_t^{(1)}, \beta$ and $d > 0$, taking $d = 1$, we see that $\mathbf{h}$ must satisfy $\mathbf{h}(\mathbf{C}_t^{(1)} + \mathbf{A}_t^{(1)}\beta) = \mathbf{h}(\mathbf{C}_t^{(1)})$ for all $\mathbf{C}_t^{(1)}$ and β. This implies that $\mathbf{h}$ must be invariant under the group of transformations

$$\mathcal{G}'' = \left\{\mathbf{g}'_\beta, \beta \in R^p : \mathbf{g}'_\beta(\mathbf{C}_t^{(1)}) = \mathbf{C}_t^{(1)} + \mathbf{A}_t^{(1)}\beta\right\}, \tag{4.10}$$

and hence must be a function of $\mathbf{K}_t\mathbf{C}_t^{(1)}$. So $\mathbf{h}(\mathbf{C}_t^{(1)}) = \mathbf{s}(\mathbf{K}_t\mathbf{C}_t^{(1)})$ where $\mathbf{s}(u \times 1)$ is an arbitrary function satisfying

$$\begin{aligned}\mathbf{s}(\mathbf{K}_t(d\mathbf{C}_t^{(1)} + \mathbf{A}_t^{(1)}\beta)) &= d\mathbf{s}(\mathbf{K}_t\mathbf{C}_t^{(1)}) \\ \text{i.e. } \mathbf{s}(d\mathbf{K}_t\mathbf{C}_t^{(1)}) &= d\mathbf{s}(\mathbf{K}_t\mathbf{C}_t^{(1)})\end{aligned} \tag{4.11}$$

for all $d > 0$. Now taking $d = (\mathbf{C}_t^{(1)^T}\mathbf{K}_t\mathbf{C}_t^{(1)})^{-\frac{1}{2}}$ for all $\mathbf{C}_t^{(1)^T}\mathbf{K}_t\mathbf{C}_t^{(1)} > 0$ we have from (4.11)

$$\begin{aligned}\mathbf{h}(\mathbf{C}_t^{(1)}) &= \mathbf{s}(\mathbf{K}_t\mathbf{C}_t^{(1)}) \\ &= \mathbf{s}\left(\left(\mathbf{C}_t^{(1)^T}\mathbf{K}_t\mathbf{C}_t^{(1)}\right)^{-\frac{1}{2}}\mathbf{K}_t\mathbf{C}_t^{(1)}\right)\left(\mathbf{C}_t^{(1)}\mathbf{K}_t\mathbf{C}_t^{(1)}\right)^{\frac{1}{2}} \\ &= \mathbf{s}\left(\mathbf{p}\left(\mathbf{K}_t\mathbf{C}_t^{(1)}\right)\right)\left(\mathbf{C}_t^{(1)}\mathbf{K}_t\mathbf{C}_t^{(1)}\right)^{\frac{1}{2}}.\end{aligned}$$

Now for $\mathbf{C}_t^{(1)}\mathbf{K}_t\mathbf{C}_t^{(1)} = 0$, if we take $\mathbf{h} = \mathbf{0}$, then (4.6) is satisfied and we can represent $\mathbf{h}$ by (4.9).

Since $\mathbf{e}_t^B(\mathbf{C}_t^{(1)})$ is an equivariant predictor of $\xi_t(\mathbf{C}_t^{(1)}, \mathbf{C}_t^{(2)})$, it follows from Lemma 4.1 and Lemma 4.2 that $\delta(\mathbf{C}_t^{(1)})$ is an equivariant predictor of $\xi_t(\mathbf{C}_t^{(1)}, \mathbf{C}_t^{(2)})$ under the group of transformations $\mathcal{G}$ if and only if

$$\delta(\mathbf{C}_t^{(1)}) = \mathbf{e}_t^B(\mathbf{C}_t^{(1)}) + (\mathbf{C}_t^{(1)^T}\mathbf{K}_t\mathbf{C}_t^{(1)})^{\frac{1}{2}}\mathbf{s}(\mathbf{p}(\mathbf{K}_t\mathbf{C}_t^{(1)})) \tag{4.12}$$

for all $\mathbf{C}_t^{(1)}$.

Definition 4.1. An equivariant predictor $\delta_0(\mathbf{C}_t^{(1)})$ of $\xi_t(\mathbf{C}_t^{(1)}, \mathbf{C}_t^{(2)})$ is said to be a best equivariant predictor under the loss L_2 if for every other equivariant predictor $\delta(\mathbf{C}_t^{(1)})$ of $\xi_t(\mathbf{C}_t^{(1)}, \mathbf{C}_t^{(2)})$, $E_\Phi\left[L_2(\xi_t, \Phi; \delta) - L_2(\xi_t, \Phi; \delta_0)\right]$ is nonnegative definite.

Remark 4.1. Note that if $\delta_0(\mathbf{C}_t^{(1)})$ is the best equivariant predictor of $\xi_t(\mathbf{C}_t^{(1)}, \mathbf{C}_t^{(2)})$ under the loss L_2, then it is so under the loss L_3 for every n.n.d. ω and vice versa.

We now establish that under the group of transformations $\mathcal{G}$, $\mathbf{e}_t^B(\mathbf{C}_t^{(1)})$ is best equivariant predictor of $\xi_t(\mathbf{C}_t^{(1)}, \mathbf{C}_t^{(2)})$ under the loss L_2.

Theorem 4.1. Consider the time domain HB model given in (2.1). Then under the group of transformations $\mathcal{G}$ and the loss function L_2, the best equivariant predictor of $\xi_t(\mathbf{C}_t^{(1)}, \mathbf{C}_t^{(2)})$ is given by $\mathbf{e}_t^B(\mathbf{C}_t^{(1)})$.

The proof of Theorem 4.1 is similar as Theorem 3 in Datta and Ghosh (1991b), and is ommitted.

Remark 4.2. Although $(\mathbf{C}_t^{(1)^T}\mathbf{K}_t\mathbf{C}_t^{(1)})^{\frac{1}{2}}\mathbf{A}_t^{(1)^T}\boldsymbol{\Sigma}_{t11}^{-1}\mathbf{C}_t^{(1)}$ is sufficient and $\mathbf{p}(\mathbf{K}_t\mathbf{C}_t^{(1)})$ is ancillary, Basu's theorem can not be applied since the sufficient statistics is not complete.

5. APPENDIX

Proof of Theorem 2.1. The joint pdf of $\mathbf{C}_t$, α and R is given by

$$\begin{aligned} f(\mathbf{c}_t, \alpha, r) \quad \propto \quad & r^{\frac{1}{2}(N_t+b_0)-1}|\boldsymbol{\Sigma}_t|^{-\frac{1}{2}} \\ & \times \exp[-\frac{r}{2}\{(\mathbf{c}_t - \mathbf{A}_t\alpha)^T\boldsymbol{\Sigma}_t^{-1}(\mathbf{c}_t - \mathbf{A}_t\alpha) + a_0\}]. \end{aligned} \quad (5.1)$$

Now

$$\begin{aligned} & (\mathbf{c}_t - \mathbf{A}_t\alpha)^T\boldsymbol{\Sigma}_t^{-1}(\mathbf{c}_t - \mathbf{A}_t\alpha) \\ = \quad & [\alpha - (\mathbf{A}_t{}^T\boldsymbol{\Sigma}_t^{-1}\mathbf{A}_t)^{-1}\mathbf{A}_t{}^T\boldsymbol{\Sigma}_t^{-1}\mathbf{c}_t]^T(\mathbf{A}_t{}^T\boldsymbol{\Sigma}_t^{-1}\mathbf{A}_t) \\ & \times[\alpha - (\mathbf{A}_t{}^T\boldsymbol{\Sigma}_t^{-1}\mathbf{A}_t)^{-1}\mathbf{A}_t{}^T\boldsymbol{\Sigma}_t^{-1}\mathbf{c}_t] + \mathbf{c}_t{}^T\mathbf{Q}_t\mathbf{c}_t. \end{aligned} \quad (5.2)$$

where

$$\mathbf{Q}_t \quad = \quad \boldsymbol{\Sigma}_t^{-1} - \boldsymbol{\Sigma}_t^{-1}\mathbf{A}_t(\mathbf{A}_t{}^T\boldsymbol{\Sigma}_t^{-1}\mathbf{A}_t)^{-1}\mathbf{A}_t^T\boldsymbol{\Sigma}_t^{-1}.$$

Integrating (5.1) w.r.t α, the joint pdf of $\mathbf{C}_t$ and R is given by

$$f(\mathbf{c}_t, r) \propto |\boldsymbol{\Sigma}_t|^{-\frac{1}{2}}|\mathbf{A}_t{}^T\boldsymbol{\Sigma}_t^{-1}\mathbf{A}_t|^{-\frac{1}{2}}r^{\frac{1}{2}(N_t+b_0-p)-1}\exp[-\frac{r}{2}(a_0 + \mathbf{c}_t{}^T\mathbf{Q}_t\mathbf{c}_t))]. \quad (5.3)$$

Integrating (5.3) w.r.t R, the pdf of $\mathbf{C}_t$ is given by

$$f(\mathbf{c}_t) \propto |\mathbf{\Sigma}_t|^{-\frac{1}{2}} |\mathbf{A}_t{}^T \mathbf{\Sigma}_t^{-1} \mathbf{A}_t|^{-\frac{1}{2}} (a_0 + \mathbf{c}_t{}^T \mathbf{Q}_t \mathbf{c}_t))^{-\frac{1}{2}(N_t + b_0 - p)}. \tag{5.4}$$

Now, using a standard formula for partitioned matrices, we have

$$\begin{aligned} \mathbf{c}_t^T \mathbf{\Sigma}_t^{-1} \mathbf{c}_t &= \mathbf{c}_t^{(1)^T} \mathbf{\Sigma}_{t11}^{-1} \mathbf{c}_t^{(1)} + (\mathbf{c}_t^{(2)} - \mathbf{\Sigma}_{t21} \mathbf{\Sigma}_{t11}^{-1} \mathbf{c}_t^{(1)})^T \mathbf{\Sigma}_{t22.1}^{-1} \\ &\quad \times (\mathbf{c}_t^{(2)} - \mathbf{\Sigma}_{t21} \mathbf{\Sigma}_{t11}^{-1} \mathbf{c}_t^{(1)}). \end{aligned} \tag{5.5}$$

Similarly

$$\begin{aligned} \mathbf{c}_t^T \mathbf{\Sigma}_t^{-1} \mathbf{A}_t &= \mathbf{c}_t{}^{(1)^T} \mathbf{\Sigma}_{t11}^{-1} \mathbf{A}_t^{(1)} + (\mathbf{c}_t^{(2)} - \mathbf{\Sigma}_{t21} \mathbf{\Sigma}_{t11}^{-1} \mathbf{A}_t^{(1)})^T \mathbf{\Sigma}_{t22.1}^{-1} \\ &\quad \times (\mathbf{c}_t^{(2)} - \mathbf{\Sigma}_{t21} \mathbf{\Sigma}_{t11}^{-1} \mathbf{A}_t^{(1)}) \\ &= \mathbf{s}_1^T + \mathbf{s}_2^T \quad \text{(say)}, \end{aligned} \tag{5.6}$$

$$\begin{aligned} \mathbf{A}_t{}^T \mathbf{\Sigma}_t^{-1} \mathbf{A}_t &= \mathbf{A}_t{}^{(1)^T} \mathbf{\Sigma}_{t11}^{-1} \mathbf{A}_t^{(1)} + (\mathbf{A}_t^{(2)} - \mathbf{\Sigma}_{t21} \mathbf{\Sigma}_{t11}^{-1} \mathbf{A}_t^{(1)})^T \mathbf{\Sigma}_{t22.1}^{-1} \\ &\quad \times (\mathbf{A}_t^{(2)} - \mathbf{\Sigma}_{t21} \mathbf{\Sigma}_{t11}^{-1} \mathbf{A}_t^{(1)}). \end{aligned} \tag{5.7}$$

Using the matrix inversion formula (see Exercise 2.9, page 33 of Rao (1973)), we have from (5.7) that

$$\begin{aligned} &(\mathbf{A}_t{}^T \mathbf{\Sigma}_t^{-1} \mathbf{A}_t)^{-1} \\ &= (\mathbf{A}_t{}^{(1)^T} \mathbf{\Sigma}_{t11}^{-1} \mathbf{A}_t^{(1)})^{-1} - (\mathbf{A}_t{}^{(1)^T} \mathbf{\Sigma}_{t11}^{-1} \mathbf{A}_t^{(1)})^{-1} (\mathbf{A}_t^{(2)} - \mathbf{\Sigma}_{t21} \mathbf{\Sigma}_{t11}^{-1} \mathbf{A}_t^{(1)})^T \\ &\quad \times \mathbf{G}_t{}^{-1} (\mathbf{A}_t^{(2)} - \mathbf{\Sigma}_{t21} \mathbf{\Sigma}_{t11}^{-1} \mathbf{A}_t^{(1)}) (\mathbf{A}_t{}^{(1)^T} \mathbf{\Sigma}_{t11}^{-1} \mathbf{A}_t^{(1)})^{-1} \\ &= \mathbf{M}_1 - \mathbf{M}_2 \quad \text{(say)} \end{aligned} \tag{5.8}$$

From (5.6), (5.8) and (2.4), we get after simplifications

$$\begin{aligned} &\mathbf{c}_t^T \mathbf{\Sigma}_t^{-1} \mathbf{A}_t (\mathbf{A}_t{}^T \mathbf{\Sigma}_t^{-1} \mathbf{A}_t)^{-1} \mathbf{A}_t{}^T \mathbf{\Sigma}_t^{-1} \mathbf{c}_t \\ &= \mathbf{s}_1^T \mathbf{M}_1 \mathbf{s}_1 - \mathbf{s}_1^T \mathbf{M}_2 \mathbf{s}_1 + \mathbf{s}_2^T \mathbf{M}_1 \mathbf{s}_2 - \mathbf{s}_2^T \mathbf{M}_2 \mathbf{s}_2 + 2\mathbf{s}_1^T (\mathbf{M}_1 - \mathbf{M}_2) \mathbf{s}_2, \end{aligned} \tag{5.9}$$

$$\mathbf{s}_1^T \mathbf{M}_1 \mathbf{s}_1 = \mathbf{c}_t^{(1)^T} (\mathbf{\Sigma}_{t11}^{-1} - \mathbf{K}_t) \mathbf{c}_t^{(1)}, \tag{5.10}$$

$$\mathbf{s}_1^T \mathbf{M}_2 \mathbf{s}_1 = (\mathbf{M}_t \mathbf{c}_t^{(1)} - \mathbf{\Sigma}_{t21} \mathbf{\Sigma}_{t11}^{-1} \mathbf{c}_t^{(1)})^T \mathbf{G}_t{}^{-1} (\mathbf{M}_t \mathbf{c}_t^{(1)} - \mathbf{\Sigma}_{t21} \mathbf{\Sigma}_{t11}^{-1} \mathbf{c}_t^{(1)}), \tag{5.11}$$

$$\mathbf{s}_2^T\mathbf{M}_1\mathbf{s}_2 = (\mathbf{c}_t^{(2)} - \boldsymbol{\Sigma}_{t21}\boldsymbol{\Sigma}_{t11}^{-1}\mathbf{c}_t^{(1)})^T \times[\boldsymbol{\Sigma}_{t22.1}^{-1}\mathbf{G}_t\boldsymbol{\Sigma}_{t22.1}^{-1} - \boldsymbol{\Sigma}_{t22.1}^{-1}](\mathbf{c}_t^{(2)} - \boldsymbol{\Sigma}_{t21}\boldsymbol{\Sigma}_{t11}^{-1}\mathbf{c}_t^{(1)}), \tag{5.12}$$

$$\mathbf{s}_2^T\mathbf{M}_2\mathbf{s}_2 = (\mathbf{c}_t^{(2)} - \boldsymbol{\Sigma}_{t21}\boldsymbol{\Sigma}_{t11}^{-1}\mathbf{c}_t^{(1)})^T \times[\boldsymbol{\Sigma}_{t22.1}^{-1}\mathbf{G}_t\boldsymbol{\Sigma}_{t22.1}^{-1} - 2\boldsymbol{\Sigma}_{t22.1}^{-1} + \mathbf{G}_t^{-1}](\mathbf{c}_t^{(2)} - \boldsymbol{\Sigma}_{t21}\boldsymbol{\Sigma}_{t11}^{-1}\mathbf{c}_t^{(1)}), \tag{5.13}$$

$$\mathbf{s}_1^T\mathbf{M}_1\mathbf{s}_2 = (\mathbf{M}_t\mathbf{c}_t^{(1)} - \boldsymbol{\Sigma}_{t21}\boldsymbol{\Sigma}_{t11}^{-1}\mathbf{c}_t^{(1)})^T\boldsymbol{\Sigma}_{t22.1}^{-1}(\mathbf{c}_t^{(2)} - \boldsymbol{\Sigma}_{t21}\boldsymbol{\Sigma}_{t11}^{-1}\mathbf{c}_t^{(1)}), \tag{5.14}$$

$$\mathbf{s}_1^T\mathbf{M}_2\mathbf{s}_2 = (\mathbf{M}_t\mathbf{c}_t^{(1)} - \boldsymbol{\Sigma}_{t21}\boldsymbol{\Sigma}_{t11}^{-1}\mathbf{c}_t^{(1)})^T \times[\boldsymbol{\Sigma}_{t22.1}^{-1} - \mathbf{G}_t^{-1}](\mathbf{c}_t^{(2)} - \boldsymbol{\Sigma}_{t21}\boldsymbol{\Sigma}_{t11}^{-1}\mathbf{c}_t^{(1)}). \tag{5.15}$$

Using the same definition of $\mathbf{Q}_t$, it follows from (5.5)-(5.15) with some algebraic manipulations that

$$\mathbf{c}_t^T\mathbf{Q}_t\mathbf{c}_t = \mathbf{c}_t^{(1)^T}\mathbf{K}_t\mathbf{c}_t^{(1)} + (\mathbf{c}_t^{(2)} - \mathbf{M}_t\mathbf{c}_t^{(1)})^T\mathbf{G}_t^{-1}(\mathbf{c}_t^{(2)} - \mathbf{M}_t\mathbf{c}_t^{(1)}). \tag{5.16}$$

Combining (5.4) and (5.16) and using the definition of multivariate t-distribution, one gets Theorem 2.1.

REFERENCES

(1) Battese, G.E., Harter, R.M. and Fuller, W.A. (1988). An Error Components Model for Prediction of County Crop Areas Using Survey and Satellite Data. *Journal of the American Statistical Association*, **83**, 28-36.

(2) Datta, G.S. and Ghosh, M. (1991a). Bayesian Prediction in Linear Model: Applications to Small Area Estimation. *The Annals of Statistics*, **19**, 1748-1770.

(3) Datta, G.S. and Ghosh, M. (1991b). On the Frequentist Properties of Some Hierarchical Bayes Predictors in Finite Population Sampling. *Calcutta Statistical Association Bulletin*, **40**, 207-217.

(4) Fay, R.E. and Herriot, R.A. (1979). Estimates of Income for Small Places: An Applixation of James-Stein Procedures to Census Data. *Journal of the American Statistical Association*, **74**, 269-277.

(5) Ghosh, M. and Lahiri, P. (1987). Robust Empirical Bayes Estimation of Means from Stratified Samples. *Journal of the American Statistical Association*, **82**, 1153-1162.

(6) Ghosh, M. and Meeden, G. (1986). Empirical Bayes Estimation in Finite Population Sampling. *Journal of the American Statistical Association*, **81**, 1058-1062.

(7) Ghosh, M., Nangia, N. and Kim, D.H. (1996). Estimation of Median Income of Four-Person Families: A Bayesian Time Series Approach. *Journal of the American Statistical Association*, **91**, 1423-1431.

(8) Hwang, J.T. (1985). Universal Domination and Stochastic Domination: Estimation Simultaneously under a Broad Class of Loss Functions. *The Annals of Statistics*, **13**, 295-314.

(9) Prasad, N.G.N. and Rao, J.N.K. (1990). On the Estimation of Mean Square Error of Small Area Predictors. *Journal of the American Statistical Association*, **85**, 163-171.

(10) Press, S.J. (1972). *Applied Multivariate Analysis.* Holt, Rinehart and Winston, New York.

(11) Rao, C.R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed. Wiley, New York.

(12) Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics.* Wiley, New York.