Available online at www.isas.org.in/jisas

JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 64(2) 2010 191-207

Bayesian Predictive Inference for Benchmarking Crop Production for Iowa Counties

Balgobin Nandram* **and Ma. Criselda S. Toto**

Department of Mathematical Sciences, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609-2280

Received 01 November 2009; Revised 12 April 2010; Accepted 12 April 2010

SUMMARY

For a long time satellite and survey data have been used to estimate crop and livestock production at county level. Typically prediction is required for counties (small areas), and parametric models have been discussed extensively. The main goal in small area estimation is to use models to 'borrow strength' from the ensemble because the direct estimates of small area parameters are generally unreliable. But such models are not completely satisfactory. We address two issues concerning these models. First, the combined estimates from all small areas do not usually match the value of the single estimate of the large area, and benchmarking is desirable. Benchmarking is done by applying a constraint that will ensure that the 'total' of the small areas matches the 'grand total'. We use a Bayesian nested error regression model to develop a method to benchmark the finite population means of small areas. Second, it is the practice to assume that the sampling variances are homogeneous, but this may not be the case. Thus, in addition to benchmarking, we also show how to study heterogeneous sampling variances. We apply our method to estimate the number of acres of corn and soybean under cultivation for twelve counties in Iowa.

Keywords: Heterogeneous variances, Monte Carlo methods, Nested-error regression model, Posterior propriety, Small area estimation.

1. INTRODUCTION

The United States Department of Agriculture has been using satellite and survey data to estimate crop and livestock production at the county level. Iwig (1993) gave a detailed description of the National Agricultural Statistics Service county estimation program. Small area estimation has been used extensively. Corn and soybeans are important crops grown in the U.S., and they are of enormous support to the U.S. economy.

The mission of the National Agricultural Statistics Service (NASS), which is an agency under the U.S. Department of Agriculture (USDA), is to provide timely, accurate, and useful statistics in service to U.S. agriculture. To accomplish this objective, NASS conducts hundreds of surveys and prepares over five hundred reports every year covering virtually every aspect of U.S. agriculture. NASS has initiated a collaborative research program with the National Institute of Statistical Sciences (NISS), which is an independent research institute dedicated to strengthening and serving the national statistics community, to investigate whether improvements can be made to their board's analysis process. One of the most important reports produced by NASS is the monthly Crop Production Report. As an example, the Crop Production Report contains forecasted corn yield from August through November and the end-of-season corn yield is published in January.

* *Corresponding author* : Balgobin Nandram
E-mail address : balnan@wpi.edu

Most of the corn grown in the United States comes from the Corn Belt, which consists of ten major states; Iowa is the largest corn-producing state in the U.S. Iowa is also among the major states for soybean production. Producers in the U.S. feed most of the corn crop to cattle, hogs, sheep and poultry. The rest is used for processed food, industrial products such as cornstarch and plastic, renewable energy and ethanol. Corn is the U.S. largest crop, both in volume and value. Iowa has produced the largest corn crop over the current decade. In an average year Iowa produces more corn than most countries (e.g., three times as much corn as Argentina). Soybean and corn are grown in rotation in Iowa. Soybean is a healthy and rich source of protein for both animals and humans (e.g., tofu). Nearly all soybeans are processed into oils, many industrial products such as lubricants, solvents, cleaners and paints. Soybeans are also used for animal feed, biodiesel, cleaning product and candles. Thus, it is important to study the production of corn and soybean in Iowa; in fact, corn and soybeans are the bread basket of the U.S. NASS has an enormous amount of data, but these data are highly confidential and are kept under strict surveillance. Thus, we are using the much-used data set reported in Battese *et al*. (1988) as surrogates to study the production of corn and soybeans.

This application concerns the estimation of areas under corn and soybeans for each of twelve counties in north-central Iowa using farm-interview data in conjunction with LANDSAT satellite data. Each county was divided into area segments, and the areas under corn and soybeans were ascertained for a sample of segments by interviewing farm operators. While the data are presented in Table 1 of BHF, we note that the sample sizes and county sizes are Cerro Gordo (1, 545), Hamilton (1, 566), Worth (1, 394), Humbolt (2, 424), Franklin (3, 564), Pocahontas (3, 570), Winnebago (3, 402), Wright (3, 567), Webster (4, 687), Hancock (5, 569), Kossuth (5, 965) and Hardin (6, 556). Auxiliary data in the form of number of pixels (a term used for "picture elements" of about 0.45 hectares) classified as corn and soybeans were also obtained for all the area segments, including the sample segments, in each county using the LANDSAT satellite readings. These data were first analyzed by Battese *et al*. (1988), and there are many other discussions of these data (e.g., Rashid and Nandram 1998). For each county we present the sample (average, standard deviation) below. For corn these summaries are

(106.8, –), (96.3, –), (76.1, –), (150.9, 40.5),
(158.6, 16.4), (102.5, 50.3),

(112.8, 10.5), (144.3, 52.0), (117.6, 23.5),
(109.4, 17.2), (110.3, 21.0), (114.8, 38.0)

and for soybeans these summaries are

(8.09, –), (106.0, –), (103.6, –), (35.2, 48.7),
(52.5, 5.7), (118.7, 43.4),

(88.6, 30.5), (97.8, 54.0), (113.0, 21.3), (117.5, 15.7),
(117.8, 12.1), (89.8, 35.3)

Thus, for both corn and soybeans the sample averages and sample standard deviations vary considerably with similar patterns for corn and soybeans. Without further evidence it may be unreasonable to assume that the sampling variances are equal (i.e., the assumption of homogeneity of sampling variances seems questionable).

The problem of small area estimation is to provide precise estimates or predictions of means or other quantities of interest from areas whose sample sizes are too small to yield reliable direct estimates. To overcome this issue, models that would define ways to "borrow strength" from the ensemble are used in estimation. But when models are used, the combined estimates from all the small areas do not often match the direct estimate on the large area obtained when the small areas are collapsed into a single area. To ensure that the combined estimate matches the direct estimate, we apply a constraint which forces the 'total' of the small areas to match the 'grand total'. This technique is called benchmarking which shifts the small area estimators to accommodate the benchmark constraint. It is true that benchmarking offers some protection against possible model failure, and it is likely to improve on the design-based bias properties of small area estimators. This makes the benchmarking technique desirable to practitioners of model-assisted small area estimation.

However, it is not clear how benchmarking affects precision of the small area estimators. In one small area model the precision can increase, and in a different model the precision can decrease; even in the same model the precision can decrease for some areas and increase for others. Nandram *et al*. (2010) have seen an increase in precision for all areas under external benchmarking. The reason for this is that under external

benchmarking new information is incorporated in the model due to the benchmarking constraint. Toto and Nandram (2010) have found that under internal benchmarking there is an increase in precision of the small area estimators when survey weights are not used, but there are increases for some areas and decreases for others when survey weights are used. There are also some theoretical results in both Nandram *et al*. (2010) and Toto and Nandram (2010) that show increases in precision when some of the model parameters are known. In any case, the change in precision is essentially small.

It is worth mentioning that benchmarking has been a long standing problem in the time series context. Hillmer and Trabelsi (1987) defined and developed a sound statistical theory to provide a generalized benchmarking procedure; for the historical development of this benchmarking problem, see the references cited in Hillmer and Trabelsi (1987); see also Trabelsi and Hillmer (1990). Recently, Pfeffermann and Tiller (2006) looked at this problem in small area estimation with state-space models subject to benchmark constraints. Specifically, they showed how to benchmark small-area estimators, produced by fitting separate state-space models, to aggregates of survey direct estimators within a group of areas. See also Park *et al*. (2006) for an interesting example on a mortality time series model in which the estimated numbers of deaths for different causes are constrained to add up to the overall number of deaths for all causes. However, we consider a single characteristic of several areas at a particular point in time.

Benchmarking in small-area estimation at a single time point is interesting in its own right. Within the hierarchical Bayes approach, You *et al*. (2004) studied benchmarked estimators for small area estimation based on unmatched sampling and linking models proposed by You and Rao (2002). They applied this approach to under coverage estimation for the ten provinces across Canada for the 1991 Canadian Census. Finally, we also note that Dick (2001) used an empirical Bayes, as opposed to a hierarchical Bayes, procedure to calibrate the direct estimates for small areas to larger areas in the 2001 Canadian Census. However, the mean squared error approximation that Dick (2001) used did not account for the calibration. Sampling variances can be studied quite easily within the hierarchical Bayes approach as is now well known.

Recently, there has been increased activity in benchmarking small area estimates. Wang *et al*. (2008) gave a characterization of the best linear unbiased predictor (BLUP) for small area means under an area level model that satisfies a benchmarking constraint and minimizes the loss function criterion that all linear unbiased predictors satisfy. They also presented an alternative way to incorporate the benchmarking constraint such that the BLUP estimator would have a self-calibrated property (discussed in You and Rao 2002). Wang *et al*. (2008) used an approach in which the weights are included in an augmented model. Their proposed self-calibrated augmented model reduces bias both at the overall and small area level. However, this work does not predict finite population means, and so their benchmarking constraint is different from ours. Datta *et al*. (2009) developed a class of constrained empirical Bayes estimators for area-level models. This is a decision-theoretic approach and uses an empirical Bayes analysis to benchmark small areas with only area-level data. It is interesting that they also showed that the standard raking procedure arises as a special case from their procedure. Again, this work does not use Bayesian predictive inference.

Motivated by the work of You and Rao (2002), we have done some research in Bayesian predictive inference for benchmarking finite population means. Nandram *et al*. (2010) benchmarked finite population means to a specified value, and Toto and Nandram (2010) benchmarked the finite population means to a Horvitz-Thompson estimator of the entire finite population mean. For the data on corn and soybeans, the design is self-weighting, and so that it is not appropriate to discuss survey weights.

The goal of this paper is to make inference about the *finite population mean of the* i^{th} county in Iowa, $\bar{Y}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} y_{ij}$, $i = 1, ..., l$, based on the observed values $\mathbf{y}_s$ under a nested-error regression model. To estimate $\bar{Y}_i$ we use a standard predictor $\hat{\bar{Y}}_i$, $i = 1, 2, ..., l$, which is to be determined. With this notation, a predictor of $\bar{Y} = \sum_{i=1}^{l} \frac{N_i \bar{Y}_i}{N}$ is $\hat{\bar{Y}} = \sum_{i=1}^{l} \frac{N_i \hat{\bar{Y}}_i}{N}$. Under

simple random sampling, a design-consistent estimator of $\bar{Y}$ is $\bar{y}_s = \sum_{i=1}^{l} \frac{n_i \bar{y}_{s_i}}{n}$. So, a practitioner might want to use $\bar{y}_s$ to predict $\bar{Y}$. It would be reasonable to expect that $\hat{\bar{Y}} = \bar{y}_s$. Hence, this is a possible benchmarking constraint. We will denote our benchmark constraint as $\bar{Y} = \bar{y}_s$, or equivalently,

$$\frac{1}{N}\sum_{i=1}^{l}\sum_{j=1}^{N_i} y_{ij} = \frac{1}{n}\sum_{i=1}^{l}\sum_{j=1}^{n_i} y_{ij} \quad (1)$$

where it should be understood that in the left-hand side (LHS) of (1), y_{ij}, $i = 1, ..., l$; $j = 1, ..., n_i$, are observed, and y_{ij}, $i = 1, ..., l$; $j = n_i + 1, ..., N_i$, are to be predicted.

Statisticians have been working on the one-way random effects model with different sampling variances for a long time; see Rao (1970) for the general fixed effects linear model with unequal sampling variances, Rao (1980) for a method to estimate the common mean of normal populations with unequal sampling variances, and Rao *et al*. (1981) for a study of many methods to estimate all parameters with unequal sampling variances. Nandram and Sedransk (1993) used a sequence of surveys to estimate the finite population mean of the final occasion. The work of Nandram and Sedransk (1993) was motivated by a similar work of Ghosh and Meeden (1986) who used equal sampling variances. Nandram and Sedransk (1993) model the unequal sampling variances using an inverse gamma distribution within the empirical Bayes framework. Other empirical Bayes methods were discussed by Kleffe and Rao (1992), Arora *et al*. (1997) and Arora and Lahiri (1997). While these authors studied the point estimators (best linear unbiased predictors) and their standard errors, Nandram (1999) discussed how to obtain empirical Bayes confidence intervals for small area means. More recently there has been a very interesting approach to this problem (Hedeker *et al*. 2008). These authors actually model mean response and the sampling variances using two correlated mixed effects models. They consider a longitudinal model with a relatively large number of time points (corresponding to the number of individuals sampled within an area). However, their model will be dificult to realize in small areas.

The rest of the paper is organized as follows. In Section 2, we describe the benchmarking Bayesian model under homogeneous sampling variances. We present key results on the joint density of the benchmarking Bayesian model. We compare the inference from the nonbenchmarking and the benchmarking models. In Section 3, we describe the benchmarking Bayesian model under heterogeneous sampling variances. We discuss the griddy Gibbs sampler to fit the nonbenchmarking model, and use the sampling importance resampling algorithm to subsample iterates from the nonbenchmarking model to fit the benchmarking model. We compare inference from the nonbenchmarking and the benchmarking models. In Section 4, we apply our methods to the LANDSAT satellite data, and we compare posterior inference for the nonbenchmarking model and benchmarking model under homogeneity and heterogeneity. Section 5 has concluding remarks.

2. HOMOGENEOUS SAMPLING VARIANCE

In this section, we describe our proposed method to estimate finite population means under the nested-error regression model using the Bayesian benchmarking approach. We present three key results. We describe the joint density of all values conditional on all parameters. As we will see, this joint density turns out to be multivariate normal, which makes it possible to compare the nonbenchmarking and benchmarking models. We also show that the joint posterior density is proper. Lastly, we describe the distribution of the finite population means under the model conditional on all parameters. Also, the proof of propriety of the joint posterior density gives a simple procedure for obtaining samples of the parameters to be used in obtaining samples of the finite population mean.

Let us begin with some basic notations used throughout the section. We will denote by $\mathbf{1}_k$ a column vector of size k with each of its elements being unity. Also, we denote the identity matrix of size k by $\mathbf{I}_k$ and the $r \times s$ matrix with each of its elements being unity by $\mathbf{J}_s^r$. If $r = s$, we write $\mathbf{J}_r$.

Assume that there are l counties (small areas). For the i^{th} small area, $i = 1, 2, ..., l$, let N_i and n_i be the fixed known population and the sample sizes, respectively. So, $f_i = \frac{n_i}{N_i}$, $i = 1, 2, ..., l$, are known

sampling fractions. Letting $n = \sum_{i=1}^{l} n_i$ be the total sample size and $N = \sum_{i=1}^{l} N_i$ be the total population size, the overall sampling fraction is given by $f = \frac{n}{N}$. Let y_{ij} denote the value for the j^{th} unit within the i^{th} area, $i = 1, ..., l; j = 1, ..., N_i$. We assume that y_{ij}, $i = 1, ..., l; j = 1, \ldots, n_i$, are observed, and write $\mathbf{y}_s = \left(\mathbf{y}'_{s_1}, ..., \mathbf{y}'_{s_l}\right)'$ with $\mathbf{y}'_{s_i} = (y_{i1}, ..., y_{in_i})'$ for $i = 1, 2, ..., l$. Since n_i values are observed, the values of the remaining $N_i - n_i$ units are unknown, and we write the vector of unobserved values as $\mathbf{y}_{ns} = \left(\mathbf{y}'_{ns_1}, ..., \mathbf{y}'_{ns_l}\right)'$ with $\mathbf{y}_{ns_i} = (y_{i, n_i+1}, ..., y_{i,N_i})'$ for $i = 1, 2, ..., l$.

Let $\mathbf{y}_{(N)} = \left(\mathbf{y}'_s, \mathbf{y}_{ns(N)'}\right)'$, where $\mathbf{y}_{ns(N)}$ is $\mathbf{y}_{ns}$ excluding y_{l,N_l} from the l^{th} area. That is, $\mathbf{y}_{ns(N)} = \left(\mathbf{y}'_{ns_1}, ..., \mathbf{y}'_{ns_l(N)}\right)'$ with $\mathbf{y}_{ns_i} = \left(y_{i,n_i+1}, ..., y_{i,N_i}\right)'$ for $i = 1, 2, ..., l - 1$, and $\mathbf{y}_{ns_l(N)} = \left(y_{l,n_l+1}, ..., y_{l,N_l-1}\right)'$. Also, let $\mu_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_i\mathbf{v}$ where $\mathbf{x}_{ij}$ is the $p \times 1$ vector of covariates for the j^{th} unit in the i^{th} area for $i = 1, 2, ..., l$, $j = 1, 2, ..., N_i$, and $\mathbf{z}_i$ is the $l \times 1$ vector with unity on the i^{th} entry for $i = 1, 2, ..., l$. Moreover, write $\boldsymbol{\mu} = \left(\boldsymbol{\mu}_{(N)'}, \mu_{l,N_l}\right)'$, where $\boldsymbol{\mu}_{(N)} = \left(\boldsymbol{\mu}'_s, \boldsymbol{\mu}_{ns(N)'}\right)'$ and $\boldsymbol{\mu}_s = \left(\boldsymbol{\mu}'_{s_1}, ..., \boldsymbol{\mu}'_{s_l}\right)'$ with $\boldsymbol{\mu}_{s_i} = \left(\mu_{i1}, \ldots, \mu_{in_i}\right)'$ for $i = 1, 2, ..., l$, and $\boldsymbol{\mu}_{ns(N)} = \left(\boldsymbol{\mu}'_{ns_1}, ..., \boldsymbol{\mu}'_{ns_l}\right)'$ with $\boldsymbol{\mu}_{ns_i} = \left(\mu_{i,n_i+1}, ..., \mu_{i,N_i}\right)'$ for $i = 1, 2, \ldots, l - 1$, and $\boldsymbol{\mu}_{ns_l} = \left(\mu_{l,n_l+1}, \ldots, \mu_{l,N_l-1}\right)'$.

Let $\mathbf{X}_s$ be the $n \times p$ design matrix of covariates for the sampled observations $\mathbf{y}_s$, and $\mathbf{Z}_s$ be the $n \times l$ block diagonal matrix with $\mathbf{1}_{n_i}$ on the i^{th} diagonal, $i = 1, 2, ..., l$. Moreover, let $\mathbf{X}_{ns}$ be the $(N - n) \times p$ matrix of covariates for the nonsampled observations $\mathbf{y}_{ns}$, and $\mathbf{Z}_{ns}$ be the $(N - n) \times l$ block diagonal matrix with $\mathbf{1}_{N_i - n_i}$ on the i^{th} diagonal, $i = 1, 2, ..., l$. Similarly, denote the $(N - 1 - n) \times p$ matrix of covariates for the nonsampled observations without y_{l,N_l} by $\mathbf{X}_{ns(N)}$, and the $(N - 1 - n) \times l$ corresponding block diagonal matrix with $\mathbf{1}_{N_i - n_i}$ on the i^{th} diagonal entry, $i = 1, 2, ..., l - 1$, and $\mathbf{1}_{N_l - n_l - 1}$ on the l^{th} diagonal entry by $\mathbf{Z}_{ns(N)}$. We write the sample mean for the i^{th} area as $\bar{y}_{s_i} = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$, and $\bar{\mathbf{x}}_{s_i} = \frac{1}{n_i}\sum_{j=1}^{n_i}\mathbf{x}_{ij}$, $\bar{\mathbf{x}}_{ns_i} = \frac{1}{N_i - n_i}\sum_{j=n_i+1}^{N_i}\mathbf{x}_{ij}$ and $\bar{\mathbf{x}}_i = \frac{1}{N_i}\sum_{j=1}^{N_i}\mathbf{x}_{ij}$ as the $p \times 1$ vectors of averages of the i^{th} sample, nonsample, and small area population covariates, respectively.

2.1 Models

Our nonbenchmarking (NBM) model is a Bayesian version of the small area estimation model of Battese *et al*. (1988) for finite population mean that accommodates covariates at the unit level. The Bayesian nonbenchmarking (NBM) model is given by

$$y_{ij}\left|\boldsymbol{\beta}, \mathbf{v}, \sigma^2 \overset{ind}{\sim} \text{Normal}\left(\mathbf{x}'_{ij}\boldsymbol{\beta} + v_i, \sigma^2\right)\right. \quad i = 1, ..., l; j = 1, ..., N_i \tag{2}$$

$$v_i\left|\rho, \sigma^2 \overset{iid}{\sim} \text{Normal}\left\{0, \left(\frac{\rho}{1-\rho}\right)\sigma^2\right\}\right. \tag{3}$$

$$p\left(\boldsymbol{\beta}, \sigma^2, \rho\right) \propto \frac{1}{\sigma^2},\ \sigma^2 > 0, \text{ and } 0 < \rho < 1 \tag{4}$$

where ρ is the intra-class correlation within areas and is the same for each area. Using Bayes' theorem in (2) - (4), the joint posterior density of the nonbenchmarking model is

$$\pi\left(\mathbf{v}, \boldsymbol{\beta}, \sigma^2, \rho|\mathbf{y}_s\right) \propto \left(\frac{1}{\sigma^2}\right)^{1+(n+l)/2}\left(\frac{1-\rho}{\rho}\right)^{l/2}$$

$$\times \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{l}\left\{\frac{n_i}{\lambda_i}\left(v_i - \lambda_i\left(\bar{y}_{s_i} - \bar{\mathbf{x}}'_{s_i}\boldsymbol{\beta}\right)\right)\right.\right.\right.^2$$

$$+\lambda_i\left(\frac{1-\rho}{\rho}\right)\left(\bar{y}_{s_i}-\bar{\mathbf{x}}_{s_i}'\boldsymbol{\beta}\right)^2$$

$$+\sum_{j=1}^{n_i}\left[\left(y_{ij}-\bar{y}_{s_i}\right)-\left(\mathbf{x}_{ij}-\bar{\mathbf{x}}_{s_i}\right)'\boldsymbol{\beta}\right]^2\Bigg\}\Bigg]\Bigg\}$$

where $\lambda_i = \dfrac{\rho n_i}{\rho n_i + (1-\rho)}$ for i =1, 2, ..., l. It can be shown that this joint posterior density is proper.

To obtain more reliable estimators, it is desirable to benchmark the estimators of the $\bar{Y}_i$, i =1, 2, ..., l. So, we now discuss the benchmarking model. It is obtained by incorporating the benchmarking constraint $\dfrac{1}{N}\sum_{i=1}^{l}\sum_{j=1}^{N_i} y_{ij} = \dfrac{1}{n}\sum_{i=1}^{l}\sum_{j=1}^{n_i} y_{ij}$ to the nonbenchmarking model. The resulting adjusted model, which we describe as the Bayesian benchmarking (BM) model, is given by:

$$y_{ij}\,|\boldsymbol{\beta},\mathbf{v},\sigma^2 \overset{ind}{\sim} \text{Normal}\left(\mathbf{x}_{ij}\boldsymbol{\beta}+v_i,\sigma^2\right),$$

$$\frac{1}{N}\sum_{i=1}^{l}\sum_{j=1}^{N_i} y_{ij} = \frac{1}{n}\sum_{i=1}^{l}\sum_{j=1}^{n_i} y_{ij} \tag{5}$$

$$v_i\,|\,\rho,\sigma^2 \overset{iid}{\sim} \text{Normal}\left(0,\left(\frac{\rho}{1-\rho}\right)\sigma^2\right),\ 0<\rho<1 \tag{6}$$

$$p\left(\boldsymbol{\beta},\sigma^2,\rho\right) \propto \frac{1}{\sigma^2},\ \sigma^2>0 \tag{7}$$

To incorporate the benchmarking constraint, throughout we condition on ϕ = 0.

Lemma 2.1 Under the benchmarking model, conditional on $\phi = \dfrac{1}{N}\sum_{i=1}^{l}\sum_{j=1}^{N_i} y_{ij} - \dfrac{1}{n}\sum_{i=1}^{l}\sum_{j=1}^{n_i} y_{ij} = 0$,

$$\mathbf{y}_{(N)}\,|\boldsymbol{\mu},\sigma^2,\phi=0 \sim \text{Normal}\Bigg(\boldsymbol{\mu}_{(N)}+f\left(\frac{1}{N-n}\right)\left(\mu_{l,N_l}+\mathbf{a}'\boldsymbol{\mu}_{(N)}\right)\mathbf{a},\ \sigma^2\left(\mathbf{I}_{N-1}-\mathbf{D}\right)\Bigg)$$

and

$$y_{l,N_l} = (N-n)\,\bar{y}_s - \mathbf{1}'_{N-n}\,\mathbf{y}_{ns}$$

Lemma 2.1 is a simple result that incorporates the benchmark constraint into the distribution of the data given the parameters. A proof is given in the Appendix. Note that from the resulting density in Lemma 2.1, with

$$\mathbf{D} = \begin{bmatrix} \dfrac{1}{n}(1-f)\,\mathbf{J}_n & -\dfrac{1}{N}\mathbf{J}^n_{(N-1-n)} \\ -\dfrac{1}{N}\mathbf{J}_n^{(N-1-n)} & f\left(\dfrac{1}{N-n}\right)\mathbf{J}_{(N-1-n)} \end{bmatrix}$$

the benchmark constraint caused the y_{ij}'s to be correlated. Henceforth, it is convenient to drop the conditioning on ϕ = 0 although it must be understood that this conditioning exists.

Applying the marginal distribution property of the multivariate normal density on Lemma 2.1, we find that $\mathbf{y}_s\,|\mathbf{v},\boldsymbol{\beta},\sigma^2$ follows a normal distribution with

$$E\left[\mathbf{y}_s\,|\mathbf{v},\boldsymbol{\beta},\sigma^2\right] = \mathbf{C}_x\boldsymbol{\beta}+\mathbf{C}_z\mathbf{v} \text{ and}$$

$$\text{Var}\left[\mathbf{y}_s\,|\mathbf{v},\boldsymbol{\beta},\sigma^2\right] = \sigma^2\left[\mathbf{I}_n-\frac{1}{n}(1-f)\,\mathbf{J}_n\right] \tag{8}$$

where $\mathbf{C}_x = \left[\mathbf{I}_n-\dfrac{1}{n}(1-f)\,\mathbf{J}_n\right]\mathbf{X}_s+\left[\dfrac{1}{N}\mathbf{J}^n_{N-n}\right]\mathbf{X}_{ns}$

and $\mathbf{C}_z = \left[\mathbf{I}_n-\dfrac{1}{n}(1-f)\,\mathbf{J}_n\right]\mathbf{Z}_s+\left[\dfrac{1}{N}\mathbf{J}^n_{N-n}\right]\mathbf{Z}_{ns}$

From (8) the likelihood function is given by

$$p\left(\mathbf{y}_s\,|\mathbf{v},\boldsymbol{\beta},\sigma^2,\rho\right)$$

$$\propto \left(\frac{1}{\sigma^2}\right)^{n/2}\exp\Bigg\{-\frac{1}{2\sigma^2}\Bigg[\left(\mathbf{y}_s-\left[\mathbf{C}_x\boldsymbol{\beta}+\mathbf{C}_z\mathbf{v}\right]\right)'$$

$$\left(\mathbf{I}_n-\frac{1}{n}(1-f)\mathbf{J}_n\right)^{-1}\left(\mathbf{y}_s-\left[\mathbf{C}_x\boldsymbol{\beta}+\mathbf{C}_z\mathbf{v}\right]\right)\Bigg]\Bigg\}$$

Now, applying Bayes' theorem on $p\left(\mathbf{y}_s\,|\mathbf{v},\boldsymbol{\beta},\sigma^2,\rho\right)$ with the distribution of $\mathbf{v}|\sigma^2$, ρ in (6), and prior

distribution given in (7), we get the joint posterior density

$$\pi\left(\mathbf{v}, \boldsymbol{\beta}, \sigma^2, \rho|\mathbf{y}_s\right) \propto \left(\frac{1}{\sigma^2}\right)^{1+(n+l)/2} \left(\frac{1-\rho}{\rho}\right)^{l/2}$$

$$\times \exp\left\{-\frac{1}{2\sigma^2}\left(\frac{1-\rho}{\rho}\right)\left[\left(\mathbf{v}-\mathbf{B}^{-1}\mathbf{C}_z'\mathbf{A}\left(\mathbf{y}_s-\mathbf{C}_x\boldsymbol{\beta}\right)\right)'\right.\right.$$

$$\left.\left.\left[\mathbf{I}_l+\mathbf{C}_z'\mathbf{A}\mathbf{C}_z\right]\left(\mathbf{v}-\mathbf{B}^{-1}\mathbf{C}_z'\mathbf{A}\left(\mathbf{y}_s-\mathbf{C}_x\boldsymbol{\beta}\right)\right)\right]\right\}$$

$$\times \exp\left\{-\frac{1}{2\sigma^2}\left(\frac{1-\rho}{\rho}\right)\left[\left(\mathbf{y}_s-\mathbf{C}_x\boldsymbol{\beta}\right)'\right.\right.$$

$$\left.\left.\left(\mathbf{A}-\mathbf{A}\mathbf{C}_z\mathbf{B}^{-1}\mathbf{C}_z'\mathbf{A}\right)\left(\mathbf{y}_s-\mathbf{C}_x\boldsymbol{\beta}\right)\right]\right\}$$

where

$$\mathbf{B}=\mathbf{I}_l+\mathbf{C}_z'\mathbf{C}\mathbf{A}_z,\ \mathbf{A}=\left[\left(\frac{1-\rho}{\rho}\right)\left(\mathbf{I}_n-\frac{1}{n}(1-f)\,\mathbf{J}_n\right)\right]^{-1},$$

and $\mathbf{C}_x$ and $\mathbf{C}_z$ are as defined in (8).

Lemma 2.2 Under the benchmarking model, $\pi\left(\mathbf{v}, \boldsymbol{\beta}, \sigma^2, \rho|\mathbf{y}_s\right)$ is proper.

Lemma 2.2 shows that the addition of the benchmarking constraint to the model does not affect the propriety of the joint posterior distribution. To prove Lemma 2.2, we use the multiplication rule

$$\pi\left(\mathbf{v}, \boldsymbol{\beta}, \sigma^2,\ \rho|\mathbf{y}_s\right) = \pi_1\left(\mathbf{v}|\boldsymbol{\beta}, \sigma^2,\ \rho, \mathbf{y}_s\right)$$

$$\pi_2\left(\boldsymbol{\beta}|\sigma^2, \rho, \mathbf{y}_s\right)\ \pi_3\left(\sigma^2|\rho, \mathbf{y}_s\right)\ \pi_4\left(\rho|\mathbf{y}_s\right)$$

and show that $\pi_1\left(\mathbf{v}|\boldsymbol{\beta}, \sigma^2, \rho, \mathbf{y}_s\right)$, $\pi_2\left(\boldsymbol{\beta}|\sigma^2, \rho, \mathbf{y}_s\right)$, $\pi_3\left(\sigma^2|\rho, \mathbf{y}_s\right)$ and $\pi_4\left(\rho|\mathbf{y}_s\right)$, are all proper densities. A proof is given in the Appendix.

To make posterior inferences about $\bar{Y}_i$ (i = 1, ..., l), given the sample observations $\mathbf{y}_s$, we take samples from the distribution $\bar{Y}_i|\mathbf{y}_s$. Theorem 2.1 below gives us the distribution of the $\bar{Y}_i|\mathbf{y}_s$.

Theorem 2.1 Under the benchmarking model,

$$p\left(\bar{Y}_i|\mathbf{y}_s\right) = \int p\left(\bar{Y}_i|\mathbf{y}_s,\ \mathbf{v}, \boldsymbol{\beta}, \sigma^2\right)\pi\left(\mathbf{v}, \boldsymbol{\beta}, \sigma^2|\mathbf{y}_s\right)$$

$$d\mathbf{v}d\boldsymbol{\beta}d\sigma^2,\ i = 1, ..., l \qquad (9)$$

where

$$\bar{Y}_i|\mathbf{y}_s, \mathbf{v}, \boldsymbol{\beta},\ \sigma^2$$

$$\sim \text{Normal}\left\{f_i\bar{y}_{s_i} + (1-f_i)\,\bar{\mathbf{x}}_{ns_i}'\boldsymbol{\beta} + (1-f_i)\,\mathbf{z}_i'\mathbf{v} + A_i,\right.$$

$$\left.\frac{\sigma^2}{N_i}(1-f_i)\,V_i\right\}$$

$$A_i = (1-f_i)\left\{\bar{y}_s - \bar{\mathbf{x}}_{ns}'\boldsymbol{\beta} - \frac{1}{N-n}(\mathbf{1}_{N-n}'\mathbf{Z}_{ns})\,\mathbf{v}\right\}$$

and $V_i = 1 - \dfrac{N_i - n_i}{N-n}$, i = 1, ..., l

Proof of Theorem 2.1

First, note that

$$p\left(\bar{Y}_i|\mathbf{y}_s, \mathbf{v}, \boldsymbol{\beta}, \sigma^2, \rho\right) = p\left(\bar{Y}_i|\mathbf{y}_s, \mathbf{v}, \boldsymbol{\beta}, \sigma^2\right)$$

and $\bar{Y}_i$ can be written as

$$\bar{Y}_i = \frac{1}{N_i}\left(n_i\bar{y}_{s_i} + \left(N_i - n_i\right)\bar{y}_{ns_i}\right)$$

$$= \frac{1}{N_i}\left(\mathbf{1}_{n_i}'\mathbf{y}_{s_i} + \mathbf{1}_{N_i-n_i}'\mathbf{y}_{ns_i}\right),\ i = 1, 2, ..., l$$

Using the conditional distribution property of the multivariate normal density in Lemma 2.1, $\mathbf{y}_{ns_{(N)}}\left|\mathbf{v}, \boldsymbol{\beta}, \sigma^2, \mathbf{y}_s\right.$ has a multivariate normal distribution with

$$E\left[\mathbf{y}_{ns_{(N)}}\ |\mathbf{v}, \boldsymbol{\beta}, \sigma^2, \mathbf{y}_s\right]$$

$$= \bar{y}_s\mathbf{1}_{N-1-n} + \left(\mathbf{X}_{ns_{(N)}} - \left(\frac{1}{N-n}\right)\mathbf{J}_{N-n}^{N-1-n}\mathbf{X}_{ns}\right)\boldsymbol{\beta}$$

$$+\left(\mathbf{Z}_{ns_{(N)}} - \left(\frac{1}{N-n}\right)\mathbf{J}_{N-n}^{N-1-n}\mathbf{Z}_{ns}\right)\mathbf{v}$$

and

$$\mathrm{Var}\left[\mathbf{y}_{ns_{(N)}} \,\middle|\, \mathbf{v}, \boldsymbol{\beta}, \sigma^2, \mathbf{y}_s\right] = \sigma^2\left(\mathbf{I}_{N-1-n} - \left(\frac{1}{N-n}\right)\mathbf{J}_{N-1-n}\right)$$

It follows that for i =1, ...,l – 1,

$$\bar{Y}_i \,|\, \mathbf{y}_s, \mathbf{v}, \boldsymbol{\beta}, \sigma^2 \sim \text{Normal}\left\{ f_i \bar{y}_{s_i} + (1-f_i)\,\bar{y}_s + \mathbf{B}_{x_i}\boldsymbol{\beta} + \mathbf{B}_{z_i}\mathbf{v},\ \frac{\sigma^2}{N_i}(1-f_i)\left(1-\frac{N_i-n_i}{N-n}\right)\right\}$$

where

$$\mathbf{B}_{x_i} = \frac{1}{N_i}\left(\mathbf{1}'_{N_i-n_i}\mathbf{X}_{ns_i}\right) - \frac{1}{N_i}\left(\frac{N_i-n_i}{N-n}\right)(\mathbf{1}'_{N-n}\mathbf{X}_{ns})$$

and

$$\mathbf{B}_{z_i} = \frac{1}{N_i}\left(\mathbf{1}'_{N_i-n_i}\mathbf{Z}_{ns_i}\right) - \frac{1}{N_i}\left(\frac{N_i-n_i}{N-n}\right)\left(\mathbf{1}'_{N-n}\mathbf{Z}_{ns}\right)$$

It is now easy to show that, using the benchmark constraint together with the distributions of $\bar{Y}_i \,|\, \mathbf{y}_s, \mathbf{v}, \boldsymbol{\beta}$, σ^2, i = 1, i = 1, ..., l – 1, $\bar{Y}_l \,|\, \mathbf{y}_s, \mathbf{v}, \boldsymbol{\beta}$, σ^2 has the same distributional form.

After some simplification, we get

$$\bar{Y}_i \,|\, \mathbf{y}_s, \mathbf{v}, \boldsymbol{\beta}, \sigma^2 \sim \text{Normal}\left\{ f_i \bar{y}_{s_i} + (1-f_i)\,\bar{\mathbf{x}}'_{ns_i}\boldsymbol{\beta} + (1-f_i)\,\mathbf{z}_i'\mathbf{v} + A_i,\ \frac{\sigma^2}{N_i}(1-f_i)\,V_i\right\}$$

where $A_i = (1-f_i)\left(\bar{y}_s - \bar{\mathbf{x}}'_{ns}\boldsymbol{\beta} - \frac{1}{N-n}\left(\mathbf{1}'_{N-n}\mathbf{Z}_{ns}\right)\mathbf{v}\right)$

and $V_i = \left(1-\left(\frac{N_i-n_i}{N-n}\right)\right)$, i = 1, ..., l.

Finally, note that this proof also shows that $p\left(\bar{Y}_i \,|\, \mathbf{y}_s\right)$ is proper because $p\left(\bar{Y}_i \,|\, \mathbf{y}_s, \mathbf{v}, \boldsymbol{\beta}, \sigma^2\right)$ is a probability density, and by Lemma 2.2 $\pi\left(\mathbf{v}, \boldsymbol{\beta}, \sigma^2 \,|\, \mathbf{y}_s\right)$ is proper.

We now discuss the effects of the benchmarking. Note as compared to Theorem 2.1 that under the nonbenchmarking model

$$\bar{Y}_i \,|\, \mathbf{y}_s, \mathbf{v}, \boldsymbol{\beta}, \sigma^2 \sim \text{Normal}\left\{ f_i \bar{y}_{s_i} + (1-f_i)\,\bar{\mathbf{x}}'_{ns_i}\boldsymbol{\beta} + (1-f_i)\,\mathbf{z}_i'\mathbf{v},\ \frac{\sigma^2}{N_i}(1-f_i)\right\}$$

First, there is an overall adjustment A_i, i =1, 2, ...,l, to the expected value in the distribution of $\bar{Y}_i \,|\, \mathbf{y}_s, \mathbf{v}, \boldsymbol{\beta}$, σ^2 under the benchmarking model, with respect to the nonbenchmarking model. This adjustment A_i is different for each small area, since f_i depends on the sample and population sizes in each area, unlike Nandram *et al*. (2010), where the overall adjustment is constant throughout all areas. Moreover, observe that in the benchmarking model, the variance in each small area gets reduced by different amounts V_i, i = 1, 2, ..., l. However, these small gains in precision might disappear for some areas when the variabilities of $\boldsymbol{\beta}, \mathbf{v}$ and σ^2 are incorporated.

2.2 Computations

To make inference about $\bar{Y}_i \,|\, \mathbf{y}_s$, our approach is to obtain samples from the posterior distribution $\mathbf{v}, \boldsymbol{\beta}, \sigma^2, \rho \,|\, \mathbf{y}_s$, and the distribution of $\bar{Y}_i \,|\, \mathbf{y}_s, \mathbf{v}, \boldsymbol{\beta}, \sigma^2$ and combine these samples to make inferences about the finite population means. The proof of propriety of the joint posterior density provided a prescription of how to draw samples from the posterior density. Note that by the multiplication rule,

$$\pi\left(\mathbf{v}, \boldsymbol{\beta}, \sigma^2, \rho \,|\, \mathbf{y}_s\right) = \pi_1\left(\mathbf{v} \,|\, \boldsymbol{\beta}, \sigma^2, \rho, \mathbf{y}_s\right) \pi_2\left(\boldsymbol{\beta} \,|\, \sigma^2, \rho, \mathbf{y}_s\right) \pi_3\left(\sigma^2 \,|\, \rho, \mathbf{y}_s\right) \pi_4\left(\rho \,|\, \mathbf{y}_s\right)$$

and since the conditional densities $\pi_1\left(\mathbf{v}|\boldsymbol{\beta}, \sigma^2, \rho, \mathbf{y}_s\right)$, $\pi_2\left(\boldsymbol{\beta}|\sigma^2, \rho, \mathbf{y}_s\right)$ and $\pi_3\left(\sigma^2|\rho, \mathbf{y}_s\right)$ are known distributions, we can draw samples from $\pi\left(\mathbf{v}, \boldsymbol{\beta}, \sigma^2, \rho|\mathbf{y}_s\right)$ using the composition method. The respective densities are shown in the proof of Lemma 2.2 in the Appendix. We would first draw samples from $\rho|\mathbf{y}_s$. There is no need for a Markov Chain Monte Carlo (MCMC) procedure to do this computation. One can simply use a grid method to draw a sample from the distribution of $\rho|\mathbf{y}_s$. It is convenient to use the grid method because ρ is bounded in the interval (0,1), and the function $\pi_4\left(\rho|\mathbf{y}_s\right)$ is easy to compute for each ρ, $0 < \rho < 1$. With each of these samples of ρ, we draw a sample from $\sigma^2|\rho, \mathbf{y}_s$. Then, with the resulting pair of values of ρ and σ^2, we draw a sample from $\boldsymbol{\beta}|\sigma^2, \rho, \mathbf{y}_s$. Also, with the sampled values of ρ, σ^2 and $\boldsymbol{\beta}$, draw values of $\mathbf{v}$ from $\mathbf{v}|\boldsymbol{\beta}, \sigma^2, \rho, \mathbf{y}_s$. We apply this algorithm to draw samples in both the NBM and BM models. Once the parameters are obtained, we use these values to draw values of the finite population means for each of the areas from the distribution of $\bar{Y}_i|\mathbf{y}_s, \mathbf{v}, \boldsymbol{\beta}, \sigma^2$, $i = 1, ..., l-1$ and the benchmarking constraint to get $\bar{Y}_l$. To facilitate inference, we have generated M =10,000 iterates using our sampling method.

3. HETEROGENEOUS SAMPLINGVARIANCES

We describe how to benchmark the finite population means when the sample variances are unequal (heterogeneity). We compare the nonbenchmarking and the benchmarking models under heterogeneity with the corresponding ones under homogeneity. The problem with heterogeneous variances is much more complex than the one with homogeneity. Thus, we fit the benchmarking model with heterogeneity using the sampling importance resampling (SIR) algorithm to subsample samples drawn from the nonbenchmarking model with homogeneity. Even the nonbenchmarking model with heterogeneity is difficult to fit; we use the griddy Gibbs sampler (Ritter and Tanner 1992).

3.1 Models

For i =1, ...,l; j =1, ..., N_i, the Bayesian nonbenchmarking (NBM) model is

$$y_{ij}|\boldsymbol{\beta}, \mathbf{v}, \sigma_i^2 \overset{ind}{\sim} \text{Normal}\left(\mathbf{x}'_{ij}\boldsymbol{\beta} + v_i, \sigma_i^2\right) \quad (10)$$

$$v_i \mid \delta^2 \overset{iid}{\sim} \text{Normal } (0, \delta^2) \quad (11)$$

$$\sigma_i^{-2}|\alpha, \delta^2 \overset{iid}{\sim} \text{Gamma } \left(\alpha, (\alpha+1)/\delta^2\right) \quad (12)$$

$$\pi\left(\alpha, \boldsymbol{\beta}, \delta^2\right) \propto \frac{1}{\delta^2(1+\alpha)^2}, \ \alpha, \ \delta^2 > 0 \quad (13)$$

In (12) we have centered the gamma density on its mode which is taken to be δ^2, the variance of the v_i (11). This parsimony is useful in small area estimation because we have essentially reduced the number of parameters by one; see Nandram and Choi (2002) for a similar centering procedure for the nonignorable nonresponse problem.

The prior distribution in (13) is obtained as follows. We take independent prior distributions for α, $\boldsymbol{\beta}$ and δ^2 with $p(\alpha) = 1/(1 + \alpha)^2$, $\alpha > 0$, a proper density, $p(\boldsymbol{\beta}) = 1$ for $\boldsymbol{\beta}$, and $p(\delta^2) = 1/\delta^2$, $\delta^2 > 0$; the two latter densities being improper. We choose the prior density for α to be proper (see Daniels 1999 and Gelman 2006) because it is the most difficult parameter to estimate. Note that $\boldsymbol{\beta}$ in (10) is closely associated with the data, and δ^2 is in (11) and (12). These features can help to make the joint posterior density of all the parameters proper. However, a practitioner can use independent proper diffuse priors for $\boldsymbol{\beta}$ and δ^2 to ensure proper posterior density. This can be accomplished as follows. Letting $\hat{\boldsymbol{\beta}}$ denote the least squares estimator of $\boldsymbol{\beta}$ in (10) with $v_i = 0$, $i = 1, ..., l$ and $\hat{\Sigma}$ its estimated covariance matrix. Then, we can take $\boldsymbol{\beta} \sim$ Normal $(\hat{\boldsymbol{\beta}}, \kappa_o\hat{\Sigma})$ with $\kappa_0 \approx 100$. Also, letting $\hat{\delta}^2 = \sum_{i=1}^{l}\hat{v}_i\Big/(l-1)$ with $\hat{v}_i = \sum_{j=1}^{n_i}\left(y_{ij} - \mathbf{x}'_{ij}\hat{\boldsymbol{\beta}}\right)\Big/n_i$, $i = 1, ..., l$, one can take $\delta^2 \sim$ Inverse-Gamma $(l_o/2, l_o\hat{\delta}^2/2)$ with $l_o \approx 1$. This is just a minor change to our algorithm.

For i =1, ...,l; j = 1, ..., N_i, the Bayesian benchmarking (BM) model is

$$y_{ij}|\boldsymbol{\beta}, \mathbf{v}, \sigma_i^2 \overset{ind}{\sim} \text{Normal}\left(\mathbf{x}'_{ij}\boldsymbol{\beta} + v_i, \sigma_i^2\right),$$

$$\frac{1}{N}\sum_{i=1}^{l}\sum_{j=1}^{N_i} y_{ij} = \frac{1}{n}\sum_{i=1}^{l}\sum_{j=1}^{n_i} y_{ij} \quad (14)$$

$$v_i \mid \delta^2 \overset{iid}{\sim} \text{Normal } (0, \delta^2) \quad (15)$$

$$\sigma_i^{-2}|\alpha, \delta^2 \overset{iid}{\sim} \text{Gamma } \left(\alpha, (\alpha+1)/\delta^2\right) \quad (16)$$

$$\pi\left(\alpha, \boldsymbol{\beta}, \delta^2\right) \propto \frac{1}{\delta^2(1+\alpha)^2},\ \alpha,\ \delta^2 > 0 \quad (17)$$

Note again that the difference between the two models is the inclusion of the benchmarking constraint $\phi = \frac{1}{N}\sum_{i=1}^{l}\sum_{j=1}^{N_i} y_{ij} - \frac{1}{n}\sum_{i=1}^{l}\sum_{j=1}^{n_i} y_{ij}$. Note specifically that the joint prior distribution $\pi(\mathbf{v}, \boldsymbol{\sigma}^2, \boldsymbol{\beta}, \delta^2, \alpha)$ of $\mathbf{v}$, σ^2, $\boldsymbol{\beta}$, δ^2, α is the same under the nonbenchmarking and the benchmarking models.

We proceed in exactly the same way as in the homogeneous models to obtain the conditional distribution of $\mathbf{y}$ given $\phi = 0$. For i =1, ..., l; j =1, ..., N_i, let $\mu_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + v_i$ and $\mathbf{D}$ = diagonal (σ_i^2, i =1, ..., l, j =1, ..., N_i). Again, we arrange $\mathbf{y}$ such that the sample part $\mathbf{y}_s$ comes first and the nonsample part $\mathbf{y}_{ns}$ comes second. For convenience, we will not distinguish between the nonsample (*ns*) with and without the observation y_{l,N_l} as it will be clear when it is written. Let $\boldsymbol{\mu}_s$ $\boldsymbol{\mu}_{ns}$, $\mathbf{D}_s$ and $\mathbf{D}_{ns}$ have similar meanings. Finally, we define three quantities. Let $\psi^2 = \sum_{i=1}^{l}[\{(N/n-1)^2 - 1\}n_i + N_i]\ \sigma_i^2/\sigma_l^2$. Also, let $\mathbf{b}_s$ be defined by $b_i = (N/n-1)(\sigma_i/\sigma_l)/\psi$, i = 1, ..., l; j = 1, ..., n_i and $\mathbf{b}_{ns}$ by $b_i = -(\sigma_i/\sigma_l)/\psi$, i = 1, ..., $l-1$; $j = n_i + 1$, ..., N_i, $i = l$, $j = n_l + 1$, ..., $N_l - 1$. Finally,

$$d = -\left\{(N/n)\sum_{i=1}^{l}\sum_{j=1}^{n_i}\mu_{ij} - \sum_{i=1}^{l}\sum_{j=1}^{N_i}\mu_{ij}\right\}\Big/\sigma_l^2$$

Now using only (14) after extensive algebraic manipulation, we have shown that

$$\mathbf{y}_s|\boldsymbol{\beta}, \boldsymbol{\sigma}^2,\ \phi = 0 \sim \text{Normal } \left\{\boldsymbol{\mu}_s + d\mathbf{D}_s\mathbf{b}_s,\ \mathbf{D}_s\left(\mathbf{I}_s - \mathbf{b}_s\mathbf{b}'_s\right)\mathbf{D}_s\right\} \quad (18)$$

and

$$\mathbf{y}_{ns}|\mathbf{y}_s, \boldsymbol{\beta}, \boldsymbol{\sigma}^2,\ \phi = 0 \sim \text{Normal } \left\{\boldsymbol{\mu}_{ns} + \frac{d - \mathbf{b}_s\mathbf{D}_s^{-1}(\mathbf{y}_s - \boldsymbol{\mu}_s)}{1 - \mathbf{b}'_s\mathbf{b}_s}\mathbf{D}_{ns}\mathbf{b}_{ns},\ \mathbf{D}_{ns}\left(\mathbf{I}_{ns} - \frac{\mathbf{b}_{ns}\mathbf{b}'_{ns}}{1-\mathbf{b}'_s\mathbf{b}_s}\right)\mathbf{D}_{ns}\right\} \quad (19)$$

To avoid confusion, note that y_{l,N_l} is not dropped from the procedure, and conditional on

$$\phi = 0,\ y_{lN_l} = N\bar{y}_s - \sum_{i=1}^{l-1}\sum_{j=1}^{N_i} y_{ij} - \sum_{j=1}^{N_{l-1}} y_{lj}$$

Thus, conditional on $\mathbf{y}_s, \boldsymbol{\beta}, \boldsymbol{\sigma}^2$ ϕ = 0, y_{l,N_l} inherits all its properties from (19).

As for the nonbenchmarking model the corresponding equations have obvious forms

$$\mathbf{y}_s|\boldsymbol{\beta}, \boldsymbol{\sigma}^2 \sim \text{Normal}\left\{\boldsymbol{\mu}_s, \mathbf{D}_s\right\} \quad (20)$$

and

$$\mathbf{y}_{ns}|\boldsymbol{\beta}, \boldsymbol{\sigma}^2 \sim \text{Normal}\left\{\boldsymbol{\mu}_{ns}, \mathbf{D}_{ns}\right\} \quad (21)$$

The distributions (20) and (21) are much simpler than (18) and (19) corresponding to the benchmarking.

Thus, under the nonbenchmarking model the joint posterior density is

$$\pi(v, \boldsymbol{\sigma}^2, \boldsymbol{\beta}, \delta^2, \alpha|\mathbf{y}_s) \propto p(\mathbf{y}_s|\mathbf{v}, \boldsymbol{\sigma}^2, \boldsymbol{\beta})\,\pi(\mathbf{v}, \boldsymbol{\sigma}^2, \boldsymbol{\beta}, \delta^2, \alpha)$$

and under the benchmarking model the joint posterior density is

$$\pi(\mathbf{v}, \boldsymbol{\sigma}^2, \boldsymbol{\beta}, \delta^2, \alpha|\mathbf{y}_s, \phi = 0) \propto p(\mathbf{y}_s|\mathbf{v}, \boldsymbol{\sigma}^2, \boldsymbol{\beta}, \phi = 0)\,\pi(\mathbf{v}, \boldsymbol{\sigma}^2, \boldsymbol{\beta}, \delta^2, \alpha)$$

It is worth noting that ratio of two posterior densities does not depend on $\pi(\mathbf{v}, \boldsymbol{\sigma}^2, \boldsymbol{\beta}, \delta^2, \alpha)$. That is, the ratio is

$$R(\mathbf{v}, \boldsymbol{\sigma}^2, \boldsymbol{\beta}) \propto \frac{p(\mathbf{y}_s | \mathbf{v}, \boldsymbol{\sigma}^2, \boldsymbol{\beta}, \phi = 0)}{p(\mathbf{y}_s | \mathbf{v}, \boldsymbol{\sigma}^2, \boldsymbol{\beta})}$$

$R(\mathbf{v}, \boldsymbol{\sigma}^2, \boldsymbol{\beta})$ is used in the computations for subsampling.

3.2 Computations

Because the joint posterior density $\pi(\mathbf{v}, \boldsymbol{\sigma}^2, \boldsymbol{\beta}, \delta^2, \alpha \mid \mathbf{y}_s, \phi = 0)$ under the benchmarking model is very complex we do not sample it directly. The complexity arises because the conditional posterior densities of the σ_i^2 are not simple. Thus, our computations have two major steps. First, we fit the much simpler posterior density under the nonbenchmarking model. Second, we use the SIR algorithm to subsample the samples drawn from the fit of the nonbenchmarking model.

Specifically, we draw 100,000 samples from the nonbenchmarking model. Then, we infer about the finite population means under the nonbenchmarking model by taking a random sample of 10,000 of these. To make inference under the benchmarking model, we subsample 10,000 samples from the 100,000 samples already drawn using $R(\mathbf{v}, \boldsymbol{\sigma}^2, \boldsymbol{\beta})$. [We take samples of size 10,000 in both cases to obtain similar numerical standard errors.] Thus, we describe these two steps in the computations.

Looking to run the Gibbs sampler to fit the joint posterior density under the nonbenchmarking model, we observe that only the conditional posterior density of α is nonstandard. Letting $A = \prod_{i=1}^{l} \sigma_i^{-2}$ and $B = \sum_{i=1}^{l} \sigma_i^{-2}$, the conditional posterior density of α is

$$p(\alpha | \boldsymbol{\sigma}^2, \delta^2) \propto \left[\frac{((\alpha+1)/\delta^2)^{\alpha}}{\Gamma(\alpha)} \right]^l A^{\alpha+1} \exp\{-B(\alpha+1)/\delta^2\} \frac{1}{(1+\alpha)^2}, \ \alpha > 0 \quad (22)$$

To draw a random value from $p(\alpha | \boldsymbol{\sigma}^2, \delta^2)$ in (22), we transform α to $\theta = \alpha/(\alpha+1)$ which brings θ, in the interval (0, 1). Then, we draw θ using a grid, dividing (0, 1) into 100 subintervals of equal widths. This allows us to approximate the probability density function by a probability mass function which is easy to sample. When all the conditional posterior densities are drawn in turn, we get a griddy Gibbs sampler (Ritter and Tanner 1992).

When we ran the griddy sampler, we use 1,000 iterates as a "burn in" and we picked every fifth iterate to remove the autocorrelation. Thus, we have a random sample of 100,000 iterates from the joint posterior density of the nonbenchmarking model. To make inference about the nonbenchmarking model, we subsample a random sample of 10,000 iterates. However, to make inference about the benchmarking model, we subsample using the SIR algorithm.

Let $\Omega = (\mathbf{v}, \boldsymbol{\sigma}^2, \boldsymbol{\beta}, \delta^2, \alpha)$ denote the set of all parameters, and let $\tilde{M}$ = 100,000. Letting $\Omega^{(h)}$, $h = 1, \ldots, \tilde{M}$ we compute $R(\mathbf{v}^{(h)}, \boldsymbol{\sigma}^{2(h)}, \boldsymbol{\beta}^{(h)})$. Then, we take a sample of size M =10,000 without replacement with probabilities proportional to the $R(\mathbf{v}^{(h)}, \boldsymbol{\sigma}^{2(h)}, \boldsymbol{\beta}^{(h)})$. This is a random sample of size M from the joint posterior density under the benchmarking model. Bayesian predictive inference of the finite population means of each county is now straight forward.

4. ANALYSIS OF LANDSAT SATELLITE DATA

We compare the nonbenchmarking model and the benchmarking model under homogeneity and heterogeneity using both corn data and soybeans data.

In this example, we will implement the two models (NBM and BM) to predict the mean corn and the mean soybean acres in 12 counties in north central Iowa. It was determined that there is a linear association between the number of hectares of corn (soybean) and the number of pixels classified as corn and soybean for each segment obtained from satellite readings. Hence, we model the number of hectares of corn (soybean) using the number of pixels of corn and soybean as the covariates. Information from 37 sampled segments in the 12 counties were obtained from survey and satellite data. The number of hectares of corn (soybean) and the

number of pixels classified as corn and soybean are available for each sample segment. Additional information such as the total number of segments in each county (the population size in each county), and the population mean pixels of corn and soybean were provided. Using these information, we estimate the mean corn and soybean hectares for each county using the two models. We summarize the results using the posterior mean (PM) and posterior standard deviation (PSD); we do not present the 95% credible intervals because the posterior distributions of the finite population means are approximately normally distributed. We present in Table 1 the comparison of the posterior inferences from the NBM and BM models for the corn and soybean hectares data.

Table 1(b) shows that the posterior means from the two models are mostly the same, with a slight shifting in the posterior means from the NBM model to the BM model. The results also show that there is a little more precision in the BM model than in the NBM model. Observe that the posterior standard deviations are all smaller under the BM model than under the NBM model in both the corn and soybean results. However, note that for Hardin county the PSD is larger for the BM model than the NBM model (5.61versus 5.73). These decreases in the PSDs are reflected in the 95% credible intervals. From Table 1, the means for the corn (soybean) data are

$$\sum_{i=1}^{l} \frac{N_i PM_{i(BM)}}{N} = 120.32(95.35)$$

but $$\sum_{i=1}^{l} \frac{N_i PM_{i(NBM)}}{N} = 119.67(96.87)$$

Notice that the value of the overall posterior mean from the BM model is equal to the sample mean of the data, but this is not true for the NBM model. This observation reflects the use of the benchmarking constraint. We have also looked at the distance between the direct estimators $\bar{y}_i$ and the PM_i from both models. As a means, we computed

$$SHR = \sqrt{\sum_{i=1}^{l} (PM_i - \bar{y}_{s_i})^2}$$

For the corn data, $SHR_{(NBM)}$ = 81.05 and $SHR_{(BM)}$ =80.84. On the other hand, for the soybean data, $SHR_{(NBM)}$ = 92.81 and $SHR_{(BM)}$ = 91.03. In both data, we have found that the BM model has a slightly lower shrinkage than the NBM model. This slight difference may again be attributed to the addition of the benchmarking constraint in the BM model.

With regards to the data on corn and soybeans (Battese *et al*. 1988), we found that it is difficult to use twelve different sampling variances (one for each county). There is instability because the sample sizes are all smaller than 6 with three of them being just one. So we decided to block the sampling variances. It is reasonable to assume that the sampling variances are proportional to the population sizes. Therefore, under a self-weighting design, we can expect that the population sizes are proportional to the sample sizes; so we choose the sample variances to be roughly proportional to the sample sizes. Thus, we have chosen four different blocks (counties 1-3, counties 4-6, counties 7-9, counties 10-12). We have run our computations according to this adjusted specification.

We provide a similar analysis as for the homogeneous scenario. From Table 1(a), the means for the corn (soybean) data are

$$\sum_{i=1}^{l} \frac{N_i PM_{i(BM)}}{N} = 120.62(96.75)$$

but $$\sum_{i=1}^{l} \frac{N_i PM_{i(NBM)}}{N} = 119.51(99.41)$$

Notice that the value of the overall posterior mean from the BM model is equal to the sample mean of the data, but this is not true for the NBM model. This observation reflects the use of the benchmarking constraint. We have also looked at the distance between the direct estimators $\bar{y}_i$ and the PM_i from both models. As a means, we computed $SHR = \sqrt{\sum_{i=1}^{l} \left(PM_i - \bar{y}_{s_i}\right)^2}$.

For the corn data, $SHR_{(NBM)}$=85.89 and $SHR_{(BM)}$ =85.26. On the other hand, for the soybean data, $SHR_{(NBM)}$ = 113.41 and $SHR_{(BM)}$ = 107.23. Also, the

Table 1. Comparisons of the nonbenchmarking (NBM) and benchmarking (BM) models using posterior means (PM) and the posterior standard deviations (PSD) for the finite population mean of corn and soybeans by county (cty) under the (a) homogeneous model and (b) heterogeneous model

	Corn				Soybean			
	NBM		BM		NBM		BM	
Cty	PM	PSD	PM	PSD	PM	PSD	PM	PSD
a. Homogeneity								
1	123.49	9.32	124.15	8.53	78.82	11.30	77.56	10.35
2	124.27	9.33	124.91	8.41	94.29	10.97	92.89	10.17
3	110.89	10.01	111.55	9.50	87.72	10.80	86.07	10.20
4	114.07	8.50	114.74	7.77	81.97	9.91	80.48	9.61
5	138.64	8.47	139.41	7.89	67.02	7.87	65.74	7.45
6	109.76	7.54	110.31	6.91	113.83	7.31	112.24	6.83
7	116.08	7.24	116.44	6.80	97.44	7.53	95.82	7.46
8	122.80	7.29	123.55	6.58	111.97	7.49	110.37	7.15
9	112.14	6.94	112.84	6.32	110.00	6.47	108.50	6.08
10	123.86	6.23	124.63	5.96	100.42	6.18	98.74	6.12
11	111.55	6.91	112.21	6.47	118.27	6.39	116.66	6.13
12	131.16	5.90	131.76	5.73	75.16	5.61	73.50	5.73
b. Heterogeneity								
1	120.98	7.94	122.18	6.68	88.46	10.52	85.26	9.36
2	122.27	7.74	123.55	6.82	94.56	10.23	91.33	9.75
3	115.50	8.22	116.81	7.76	96.19	10.12	93.02	9.58
4	116.41	7.19	118.12	6.57	101.04	11.31	95.23	10.74
5	131.70	7.92	133.40	6.99	86.24	11.10	80.27	10.21
6	105.72	7.00	107.38	6.14	116.31	8.78	110.46	9.28
7	117.68	6.34	118.66	5.45	91.03	8.69	89.76	7.84
8	121.85	6.25	122.82	5.09	108.02	8.21	106.72	7.03
9	108.40	6.59	109.17	5.74	114.96	7.39	113.47	6.69
10	125.74	5.99	126.46	4.83	97.72	7.72	96.94	6.45
11	117.65	7.09	118.46	6.05	104.70	10.57	103.84	9.71
12	131.67	5.74	132.35	4.74	82.64	8.05	81.69	6.92

PSDs are slightly smaller for the BM model with the PSD for soybeans in Pocahontas county under the BM model larger than under the NBM model (8.78 versus 9.28).

Again the differences between the BM and NBM models are similar under heterogeneity. There is an increase in PM from NBM to BM, and there is a decrease in PSD from NBM to BM. As for the homogeneous models, this is true for both corn and soybeans. There are diûerences in the homogeneous and heterogeneous models. The PSD's for corn are almost always smaller under the heterogeneous model. However, for soybeans there are some counties with larger PSDs under the heterogeneous model. For soybeans we observe an important difference between the PMs for the homogeneous and heterogeneous

benchmarking models. The posterior means are higher for some counties and lower for others by quite a bit. We have calculated the percent changes for the heterogeneous model over the homogeneous model for each county. There are increases for counties 1, 2, 3, 4, 5, 6, 9, 12 which are 12.2, 0.3, 9.7, 23.3, 28.7, 2.2, 4.5, 10.0 and decreases for counties 8, 9, 10, 11 which are 6.6, 3.5, 2.7, 11.5. The changes for corn are relatively small.

5. SUMMARY AND CONCLUSION

We have developed a Bayesian benchmarking approach to estimate finite population means of small areas under a nested-error regression model. We have discussed two important scenarios, one in which there is a common sampling variance (homogeneity) and the other in which there are unequal sampling variances (heterogeneity). It is important to consider a benchmarking model with heterogeneity because unknown to the investigator it could very well be the case for small areas that the sampling variances are unequal, which can make a difference in inference about the finite population means for both corn and soybeans.

We have obtained some theoretical results for the homogeneous scenario. We also found a closed form for the distribution of the finite population means, which makes it easy to generate samples from its distribution. The proof of the propriety of the posterior distribution also showed a simple procedure to obtain samples from the posterior distribution of the parameters. Because of this, we do not have to use a Markov chain Monte Carlo method to make inferences; we simply use the composition method which provides random samples. Relative to the nonbenchmarking model, we found that there is a shift in the posterior densities of the finite population means. Moreover, the results showed a slight increase in the precision for the estimate of the finite population mean in each area under the BM model; this increased precision is observed in the slightly narrower 95% credible intervals in the benchmarking model.

The models with heterogeneity of sampling variances are more complex. Unlike the scenario with homogeneous variance, we were not able to use random samples to fit these models. Instead we have used the griddy Gibbs sampler to fit the nonbenchmarking model and coupled with the SIR algorithm to fit the benchmarking model. We have considered blocking the sampling variances for the application (i.e., partition the counties so that within the same partition set there is a single sampling variance). Blocking does help to reduce the errors of the predictions for corn and not so much for soybeans especially for the larger samples. The predictions for corn are more similar to the homogeneous scenario but those for soybeans are more different.

Our research on the nested error regression model currently looks atrobustness. It is necessary to further study the heterogeneity of the sampling variances; this has been a long standing problem. Future work on the nested error regression model to address the assumptions of linearity of the expected response and the normality of the sampling errors and random effects are important goals in small area estimation. These works will be useful to predict crop productions for small counties.

ACKNOWLEDGEMENT

The authors thank the referee for some useful comments and the editor for encouragement.

REFERENCES

Arora, V. and Lahiri, P. (1997). On the superiority of the Bayesian method over the BLUP in small area estimation problems. *Statist. Sin.*, **7**, 1053-1063.

Arora, V., Lahiri, P. and Mukherjee, K. (1997). Empirical Bayes estimation of finite population means from complex surveys. *J. Amer. Statist. Assoc.*, **92**, 1555-1562.

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

Datta, G. and Ghosh, M. (1991). Bayesian prediction in linear models: Applications to small area estimation. *Ann. Statist*., **19**, 1748-1770.

Datta, G., Ghosh, M., Steorts, R. and Maples, J. (2009). Bayesian benchmarking with applications to small area estimation. *Statistical Research Division Research Report Series (Statistics 2009). U.S. Census Bureau.*

Daniells, M.J. (1999). A prior for the variance in hierarchical models. *Can. J. Statist.*, **27**, 567-578.

Dick, J.P. (2001). Small domain estimation of missed persons in the 2001 census. In: *Proce. Sur. Met. Sec.,* Statistic Society of Canada, 37-46.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Anal.*, **1**, 515-533.

Ghosh, M. and Meeden, G. (1986). Empirical Bayes estimation in finite population sampling. *J. Amer. Statist. Assoc.*, **81**, 1058-1062.

Hedeker, D., Mermelstein, R.J. and Demirtas, H. (2008). An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. *Biometrics*, **64**, 627-634.

Hillmer, S.C. and Trabelsi, A. (1987). Benchmarking of economic time series. *J. Amer. Statist. Assoc.*, **82**, 1064-1071.

Iwig, W.C. (1993). The National Agricultural Statistics Service County Estimates Program. In: *Indirect Estimators in Federal Programs*, Statistical Policy Working Paper 21, Report of the Federal Committee on Statistical Methodology, Subcommittee on Small Area Estimation, Washington DC, **7**, 1-7.15.

Kleffe, J. and Rao, J.N.K. (1992). Estimation of mean squared error of empirical best linear unbiased predictors under random error variance linear model. *J. Mult. Anal*., **43**, 1-15.

Nandram, B. (1999). An empirical Bayes prediction interval for the finite population mean of a small area. *Statist. Sin*., **9**, 325-343.

Nandram, B. and Choi, J.W. (2002). A hierarchical Bayesian nonresponse model for binary data with uncertainty about ignorability. *J. Amer. Statist. Assoc.*, **97**, 381-388.

Nandram, B. and Sedransk, J. (1993). Empirical Bayes estimation for the finite population mean on the current occasion. *J. Amer. Statist. Assoc.*, **88**, 994-1000.

Nandram, B., Toto, M.C.S. and Choi, J.W. (2010). A Bayesian benchmarking of the Scott-Smith model for small areas. *J. Statist. Compu. Simul.* In press.

Park, Y., Choi, J.W. and Kim, H.Y. (2006). Forecasting Cause-age-specific mortality using two random processes. *J. Amer. Statist. Assoc.*, **101**, 472 -483.

Pfefferman, D. and Tiller, R. (2006). Small-area estimation with state-space models subject to benchmark constraints. *J. Amer. Statist. Assoc.*, **64**, 830-840.

Rao, C.R. (1970). Estimation of heteroscedastic variances in random linear models. *J. Amer. Statist. Assoc.*, **65**, 161-172.

Rao, J.N.K. (1980). Estimating the common mean of possibly different normal populations. *J. Amer. Statist. Assoc.*, **75**, 447-453.

Rao, P.S.R.S., Kaplan, J. and Cochran, W.G. (1981). Estimators for the one-way random effects model with unequal error variances. *J. Amer. Statist. Assoc.*, **76**, 89-97.

Rashid, M.M. and Nandram, B. (1998). A rank based predictor for the finite population mean of a small area: An application to crop production. *J. Ag. Biol. Environ. Statist.*, **3**, 201-222.

Ritter, C. and Tanner, M.A. (1992). The Gibbs stopper and the griddy Gibbs sampler. *J. Amer. Statist. Assoc.*, **87**, 861-868.

Toto, M.C.S and Nandram, B. (2010). A Bayesian predictive inference for small area means incorporating covariates and sampling weights. *J. Statist. Plann. Inf.*, **140**, 2963-2979.

Trabelsi, A. and Hillmer, S.C. (1990). Benchmarking time series with reliable benchmarks. *Appl. Statist.*, **39**, 367-379.

Wang, J., Fuller, W.A. and Qu, Y. (2008). Small area estimation under a restriction. *Survey Methodology*, **34**, 29-36.

You, Y. and Rao, J.N.K. (2002). Small area estimation using unmatched sampling and linking models. *Canad. J. Statist.*, **30**, 3-15.

You, Y., Rao, J.N.K. and Dick, J.P. (2004). Benchmarking hierarchical Bayes small area estimators in the Canadian Census under coverage estimation. *Statist. Trans.*, **6**, 631-640.

APPENDIX

Proofs of Lemma 2.1 and Lemma 2.2

Lemma 2.1

First, consider the slightly simpler notation in which

$$Z_k \,|\alpha_k, \sigma^2 \overset{ind}{\sim} \text{Normal}\ (\alpha_k,\ \sigma^2),\ k = 1, 2, ..., N$$

subject to the constraint $\frac{1}{N}\sum_{k=1}^{N} z_k = \frac{1}{n}\sum_{k=1}^{n} z_k$. Make the transformation $\phi = \frac{1}{N}\sum_{k=1}^{N} z_k - \frac{1}{n}\sum_{k=1}^{n} z_k$ with Z_1, ..., Z_{N-1} untransformed. We can show that the jacobian is N, and

$$p(Z_1, ..., Z_{N-1}, \phi = 0|\boldsymbol{\alpha}, \sigma^2) = N\left(\frac{1}{2\pi\sigma^2}\right)^{N/2} \times \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{k=1}^{N-1}(z_k-\alpha_k)^2 + \left(\sum_{k=1}^{n}\left(\frac{N}{n}-1\right)z_k - \sum_{k=n+1}^{N-1} z_k - \alpha_N\right)^2\right]\right\}$$

where $Z_N = \left(\sum_{k=1}^{n}\left(\frac{N}{n}-1\right)z_k\right) - \left(\sum_{k=n+1}^{N-1} z_k\right)$

After some algebraic manipulation, we have $\mathbf{Z}_{(N)}|\boldsymbol{\alpha}\,\sigma^2,\ \phi = 0 \sim$ Normal, $(\boldsymbol{\theta},\ \sigma^2(\mathbf{I}_{N-1} - \mathbf{D}))$, where $\boldsymbol{\theta} = \boldsymbol{\alpha}_{(N)} + f\left(\frac{1}{N-n}\right)\left(\alpha_N - \mathbf{a}'\boldsymbol{\alpha}_{(N)}\right)\mathbf{a}$ with $\boldsymbol{\alpha}_{(N)} = (\alpha_1, ..., \alpha_{N-1})'$ and $\mathbf{a} = \left[\left(\frac{N-n}{n}\right)\mathbf{1}_n', -\mathbf{1}_{N-1-n}'\right]'$. Now, match the original vector $\mathbf{y} = \left(\mathbf{y}_s', \mathbf{y}_{ns}'\right)'$ with $(z_1, ..., z_n, z_{n+1}, ..., z_N)'$, as well as $\boldsymbol{\mu}$ with $\boldsymbol{\alpha}$, to obtain the desired result.

Lemma 2.2

Omitting the conditioning on $\phi = 0$, the joint posterior density

$$\pi\left(\mathbf{v}, \boldsymbol{\beta}, \sigma^2, \rho|\mathbf{y}_s\right) \propto \left(\frac{1}{\sigma^2}\right)^{1+(n+l)/2}\left(\frac{1-\rho}{\rho}\right)^{l/2} \times \exp\left\{-\frac{1}{2\sigma^2}\left(\frac{1-\rho}{\rho}\right)\left[\left(\mathbf{v} - \mathbf{B}^{-1}\mathbf{C}_z'\mathbf{A}\left(\mathbf{y}_s - \mathbf{C}_x\boldsymbol{\beta}\right)\right)'\right.\right.$$
$$\left.\left.[\mathbf{I}_l + \mathbf{C}_z'\mathbf{A}\mathbf{C}_z]\left(\mathbf{v} - \mathbf{B}^{-1}\mathbf{C}_z'\mathbf{A}\left(\mathbf{y}_s - \mathbf{C}_x\boldsymbol{\beta}\right)\right)\right]\right\}$$
$$\times \exp\left\{-\frac{1}{2\sigma^2}\left(\frac{1-\rho}{\rho}\right)\left[\left(\mathbf{y}_s - \mathbf{C}_x\boldsymbol{\beta}\right)'\left(\mathbf{A} - \mathbf{A}\mathbf{C}_z\mathbf{B}^{-1}\mathbf{C}_z'\mathbf{A}\right)\left(\mathbf{y}_s - \mathbf{C}_x\boldsymbol{\beta}\right)\right]\right\}$$

where

$$\mathbf{B} = \mathbf{I}_l + \mathbf{C}_z'\mathbf{A}\mathbf{C}_z,\ \mathbf{A} = \left[\left(\frac{1-\rho}{\rho}\right)\left(\mathbf{I}_n - \frac{1}{n}(1-f)\mathbf{J}_n\right)\right]^{-1},$$

and $\mathbf{C}_x$ and $\mathbf{C}_z$ are as defined in (8).

First, it can be shown that

$$\mathbf{v}|\boldsymbol{\beta},\ \sigma^2,\ \rho,\ \mathbf{y}_s \sim \text{Normal}\left(\mathbf{B}^{-1}\mathbf{C}_z'\mathbf{A}\left(\mathbf{y}_s - \mathbf{C}_x\boldsymbol{\beta}\right),\ \sigma^2\left(\frac{\rho}{1-\rho}\right)\mathbf{B}^{-1}\right)$$

Integrating out $\mathbf{v}$ from $\pi\left(\mathbf{v}, \boldsymbol{\beta}, \sigma^2, \rho|\mathbf{y}_s\right)$, we get

$$\pi\left(\boldsymbol{\beta}, \sigma^2, \rho|\mathbf{y}_s\right) \propto \left(\frac{1}{\sigma^2}\right)^{1+n/2}\left(\frac{1}{|\mathbf{B}|}\right)^{1/2} \times \exp\left\{-\frac{1}{2\sigma^2}\left(\frac{1-\rho}{\rho}\right)\left[\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)'\mathbf{V}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)\right]\right\}$$
$$\times \exp\left\{-\frac{1}{2\sigma^2}\left(\frac{1-\rho}{\rho}\right)\left[\mathbf{y}_s'\left(\mathbf{D} - \mathbf{D}\mathbf{C}_x\mathbf{V}^{-1}\mathbf{C}_x'\mathbf{D}\right)\mathbf{y}_s\right]\right\}$$

and it can be shown that

$$\boldsymbol{\beta}|\sigma^2,\rho,\mathbf{y}_s \sim \text{Normal}\left(\hat{\boldsymbol{\beta}}, \sigma^2\left(\frac{\rho}{1-\rho}\right)\mathbf{V}^{-1}\right)$$

where $= \hat{\boldsymbol{\beta}} = \mathbf{V}^{-1}\mathbf{C}_x'\mathbf{D}y_s$, $\mathbf{D} = \mathbf{A} - \mathbf{A}\mathbf{C}_z\mathbf{B}^{-1}\mathbf{C}_z'\mathbf{A}$, and $\mathbf{V} = \mathbf{C}_x'\mathbf{D}\mathbf{C}_x$.

Then, integrating out $\boldsymbol{\beta}$ from $\pi\left(\boldsymbol{\beta}, \sigma^2, \rho|\mathbf{y}_s\right)$, we have

$$\pi\left(\sigma^2,\rho|\mathbf{y}_s\right) \propto \left(\frac{1}{\sigma^2}\right)^{1+(n+p)/2}$$

$$\times \left(\frac{\rho}{1-\rho}\right)^{p/2}\left(\frac{1}{|\mathbf{V}|}\right)^{1/2}\left(\frac{1}{|\mathbf{B}|}\right)^{1/2}$$

$$\times \exp\left\{-\frac{1}{2\sigma^2}\left(\frac{1-\rho}{\rho}\right)\left[\mathbf{y}_s'\left(\mathbf{D}-\mathbf{D}\mathbf{C}_x\mathbf{V}^{-1}\mathbf{C}_x'\mathbf{D}\right)\mathbf{y}_s\right]\right\}$$

and it can be shown that, provided $n > p$,

$$\sigma^{-2}|\rho,\mathbf{y}_s \sim \text{Gamma}\left(\frac{n-p}{2}, \frac{\mathbf{G}}{2}\right)$$

where $\mathbf{G} = \left(\frac{1-\rho}{\rho}\right)\left[\mathbf{y}_s'\left(\mathbf{D}-\mathbf{D}\mathbf{C}_x(\mathbf{V})^{-1}\mathbf{C}_x'\mathbf{D}\right)\mathbf{y}_s\right]$

Finally, integrating out σ^2 from $\pi\left(\sigma^2,\rho|\mathbf{y}_s\right)$, we get

$$\pi_4\left(\rho|\mathbf{y}_s\right) \propto \left(\frac{\rho}{1-\rho}\right)^{n/2}(|\mathbf{V}|)^{-1/2}(|\mathbf{B}|)^{-1/2}$$

$$\times\left(\mathbf{y}_s'\left(\mathbf{D}-\mathbf{D}\mathbf{C}_x(\mathbf{V})^{-1}\mathbf{C}_x'\mathbf{D}\right)\mathbf{y}_s\right)^{-\left(\frac{n-p}{2}\right)} \qquad (23)$$

To complete the proof, we showed that $\pi_4\left(\rho|\mathbf{y}_s\right)$ is proper for all ρ, $0 < \rho < 1$. It is worth noting that **B**, **V** and **D** depend on ρ through **A** in which ρ appears in the function $(1 - \rho)/\rho$. Thus, once $(1 - \rho)/\rho$ is finite, A will be finite and so are **B**, **V** and **D**. This is true if we assume that ρ lies in a closed interval in (0, 1) [e.g., (.0001, .9999)]. Therefore, the right side of (23) is finite; and so the inegral over (.0001, .9999) is finite. Thus, $\pi_4\left(\rho|\mathbf{y}_s\right)$ is proper.