doi: 10.1111/j.1467-9469.2006.00492.x
 Vol 33: 591–608, 2006

# Empirical and Hierarchical Bayesian Estimation in Finite Population Sampling under Structural Measurement Error Models

MALAY GHOSH

*Department of Statistics, University of Florida*

KARABI SINHA

*Division of Epidemiology and Biostatistics, University of Illinois at Chicago*

DALHO KIM

*Department of Statistics Kyungpook National University*

**ABSTRACT. This paper considers simultaneous estimation of means from several strata. A model-based approach is taken, where the covariates in the superpopulation model are subject to measurement errors. Empirical Bayes (EB) and Hierarchical Bayes estimators of the strata means are developed and asymptotic optimality of EB estimators is proved. Their performances are examined and compared with that of the sample mean in a simulation study as well as in data analysis.**

*Key words:* asymptotic optimality, Bayes risk, consistency

## 1. Introduction

Empirical Bayes (EB) and Hierarchical Bayes (HB) estimators are widely used in these days for simultaneous estimation or prediction. The biggest advantage of these methods is their ability to enhance the precision of individual estimators by 'borrowing strength' from similar other estimators. In particular, these methods are very well suited for finite population sampling where the target is to estimate simultaneously several strata parameters, e.g. the strata means. This is especially so for small area estimation where each individual stratum often contains very few observations, and direct estimators are usually subject to large standard errors and coefficients of variation.

Ghosh & Meeden (1986) considered EB estimation of finite population strata means using a model-based approach. They used a simple one-way random effects ANOVA model for this purpose. The results can be extended by inclusion of covariates, and such procedures have been discussed in Ghosh & Meeden (1996).

Often, however, it is not possible to obtain exact measurements of these covariates. We provide a few examples to illustrate this.

**1** This example is taken from Fuller (1987, p. 2). Suppose, we want to predict the yield of corn in several counties in Iowa and the covariate used is available nitrogen in the soil. To estimate the available soil nitrogen, it is necessary to sample the soil of the experimental plot, and to perform a laboratory analysis on the selected sample. As a result of the sampling and of the laboratory analysis, we do not observe the true available nitrogen, but only its estimate.

**2** Suppose we have several strata obtained after stratifying by gender, ethnicity, age and region. We take measures on the blood pressure (bp), body weight (bw), body mass index

(BMI) and waist-hip ratio (WHR) for each subject within each stratum and we want to model the average bp in each stratum as a function of the other three variables. It seems likely that bw, BMI and WHR would be measured with error.

**3** Suppose we are looking at patients undergoing a rare surgical procedure in different hospitals. As this is a rare surgery, the number of observed cases in each hospital will be very small. Now, suppose, we are interested in modelling the time to recovery by bp, heart rate (hr) and other such measurements; then, again, it seems likely that the measured covariates are affected by some error.

**4** Suppose we are interested in estimating the volume of trees for several areas. We have data (for smaller subareas) on volume of stem wood for the current year and measures of the diameter and height of stem wood from a previous year. Here, we may model the volume of trees by their diameter and height. However, the latter two are likely to be measured with error.

All the above instances indicate why measurement error models are so suitable for simultaneous estimation of strata parameters.

In this study, we develop EB and HB procedures for simultaneous estimation of finite population strata means as well as the superpopulation parameters when the covariates, say $\boldsymbol{x}$, are measured with error. Our work is a natural extension of the work of Ghosh & Meeden (1986, 1996) in that we too consider a simple one-way random effects model (with a single covariate) with the additional assumption that the covariate is measured with error. We also assume that $\boldsymbol{x}$ is stochastic. In the common terminology (cf. Fuller, 1987, p. 2; Carroll *et al.*, 1995, p. 6) this is the so-called structural measurement error model. This is in contrast to the functional measurement error model where $\boldsymbol{x}$ is non-stochastic. The latter is considered in Ghosh & Sinha (2004).

EB estimators (or more appropriately predictors) for strata means are developed in section 2. This section also contains estimation of the superpopulation parameters taking into account the possibility that the covariates are measured with error. Section 3 proves the 'asymptotic optimality' of EB estimators in the sense of Robbins (1956). The HB estimators are developed in section 4. Also, in this section, we have established the propriety of the posteriors, and have discussed the Markov chain Monte Carlo (MCMC) implementation of the proposed hierarchical Bayes procedure. A simulation study is conducted in section 5 to compare the performances of the EB and HB estimators. Analysis of some real-life data is undertaken in section 6. Some concluding remarks are stated in section 7. The proofs of certain technical results are deferred to the appendix.

In a single stratum, Bolfarine & Zacks (1992) considered a measurement error model similar to ours. However, their main objective was Bayesian estimation of the superpopulation parameters. Also, they assumed the variance components to be known, and provided a normal approximation of the posterior distribution. Our aim was to estimate instead the strata population means. More importantly, the EB procedure as introduced here, estimates all the hyperparameters including all the variance components, and does not require any approximation of the posterior. The HB procedure also does not rely any normal approximation. Bolfarine & Sandoval (1990) provided Bayesian estimates of the finite population mean when the superpopulation mean is measured without error, and a certain variance ratio is known. Also, they used a non-informative prior which is different from ours.

## 2. EB estimators

Suppose there are $m$ strata labelled $1,\dots,m$ and let $N_i$ denote the known population size for the $i$th stratum. We denote by $y_{ij}$ the response of the $j$th unit in the $i$th stratum ($j=1,\dots,$

$N_i; i=1,\ldots,m$). A sample of size $n_i$ is drawn from the $i$th stratum. Without loss of generality, we denote the sampled units by $1,2,\ldots,n_i$ $(i=1,\ldots,m)$. Throughout, we will use the notations $\boldsymbol{y}_i^{(1)}=(y_{i1},\ldots,y_{in_i})^T$, $\boldsymbol{y}_i^{(2)}=(y_{in_{i+1}},\ldots,y_{iN_i})^T$, $\boldsymbol{y}_i^T=(\boldsymbol{y}_i{}^{(1)T},\boldsymbol{y}_i{}^{(2)T})$, $\boldsymbol{y}^{(1)T}=(\boldsymbol{y}_1{}^{(1)T},\ldots,\boldsymbol{y}_m{}^{(1)T})$, $\boldsymbol{y}^{(2)T}=(\boldsymbol{y}_1{}^{(2)T},\ldots,\boldsymbol{y}_m{}^{(2)T})$ and $\boldsymbol{y}^T=(\boldsymbol{y}^{(1)T},\boldsymbol{y}^{(2)T})$. The basic problem in finite population sampling is inference about $\boldsymbol{y}^{(2)}$ conditional on $\boldsymbol{y}^{(1)}$. More specifically, we are interested in the estimation (more appropriately prediction) of finite population means

$$\gamma_i=N_i^{-1}\sum_{j=1}^{N_i}y_{ij}\quad(i=1,\ldots,m)$$

given the data.

We assume the superpopulation model

$$y_{ij}=b_0+b_1x_i+u_i+e_{ij}\quad(j=1,\ldots,N_i;i=1,\ldots,m),\tag{1}$$

$$X_{ij}=x_i+\eta_{ij}\quad(j=1,\ldots,N_i;i=1,\ldots,m).\tag{2}$$

It is assumed that the $x_i,u_i,e_{ij}$ and $\eta_{ij}$ are mutually independent with $x_i\sim N(\mu_x,\sigma_x^2)$, $u_i\sim N(0,\sigma_u^2)$, $e_{ij}\sim N(0,\sigma_e^2)$ and $\eta_{ij}\sim N(0,\sigma_\eta^2)$. The available data consist of $(y_{ij},X_{ij}),(j=1,\ldots,n_i;$ $i=1,\ldots,m)$. Also, we write $\boldsymbol{\phi}=(b_0,b_1,\mu_x,\sigma_u^2,\sigma_e^2,\sigma_\eta^2,\sigma_x^2)^T$.

Clearly (1) is a random effects model. An alternative way of expressing the same is

$$y_{ij}=\theta_i+e_{ij};\theta_i=b_0+b_1x_i+u_i,\quad j=1,\ldots,N_i,\ \ i=1,\ldots,m.$$

In this way, it is possible to identify (1) as a Bayesian model. Throughout this study, we use the Bayesian terminology, although the EB estimators to be developed in this section can also be viewed as empirical best linear unbiased predictors.

Now, writing $\mathbf{1}_{n_i}$ as the $n_i$ dimensional column vector of $\mathbf{1}$, $\boldsymbol{J}_{n_i}=\mathbf{1}_{n_i}\mathbf{1}_{n_i}^T$ and $\boldsymbol{I}_{n_i}$ as the identity matrix of order $n_i$, $(\boldsymbol{y}_i^{(1)}\boldsymbol{y}_i^{(2)}|\boldsymbol{\phi})^T$ follows bivariate normal distribution with parameters

$$\left[\begin{pmatrix}(b_0+b_1x_i)\mathbf{1}_{n_i}\\(b_0+b_1x_i)\mathbf{1}_{N_i-n_i}\end{pmatrix},\begin{pmatrix}\sigma_e^2\boldsymbol{I}_{n_i}+(\sigma_u^2+b_1^2\sigma_x^2)\boldsymbol{J}_{n_i} & (\sigma_u^2+b_1^2\sigma_x^2)\mathbf{1}_{n_i}\mathbf{1}_{N_i-n_i}^T\\(\sigma_u^2+b_1^2\sigma_x^2)\mathbf{1}_{N_i-n_i}\mathbf{1}_{n_i}^T & \sigma_e^2\boldsymbol{I}_{N_i-n_i}+(\sigma_u^2+b_1^2\sigma_x^2)\boldsymbol{J}_{N_i-n_i}\end{pmatrix}\right].\tag{3}$$

Then the Bayes predictor of $\boldsymbol{y}_i^{(2)}$ given $\boldsymbol{y}_i^{(1)}$ and $\phi$ reduces after some simplification to

$$\begin{aligned}E(\boldsymbol{y}_i^{(2)}|\boldsymbol{y}_i^{(1)},\phi)&=(b_0+b_1\mu_x)\mathbf{1}_{N_i-n_i}+(\sigma_u^2+b_1^2\sigma_x^2)\mathbf{1}_{N_i-n_i}\mathbf{1}_{n_i}^T\\&\quad\times[\sigma_e^2\boldsymbol{I}_{n_i}+(\sigma_u^2+b_1^2\sigma_x^2)\boldsymbol{J}_{n_i}]^{-1}[\boldsymbol{y}_i^{(1)}-(b_0+b_1\mu_x)\mathbf{1}_{n_i}]\\&=[(1-B_i)\bar{y}_i^{(1)}+B_i(b_0+b_1\mu_x)]\mathbf{1}_{N_i-n_i},\end{aligned}\tag{4}$$

where

$$\bar{y}_i^{(1)}=n_i^{-1}\sum_{j=1}^{n_i}y_{ij}\quad\text{and}\quad B_i=\frac{\sigma_e^2}{\sigma_e^2+n_i(\sigma_u^2+b_1^2\sigma_x^2)}\ \ (i=1,\ldots,m).$$

The above predictor can be viewed also as the best linear unbiased predictor of $\boldsymbol{y}_i^{(2)}$ given $\boldsymbol{y}_i^{(1)}$.

This leads to the Bayes predictor of $\gamma_i$ as

$$\hat{\gamma}_i^B=E[\gamma_i|\boldsymbol{y}_i^{(1)}]=(1-f_iB_i)\bar{y}_i^{(1)}+f_iB_i(b_0+b_1\mu_x),\tag{5}$$

where $f_i=(N_i-n_i)/N_i$ is the finite population correction fraction. For simplicity, henceforth, we will denote $\bar{y}_i^{(1)}$ by $\bar{y}_i$.

In an EB scenario, the components of $\phi$ are unknown and need to be estimated from the data. We write

$$\bar{X}_i = n_i^{-1}\sum_{j=1}^{n_i} X_{ij}, \quad \mathrm{SSW}_X = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_i)^2, \quad \mathrm{SSW}_y = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2.$$

Then the minimal sufficient statistic is $(\bar{y}_1,\ldots,\bar{y}_m,\mathrm{SSW}_y,\bar{X}_1,\ldots,\bar{X}_m,\mathrm{SSW}_X)$. Also, $\bar{y}_i$ are independent $N(b_1\mu_x,\sigma_e^2/n_i+\sigma_u^2+b_1^2\sigma_x^2)$, $\bar{X}_i$ are independent $N(\mu_x,\sigma_\eta^2/n_i+\sigma_x^2)$, $\mathrm{SSW}_y\sim\sigma_e^2\chi^2_{n_T-m}$ and $\mathrm{SSW}_X\sim\sigma_\eta^2\chi^2_{n_T-m}$, where $n_T=\sum_{i=1}^m n_i$ is assumed to be bigger than $m$. A simple-minded initial estimator of $b_1$ is given by

$$\tilde{b}_1 = \frac{\sum_{i=1}^{m} n_i\bar{y}_i(\bar{X}_i-\bar{X})}{\sum_{i=1}^{m} n_i(\bar{X}_i-\bar{X})^2}.$$

The following theorem shows that $\tilde{b}_1$ is typically an inconsistent estimator of $b_1$. Let

$$g_m = n_T - n_T^{-1}\sum_{i=1}^{m} n_i^2.$$

**Theorem 1**

*Assume (i)* $\max_{1\le i\le x} n_i \le K<\infty$ *and (ii)* $g_m/(m-1)\to c$ *as* $m\to\infty$. *Then* $E(\tilde{b}_1)\to b_1 c\sigma_x^2/(\sigma_\eta^2+c\sigma_x^2)$.

*Remark 1.* Assumption (i) is very natural in a small area context. Assumption (ii) also holds when the $n_i$ do not differ significantly. In particular, when $n_1=\cdots=n_m=n$, then $g_m/(m-1)=n$. Also, due to assumptions (i) and (ii), $n_T/m\to c$ as $m\to\infty$.

The proof of the theorem is deferred to the appendix. The theorem says that the regression coefficient $\tilde{b}_1$ converges in probability to a fraction multiplier of $b_1$. Thus, the regression coefficient is attenuated by measurement error. The theorem also implies that a consistent estimator of $b_1$ is obtained when one multiplies $\tilde{b}_1$ by a consistent estimator of $(\sigma_\eta^2+c\sigma_x^2)/c\sigma_x^2$.

To this end, we observe that as $\mathrm{SSW}_X\sim\sigma_\eta^2\chi^2_{n_T-m}$, writing $\mathrm{MSW}_X=\mathrm{SSW}_X/(n_T-m)$, $E(\mathrm{MSW}_X)=\sigma_\eta^2$ and $V(\mathrm{MSW}_X)=2\sigma_\eta^4/(n_T-m)=2[\sigma_\eta^4/m]\cdot[m/(n_T-m)]\to 0$ as $m\to\infty$ by remark 1. Hence, $\mathrm{MSW}_X\xrightarrow{P}\sigma_\eta^2$ as $m\to\infty$. Next, writing

$$\mathrm{SSB}_X = \sum_{i=1}^{m} n_i(\bar{X}_i-\bar{X})^2,$$

one can write

$$\mathrm{SSB}_X = \boldsymbol{Z}_m^T(\boldsymbol{I}_m-\boldsymbol{u}_m\boldsymbol{u}_m^T)\boldsymbol{Z}_m,$$

where

$$\boldsymbol{Z}_m^T=(\sqrt{n_1}\bar{X}_1,\ldots,\sqrt{n_m}\bar{X}_m), \quad \boldsymbol{u}_m^T=\left(\frac{\sqrt{n_1}}{\sqrt{n_T}},\ldots,\frac{\sqrt{n_m}}{\sqrt{n_T}}\right)$$

and $\boldsymbol{I}_m$ is the identity matrix of order $m$. Now, conditional on $\boldsymbol{x}=(x_1,\ldots,x_m)^T$, $\boldsymbol{Z}_m$ has the mean vector $(\sqrt{n_1}x_1,\ldots,\sqrt{n_m}x_m)^T$ and the variance–covariance matrix $\sigma_\eta^2\boldsymbol{I}_m$. By the symmetry and idempotency of $\boldsymbol{I}_m-\boldsymbol{u}_m\boldsymbol{u}_m^T$, $\mathrm{SSB}_X|\boldsymbol{x}\sim\sigma_\eta^2\chi^2_{m-1}(\xi_m)$, where

$$\xi_m = \frac{1}{2\sigma_\eta^2}(\sqrt{n_1}x_1, \dots, \sqrt{n_m}x_m)^T(\boldsymbol{I}_m - \boldsymbol{u}_m\boldsymbol{u}_m^T)(\sqrt{n_1}x_1, \dots, \sqrt{n_m}x_m)$$
$$= \frac{1}{2\sigma_\eta^2}\sum_1^m n_i(x_i - \bar{x})^2,$$

we now state the following lemma, the proof of which also is deferred to the appendix.

**Lemma 1**
*Assume conditions (i) and (ii) of theorem 1. Then*

$$E\left[\sum_{i=1}^m \frac{n_i(x_i-\bar{x})^2}{(m-1)}\right] \to c\sigma_x^2 \quad \textit{and} \quad V\left[\sum_{i=1}^m n_i(x_i-\bar{x})^2/(m-1)\right] \to 0.$$

In view of lemma 1, writing

$$\mathrm{MSB}_X = \frac{\mathrm{SSB}_X}{(m-1)},$$

$$E(\mathrm{MSB}_X) = EE(\mathrm{MSB}_X|\boldsymbol{x})$$
$$= \sigma_\eta^2 \frac{\left[m-1+\sum_{i=1}^m n_i(x_i-\bar{x})^2/\sigma_\eta^2\right]}{(m-1)} \to \sigma_\eta^2 + c\sigma_x^2 \quad \text{as } m\to\infty,$$

while

$$V(\mathrm{MSB}_X) = E[V(\mathrm{MSB}_X|\boldsymbol{x})] + V[E(\mathrm{MSB}_X|\boldsymbol{x})]$$
$$= \sigma_\eta^4 \frac{E\left[2(m-1)+4\sum_{i=1}^m n_i(x_i-\bar{x})^2/\sigma_\eta^2\right]}{(m-1)^2}$$
$$+ \sigma_\eta^4 V\left[1 + \sum_{i=1}^m \frac{n_i(x_i-\bar{x})^2}{(m-1)}\right] \to 0+0=0 \quad \text{as } m\to\infty.$$

Thus, under the assumptions of theorem 1, as $m\to\infty$, $(\sigma_\eta^2 + c\sigma_x^2)/c\sigma_x^2$ is consistently estimated by $\mathrm{MSB}_X/(\mathrm{MSB}_X - \mathrm{MSW}_X) = (1-\mathrm{MSW}_X/\mathrm{MSB}_X)^{-1}$. Thus, $b_1$ and $b_0$ are consistently estimated by

$$\hat{b}_1 = (1-\mathrm{MSW}_X/\mathrm{MSB}_X)^{-1}\tilde{b}_1, \quad \hat{b}_0 = \bar{y} - \tilde{b}_0\bar{X}. \tag{6}$$

Next, we need consistent estimators of the $B_i$, where $B_i$ is defined after (4). First, writing

$$\mathrm{MSW}_y = \mathrm{SSW}_y/(n_T-m), E(\mathrm{MSW}_y) = \sigma_e^2 \quad \text{and} \quad V(\mathrm{MSW}_y) = \frac{2\sigma_e^4}{(n_T-m)} \to 0 \quad \text{as } m\to\infty.$$

Thus, $\sigma_e^2$ is consistently estimated by $\mathrm{MSE}_y$ as $m\to\infty$. Next, defining

$$\mathrm{MSB}_y = \sum_{i=1}^m \frac{n_i(\bar{y}_i-\bar{y})^2}{(m-1)},$$

we calculate

$$E(\mathrm{MSB}_y) = E[E(\mathrm{MSB}_y|\boldsymbol{x})]$$
$$= E\left[\sigma_e^2 + b_1^2\sum_1^m \frac{n_i(x_i-\bar{x})^2}{(m-1)} + \sigma_u^2\frac{g_m}{(m-1)}\right] \to \sigma_e^2 + c(b_1^2\sigma_x^2 + \sigma_u^2), \tag{7}$$

by lemma 1. We will prove next the following theorem.

**Theorem 2**

*Assume conditions (i) and (ii) of theorem 1. Then $V(MSB_y) \to 0$ as $m \to \infty$.*

The proof of this theorem is deferred to the appendix. In view of this theorem and (7), we estimate $b_1^2\sigma_x^2 + \sigma_u^2 = \zeta$ (say) consistently by $\hat{\zeta}_m = \max[0, (\text{MSB}_y - \text{MSW}_y)(m-1)/g_m]$. The introduction of 0 is simply to overcome the fact that $\text{MSB}_y - \text{MSW}_y$ can assume negative values with positive probability. Now $B_i$ is estimated consistently by

$$\hat{B}_i = \frac{\text{MSW}_y}{(\text{MSW}_y + n_i\hat{\zeta}_m)}, \quad (i=1,\ldots,m). \tag{8}$$

The EB predictor of $\gamma_i$ is thus given by

$$\hat{\gamma}_i^{\text{EB}} = (1 - f_i\hat{B}_i)\bar{y}_i + f_i\hat{B}_i(\hat{b}_0 + \hat{b}_1\bar{X}) = (1 - f_i\hat{B}_i)\bar{y}_i + f_i\hat{B}_i\bar{y}, \quad i=1,\ldots,m. \tag{9}$$

*Remark 2.* It may be noted that with the present method of moments estimators of the superpopulation and prior parameters, the EB estimator of $\boldsymbol{\gamma}$ is the same as in the situation when the covariates are measured without error. One may view this as a deficiency of the proposed EB estimator, but it seems possible to overcome this by considering the MLEs of $b_0$, $b_1$ and the variance components. However, then one may have to give up direct analytic evaluations, and rely instead only on numerical findings. Also (9) demonstrates the robustness of the EB estimators developed earlier with known covariates.

## 3. Asymptotic optimality of the EB predictor

We first compute the Bayes risk of the EB predictor $\hat{\boldsymbol{\gamma}}^{\text{EB}} = (\hat{\gamma}_1^{\text{EB}}, \ldots, \hat{\gamma}_m^{\text{EB}})^T$ of $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)^T$, the vector of population strata means, i.e. we compute $m^{-1}\sum_{i=1}^m E(\hat{\gamma}_i^{\text{EB}} - \gamma_i)^2$. For this, we begin with the identity

$$m^{-1}\sum_{i=1}^m E(\hat{\gamma}_i^{\text{EB}} - \gamma_i)^2 = m^{-1}\sum_{i=1}^m E(\hat{\gamma}_i^B - \gamma_i)^2 + m^{-1}\sum_{i=1}^m E(\hat{\gamma}_i^{\text{EB}} - \hat{\gamma}_i^B)^2, \tag{10}$$

which holds as

$$E[(\hat{\gamma}_i^{\text{EB}} - \hat{\gamma}_i^B)(\hat{\gamma}_i^B - \gamma_i)] = E[(\hat{\gamma}_i^{\text{EB}} - \hat{\gamma}_i^B)E(\hat{\gamma}_i^B - \gamma_i|\boldsymbol{y})] = 0.$$

We first note that

$$m^{-1}\sum_{i=1}^m E(\hat{\gamma}_i^B - \gamma_i)^2 = m^{-1}\sum_{i=1}^m V(\gamma_i|\boldsymbol{y}_1^{(1)}) = m^{-1}\sum_{i=1}^m N_i^{-2}V(\mathbf{1}_{N_i-n_i}^T\boldsymbol{y}_i^{(2)}|\boldsymbol{y}_i^{(1)}). \tag{11}$$

Straightforward but tedious calculations yield

$$V(\boldsymbol{y}_i^{(2)}|\boldsymbol{y}_i^{(1)}) = \frac{\sigma_e^2\boldsymbol{I}_{N_i-n_i} + \sigma_e^2(\sigma_u^2 + b_1^2\sigma_x^2)}{[\sigma_e^2 + n_i(\sigma_u^2 + b_1^2\sigma_x^2)]\boldsymbol{J}_{N_i-n_i}}. \tag{12}$$

Hence, from (10) and (11),

$$m^{-1}\sum_{i=1}^m E(\hat{\gamma}_i^B - \gamma_i)^2 = m^{-1}\sum_{i=1}^m N_i^{-2}\left[\frac{(N_i - n_i)\sigma_e^2 + (N_i - n_i)^2\sigma_e^2(\sigma_u^2 + b_1^2\sigma_x^2)}{\sigma_e^2 + n_i(\sigma_u^2 + b_1^2\sigma_x^2)}\right]$$

$$= m^{-1}\sigma_e^2\sum_{i=1}^m f_i\left[\frac{N_i^{-1} + f_i(\sigma_u^2 + b_1^2\sigma_x^2)}{\sigma_e^2 + n_i(\sigma_u^2 + b_1^2\sigma_x^2)}\right]. \tag{13}$$

This leads to

$$m^{-1}\sum_{i=1}^{m} E(\hat{\gamma}_i^{\mathrm{EB}}-\gamma_i)^2 = m^{-1}\sigma_e^2\sum_{i=1}^{m} f_i\left[\frac{N_i^{-1}+f_i(\sigma_u^2+b_1^2\sigma_x^2)}{\sigma_e^2+n_i(\sigma_u^2+b_1^2\sigma_x^2)}\right] + m^{-1}\sum_{i=1}^{m} E(\hat{\gamma}_i^{\mathrm{EB}}-\hat{\gamma}_i^{B})^2. \tag{14}$$

We will next prove the following theorem which establishes the asymptotic optimality (cf. Robbins, 1956) of EB estimators under certain conditions.

**Theorem 3**

*Assume conditions (i) and (ii) of theorem 1. Then* $m^{-1}\sum_{i=1}^{m} E(\hat{\gamma}_i^{EB}-\hat{\gamma}_i^{B})^2 \to 0$ *as* $m \to \infty$.

The proof of this theorem is deferred to the appendix.

## 4. HB predictors

Next we consider a hierarchical Bayesian framework to predict the population strata means $\gamma_i\,(i=1,\ldots,m)$. To this end, we begin with the following model:

Stage 1. $y_{ij}=\theta_i+e_{ij}\,(j=1,\ldots,n_i; i=1,\ldots,m)$ where $e_{ij}$ are i.i.d. $N(0,\sigma_e^2)$.

Stage 2. $\theta_i=b_0+b_1x_i+u_i\,(i=1,\ldots,m)$ where $u_i$ are i.i.d. $N(0,\sigma_u^2)$ $X_{ij}=x_i+\eta_{ij}\,(j=1,\ldots,n_i; i=1,\ldots,m)$, where $\eta_{ij}$ are i.i.d. $N(0,\sigma_\eta^2)$.

Stage 3. $x_i \sim N(\mu_x,\sigma_x^2)$

Stage 4. $b_0, b_1, \mu_x, \sigma_e^2, \sigma_u^2, \sigma_x^2, \sigma_\eta^2$ are mutually independent with $b_0, b_1$ and $\mu_x$ i.i.d. uniform$(-\infty,\infty)$; $\sigma_e^2 \sim IG(a_e/2, b_e/2)$, $\sigma_u^2 \sim IG(a_u/2, b_u/2)$, $\sigma_\eta^2 \sim IG(a_\eta/2, b_\eta/2)$, $\sigma_x^2 \sim IG(a_x/2, b_x/2)$, where $IG(\alpha,\beta)$ denotes an inverse gamma distribution with pdf $f_{\alpha,\beta}(z) \propto \exp(-\alpha/z)z^{-\beta-1}I_{[z>0]}$.

First check the propriety of the posterior under the given prior. The following theorem is proved. We will write $\boldsymbol{b}=(b_0,b_1)^T$; $\boldsymbol{z}_i^T=(1,x_i)$.

**Theorem 4**

*Assume* $a_e, a_u, a_\eta, a_x$ *all positive. Also, let* $b_e+n_T-m>0$, $b_u+m-p>0$ *and* $b_x+m-1>0$. *Then the joint posterior is proper.*

The proof of the theorem is deferred to the appendix.

The implementation of the Bayesian procedure is greatly facilitated by the MCMC numerical integration technique, in particular the Gibbs sampler. This requires generating samples from the full conditionals of each of $\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{b}, \mu_x, \sigma_e^2, \sigma_u^2, \sigma_\eta^2$ and $\sigma_x^2$ given the remaining parameters and the data. The details are given below.

(i) $[\theta_i|\boldsymbol{x},\boldsymbol{b},\mu_x,\sigma_e^2,\sigma_u^2,\sigma_\eta^2,\sigma_x^2,\boldsymbol{y},\boldsymbol{X}]$ are independently $N((1-B_i)\bar{y}_i+B_i\boldsymbol{x}_i^T\boldsymbol{b},(\sigma_e^2/n_i)(1-B_i))$, where $B_i=(\sigma_e^2/n_i)/(\sigma_e^2/n_i+\sigma_u^2)$;

(ii) $[x_i|\boldsymbol{\theta},\boldsymbol{b},\mu_x,\sigma_e^2,\sigma_u^2,\sigma_\eta^2,\sigma_x^2,\boldsymbol{y},\boldsymbol{X}]$ are independently $N((\sigma_u^{-2}b_1^2+\sigma_\eta^{-2}n_i+\sigma_x^{-2})^{-1}(\sigma_u^{-2}b_1(\theta_i-b_0)+\sigma_\eta^{-2}n_i x_i+\sigma_x^{-2}\mu_x),\sigma_u^{-2}b_1^2+\sigma_\eta^{-2}n_i+\sigma_x^{-2})$;

(iii) $[\boldsymbol{b}|\boldsymbol{\theta},\boldsymbol{x},\mu_x,\sigma_e^2,\sigma_u^2,\sigma_\eta^2,\sigma_x^2,\boldsymbol{y},\boldsymbol{X}] \sim N((\boldsymbol{X}_*^T\boldsymbol{X}_*)^{-1}\boldsymbol{X}_*^T\boldsymbol{\theta},\sigma_u^2(\boldsymbol{X}_*^T\boldsymbol{X}_*)^{-1})$;

(iv) $[\mu_x|\boldsymbol{\theta},\boldsymbol{b},\boldsymbol{x},\sigma_e^2,\sigma_u^2,\sigma_\eta^2,\sigma_x^2,\boldsymbol{y},\boldsymbol{X}] \sim N(\bar{x},\sigma_x^2/m)$;

(v) $[\sigma_e^2|\boldsymbol{\theta},\boldsymbol{b},\boldsymbol{x},\mu_x,\sigma_u^2,\sigma_\eta^2,\sigma_x^2,\boldsymbol{y},\boldsymbol{X}] \sim \mathrm{IG}((1/2)(n_T+b_e),(1/2)(\sum_{i=1}^m n_i(\bar{y}_i-\theta_i)^2+\mathrm{SSW}_y+a_e))$;

(vi) $[\sigma_u^2|\boldsymbol{\theta},\boldsymbol{b},\boldsymbol{x},\mu_x,\sigma_e^2,\sigma_\eta^2,\sigma_x^2,\boldsymbol{y},\boldsymbol{X}] \sim \mathrm{IG}((1/2)(m+b_u),(1/2)(\sum_{i=1}^m(\theta_i-z_i^T\boldsymbol{b})^2+a_u)$;

(vii) $[\sigma_\eta^2|\boldsymbol{\theta},\boldsymbol{b},\boldsymbol{x},\mu_x,\sigma_e^2,\sigma_u^2,\sigma_x^2,\boldsymbol{y},\boldsymbol{X}] \sim \mathrm{IG}((1/2)(n_T+b_\eta),(1/2)(\sum_{i=1}^m n_i(\bar{X}_i-x_i)^2+a_\eta)$;

(viii) $[\sigma_x^2|\boldsymbol{\theta},\boldsymbol{b},\boldsymbol{x},\mu_x,\sigma_e^2,\sigma_u^2,\sigma_\eta^2,\boldsymbol{y},\boldsymbol{X}] \sim \mathrm{IG}((1/2)(m+b_x),(1/2)(\sum_{i=1}^m(x_i-\mu_x)^2+a_x))$.

We generate several sets of these samples. For the $G$th generated set, we obtain the HB estimate: $\hat{\gamma}_i^{\text{HB}} = (1 - f_i B_{iG})\bar{y}_i + f_i B_{iG}(b_{0G} + b_{1G}\mu_{xG})$ and the corresponding Bayes risk: $(\sigma_{eG}^2)(f_i/N_i + f_i(1 - B_{iG})/n_i)$. After burning out the first half (to eliminate any possible instability in the initial generated samples), we use the averaging principle and take the average of the HB estimates over all the remaining sets to obtain the final HB estimate. The same method is applied to calculate the Bayes risk.

## 5. Simulation study

We conducted a simulation study to compare the performance of the HB and the EB estimators in comparison with the sample mean. To this end, we created a finite population of size 1400 spread across 12 strata of sizes 50, 250, 50, 100, 200, 150, 50, 150, 100, 150, 100 and 50. We considered several starting values of the parameters $b_0, b_1, \sigma_e^2, \sigma_u^2, \sigma_\eta^2, \sigma_x^2,$ and obtained similar results. Here, we reported the results when the responses $y_{ij}$ are generated under the superpopulation model as considered in this study and with $b_0 = 100$, $b_1 = 2$, $\sigma_e^2 = 100$, $\sigma_u^2 = 16$, $\sigma_\eta^2 = 25$, $\mu_x = 194$ and $\sigma_x^2 = 2,737$. A 2% simple random sample was used to generate samples from each stratum. Accordingly, the sample sizes for the 12 strata are given, respectively, by 1, 5, 1, 2, 4, 3, 1, 3, 2, 3, 2 and 1.

We drew 400 independent samples $(y_{ij}, X_{ij})$ $(j = 1, \ldots, n_i; i = 1, \ldots, 12)$ from this population, and found for each sample the sample mean, the EB and the HB estimators. To obtain the HB estimators, we ran a Gibbs chain of size 10,000 with a burn-in of the first 5000. The HB estimators of the population means $\gamma_i$ are the averages over the remaining 5000 Gibbs samples generated. Moreover, we took the average of the squared differences of the estimators from the true means (TM) over the 400 simulations and took their squared roots to obtain the root mean squared errors (RMSE). The resulting values for the sample means (SM), EB and HB estimators are reported as RMSE(SM), RMSE(EB) and RMSE(HB) respectively. We considered uniform $(-\infty, \infty)$ priors for $b_0, b_1$ and $\mu_x$. We also considered several choices of $a_e, b_e, a_u, b_u, a_\eta, b_\eta, a_x, b_x,$ the parameters of the inverse gamma distributions of the variance components. We report here the results when they are all equal to 0.002.

Table 1 reports the sample sizes, the TM, the SM, the EB and the HB estimators as well as RMSE(SM), RMSE(EB) and RMSE(HB) for the 12 counties. The counties are denoted by $i = 1, \ldots, 12$ in the table. It follows from the above table that according to the RMSE

Table 1. *The sample sizes, means and RMSEs for the 12 counties*

| $i$ | $n_i$ | TM | SM | EB | HB | RMSE(SM) | RMSE(EB) | RMSE(HB) |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 127.32 | 127.16 | 130.14 | 131.90 | 8.76 | 6.99 | 6.88 |
| 2 | 5 | 117.41 | 117.25 | 118.99 | 121.53 | 4.46 | 4.77 | 4.13 |
| 3 | 1 | 152.10 | 152.12 | 147.32 | 145.30 | 9.01 | 8.73 | 8.24 |
| 4 | 2 | 160.24 | 160.05 | 155.54 | 151.39 | 6.82 | 7.88 | 10.62 |
| 5 | 4 | 127.41 | 127.57 | 128.57 | 130.75 | 4.61 | 4.43 | 5.08 |
| 6 | 3 | 140.51 | 140.62 | 140.11 | 139.89 | 5.20 | 4.61 | 3.92 |
| 7 | 1 | 141.34 | 141.15 | 139.77 | 139.59 | 9.95 | 7.21 | 5.65 |
| 8 | 3 | 155.97 | 156.40 | 153.69 | 151.47 | 5.77 | 6.01 | 6.94 |
| 9 | 2 | 131.61 | 131.46 | 132.50 | 134.50 | 6.59 | 5.48 | 5.25 |
| 10 | 3 | 137.18 | 136.21 | 136.29 | 136.82 | 5.58 | 4.93 | 4.09 |
| 11 | 2 | 148.14 | 148.01 | 145.86 | 143.94 | 6.62 | 6.16 | 6.46 |
| 12 | 1 | 133.43 | 133.99 | 134.92 | 135.58 | 8.82 | 6.45 | 5.27 |

TM, true means; SM, sample means; RMSE, root mean squared error; EB, empirical Bayes; HB, hierarchical Bayes.

criterion, the HB estimator is doing better than both the sample mean and the EB estimator in eight of the 12 counties. In most of these counties, the EB estimator is second best and the sample mean is the worse, having the largest RMSE. The exceptions are counties 4, 5 and 8. The sample mean has the smallest RMSE in counties 4 and 8 while the EB estimator has the smallest RMSE in counties 5 and 11. Thus, based on our simulation, it appears that the HB method has overall better performance than the other two estimators.

In the following section, we conduct a data analysis and compare the performance of the HB estimator with that of other standard estimators.

## 6. Data analysis

We use the data used by Battese *et al.* (1988), hereafter BHF, for analysis. Knowledge of the area under different crops is important to the US Department of Agriculture (USDA). Sample surveys have designed to estimate crop areas for large regions, such as crop-reporting districts, individual states, and the USA as a whole. Predicting crop areas for small areas such as counties has generally not been attempted, due to a lack of availability of data from farm surveys for these areas. The use of satellite data in association with farm-land survey observations has been the subject of considerable research over the years. In their paper, BHF considered data for 12 counties in Iowa, obtained from the 1978 June Enumerative Survey of the USDA as well as from the satellite LANDSAT during the 1978 growing season. The purpose was to predict the area under soya bean and corn in these counties. As we have discussed before, BHF developed a variance components model for small area estimation and they provided analysis of the soya bean data (reported by farmers) using two covariates, corn and soya bean (reported by satellite). The actual data are provided in BHF (1988).

We consider prediction of soya bean data using soya bean pixels only as covariates. BHF observed that based on the p-values of the slopes in their model, ‘only the coefficient of soya bean pixels is significantly different from 0 for the soya bean function’. This motivated us to develop our model and data analysis taking only one covariate viz, soya bean pixels for prediction of soya bean hectares.

In order to make a valid comparison between our procedure and the one due to BHF, we consider a variation of their model in the sense that we incorporate possible measurement errors in the values of the covariates. Thus, before we apply any method towards analysing the data, we note that the $x$-observations are random due to measurement error. We generate ‘copies’ of these values, and then apply both our method and the one due to BHF in finding the predicted hectares. The procedure is repeated over and over again and finally the estimates and their standard errors are computed using the averaging principle.

Table 2 provides the predicted hectares and estimated standard errors of the predicted hectares for each county according to the HB procedure and compare the same with those of the sample mean and BHF. It appears from the table that in the presence of measurement errors, the proposed approach of this paper results in a RMSE reduction in the 9–26% range for all the counties. Thus, based on our data analysis as well, we recommend the HB method for simultaneous of strata means when the covariates are subject to measurement error. Both our method and the BHF method outperform the direct estimates, namely, the SM.

## 7. Summary and conclusion

This paper initiates the study of EB and HB estimators of strata means when the covariates are potentially measured with error. As revealed in our simulation study as well as data analysis, the HB estimators turn out to be the winner according to the RMSE criterion. We want

Table 2. *Predicted hectares of soya bean (HB and BHF) with corresponding standard errors, using soya bean pixels as the only covariate*

| County | Sample segments | Predicted hectares (BHF) | Predicted hectares (HB) | Standard errors | | |
|---|---|---|---|---|---|---|
| | | | | BHF | HB | Sample mean |
| Cerro Gordo | 1 | 52.8 | 32.1 | 22.5 | 20.5 | 29.1 |
| Hamilton | 1 | 97.3 | 100.9 | 22.5 | 20.5 | 29.1 |
| Worth | 1 | 98.0 | 99.2 | 22.5 | 20.5 | 29.1 |
| Humboldt | 2 | 57.1 | 44.7 | 18.6 | 15.8 | 20.6 |
| Franklin | 3 | 63.0 | 57.1 | 16.0 | 13.3 | 16.8 |
| Pocahontas | 3 | 111.4 | 114.9 | 16.0 | 13.3 | 16.8 |
| Winnebago | 3 | 88.8 | 88.6 | 16.0 | 13.3 | 16.8 |
| Wright | 3 | 95.6 | 96.6 | 16.1 | 13.3 | 16.8 |
| Webster | 4 | 107.2 | 110.6 | 14.3 | 11.7 | 14.6 |
| Hancock | 5 | 113.4 | 115.1 | 13.0 | 10.6 | 13.0 |
| Kossuth | 5 | 113.1 | 116.9 | 13.1 | 10.6 | 13.0 |
| Hardin | 5 | 99.5 | 89.7 | 12.9 | 10.6 | 13.0 |

HB, hierarchical Bayes; BHF, Battese, Harter & Fuller (1988).

to pursue further the HB analysis in our future work. In particular, we plan to develop HB estimators in multiple regression superpopulation models where the covariates are subject to measurement error.

We have shown the convergence of the Bayes risk of the EB estimator to the optimal Bayes risk in theorem 3. This shows the first order correctness of the EB estimator. A natural extension is to expand the Bayes risk of the EB estimator which is correct up to order $m^{-1}$ as in Prasad & Rao (1990). This requires Taylor expansion of the Bayes risk $m^{-1}\sum_{i=1}^{m} E(\hat{\gamma}_i^{\mathrm{EB}} - \gamma_i)^2$.

The HB approach provides credible intervals for the $\gamma_i$ based on the MCMC generated samples. In contrast, there is no direct way of constructing confidence intervals based on the EB estimators. As is well known, the naïve plug-in approach, namely $\hat{v}_i = V(\gamma_i|\boldsymbol{y}_i^{(1)}, \hat{\phi})$ are typically underestimates and consequently, confidence intervals of the form $\hat{\gamma}_i^{\mathrm{EB}} \pm z_{\alpha/2}$, where $z_{\alpha/2}$ is the upper $100(1-\alpha)\%$ point of the standard normal curve, have coverage probabilities falling short of the target $(1-\alpha)$. This can be partially rectified by finding MSEs of $\hat{\gamma}_i^{\mathrm{EB}}$ which are correct up to $O(m^{-1})$, but even then there is no guarantee that confidence intervals $\hat{\gamma}_i^{\mathrm{EB}} \pm z_{\alpha/2}\sqrt{\mathrm{MSE}(\hat{\gamma}_i^{\mathrm{EB}})}$ will meet the target because of the skewness of the unknown distribution of $\hat{\gamma}_i^{\mathrm{EB}}$.

## Acknowledgements

The authors are very grateful to the Associate Editor and two referees for their helpful comments on an earlier version of this paper. They would also like to acknowledge partial support of their research to NSF grants.

## References

Battese, G., Harter, R. & Fuller, W. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *J. Amer. Statist. Assoc.* **83**, 28–36.

Bolfarine, H. & Sandoval, M. C. (1990). *Finite population prediction under error in variables models with known reliability ratio*. Technical Report. Department of Statistics, University of Sao Paulo, Sao Paulo.

Bolfarine, H. & Zacks, S. (1992). *Prediction theory for finite populations*. Springer-Verlag, New York, NY.

Carroll, R. J., Ruppert, D. & Stefanski, L. A. (1995). *Measurement error in nonlinear models*. Chapman & Hall, New York, NY.

Fuller, W. (1987). *Measurement error models*. Wiley, New York, NY.

Ghosh, M. & Meeden, G. (1986). Empirical Bayes estimation in finite population sampling. *J. Amer. Statist. Assoc.* **81**, 1058–1062.
Ghosh, M. & Meeden, G. (1996). *Bayesian methods for finite population sampling*. Chapman & Hall, New York, NY.
Ghosh, M. & Sinha, K. (2004). *Empirical Bayes estimation in finite population sampling under functional measurement error models*. Technical Report No. 7. Department of Statistics, University of Florida, Gainesville, FL.
Prasad, N. G. N. & Rao, J. N. K. (1990). The estimation of mean squared errors of small area estimators. *J. Amer. Statist. Assoc.* **85**, 163–171.
Robbins, H. (1956). An empirical Bayes approach to statistics. In *Proceedings of the third Berkeley symposium on mathematical statistics and probability, 1954–1955*, Vol. I (eds J. Neyman & L. LeCam), 157–163. University of California Press, Berkeley and Los Angeles, CA.
Searle, S. R. (1971). *Linear models*. Wiley, New York, NY.

*Received August 2004, in final form November 2005*

Malay Ghosh, Department of Statistics, University of Florida, Gainesville, FL 32611-8545, USA.
E-mail: ghoshm@stat.ufl.edu

## Appendix

### Proofs

*Proof of lemma 1.* Let $H_i=(x_i-\mu_x)/\sigma_x, i=1,\dots,m$. Then $H_i$ are i.i.d. $N(0,1)$ and

$$\sum_{i=1}^{m} n_i(x_i-\bar{x})^2=\sigma_x^2\sum_{i=1}^{m} n_i(H_i-\bar{H})^2,$$

where $\bar{H}=\sum_{i=1}^{m} n_iH_i/n_T$. We write

$$\sum n_i(H_i-\bar{H})^2=\boldsymbol{H}^T\left(\boldsymbol{D}_m-\boldsymbol{d}_m\frac{\boldsymbol{d}_m^T}{n_T}\right)\boldsymbol{H},$$

where $\boldsymbol{H}=(H_1,\dots,H_m)^T$, $\boldsymbol{D}_m=$ Diag$(n_1,\dots,n_m)$ and $\boldsymbol{d}_m=(n_1,\dots,n_m)^T$. Then

$$\begin{aligned}E\left[\sum_{i=1}^{m}\frac{n_i(x_i-\bar{x})^2}{(m-1)}\right]&=\sigma_x^2 tr\frac{\left(\boldsymbol{D}_m-\frac{\boldsymbol{d}_m\boldsymbol{d}_m^T}{n_T}\right)}{(m-1)}=\sigma_x^2\frac{\left(n_T-n_T^{-1}\sum_{i=1}^{m}n_i^2\right)}{(m-1)}\\&=\sigma_x^2\frac{\left(n_T-n_T^{-1}\sum_{i=1}^{m}n_i^2\right)}{(m-1)}=\sigma_x^2\frac{g_m}{(m-1)}\to c\sigma_x^2\quad\text{as } m\to\infty,\end{aligned}\tag{A.1}$$

by assumption (ii) of theorem 1. Also,

$$\begin{aligned}V\left[\sum_{i=1}^{m}\frac{n_i(x_i-\bar{x})^2}{(m-1)}\right]&=\sigma_x^4(m-1)^{-2}V\left[\boldsymbol{H}^T\left(\boldsymbol{D}_m-\boldsymbol{d}_m\frac{\boldsymbol{d}_m^T}{n_T}\right)\boldsymbol{H}\right]\\&=2\sigma_x^4(m-1)^{-2}tr\left[\left(\boldsymbol{D}_m-\boldsymbol{d}_m\frac{\boldsymbol{d}_m^T}{n_T}\right)^2\right]\\&=2\sigma_x^4(m-1)^{-2}\left[\sum_{i=1}^{m}n_i^2-2\sum_{i=1}^{m}\frac{n_i^3}{n_T}+\sum_{i=1}^{m}\frac{n_i^4}{n_T^2}\right]\\&=2\sigma_x^4(m-1)^{-2}\sum_{i=1}^{m}n_i^2\left(1-\frac{n_i}{n_T}\right)^2\\&\le 2\sigma_x^4(m-1)^{-2}mK^2\to 0\quad\text{as } m\to\infty,\end{aligned}\tag{A.2}$$

by assumption (i) of theorem 1. This completes the proof of lemma 1.

The following lemma will also be very useful in the sequel.

**Lemma 2**

*Let $T_1, T_2, \dots, T_s$ be independently distributed with $T_1 \sim N(\mu, 1)$, and $T_2, \dots, T_s \sim N(0, 1)$. Then writing $\lambda = 1/2\mu^2$,*

*(i)* $E(T_1^2/\sum_1^s T_i^2) = E[(2K+1)/(2K+s)]$;
*(ii)* $E(T_1/\sum_1^s T_i^2) = \mu E[(2K+s)^{-1}]$;
*(iii)* $E(T_1/(\sum_1^s T_i^2)^{1/2}) = \mu E(\chi^2_{2K+2+s})^{-1/2}$;
*(iv)* $E[T_1^2/(\sum_1^s T_i^2)^2] = E[(2K+1)(2K+s)^{-1}(2K+s-2)^{-1}]$,

*where $K \sim$ Poisson $(\lambda)$.*

*Proof.*

(i) Conditional on $K = k, T_1^2, T_2^2, \dots, T_s^2$ are mutually independent with $T_1^2 \sim \chi^2_{2K+1}$ and $T_2^2, \dots, T_s^2$ i.i.d. $\chi_1^2$ and $K \sim$ Poisson $(\lambda)$. Hence, conditional on $K = k$, $T_1^2/\Sigma_1^s T_i^2 \sim$ Beta$(1/2(2k+1), 1/2(s-1))$. Hence,

$$E\left(\frac{T_1^2}{\sum_1^s T_i^2 | K = k}\right) = \frac{(2k+1)}{(2k+s)}.$$

The result follows.

(ii) Conditional on $K = k$, $\sum_1^s T_i^2 \sim \chi^2_{2k+s}$ and $K \sim$ Poisson $(\lambda)$. This leads to the identity

$$\sum_{k=0}^{\infty} \frac{\exp(-\lambda)\lambda^k}{k!(2k-2+s)} = E\left(\sum_i^k T_i^2\right)^{-1}$$
$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{\exp[-(T_1-\mu)^2/2 - \sum_2^s T_i^2/2]}{(\sum_1^s T_i^2)(2\pi)^{s/2}} \prod_{i=1}^{s} \mathrm{d}T_i. \tag{A.3}$$

As $\lambda = 1/2\mu^2$, differentiating both sides of (A.2.1) with respect to $\mu$, one gets

$$\sum_{k=0}^{\infty} \frac{\exp(-\lambda)\lambda^k}{k!(2k+s)} - \mu E\left(\sum_1^k T_i^2\right)^{-1} = E\left[\frac{(T_1-\mu)}{\sum_2^s T_i^2}\right]$$
$$= E\left(\frac{T_1}{\sum_2^s T_i^2}\right) - \mu E\left(\sum_1^k T_i^2\right)^{-1}. \tag{A.4}$$

The result follows now from (A.4).

(iii) Similar to (A.3), we begin with the identity

$$\sum_{k=0}^{\infty} \frac{\exp(-\lambda)\lambda^k}{k!} E(\chi^2_{2k+s})^{-1/2} = E\left(\sum_1^s T_i^2\right)^{-1/2}$$
$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{\exp[-(T_1-\mu)^2/2 - \sum_2^s T_i^2/2]}{(\sum_1^s T_i^2)^{1/2}(2\pi)^{s/2}} \prod_{i=1}^{k} \mathrm{d}T_i. \tag{A.5}$$

Differentiating both sides of (A.5) with respect to $\mu$,

$$\sum_{k=0}^{\infty} \frac{\exp(-\lambda)\lambda^k}{k!} E(\chi^2_{2k+2+s})^{-1/2} - \mu E\left(\sum_1^k T_i^2\right)^{-1/2}$$
$$= E\left[\frac{T_1-\mu}{(\sum_1^s T_i^2)^{1/2}}\right] = E\left[\frac{T_1}{(\sum_1^s T_i^2)^{1/2}}\right] - \mu E\left(\sum_1^k T_i^2\right)^{-1/2}. \tag{A.6}$$

The result follows now from (A.6).

(iv) Conditional on $K=k$, $T_1^2/\sum_1^s T_i^2$ and $\sum_1^s T_i^2$ are mutually independent with $T_1^2/\sum_1^s T_i^2 \sim \text{Beta}((2k+1)/2, (s-1)/2)$ and $\sum_1^s T_i^2 \sim \chi^2_{2k+s}$. Hence, $E(T_1^2/(\sum_1^s T_i^2)^2)=[(2k+1)/(2k+s)](2k+s-2)^{-1}$. This proves the result.

*Proof of theorem 1.* Let $\boldsymbol{L}=(\bar{X}_1,\ldots,\bar{X}_m)^T$ and $\boldsymbol{x}=(x_1,\ldots,x_m)$. By the independence of $(\bar{y}_1,\ldots,\bar{y}_m)$ with $(\bar{X}_1,\ldots,\bar{X}_m)$

$$\begin{aligned}
E(\tilde{b}_1) &= EE\left[\frac{\sum_{i=1}^m n_i\bar{y}_i(\bar{X}_i-\bar{X})}{\sum_{i=1}^m n_i(\bar{X}_i-\bar{X})^2|\boldsymbol{L},\boldsymbol{x}}\right] \\
&= EE\left[\frac{\sum_{i=1}^m n_i(b_0+b_1x_i)(\bar{X}_i-\bar{X})}{\sum_{i=1}^m n_i(\bar{X}_i-\bar{X})^2|\boldsymbol{L},\boldsymbol{x}}\right] = b_1E\left[\frac{\sum_{i=1}^m n_ix_i(\bar{X}_i-\bar{X})}{\sum_{i=1}^m n_i(\bar{X}_i-\bar{X})^2}\right] \\
&= b_1E\left[\frac{\sum_{i=1}^m n_i\bar{X}_i(\bar{x}_i-\bar{x})}{\sum_{i=1}^m n_i(\bar{X}_i-\bar{X})^2}\right] = b_1EE\left[\frac{\sum_{i=1}^m n_i\bar{X}_i(\bar{x}_i-\bar{x})}{\sum_{i=1}^m n_i(\bar{X}_i-\bar{X})^2|\boldsymbol{x}}\right].
\end{aligned} \tag{A.7}$$

For fixed $\boldsymbol{x}$, we introduce the orthogonal transformation $(z_1,\ldots,z_m)^T=\boldsymbol{C}(\sqrt{n_1}\bar{X}_1,\ldots,\sqrt{n_1}\bar{X}_1)^T$, where $\boldsymbol{C}$ is an orthogonal matrix with the first two rows given by $(\sqrt{n_1}/\sqrt{n_T},\ldots,\sqrt{n_m}/\sqrt{n_T})$ and

$$\frac{\sqrt{n_1}(x_1-\bar{x})}{(\sum n_i(x_1-\bar{x})^2)^{1/2}},\ldots,\frac{\sqrt{n_m}(x_m-\bar{x})}{(\sum n_i(x_1-\bar{x})^2)^{1/2}}.$$

Then

$$\sum_{i=1}^m n_i\bar{X}_i(x_i-\bar{x})=Z_2\left[\sum_{i=1}^m n_i(x_i-\bar{x})^2\right]^{1/2}$$

and

$$\sum_{i=1}^m n_i(\bar{X}_i-\bar{X})^2=\sum_{i=1}^m n_i\bar{X}_i^2-\frac{(\sum n_i\bar{X}_i)^2}{n_T}=\sum_{i=1}^m Z_i^2-Z_1^2=\sum_{i=2}^m Z_i^2.$$

Hence,

$$E\left[\frac{\sum_{i=1}^m n_i\bar{X}_i(x_i-\bar{x})}{\sum_{i=1}^m n_i(\bar{X}_i-\bar{X})^2|\boldsymbol{x}}\right]=\left[\sum_{i=1}^m n_i(x_i-\bar{x})^2\right]^{1/2}E\left[Z_2\Big/\left(\sum_{i=2}^m Z_i^2\right)^{1/2}\right],$$

where $Z_2\sim N\left((\sum_{i=1}^m n_i(x_i-\bar{x})^2)^{1/2},\sigma_\eta^2\right)$ and is distributed independently of $(Z_3,\ldots,Z_m)$ conditional on $\boldsymbol{x}$. Next for every $3\le i\le m$, writing the $i$th row of $\boldsymbol{C}$ as $(c_{i1},\ldots,c_{im})$, one gets the identities

$$\frac{\sum_{k=1}^m c_{ik}\sqrt{n_k}}{\sqrt{n_T}}=0 \quad\text{and}\quad \sum_{k=1}^m c_{ik}n_k^{1/2}(x_k-\bar{x})=0.$$

Together, they imply $\sum_{k=1}^m c_{ik}n_k^{1/2}x_k=0$ which is equivalent to $E(Z_i)=0, 3\le i\le m$.

Next applying part (ii) of lemma 2. and (A.7),

$$\begin{aligned}
E(\tilde{b}_1|\boldsymbol{x}) &= b_1\left(\sum_{i=1}^m n_i(x_i-\bar{x})^2\right)^{1/2}\sigma_\eta^{-1}\left(\sum_{i=1}^m n_i(x_i-\bar{x})^2\right)^{1/2}\sigma_\eta^{-1}E[(2K_m+m-1)^{-1}|\boldsymbol{x}] \\
&= b_1\left(\sum_{i=1}^m n_i(x_i-\bar{x})^2\Big/\sigma_\eta^2\right)E[(2K_m+m-1)^{-1}|\boldsymbol{x}] \\
&= b_1\sigma_\eta^{-2}\left(\sum_{i=1}^m n_i(x_i-\bar{x})^2\Big/(m-1)\right)E[(1+2(m-1)^{-1}K_m)^{-1}|\boldsymbol{x}],
\end{aligned} \tag{A.8}$$

where $K_m|\boldsymbol{x}\sim$ Poisson $\left(\frac{1}{2}\sum_{i=1}^m n_i(x_i-\bar{x})^2/\sigma_\eta^2\right)$. By lemma 1,

$$\sum_{i=1}^{m} \frac{n_i(x_i-\bar{x})^2}{(m-1)} \xrightarrow{\mathrm{P}} c\sigma_x^2 \quad \text{as } m \to \infty$$

and

$$E[(m-1)^{-1}K_m] = E\left[\sum_{i=1}^{m} \frac{n_i(x_i-\bar{x})^2}{(m-1)}\right](2\sigma_\eta^2)^{-1} \to c\sigma_x^2(2\sigma_\eta^2)^{-1} \quad \text{as } m \to \infty; \tag{A.9}$$

$$\begin{aligned} V[(m-1)^{-1}K_m] &= V\left[\sum_{i=1}^{m} \frac{n_i(x_i-\bar{x})^2}{(m-1)}\right](2\sigma_\eta^2)^{-2} \\ &\quad + E\left[(m-1)^{-2}\sum_{i=1}^{m} n_i(x_i-\bar{x})^2\right](2\sigma_\eta^2)^{-1} \\ &\to 0+0=0 \quad \text{as } m \to \infty. \end{aligned} \tag{A.10}$$

Hence, by (A.9) and (A.10), $(m-1)^{-1}K_m \xrightarrow{\mathrm{P}} c\sigma_x^2(2\sigma_\eta^2)^{-1}$ as $m \to \infty$. Thus, $(1+2K_m/(m-1))^{-1} \xrightarrow{\mathrm{P}} (1+c\sigma_x^2/\sigma_\eta^2)^{-1}$, and now by the dominated convergence theorem,

$$E\left[1+\frac{2K_m}{(m-1)}\right]^{-1} \to (1+c\sigma_x^2/\sigma_\eta^2)^{-1} \quad \text{and} \quad V\left[1+\frac{2K_m}{(m-1)}\right]^{-1} \to 0 \quad \text{as } m \to \infty.$$

Now writing

$$J_m(\boldsymbol{x}) = E\left[\left(1+\frac{2K_m}{m-1}\right)^{-1} |\boldsymbol{x}\right], E[J_m(\boldsymbol{x})] = E\left[\left(1+\frac{2K_m}{(m-1)}\right)^{-1}\right] \to (1+c\sigma_x^2/\sigma_\eta^2)^{-1}$$

while

$$V[J_m(\boldsymbol{x})] \le V\left[\left(1+\frac{2K_m}{(m-1)}\right)^{-1}\right] \to 0 \quad \text{as } m \to \infty.$$

Thus, $J_m(\boldsymbol{x}) \xrightarrow{\mathrm{P}} (1+c\sigma_x^2/\sigma_\eta^2)^{-1}$. Hence,

$$E(\tilde{b}_1|\boldsymbol{x}) \xrightarrow{\mathrm{P}} b_1\sigma_\eta^2 c\sigma_x^2\left(1+\frac{c\sigma_x^2}{\sigma_\eta^2}\right)^{-1} = \frac{b_1(c\sigma_x^2)}{(c\sigma_x^2+\sigma_\eta^2)}. \tag{A.11}$$

Again, by assumption (i) of theorem 1,

$$|E(\tilde{b}_1|\boldsymbol{x})| \le |b_1|\sigma_\eta^{-2}\sum \frac{n_i(x_i-\bar{x})^2}{(m-1)} \le K|b_1|\sigma_\eta^{-2}\sum_1^m \frac{(x_i-\bar{x})^2}{(m-1)}.$$

As

$$\sum_1^m \frac{(x_i-\bar{x})^2}{(m-1)} \sim \frac{\sigma_x^2\chi_{m-1}^2}{(m-1)}$$

and

$$E\left[\frac{\chi_{m-1}^2}{(m-1)}\right]^2 = 2(m-1)^{-1}+1 \le 3$$

for all $m \ge 2$, it follows that $E(\tilde{b}_1|\boldsymbol{x})$ is uniformly integrable in $m \ge 2$. Hence,

$$E(\tilde{b}_1) = E[E(\tilde{b}_1|\boldsymbol{x})] \to \frac{b_1(c\sigma_x^2)}{(c\sigma_x^2+\sigma_\eta^2)} \quad \text{as } m \to \infty.$$

This completes the proof of theorem 1.

*Proof of theorem 2.* Let $r_i = n_i^{1/2}\bar{y}_i, i = 1, \ldots, m$. Then $E(r_i) = \sqrt{n_i}(b_0 + b_1 x_i) = \sqrt{n_i}[b_0 + b_1\bar{x} + b_1(x_i - \bar{x})]$ and $V(r_i) = \sigma_e^2 + n_i\sigma_u^2$. Thus, writing $\boldsymbol{r} = (r_1, \ldots, r_m)^T$, $\boldsymbol{q}_m = [\sqrt{n_1}(x_1 - \bar{x}), \ldots, \sqrt{n_m}(x_m - \bar{x})]$, $\boldsymbol{D}_m = \mathrm{Diag}(n_1, \ldots, n_m)$,

$$E(\boldsymbol{r}) = \sqrt{n_T}(b_0 + b_1\bar{x})\boldsymbol{u}_m + \boldsymbol{q}_m, \quad V(\boldsymbol{r}) = \sigma_e^2\boldsymbol{I}_m + \sigma_u^2\boldsymbol{D}_m, \tag{A.12}$$

where we recall that $\boldsymbol{u}_m^T = (\sqrt{n_1}/\sqrt{n_T}, \ldots, \sqrt{n_m}/\sqrt{n_T})$. Now, by part (ii) of Theorem 1 of Searle (1971, p. 55),

$$\begin{aligned} V(\mathrm{MSB}_y) = 2(m-1)^{-2}\Big[ & t_r\{(\sigma_e^2\boldsymbol{I}_m + \sigma_u^2\boldsymbol{D}_m)(\boldsymbol{I}_m - \boldsymbol{u}_m\boldsymbol{u}_m^T)\}^2 \\ & + 2n_T[(b_0 + b_1\bar{x})\boldsymbol{u}_m^T + \boldsymbol{q}_m^T](\boldsymbol{I}_m - \boldsymbol{u}_m\boldsymbol{u}_m^T)\boldsymbol{D}_m(\boldsymbol{I}_m - \boldsymbol{u}_m\boldsymbol{u}_m^T)[(b_0 + b_1\bar{x})\boldsymbol{u}_m + \boldsymbol{q}_m]\Big]. \end{aligned} \tag{A.13}$$

Noting that $\boldsymbol{u}_m^T\boldsymbol{u}_m = 1$ and $\boldsymbol{q}_m^T\boldsymbol{u}_m = 0$, it follows that $[(b_0 + b_1\bar{x})\boldsymbol{u}_m^T + \boldsymbol{q}_m^T](\boldsymbol{I}_m - \boldsymbol{u}_m\boldsymbol{u}_m^T) = \boldsymbol{0}^T$. Further, by idempotency of $\boldsymbol{I}_m - \boldsymbol{u}_m\boldsymbol{u}_m^T$,

$$\begin{aligned} & tr[(\sigma_e^2\boldsymbol{I}_m + \sigma_u^2\boldsymbol{D}_m)(\boldsymbol{I}_m - \boldsymbol{u}_m\boldsymbol{u}_m^T)]^2 \\ & = \sigma_e^4 tr(\boldsymbol{I}_m - \boldsymbol{u}_m\boldsymbol{u}_m^T) + \sigma_u^4 tr\{\boldsymbol{D}_m(\boldsymbol{I}_m - \boldsymbol{u}_m\boldsymbol{u}_m^T)\}^2 + 2\sigma_e^2\sigma_u^2 tr\{\boldsymbol{D}_m(\boldsymbol{I}_m - \boldsymbol{u}_m\boldsymbol{u}_m^T)\} \\ & = \sigma_e^4(m-1) + \sigma_u^4 t_r[\boldsymbol{D}2_m + \boldsymbol{D}_m\boldsymbol{u}_m\boldsymbol{u}_m^T\boldsymbol{D}_m\boldsymbol{u}_m\boldsymbol{u}_m^T - 2\boldsymbol{D}2_m\boldsymbol{u}_m\boldsymbol{u}_m^T] \\ & \quad + 2\sigma_e^2\sigma_u^2[t_r(\boldsymbol{D}_m) - \boldsymbol{u}_m^T\boldsymbol{D}_m\boldsymbol{u}_m] \\ & = \sigma_e^4(m-1) + \sigma_u^4 tr[\boldsymbol{D}_m 2 + (\boldsymbol{u}_m^T\boldsymbol{D}_m\boldsymbol{u}_m)^2 - 2\boldsymbol{u}_m^T\boldsymbol{D}_m 2\boldsymbol{u}_m] \\ & = \sigma_e^4(m-1) + \sigma_u^2\left[\sum n_i^2 + \left(\sum_1^m \frac{n_i^2}{n_T}\right)^2 - 2\sum_1^m \frac{n_i^3}{n_T}\right] + 2\sigma_e^2\sigma_u^2\left(n_T - \sum\frac{n_i^2}{n_T}\right). \end{aligned} \tag{A.14}$$

As $n_i \le K$ for all $1 \le i \le m$, $\sum n_i^2 + \left(\sum \frac{n_i^2}{n_T}\right) - 2\sum\frac{n_i^3}{n_T} \le K^2 m + K^2 + 2K^2 = K^2(m+3)$. Hence, from (A.13) and (A.14),

$$V(\mathrm{MSB}_y) \le 2(m-1)^{-2}[\sigma_e^4(m-1) + K^2(m+3)\sigma_u 4 + 2Km\sigma_e^2\sigma_u^2] = 0(m^{-1}) \to 0 \quad \text{as} \quad m \to \infty.$$

This completes the proof of theorem 2.

*Proof of theorem 3.* We rewrite

$$\begin{aligned} \hat{\gamma}_i^B - \hat{\gamma}_i^{\mathrm{EB}} & = (1 - f_i B_i)\bar{y}_i + f_i B_i(b_0 + b_1\mu_x) - (1 - f_i\hat{B}_i)\bar{y}_i - f_i\hat{B}_i\bar{y} \\ & = f_i\left[(\hat{B}_i - B_i)(\bar{y}_i - \bar{y}) - B_i(\bar{y} - b_0 - b_1\mu_x)\right]. \end{aligned}$$

Hence, by the elementary inequality $(a+b)^2 \le 2(a2 + b2)$ and $f_i^2 \le 1$,

$$\begin{aligned} m^{-1}\sum_{i=1}^m(\hat{\gamma}_i^B - \hat{\gamma}_i^{\mathrm{EB}})^2 & \le 2m^{-1}\sum_{i=1}^m\left[(\hat{B}_i - B_i)^2(\bar{y}_i - \bar{y})^2 + B_i^2(\bar{y} - b_0 - b_1\mu_x)^2\right] \\ & \le 2\left[m^{-1}\sum_{i=1}^m(\hat{B}_i - B_i)^2(\bar{y}_i - \bar{y})^2 + (\bar{y} - b_0 - b_1\mu_x)^2\right]. \end{aligned} \tag{A.15}$$

But, under assumption (i) of theorem 1,

$$\begin{aligned} E(\bar{y} - b_0 - b_1\mu_x)^2 = V(\bar{y}) & = n_T^{-2}\sum_{i=1}^m n_i^2\left(\frac{\sigma_e^2}{n_i} + \sigma_u^2 + b_1^2\sigma_x^2\right) \\ & \le \frac{[\sigma_e^2 + K(\sigma_u^2 + b_1^2\sigma_x^2)]}{n_T} \to 0 \quad \text{as } m \to \infty. \end{aligned} \tag{A.16}$$

Next, writing $M=\sigma_u^2/\sigma_e^2$, $\hat{M}=\hat{\sigma}_u2/\hat{\sigma}_e2$, we observe that

$$\begin{aligned}\max_{1\le i\le m}|\hat{B}_i-B_i| &= \max_{1\le i\le m}\left|\frac{\hat{\sigma}_e2}{\hat{\sigma}_e2+n_i\hat{\sigma}_u2}-\frac{\sigma_e^2}{\sigma_e^2+n_i\sigma_u^2}\right| \\ &= \max_{1\le i\le m}\left|\frac{1}{1+n_i\hat{M}}-\frac{1}{1+n_iM}\right| \\ &= \max_{1\le i\le m}\frac{n_i|\hat{M}-M|}{(1+n_i\hat{M})(1+n_iM)}\le n_i|\hat{M}-M|. \qquad \text{(A.17)}\end{aligned}$$

Note that $\hat{M}\xrightarrow{\mathrm{P}}M$. Also, as $m^{-1}\sum_{i=1}^m n_i(\bar{y}_i-\bar{y})^2=[(m-1)/m]\mathrm{MSB}_y$ and $\mathrm{MSB}_y\xrightarrow{\mathrm{P}}\sigma_e^2+c(\sigma_u^2+b_1^2\sigma_x^2)$, it follows that $m^{-1}\sum_{i=1}^m(\hat{B}_i-B_i)^2(\bar{y}_i-\bar{y})^2\xrightarrow{\mathrm{P}}0$. Moreover, $m^{-1}\sum_{i=1}^m(\hat{B}_i-B_i)^2\times(\bar{y}_i-\bar{y})^2\le m^{-1}\sum_{i=1}^m n_i(\bar{y}_i-\bar{y})^2=m^{-1}\boldsymbol{q}_m^T(\boldsymbol{I}_m-n_T^{-1}\boldsymbol{d}_m\boldsymbol{d}_m^T)\boldsymbol{q}_m$, where $\boldsymbol{q}_m^T=(\bar{y}_1-b_0-b_1\mu_x,\ldots,\bar{y}_m-b_0-b_1\mu_x)$, $\boldsymbol{D}_m=\mathrm{Diag}(n_1,\ldots,n_m)$, $\boldsymbol{d}_m^T=(n_1,\ldots,n_m)$. Hence, $E[m^{-1}\sum_{i=1}^m n_i(\bar{y}_i-\bar{y})^2]=2m^{-2}tr(\boldsymbol{I}_m-n_T^{-1}\boldsymbol{d}_m\boldsymbol{d}_m^T)^2=O(m^{-1})$. Thus, $\sup_{m\ge1}m^{-1}\sum_{i=1}^m(\hat{B}_i-B_i)^2(\bar{y}_i-\bar{y})^2$ is uniformly integrable in $m$ and $E\left[m^{-1}\sum_{i=1}^m(\hat{B}_i-B_i)^2(\bar{y}_i-\bar{y})^2\right]\to0$ as $m\to\infty$.

This establishes theorem 3.

*Proof of theorem 4.* The joint posterior is given by

$$\begin{aligned}&\pi(\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{b},\boldsymbol{\mu}_x,\sigma_e^2,\sigma_u^2,\sigma_\eta^2,\sigma_x^2|\boldsymbol{y},\boldsymbol{X})\\ &\quad\propto(\sigma_e^2)^{-n_T/2}\exp\left[-\frac{1}{2\sigma_e^2}\left(\sum_{i=1}^m n_i(\bar{y}_i-\theta_i)^2+\mathrm{SSW}_y\right)\right]\\ &\quad\times(\sigma_u^2)^{-m/2}\exp\left[-\frac{1}{2\sigma_u^2}\sum_{i=1}^m(\theta_i-\boldsymbol{z}_i^T\boldsymbol{b})^2\right]\\ &\quad\times(\sigma_\eta^2)^{-n_T/2}\exp\left[-\frac{1}{2\sigma_\eta^2}\left(\sum_{i=1}^m n_i(\bar{X}_i-x_i)^2+\mathrm{SSW}_X\right)\right]\\ &\quad\times(\sigma_x^2)^{-m/2}\exp\left[-\frac{1}{2\sigma_x^2}\sum_{i=1}^m(x_i-\mu_x)^2\right]\\ &\quad\times(\sigma_e^2)^{-(b_e/2)-1}\exp\left[-\frac{a_e}{2\sigma_e^2}\right](\sigma_u^2)^{-(b_u/2)-1}\exp\left[-\frac{a_u}{2\sigma_u^2}\right]\\ &\quad\times(\sigma_\eta^2)^{-(b_\eta/2)-1}\exp\left[-\frac{a_\eta}{2\sigma_\eta^2}\right](\sigma_x^2)^{-(b_x/2)-1}\exp\left[-\frac{a_x}{2\sigma_x^2}\right]. \qquad \text{(A.18)}\end{aligned}$$

First integrating out with respect to $\mu_x$, and noting that $\exp[-1/2\sigma_e^2\sum_{i=1}^m(x_i-\bar{x})^2]\le1$,

$$\begin{aligned}&\pi(\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{b},\sigma_e^2,\sigma_u^2,\sigma_{\boldsymbol{\eta}}^2,\sigma_x^2|\boldsymbol{y},\boldsymbol{X})\\ &\quad\le K\exp\left[-\frac{1}{2\sigma_e^2}\left(\sum_{i=1}^m n_i(\bar{y}_i-\theta_i)^2+\mathrm{SSW}_y\right)\right]\\ &\quad\times\exp\left[-\frac{1}{2\sigma_u^2}\sum_{i=1}^m(\theta_i-\boldsymbol{z}_i^T\boldsymbol{b})^2\right]\times\exp\left[-\frac{1}{2\sigma_\eta^2}\left(\sum_{i=1}^m n_i(\bar{X}_i-\bar{x}_i)^2+\mathrm{SSW}_X\right)\right]\\ &\quad\times(\sigma_e^2)^{-(b_e+n_T/2)-1}\exp\left[-\frac{a_e}{2\sigma_e^2}\right](\sigma_u^2)^{-(b_u+m/2)-1}\exp\left[-\frac{a_u}{2\sigma_u^2}\right]\\ &\quad\times(\sigma_\eta^2)^{-(b_\eta+n_T/2)-1}\exp\left[-\frac{a_\eta}{2\sigma_\eta^2}\right](\sigma_x^2)^{-(b_x+m-1/2)-1}\exp\left[-\frac{a_x}{2\sigma_x^2}\right], \qquad \text{(A.19)}\end{aligned}$$

where in the above and in what follows, $K(>0)$ is a generic constant.

Next, writing $\mathbf{X}_*^T = (\mathbf{1}_m, \boldsymbol{x})$,

$$\sum_{i=1}^{m} (\theta_i - \boldsymbol{x}_i^T \boldsymbol{b})^2 = \boldsymbol{b}^T (\boldsymbol{X}_*^T \boldsymbol{X}_*)^{-1} \boldsymbol{b} - 2\boldsymbol{b}^T \boldsymbol{x}^T \boldsymbol{\theta} + \boldsymbol{\theta}^T \boldsymbol{\theta}$$
$$= [\boldsymbol{b} - (\boldsymbol{X}_*^T \boldsymbol{X}_*)^{-1} \boldsymbol{X}_*^T \boldsymbol{\theta}]^T (\boldsymbol{X}_*^T \boldsymbol{X}_*)^{-1} [\boldsymbol{b} - (\boldsymbol{X}_*^T \boldsymbol{X}_*)^{-1} \boldsymbol{X}_*^T \boldsymbol{\theta}]$$
$$+ \boldsymbol{\theta}^T (\boldsymbol{I} - \boldsymbol{P}_{X_*}) \boldsymbol{\theta},$$

where $\boldsymbol{P}_{X_*} = \boldsymbol{X}_* (\boldsymbol{X}_*^T \boldsymbol{X}_*)^{-1} \boldsymbol{X}_*^T$. Now, integrating with respect to $\boldsymbol{b}$ and using $\boldsymbol{\theta}^T (\boldsymbol{I} - \boldsymbol{P}_{X_*}) \boldsymbol{\theta} \geq 0$,

$$\pi(\boldsymbol{\theta}, \boldsymbol{x}, \sigma_e^2, \sigma_u^2, \sigma_{\boldsymbol{\eta}}^2, \sigma_x^2 | \boldsymbol{y}, \boldsymbol{X})$$
$$\leq K \exp\left[ -\frac{1}{2\sigma_e^2} \left( \sum_{i=1}^{m} n_i (\bar{y}_i - \theta_i)^2 + \mathrm{SSW}_y \right) \right] |\boldsymbol{X}_*^T \boldsymbol{X}_*|^{-1/2}$$
$$\times \exp\left[ -\frac{1}{2\sigma_\eta^2} \left( \sum_{i=1}^{m} n_i (\bar{X}_i - \bar{x}_i)^2 + \mathrm{SSW}_X \right) \right] (\sigma_e^2)^{-(b_e + n_T/2) - 1} \exp\left[ -\frac{a_e}{2\sigma_e^2} \right] (\sigma_u^2)^{-(b_u + m - 2/2) - 1}$$
$$\times \exp\left[ -\frac{a_u}{2\sigma_u^2} \right] (\sigma_\eta^2)^{-(b_\eta + n_T/2) - 1} \exp\left[ -\frac{a_\eta}{2\sigma_\eta^2} \right] (\sigma_x^2)^{-(b_x + m - 1/2) - 1} \exp\left[ -\frac{a_x}{2\sigma_x^2} \right]. \tag{A.20}$$

Next, integrating with respect to $\boldsymbol{\theta}$, it follows from (A.20) that

$$\pi(\boldsymbol{x}, \sigma_e^2, \sigma_u^2, \sigma_{\boldsymbol{\eta}}^2, \sigma_x^2 | \boldsymbol{y}, \boldsymbol{X}) \leq K |\boldsymbol{X}_*^T \boldsymbol{X}_*|^{-1/2}$$
$$\times \exp\left[ -\frac{1}{2\sigma_\eta^2} \left( \sum_{i=1}^{m} n_i (\bar{X}_i - \bar{x}_i)^2 + \mathrm{SSW}_X \right) \right] (\sigma_e^2)^{-(b_e + n_T - m/2) - 1}$$
$$\times \exp\left[ -\frac{a_e}{2\sigma_e^2} \right] (\sigma_u^2)^{-(b_u + m - 2/2) - 1} \exp\left[ -\frac{a_u}{2\sigma_u^2} \right] (\sigma_\eta^2)^{-(b_\eta + n_T/2) - 1}$$
$$\times \exp\left[ -\frac{a_\eta}{2\sigma_\eta^2} \right] (\sigma_x^2)^{-(b_x + m - 1/2) - 1} \exp\left[ -\frac{a_x}{2\sigma_x^2} \right]. \tag{A.21}$$

Next, as $a_e > 0$, $a_u > 0$ and $a_x > 0$, integrating out with respect to $\sigma_e^2$, $\sigma_u^2$ and $\sigma_x^2$, one gets

$$\pi(\boldsymbol{x}, \sigma_{\boldsymbol{\eta}}^2, | \boldsymbol{y}, \boldsymbol{X}) \leq K |\boldsymbol{X}_*^T \boldsymbol{X}_*|^{-1/2} \times \exp\left[ -\frac{1}{2\sigma_\eta^2} \left( \sum_{i=1}^{m} n_i (\bar{X}_i - \bar{x}_i)^2 + \mathrm{SSW}_X \right) \right] (\sigma_\eta^2)^{-(b_\eta + n_T/2) - 1}. \tag{A.22}$$

We now observe that $|\boldsymbol{X}_*^T \boldsymbol{X}_*| = m \sum_{i=1}^{m} (x_i - \bar{x})^2$. Also, from (A.21), conditional on $\sigma_\eta^2, \boldsymbol{y}$ and $\boldsymbol{X}$, $x_i \overset{\text{ind}}{\sim} N(\bar{X}_i, \sigma_\eta^2 / n_i)\,(i = 1, \ldots, m)$. Now, writing $\boldsymbol{u} = (\sqrt{n_1} x_1, \ldots, \sqrt{n_m} x_m)^T$ and $\boldsymbol{U} = (\sqrt{n_1} \bar{X}_1, \ldots, \sqrt{n_m} \bar{X}_m)^T$, $\boldsymbol{u} \sim N(\boldsymbol{U}, \sigma_\eta^2 \boldsymbol{I}_m)$. Then, if $\boldsymbol{D}^{-1} = \mathrm{Diag}(\sqrt{n}_1, \ldots, \sqrt{n}_m)$, $\sum_{i=1}^{m} (x_i - \bar{x})^2 = \boldsymbol{x}^T (\boldsymbol{I}_m - m^{-1} \boldsymbol{J}_m) \boldsymbol{x} = \boldsymbol{u}^T \boldsymbol{A} \boldsymbol{u}$, where $\boldsymbol{A} = \boldsymbol{D}^{1/2} (\mathbf{I}_m - m^{-1} \mathbf{J}_m) \boldsymbol{D}^{1/2}$. As $\mathrm{rank}(\boldsymbol{A}) = m - 1$, by the spectral decomposition theorem, we can write $\boldsymbol{A} = \sum_{i=1}^{m-1} \lambda_i \xi_i \xi_i^T$, where $\lambda_i$ are the non-zero eigen values of $\boldsymbol{A}$ and $\xi_i, \ldots, \xi_{m-1}$ are the corresponding orthonormal eigen vectors. Now, $\boldsymbol{u}^T \boldsymbol{A} \boldsymbol{u} = \sum_{i=1}^{m-1} \lambda_i (\xi_i^T \boldsymbol{u})^2$ where $\xi_i^T \overset{\text{ind}}{\sim} N(\xi_i^T \boldsymbol{U}, \sigma_\eta^2)$. Hence,

$$\int \cdots \int |\boldsymbol{X}_*^T \boldsymbol{X}_*|^{-\frac{1}{2}} \exp\left[ -\frac{1}{2\sigma_\eta^2} \sum_{i=1}^{m} n_i (\bar{X}_i - x_i)^2 \right]$$
$$= E\left[ m \sum_{i=1}^{m-1} \lambda_i \chi_1^2 \left( \frac{1}{2} (\xi_i^T \boldsymbol{U})^2 \right) \right]^{-1/2}$$
$$\leq m^{-\frac{1}{2}} \lambda_{\min}^{-1/2} E\left[ \chi_{m-1}^{-1} \left( \frac{1}{2} \boldsymbol{U}^T \boldsymbol{U} \right) \right] \leq m^{-\frac{1}{2}} \lambda_{\min}^{-\frac{1}{2}} E(\chi_{m-1}^{-1}), \tag{A.23}$$

where $\lambda_{\min} = \min(\lambda_1, \ldots, \lambda_{m-1})$. Hence, by (A.22) and (A.23), integrating with respect to $\boldsymbol{x}$, $\pi(\sigma_\eta^2, |\boldsymbol{y}, \boldsymbol{X}) \leq K(\sigma_\eta^2)^{-(b_\eta + n_T - m)/2-1} \exp(-\mathrm{SSW}_X/2\sigma_\eta^2)$.

It is clear that $\int_0^\infty \pi(\sigma_\eta^2|\boldsymbol{y}, \boldsymbol{X})\, \mathrm{d}\sigma_\eta^2 < \infty$ as $b_\eta + n_T - m >> 0$.

The proof of the theorem is thus complete.