A Hierarchical Bayes Unit-Level Small Area Estimation Model for Normal Mixture Populations

Shuchi Goyal, Gauri Sankar Datta & Abhyuday Mandal

Springer

Springer

Check for updates

A Hierarchical Bayes Unit-Level Small Area Estimation Model for Normal Mixture Populations

Shuchi Goyal
University of California at Los Angeles, Los Angeles, USA
Gauri Sankar Datta and Abhyuday Mandal
University of Georgia, Athens, USA

Abstract

National statistical agencies are regularly required to produce estimates about various subpopulations, formed by demographic and/or geographic classifications, based on a limited number of samples. Traditional direct estimates computed using only sampled data from individual subpopulations are usually unreliable due to small sample sizes. Subpopulations with small samples are termed small areas or small domains. To improve on the less reliable direct estimates, model-based estimates, which borrow information from suitable auxiliary variables, have been extensively proposed in the literature. However, standard model-based estimates rely on the normality assumptions of the error terms. In this research we propose a hierarchical Bayesian (HB) method for the unit-level nested error regression model based on a normal mixture for the unit-level error distribution. Our method proposed here is applicable to model cases with unit-level error outliers as well as cases where each small area population is comprised of two subgroups, neither of which can be treated as an outlier. Our proposed method is more robust than the normality based standard HB method (Datta and Ghosh, Annals Stat. **19**, 1748–1770, 1991) to handle outliers or multiple subgroups in the population. Our proposal assumes two subgroups and the two-component mixture model that has been recently proposed by Chakraborty et al. (Int. Stat. Rev. **87**, 158–176, 2019) to address outliers. To implement our proposal we use a uniform prior for the regression parameters, random effects variance parameter, and the mixing proportion, and we use a partially proper non-informative prior distribution for the two unit-level error variance components in the mixture. We apply our method to two examples to predict summary characteristics of farm products at the small area level. One of the examples is prediction of twelve county-level crop areas cultivated for corn in some Iowa counties. The other example involves total cash associated in farm operations in twenty-seven farming regions in Australia. We compare predictions of small area characteristics based on the proposed method with those obtained by applying the Datta and Ghosh (Annals Stat. **19**, 1748–1770, 1991)

and the Chakraborty et al. (Int. Stat. Rev. **87**, 158–176, 2019) HB meth-
ods. Our simulation study comparing these three Bayesian methods, when
the unit-level error distribution is normal, or t, or two-component normal
mixture, showed the superiority of our proposed method, measured by pre-
diction mean squared error, coverage probabilities and lengths of credible
intervals for the small area means.

AMS (2000) *subject classification.* Primary 62F15; Secondary 62D05.
Keywords and phrases. Nested error regression, Outliers, Prediction intervals
and uncertainty, Robust empirical best linear unbiased prediction

1   Introduction

National statistical offices around the world have been mandated for
many years to produce reliable statistics for important variables such as
population, income, unemployment, and health outcomes for various geo-
graphic domains (e.g., states, counties) and/or demographic domains (e.g.,
age, race, gender). However, the sample available from many of these do-
mains are often small to produce direct estimates of adequate accuracy. This
situation is known as small area estimation. To develop estimates that are
more reliable than the direct estimates, data from the entire sample (that is,
a sample covering all small areas) is used and combined with other appro-
priate auxiliary variables to produce indirect estimates of the small domain
characteristics. Model-based approaches have been shown to be useful in
producing reliable small area or small domain estimates.

The earliest important application of model based small area estima-
tion is by Fay and Herriot (1979). They adopted shrinkage estimation of
Stein (1955), popularized as empirical Bayes estimation by Efron and Mor-
ris (1973). Using empirical Bayes method, Fay and Herriot (1979) proposed
shrinkage of a direct estimator of a small-area mean to a suitable regression
plane determined jointly by the direct estimators and auxiliary variables
from the small areas. This approach is based on modeling of small area level
summary statistics, often sample means.

Battese et al. (1988) proposed the popular nested-error regression (NER)
model to develop small area estimates based on data available on the indi-
vidual sampled units. Battese et al. (1988) proposed a regression model
for a response for all the units in the population based on some auxiliary
variables. The NER model, aptly called unit-level model, is developed un-
der the normality assumption of small area random effects and unit-level
random errors. For unit-level data, the NER model has been the basis for
producing reliable small-area estimates either by a frequentist or a Bayesian

approach. Datta and Ghosh (1991) used the NER model, in conjunction with suitable noninformative priors for the regression coefficients and variance parameters, to develop hierarchical Bayes estimates of finite population small area means. Prasad and Rao (1990) and Datta and Lahiri (2000) used a frequentist approach for the NER model to develop empirical best linear unbiased prediction (EBLUP) of the finite population means. To facilitate our discussion of robust HB method of small area estimation we reviewed the existing HB models in the next section.

It is desirable to have a model that is robust in the presence of random errors prone to outliers. To address the specific case where outliers are present, Chakraborty et al. (2019) proposed an HB alternative to Datta and Ghosh's method (1991). By using a two-component mixture of normal distribution, this model accommodates populations where a small portion of unit-level errors come from a secondary distribution with a larger variance than the primary distribution. Chakraborty et al. (2019) showed that their model consistently performs as well as or better than that of Datta and Ghosh (1991), including in the special case of no outliers (i.e. "no contamination").

We note that the model proposed by Chakraborty et al. (2019) is most effective when only a small portion of the population comes from the secondary distribution. In this paper we suggest an HB method built from the NER model to handle more general cases of two-component mixture populations, where the proportion of members from the secondary distribution may be as high as 50 percent.

2 Existing Unit-Level HB Small Area Models

The NER model of Battese et al. (1988) is immensely popular in unit-level modeling for small area estimation. This model supposes that a population is partitioned in m small areas with N_i units in the ith small area. The value of the response variable for the jth unit in the ith small area Y_{ij} satisfies

$$Y_{ij} = \boldsymbol{x}_{ij}^T \boldsymbol{\beta} + v_i + e_{ij}, j = 1, \ldots, N_i, i = 1, \ldots, m, \qquad (2.1)$$

where $\boldsymbol{x_{ij}} = (x_{ij1}, \ldots, x_{ijq})^T$ is a $q \times 1$ vector of values for predictor variables for that observation. Here $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_q)^T$ denotes the vector of regression coefficients. The zero mean random variables v_i and e_{ij} account for area- and unit-level errors, respectively, and are assumed to be independent of each other. We further assume that v_i's are i.i.d. $N(0, \sigma_v^2)$. As in Battese et al. (1988), under appropriate distributional assumptions for the e_{ij}'s, our goal is to predict the population mean θ_i in the ith county defined as the conditional mean of the response given the realized random effect v_i, where

$\theta_i = \bar{\boldsymbol{x}}_{i(p)}^T \boldsymbol{\beta} + v_i$, and $\bar{\boldsymbol{x}}_{i(p)} = \frac{1}{N_i} \sum_{j=1}^{N_i} x_{ij}$. The $\bar{\boldsymbol{x}}_{i(p)}$'s are known for all the small areas.

A special case of an HB model introduced by Datta and Ghosh (1991) includes the following HB version of the NER model. We denote this by DG HB model.

(I) Conditional on $\boldsymbol{\beta}, \boldsymbol{v} = (v_1, \ldots, v_m)^T, \sigma_e^2$, and σ_v^2,

$$Y_{ij} \overset{ind}{\sim} N(\boldsymbol{x}_{ij}^T \boldsymbol{\beta} + v_i, \sigma_e^2)$$

for $j = 1, \ldots, N_i, i = 1, \ldots, m$.

(II) Conditional on $\boldsymbol{\beta}, \sigma_e^2$ and σ_v^2, $v_i \overset{iid}{\sim} N(0, \sigma_v^2)$ for all i.

(III) Model parameters $\boldsymbol{\beta}, \sigma_e^2$ and σ_v^2 are assigned the improper prior

$$\pi(\boldsymbol{\beta}, \sigma_v^2, \sigma_e^2) \propto \frac{1}{\sigma_e^2}. \tag{2.2}$$

Based on a random sample of size $n_i, i = 1, \ldots, m$, from all the small areas Datta and Ghosh (1991) used the above model to develop HB predictors of small area finite population means $\bar{Y}_i$'s, $i = 1, \ldots, m$. This model can also be used to develop Bayes predictors of θ_i's. For small n_i/N_i, the two quantities $\bar{Y}_i$ and θ_i's are approximately the same.

While the HB estimates developed by Datta and Ghosh (1991) are effective for populations in which the unit-level random errors follow a normal distribution, they are less effective when the errors follow a mixture of normal distributions. This scenario can be formulated by a two-component normal mixture model for the unit-level errors, which have the same mean but different variances. Another example of this situation is a population with "representative outliers" (Chambers, 1986). In this case, the underlying distribution of outliers is assumed to have the same zero mean, but a larger variance than that of the non-outliers.

Chakraborty et al. (2019) proposed a two-component normal mixture for the unit-level error distribution. This model, referred to as the CDM model hereafter, specifically facilitates small area estimation for populations which are suspected to contain representative outliers. Chambers (1986) defines a representative outlier as a value which is non-unique in the population and influences the estimates of finite population means $\bar{Y}_i$'s from the model. The CDM HB model, which modifies the DG HB model, is given below:

(I) Conditional on $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_q)^T, v_i, z_{ij}, p_e, \sigma_1^2, \sigma_2^2,$ and $\sigma_v^2,$

$$Y_{ij} \sim z_{ij} N(\boldsymbol{x}_{ij}^T \boldsymbol{\beta} + v_i, \sigma_1^2) + (1 - z_{ij}) N(\boldsymbol{x}_{ij}^T \boldsymbol{\beta} + v_i, \sigma_2^2)$$

for $j = 1, \ldots, N_i, i = 1, \ldots, m.$

(II) The indicator variables z_{ij} are i.i.d. with $P(z_{ij} = 1|p_e) = p_e$ for all i, j. Also, z_{ij}'s are independent of v_i's, $\boldsymbol{\beta}$, σ_1^2, σ_2^2, and σ_v^2.

(III) Conditional on $\boldsymbol{\beta}, z, p_e, \sigma_1^2, \sigma_2^2,$ and $\sigma_v^2, v_i \overset{\text{iid}}{\sim} N(0, \sigma_v^2)$ for all i.

The DG HB model is a limiting version of the CDM HB model when p_e is on the boundary. A key assumption of the CDM HB model is that outlier observations come from a distribution which has the same mean $\boldsymbol{x}_{ij}^T \boldsymbol{\beta} + v_i$ (conditional on random effects) as the distribution of non-outliers but a larger variance. The variances for non-outliers and outliers are denoted as σ_1^2 and σ_2^2, respectively, with $\sigma_1^2 < \sigma_2^2$. A priori outliers are assumed to occur in the various small areas with equal probability $(1 - p_e)$. The CDM HB model is completed by assigning independent noninformative priors for $\boldsymbol{\beta}, \sigma_1^2, \sigma_2^2, p_e,$ and σ_v^2, with $\boldsymbol{\beta} \sim Uniform(R^q), \sigma_v^2 \sim Uniform(R^+),$ $\pi(\sigma_1^2, \sigma_2^2) \propto \dfrac{1}{(\sigma_2^2)^2} I(\sigma_1^2 < \sigma_2^2),$ and $p_e \sim Uniform(0, 1).$

In the frequentist approach Prasad and Rao (1990) also used the NER model to derive the EBLUPs of θ_i and $\bar{Y}_i$ and estimators of their mean squared errors (MSE). In a subsequent article, Sinha and Rao (2009) investigated robustness of EBLUPs and the estimates of MSE in the presence of outliers. Their investigation showed that while the departure of random small area effects from the normality does not severely affect the EBLUPs and MSE estimates, a departure of normality assumption of the unit-level error terms adversely impacts the EBLUPs and their MSE estimates. Sinha and Rao (2009) proposed a robust empirical best linear unbiased prediction (REBLUP) approach to mitigate the impact of outliers in the unit-level error and/or in the area-level random effects on the EBLUPs.

We note that Chakraborty et al. (2019) used a Bayesian version of a popular contamination model to accommodate a small fraction of outliers in the sample. There are applications where the population is actually a mixture of multiple component distributions, where each component is a significant minority. To address such applications, we propose an HB model built from the NER model to handle more general cases of two-component mixture populations, where the proportion of members from the secondary distribution may be nearly half. The proposed new model is more appropriate to

deal with mixture populations, comprised of two sub-groups, differentiated by their variances. We propose our new model in Section 3. We apply this new HB model as well as the DG HB and CDM HB models in Section 4 to two examples to predict summary characteristics of farm products at the small area level. One of the examples is prediction of county-level crop areas cultivated for corn in twelve Iowa counties (cf. Battese et al. (1988)). The other example involves total cash associated in farm operations in twenty-seven farming regions in Australia based on a dataset by Chambers et al. (2011).

Chakraborty et al. (2019) showed via simulations that the CDM model performs as well as the DG HB model when the data set contains no outliers but has smaller empirical bias and mean squared error when the data contain outliers. The performance of CDM HB predictors is comparable to the REBLUP method predictors that was proposed by Sinha and Rao (2009). Due to a lack of space we exclude the Sinha-Rao frequentist method here to focus only on Bayesian techniques. For more details on the comparative performance of these robust methods and the DG HB method, we refer the reader to Section 6 of Chakraborty et al. (2019). In the setup of the corn data, we compare in Section 5 our proposed method with two competing Bayesian methods via simulation studies. Two data analyses and the simulation studies demonstrated the superiority of the new proposed HB model. Concluding comments are provided in Section 6, and relevant proofs and in-depth details are relegated to the Appendix and Supplementary Information sections.

3   An HB Normal Mixture Model for Unit-Level Error

We propose here a model that is a mixture extension of the nested-error regression model which accounts for unit-level error terms coming from two different normal distributions. To extend the basic NER model to account for observations from a mixture of two underlying distributions, we rely on the same assumptions (I) to (III) of the CDM model. The CDM model is a contamination model frequently used in the literature to accommodate a handful of outlying observations. In some applications, however, there may be a larger proportion of observations which may be different from the rest of the data. In these cases, since this group of observations is not really outliers, a mixture model, which we propose below, will be better suited than the contamination model. However, the proposed mixture model is also flexible enough to explain a small fraction of outliers in a dataset.

In our new formulation of the two-component mixture model for the unit-level error component, we treat the unit-level variances σ_1^2 and σ_2^2 symmetrically, and consequently assign the prior

$$\pi(\sigma_1^2, \sigma_2^2) \propto \frac{1}{(\sigma_1^2 + \sigma_2^2)^2}.$$

It is a key difference in the priors assumed in our proposed model, which we refer as GDM hereafter, and those used for CDM. To ensure the identifiability of all the parameters in the likelihood of the mixture model described by the hierarchy (I) to (III), we assume $p_e > 2^{-1}$, that avoids the label-switching problem. To complete specification of the prior distribution for the remaining parameters, we assign the same independent noninformative uniform prior to $\boldsymbol{\beta}$, σ_v^2 and p_e given by

$$\pi(\boldsymbol{\beta}, \sigma_v^2, p_e) \propto I_{(p_e \geq 2^{-1})}.$$

We establish the propriety of the posterior distribution in Appendix A.1. We used Gibbs sampling for our computation. Full conditional distributions for Gibbs sampling are listed in the Supplementary Material.

4 Data Analysis

4.1. Prediction of County Means of Crop Areas Battese et al. (1988) proposed the NER model to compute EBLUP prediction of mean hectares of corn grown in twelve counties of Iowa based on auxiliary variables provided by LANDSAT satellite data from the U.S. Department of Agriculture. The two auxiliary variables considered are mean number of pixels of corn and soybeans in sample segments satellite imaging. Of 37 measurements of hectares of corn sample, one observation from Hardin County was suspected as an outlier. The reported hectares of corn in this segment seems to be very low relative to the pixels of corn observed there, relative to other segments in the same county. Battese et al. (1988) suggested removing the suspected outlier altogether from the data set to improve the fit of the basic nested-error regression model. Datta and Ghosh (1991) subsequently used this reduced data to develop their HB prediction.

It is well-known that discarding suspected outliers can lead to loss of valuable information about the data set. By including the outlier from Hardin County when fitting a robust model, it would make sense that the estimated mean corn hectares would be higher than in the non-robust DG model. Chakraborty et al. (2019) demonstrated that when the full set of sampled observations is used, their HB prediction (CDM HB) of mean in Hardin

County is closer to REBLUP estimate of Sinha and Rao (2009), than the prediction obtained from the DG HB model. When applied to the reduced data set ($n = 36$), where the suspected outlier is discarded, the CDM HB model performs similarly to the DG HB model, indicating no loss in applying the CDM model to data which may not have any outliers. We apply the proposed model to calculate point estimates (posterior means) and standard errors (posterior standard deviations) of mean corn production in each county and compare our results to the predictions obtained from DG and CDM models. The results are shown in Table 1.

Our proposed model performs as well as the CDM model in the presence of a suspected outlier. The point estimates and standard errors calculated based on the proposed model, with the exception of one county, are very close to those produced by the CDM method. While there is considerable agreement in the estimates from the two robust Bayesian methods, the estimate for the Hardin county from the non-robust DG HB method is substantially different from the others.

We present in Fig. 1 a graphical display of posterior probability that an observation's unit-level error comes from subpopulation 2. The horizontal axes of the plots in Fig. 1 represent the standardized versions of the reported hectares of corn in a surveyed segment y_{ij}, defined by $E(y_{ij} - \theta_i|\boldsymbol{y})/\sqrt{var(y_{ij} - \theta_i|\boldsymbol{y})}$ where $\boldsymbol{y} = \{y_{ij}; j = 1, \ldots, n_i, i = 1, \ldots, m\}$, $E(\cdot|\boldsymbol{y})$ and $var(\cdot|\boldsymbol{y})$ represent the posterior mean and posterior variance operators.

Table 1: Various HB point estimates and standard errors of county hectares of corn (Full)

| County	n_i	DG		CDM		GDM	
		Mean	SD	Mean	SD	Mean	SD
Cerro Gordo	1	123.8	11.7	123.4	9.8	123.6	11.3
Hamilton	1	124.9	11.4	126.6	10.3	125.8	10.2
Worth	1	110.0	12.3	108.0	11.3	107.7	11.7
Humboldt	2	114.2	10.7	112.3	10.2	112.0	10.7
Franklin	3	140.3	10.8	142.1	8.1	142.4	8.4
Pocahontas	3	110.0	9.6	111.4	7.6	111.6	7.3
Winnebago	3	116.0	9.7	114.3	7.6	113.7	7.9
Wright	3	123.2	9.5	122.7	7.9	122.3	7.7
Webster	4	112.6	9.9	113.9	6.9	114.3	6.8
Hancock	5	124.4	8.9	123.5	6.1	123.6	6.1
Kossuth	5	111.3	8.9	108.2	6.8	108.1	6.9
Hardin	6	130.7	8.3	135.3	7.5	136.5	7.4
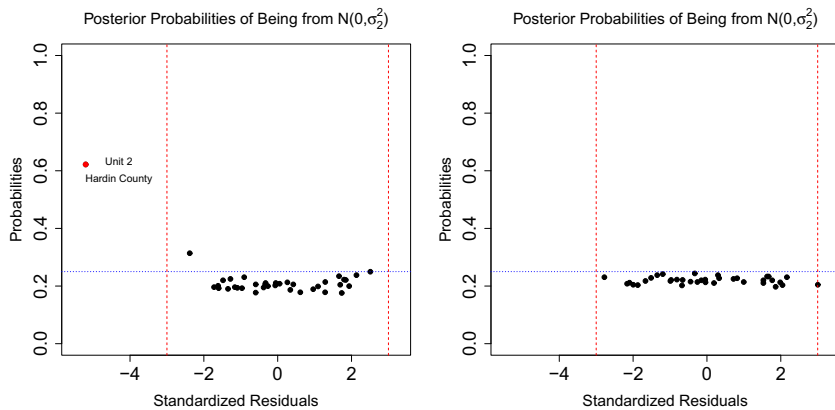
Figure 1: Posterior probabilities of observations coming from subpopulation 2 in *full* and *reduced* corn data

The vertical axes represent the posterior probability of an observation coming from the subpopulation 2. The GDM model identifies the second Hardin County observation, which has the most extreme standardized residual, as a likely member of a secondary subpopulation when analyzing the full data, as shown in the left panel of Fig. 1. The posterior probability that this observation may belong to the secondary population is 0.62, which is about 2.5 times the corresponding prior probability 0.25. For the other observations, most of their posterior probabilities are near 0.25, not much different from their prior values. We note here that in the right panel of Fig. 1, we plotted the same posterior probabilities for the *reduced* data, after removing the second observation from Hardin county. Interestingly, for none of these observations, the posterior probabilities are greater than 0.25.

We also compare model estimates for the data set after removing the outlier. The point estimates and posterior standard deviations given by each method for the reduced data are given in Table 2. For the first 11 counties listed, the estimates produced by each model change only slightly (smaller changes for the robust methods) from those calculated using the full data set. As expected, the estimate for Hardin County changes the most significantly. With the outlier removed, the point estimates for Hardin County increase in all three models but the change is less for the two mixture models and is the most substantial for the DG model. A similar conclusion was reached by Chakraborty et al. (2019) for Sinha-Rao REBLUP, see Table 1 of Chakraborty et al. (2019). This makes sense, as the estimates from the mixture and robust models should be less sensitive to outliers. The corn hectare

Table 2: Various HB estimates and standard errors of county hectares of corn (Reduced)

SA	n_i	DG		CDM		GDM	
		Mean	SD	Mean	SD	Mean	SD
Cerro Gordo	1	122.0	11.6	121.7	9.7	121.9	10.2
Hamilton	1	126.4	10.9	127.2	9.7	126.3	9.8
Worth	1	107.6	12.4	105.6	10.1	105.3	10.9
Humboldt	2	108.9	10.5	108.2	8.7	108.0	9.3
Franklin	3	143.6	9.7	144.1	7.0	144.3	7.0
Pocahontas	3	112.3	9.7	112.5	6.5	112.3	6.7
Winnebago	3	113.4	9.1	112.5	6.8	111.5	7.4
Wright	3	121.9	8.8	121.9	6.6	121.8	6.7
Webster	4	115.5	9.2	115.7	5.7	115.8	6.1
Hancock	5	124.8	8.4	124.4	5.4	124.6	5.5
Kossuth	5	107.7	8.5	106.3	5.7	106.0	5.6
Hardin	5	142.6	9.0	143.5	5.9	143.6	5.6

estimates in Hardin County given by the three models are also much closer in value to each other relative to the full data set. (Again, these estimates agree very closely with the REBLUP estimate; see Table 1 of Chakraborty et al. (2019)).

For the reduced data set, a comparison of posterior standard deviations associated with the HB estimates shows that the standard deviations from the mixture models are consistently lower than those given by the DG HB model. We also compare posterior standard deviations between the full data analysis and the reduced data analysis. Intuitively, the presence of an outlier will cause an increase in unit-level variances, and therefore may also cause an increase in posterior variances of small area means. While the standard deviations produced by the robust CDM and GDM HB models appear to be higher for the full data than for the reduced data, the standard deviations given by the DG HB model seem to change only moderately. Standard deviations for the non-robust DG HB model are the highest.

Tables 3 and 4 show posterior means, posterior standard deviations and posterior medians, respectively, for $\beta_0, \beta_1, \beta_2, p_e, \sigma_v^2, \sigma_1^2$, and σ_2^2. The estimated values of β_0, β_1, and β_2 found from various methods appear to be similar, despite the difference in priors for (σ_1^2, σ_2^2) and p_e. We note that the estimate of p_e is higher when using the proposed HB model, which constrains p_e between 2^{-1} and 1, than when using the CDM model, which does not constrain p_e but constrains $\sigma_1^2 < \sigma_2^2$. In the proposed method, we define

Table 3: Posterior means and standard deviations for relevant parameters in various models with and without the suspected outlier for corn data

Estimates	DG Full Data		DG Reduced Data		CDM Full Data		CDM Reduced Data		GDM Full Data		GDM Reduced Data	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
$\hat{\beta}_0$	17.29	36.48	50.46	28.43	30.55	31.67	51.17	25.14	33.26	30.98	50.98	27.28
$\hat{\beta}_1$	0.37	0.08	0.33	0.06	0.35	0.06	0.33	0.05	0.35	0.06	0.33	0.06
$\hat{\beta}_2$	−0.03	0.08	−0.13	0.06	−0.07	0.08	−0.14	0.06	−0.08	0.07	−0.14	0.06
$\hat{p}_e$	—	—	—	—	0.60	0.27	0.47	0.29	0.77	0.14	0.78	0.15
$\hat{\sigma}^2_v$	175.08	193.42	228.88	178.87	203.53	178.55	257.32	183.64	205.71	176.27	266.17	202.47
$\hat{\sigma}^2_1$	364.47	144.71	210.48	89.81	186.95	100.36	118.96	56.58	246.33	158.54	166.58	68.51
$\hat{\sigma}^2_2$	—	—	—	—	842.25	2090.03	364.21	1580.03	1059.20	2487.99	233.30	432.57

Table 4: Posterior medians and interquartile ranges for relevant parameters in various models with and without the suspected outlier for corn data

Estimates	DG Full Data Median	IQR	DG Reduced Data Median	IQR	CDM Full Data Median	IQR	CDM Reduced Data Median	IQR	GDM Full Data Median	IQR	GDM Reduced Data Median	IQR
$\hat{\beta}_0$	18.87	43.37	50.22	36.96	30.59	42.12	53.40	31.34	34.03	42.54	51.19	34.78
$\hat{\beta}_1$	0.37	0.10	0.33	0.07	0.35	0.09	0.33	0.06	0.35	0.08	0.33	0.07
$\hat{\beta}_2$	−0.03	0.10	−0.13	0.08	−0.07	0.10	−0.14	0.08	−0.08	0.09	−0.14	0.08
$\hat{p}_e$	—	—	—	—	0.66	0.44	0.45	0.53	0.79	0.24	0.79	0.25
$\hat{\sigma}_v^2$	123.56	153.47	188.54	169.17	157.17	170.35	208.56	183.63	151.16	171.94	206.04	199.12
$\hat{\sigma}_1^2$	334.70	152.13	191.44	86.37	173.04	120.84	114.18	65.50	203.78	180.15	155.76	72.71
$\hat{\sigma}_2^2$	—	—	—	—	480.28	447.43	191.71	93.85	533.24	852.00	136.73	165.04

the primary variance σ_1^2 as the variance of the distribution from which more than 50% of observations occur and the secondary variance σ_2^2 for the distribution of the remaining observations. When examining the full data, we calculate the posterior mean and median estimates of σ_1^2 to be 246.33 and 203.78 respectively, while those for σ_2^2 are 1059.20 and 533.24 respectively. We can compare these values to the estimates produced using the CDM HB approach, where the primary distribution is defined as the one with the smaller variance. Using the CDM method and the full set of data, we find the posterior mean and median of σ_1^2 to be 186.95 and 173.04 respectively, and those of σ_2^2 as 842.25 and 480.48 respectively. Notably, in both methods, the primary population occurs with $p_e > 2^{-1}$ and has the smaller variance σ_1^2.

    4.2. AAGIS Farm Data Analysis Chambers et al. (2011) considered data from the Australian Agricultural and Grazing Industries Survey (AAGIS) to provide at the regional level the estimated Total Cash Costs (TCC) associated with operation of a farm based on the farm area covariate. In our illustration we treated their sampled data of 1,652 farms as the *finite* population with 27 small areas. In the original dataset, there were 29 small areas. We merged two small areas which had small values of N_i with the neighboring ones. From this population we considered a random sample of 50 units to create our working sample. We drew a sample of 50 units with probabilities proportional to the sizes of the small areas. These 50 data points, along with the identification codes of the 27 small areas are given in Table 5. Here the response Y is the total cash costs associated with operation of the farms, and we consider the farm area as the predictor variable x. A preliminary analysis of the data indicated a long right-tail for the response. To address this excessive skewness, we consider a logarithm transformation of the original response. We also use a similar transformation for the covariate farm area.

    Following Eq. 2.1, we fit a model $Y_{ij}^* = \beta_0 + \beta_1 x_{ij}^* + v_i + e_{ij}$ to predict the $m = 27$ small area means $\theta_i^* = \beta_0 + \beta_1 \bar{x}_{i(p)}^* + v_i$ of Y_{ij}^*'s, for $i = 1, \ldots, 27$, where $x_{ij}^* = \log(x_{ij})$ and $Y_{ij}^* = \log(Y_{ij})$. We use the HB model to predict $\theta_i = \exp(\theta_i^*)$, as prediction in the original scale of the response is preferable. Here θ_i is unknown but the finite population is known, so we approximate θ_i by $\bar{Y}_{iG} = \left(\prod_{j=1}^{N_i} Y_{ij}\right)^{1/N_i}$, the geometric mean of responses of all units in the ith small area.

    The predictors $\hat{\theta}_i$'s are calculated for DG, CDM and GDM HB models, and compared with $\bar{Y}_{iG}$'s. Since the posterior distributions are long-tailed (to the right), we use the median of the $\hat{\theta}_{i,k}$ values, given by $\exp(\beta_{0,k} +$

Table 5: Small Areas and Samples from the AAGIS Farm Data

Small Area	y_{ij}	X_{ij}	Small Area	y_{ij}	X_{ij}	Small Area	y_{ij}	X_{ij}
111	453006	48583.0	223	31913070[a]	260.1	411	47169	2985.0
121	144606	647.7	223	18592	40.5	421	80999	838.0
121	1212066	11660.0	231	108257	744.7	421	121788	2886.6
121	16695291[a]	445.5	231	145922	279.0	422	63476	362.3
122	140520	1042.0	312	410995	48526.0	422	54554	288.0
122	137756	2063.9	313	21792	3200.0	431	123407	1135.7
122	198754	1978.0	314	307842	12040.0	431	55208	500.0
123	83055	628.0	321	50352	1251.0	512	216138	176732.0
123	245025	1205.3	321	140634	3989.0	521	227858	2682.0
123	106124	491.0	322	149343	1537.9	521	147555	1403.6
131	167385	1021.0	322	38283	8461.5	521	49280	354.1
131	335802	1807.0	322	188839	2443.3	522	157571	3152.3
132	134251	2332.0	322	254143	1603.0	531	82563	151.0
221	47380	652.3	331	96744	1862.0	531	220028	40.0
221	231261	2630.0	331	269170	25101.2	631	599960	1126.4
222	68023	683.8	332	216304	23083.9	631	263680	775.3
222	60066	1881.0				711	173869	120800.0

[a] suspected with high unit error variance from a subpopulation

$\beta_{1,k}\bar{x}_i^* + v_{i,k})$, to estimate θ_i. To evaluate the effectiveness of an estimator $\hat{\theta}_i$, we computed the following four deviation measures for the estimator from the "truth"; the average absolute deviation (AAD), the average squared deviation (ASD), average absolute relative deviation (AARD) and the average squared relative deviation (ASRD):

$$\text{AAD}(\hat{\theta}) = \frac{1}{m}\sum_{i=1}^{m}|\hat{\theta}_i - \bar{Y}_{iG}|, \qquad \text{ASD}(\hat{\theta}) = \frac{1}{m}\sum_{i=1}^{m}(\hat{\theta}_i - \bar{Y}_{iG})^2,$$

$$\text{AARD}(\hat{\theta}) = \frac{1}{m}\sum_{i=1}^{m}\frac{|\hat{\theta}_i - \bar{Y}_{iG}|}{\bar{Y}_{iG}}, \qquad \text{ASRD}(\hat{\theta}) = \frac{1}{m}\sum_{i=1}^{m}\frac{(\hat{\theta}_i - \bar{Y}_{iG})^2}{\bar{Y}_{iG}^2}.$$

These summary measures for the three competing methods are given in Table 6.

We also calculated 90% credible intervals (CrI) for θ_i under the DG, CDM and GDM methods, and reported the ratios of their lengths in Table 7. In Fig. 2 we plotted the posterior probabilities of each sampled observation

Table 6: Performance of competing methods

	AAD	ASD	AARD	ASRD
DG	50168	4865362824	0.37	0.34
CDM	49059	4413325890	0.38	0.36
GDM	36857	2592492269	0.22	0.09

coming from the subpopulation 2. We noticed that the GDM method correctly identifies the observations that are believed to have unit-level error distribution from subpopulation 2.

5   Simulation Study

Sinha and Rao (2009) and Chakraborty et al. (2019) employed a simulation study to evaluate and compare performances of suggested extensions of NER models. We follow their example by first assuming a population with $m = 40$ small areas, where each small area has $N_i = 200$ units. We assume a single auxiliary variable x_{ij} for each unit in the population, drawn independently from $N(1, 1)$. The set of auxiliary variables $\mathbf{X}$ is kept fixed for all simulations.

For each simulation, we independently generate sets of area-level random effects v_i for $i = 1, \ldots, m$ from $N(0, 1)$. In the first four simulation setups, we generate e_{ij} such that the mean of the unit-level errors is centered around 0. In these scenarios, we generate e_{ij} from one of the four possible distributions: (i) all e_{ij} are generated independently from $N(0, 1)$; (ii) each e_{ij} is drawn from $N(0, 1)$ with probability $p_e = 0.90$ and from the secondary population with distribution $N(0, 5^2)$ otherwise; (iii) each e_{ij} is drawn from $N(0, 1)$ with probability $p_e = 0.60$ and from $N(0, 5^2)$ otherwise; (iv) e_{ij} are iid from a t-distribution with 4 degrees of freedom. We also perform a fifth simulation motivated by an example in Chambers et al. (2014) in which a very small fraction of e_{ij}'s come from a secondary distribution with a non-zero mean. Here, each e_{ij} is drawn from $N(0, 1)$ with probability $p_e = 0.97$ and from $N(5, 5^2)$ otherwise. Setting $\beta_0 = 1$ and $\beta_1 = 1$ as in Sinha and Rao (2009) for each simulation method, we generate m small area finite populations of $Y_{ij} = \beta_0 + \beta_1 x_{ij} + v_i + e_{ij}$ based on Eq. 2.1.

We compute a summary of auxiliary information for each small area as $\bar{X}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} x_{ij}$ for $i = 1, \ldots, m$. We then take a sample of size $n_i = 4$ from each small area. Using auxiliary information, our goal is prediction of small area means $\bar{Y}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} Y_{ij}, i = 1, \cdots, m$. From each sample, we derive HB predictors from the DG model and robust HB predictors from the outlier-accommodating CDM model and the more general proposed mixture

Table 7: Summary results of full AAGIS data analysis under GDM, CDM and DG methods

Small Area$_i$	N_i	n_i	$\bar{x}_i^*$	$\bar{Y}_{iG}$	Median estimate			90% DG CrI		90% CDM CrI		90% GDM CrI		Lengths of CrIs	
					$\hat{\theta}_{i,DG}$	$\hat{\theta}_{i,CDM}$	$\hat{\theta}_{i,GDM}$	Lower	Upper	Lower	Upper	Lower	Upper	$\frac{DG}{GDM}$	$\frac{CDM}{GDM}$
111	30	1	9.89	201370	226960	234020	274489	89632	659957	97726	636638	148199	510486	1.57	1.49
121	95	3	7.55	185680	299834	288577	207637	136725	995215	139375	883369	111519	424164	2.75	2.38
122	103	3	7.06	129304	153431	155675	134176	72126	328112	73394	308858	88888	201792	2.27	2.09
123	108	3	7.02	161197	148127	150148	133846	65722	318273	71636	303909	86265	203250	2.16	1.99
131	81	2	6.83	122631	161584	164871	155646	73644	388526	79466	378048	94019	266281	1.83	1.73
132	34	1	5.93	43616	133227	133394	87551	49643	318228	49162	308223	46084	156649	2.43	2.34
221	55	2	6.93	108188	138689	146352	108314	58761	312844	62624	304393	62320	186072	2.05	1.95
222	60	2	6.74	100614	123660	123288	80202	48018	268881	49547	243273	45724	136427	2.44	2.14
223	73	2	6.23	80062	207196	220491	76694	97501	676351	103365	642831	36630	156016	4.85	4.52
231	77	2	6.13	87463	133833	141195	109808	56316	318677	61254	304479	68852	180173	2.36	2.18
312	46	1	11.43	327596	264692	285847	400303	87422	905536	99404	854117	198483	812195	1.33	1.23
313	30	1	10.02	218926	161457	179806	180586	50375	408233	53825	435024	53637	473588	0.85	0.91
314	40	1	9.94	255903	218557	228872	279868	86875	642940	91385	600219	152722	546724	1.41	1.29
321	79	2	6.81	103095	129885	133642	85908	50192	281050	52393	272303	49185	146315	2.38	2.26
322	117	4	8.18	198718	161917	160211	182018	73386	332377	78158	307968	107870	296511	1.37	1.22
331	51	2	7.31	87874	153162	153655	115249	65710	358745	67244	336329	69382	189235	2.44	2.25
332	19	1	8.61	178834	183179	187933	161215	77278	464113	79663	437318	90782	280135	2.04	1.89
411	36	1	10.81	206493	193958	204356	213142	64162	565956	65898	566532	97021	454788	1.40	1.40
421	51	2	7.64	135527	150222	151062	121699	61357	312654	66887	324839	74755	192536	2.13	2.19
422	80	2	6.87	95703	128930	127823	94119	48322	269358	52539	256305	56972	148261	2.42	2.23
431	74	2	6.81	120257	133921	134183	98098	52504	292267	54869	281025	59034	158215	2.42	2.28
512	31	1	12.29	256312	261533	280981	344572	75382	965835	83982	878019	170695	761121	1.51	1.34

Table 7: Continued

Small					Median estimate			90% DG CrI		90% CDM CrI		90% GDM CrI		Lengths of CrIs	
Area$_i$	N_i	n_i	$\bar{x}_i^*$	$\bar{Y}_{iG}$	$\hat{\theta}_{i,DG}$	$\hat{\theta}_{i,CDM}$	$\hat{\theta}_{i,GDM}$	Lower	Upper	Lower	Upper	Lower	Upper	$\frac{DG}{GDM}$	$\frac{CDM}{GDM}$
521	83	3	7.60	215403	151586	155214	138008	69269	319690	72285	303267	91402	208303	2.14	1.98
522	47	1	8.19	245206	171282	176102	158791	72338	410603	74490	419638	91303	282574	1.77	1.80
531	60	2	6.53	124681	149993	151891	139569	64711	352785	70671	332599	78966	280841	1.43	1.30
631	62	2	6.62	133561	180222	182837	197633	84495	461879	86294	450867	96456	391833	1.28	1.23
711	30	1	12.47	503157	261186	284549	347653	72572	984729	84398	927491	165286	788502	1.46	1.35

Shuchi Goyal et al.

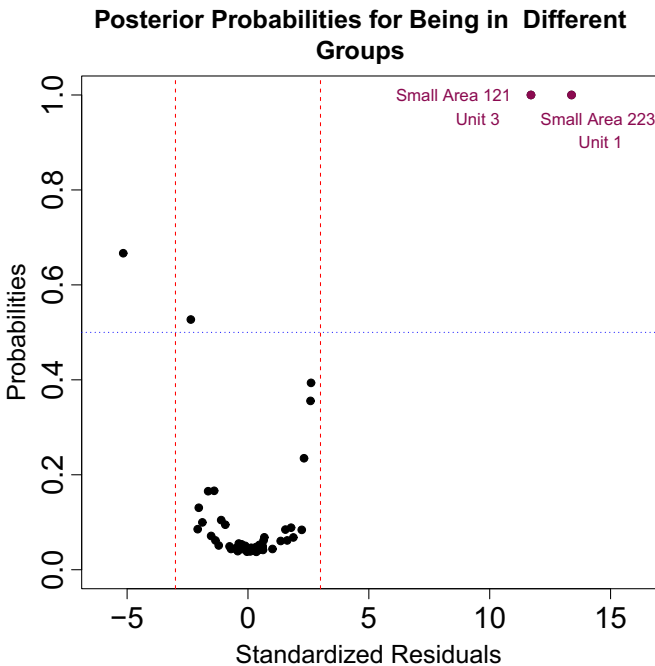**Posterior Probabilities for Being in Different Groups**

Figure 2: Posterior probabilities of observations coming from subpopulation 2

model. These predictors are denoted as DG, CDM, and GDM, respectively, in subsequent data visualizations included in this paper. Since all three HB methods perform equally well when the unit-level errors contain no contamination, the plots for this simulation setup are relegated to Appendix A.2. We visualize the results of the other four simulation methods in Figs. 3, 4 and 5.

For each simulation setup, we simulate $S = 100$ populations. For the sth simulated population, where $s = 1, \ldots, S$, we compute the true small area means $\theta_i^{(s)}$. We denote the predictors of small area means calculated using HB methods as $\hat{\theta}_i^{(s)}$ and the variances of those predictors as $V_i^{(s)}$. For each HB method, given the predicted small area means $\hat{\theta}_i^{(s)}$, we calculate empirical biases as $eB_i = \frac{1}{S} \sum_{s=1}^{S} \left(\hat{\theta}_i^{(s)} - \theta_i^{(s)} \right)$ and empirical mean squared errors as $eM_i = \frac{1}{S} \sum_{s=1}^{S} \left(\hat{\theta}_i^{(s)} - \theta_i^{(s)} \right)^2$.

In Fig. 3, we provide plots of empirical biases and empirical mean squared errors (MSEs) for HB predictors considered. None of the HB predictors shows signs of systematic bias. However, in the simulation setup where
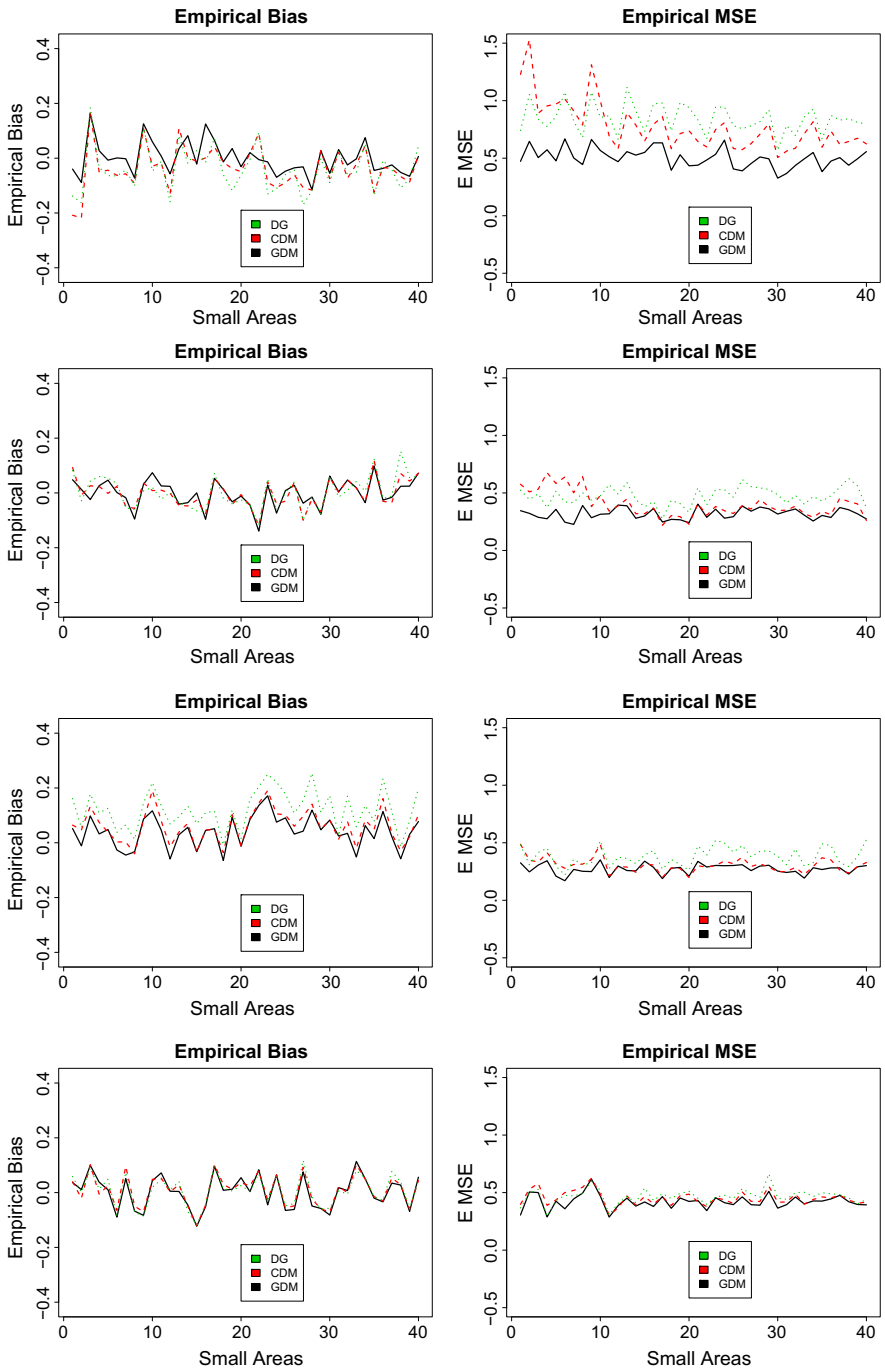
Figure 3: Plot of empirical biases and empirical MSEs of $\hat{\theta}$s
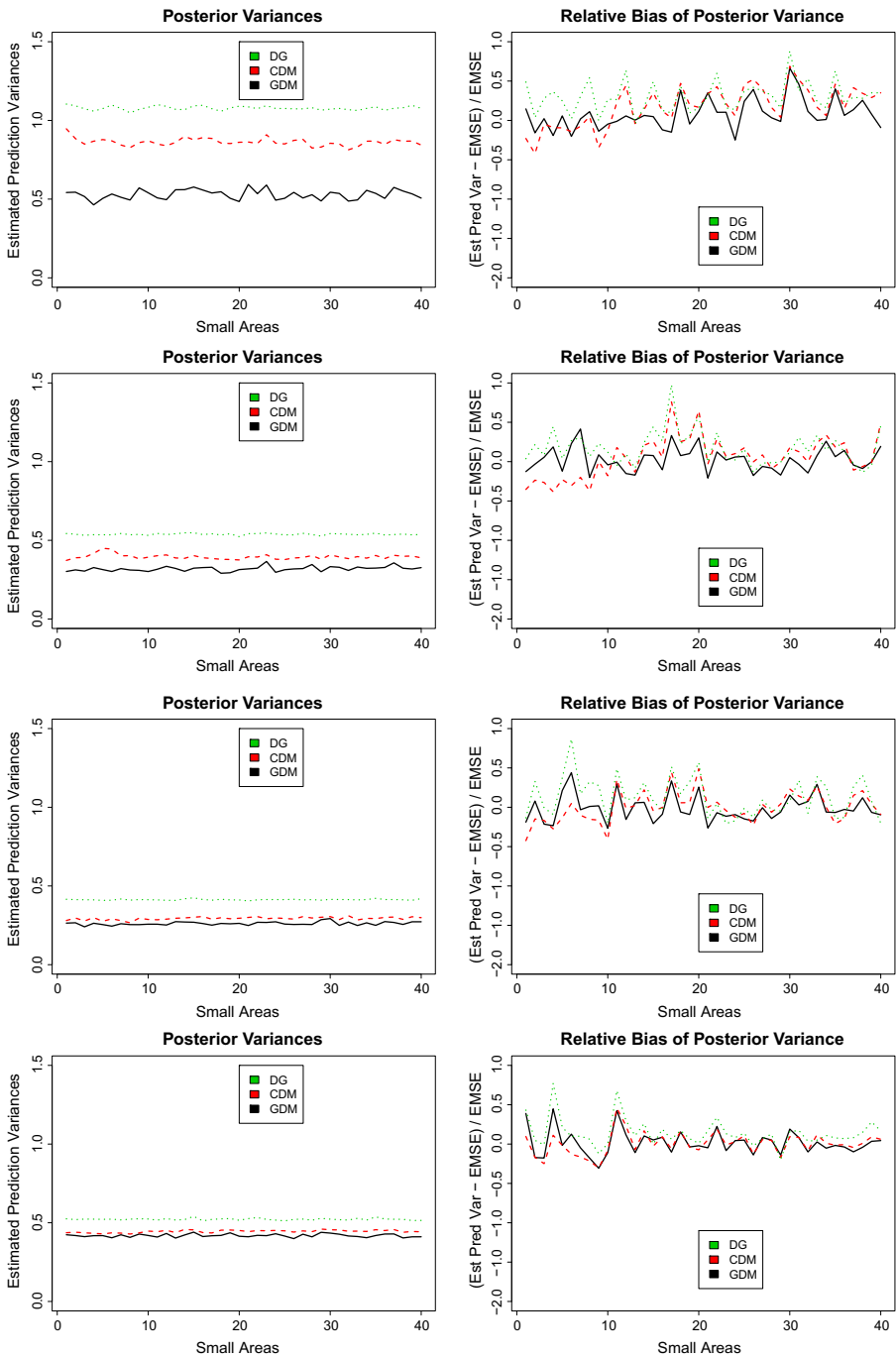
Shuchi Goyal et al.

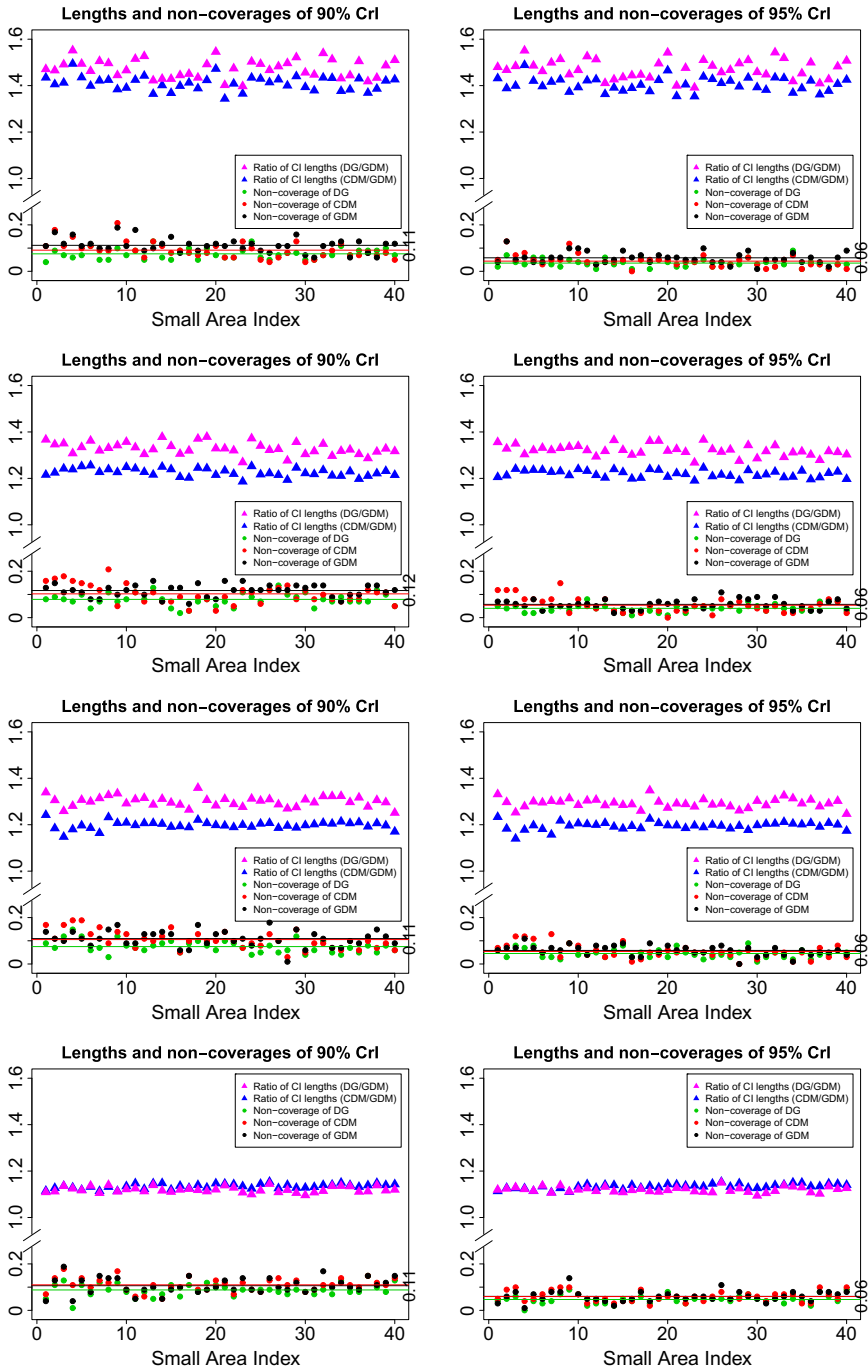Figure 4: Plot of posterior variances and their empirical relative biases

Figure 5: Plot of lengths and non-coverages of credible intervals (CrI)

$p_e = 0.6$, the empirical biases of the GDM HB predictors seem to have smaller variability than the empirical biases of the other two HB predictors. In the case of 3% contamination in e_{ij} or where e_{ij} is determined by a t-distribution, the three models perform equally well in producing small MSEs. In the case of 10% contamination, the MSEs of the CDM and GDM HB predictors are approximately equal for most of the small areas but smaller than the DG model prediction. The most substantial difference among the three models results in the case where $p_e = 0.6$. Here, the GDM predictor has the lowest MSEs overall of the three methods, followed by the CDM predictor and then by the DG predictor. Moreover, the GDM model MSEs stay generally stable across all small areas.

Next, Fig. 4 shows posterior variances $V_i^{(s)}$ for 40 small areas and the relative biases of those variances, calculated as $RE_V = \{(1/S)\sum_{s=1}^{S} V_i^{(s)} - eM_i\}/eM_i$. The CDM and GDM predictors seem to enjoy lower posterior variances than the DG model. Furthermore, as the amount of contamination increases, the GDM model also produces lower posterior variances than the CDM model. The differences between the three models become more pronounced as contamination increases. The DG model also displays a mild tendency toward positive relative bias when calculating posterior variance. The CDM and GDM models do not show systematic bias in calculations of posterior variance and overall perform equally well.

Figure 5 shows empirical non-coverage probabilities of 90% and 95% credible intervals of small area means θ_i. For each simulation setup, we also use solid horizontal lines to show the mean non-coverage probability of the credible intervals produced by each method. For a Bayesian method, we compute our 90% credible interval $I_{i,90}^{(s)}$ for θ_i by the 5th and 95th percentiles of the posterior distribution of θ_i. We then calculate the non-coverage probabilty of this credible interval as $eC_{i,90} = \frac{1}{S}\sum_{s=1}^{S} I[\theta_i^{(s)} \notin I_{i,90}^{(s)}]$. The same calculations are done for 95% credible intervals using the 2.5th and 97.5th percentiles. The label on the right axis of each plot is the non-coverage probability of the GDM HB credible intervals. The plots also show two ratios which compare the lengths of the DG HB and CDM HB credible intervals to those of the GDM HB credible intervals. We denote the length of the 90% credible interval $I_{i,90}^{(s)}$ as $L_{i,90}^{(s)}$. The empirical average length of the credible interval of θ_i for a specific HB method is then computed as $\bar{L}_{i,90} = \frac{1}{S}\sum_{s=1}^{S} L_{i,90}^{(s)}$. Again, this calculation is repeated for the 95% credible intervals. We see the credible intervals produced by the DG HB method consistently have the lowest non-coverage probabilities for each simulation, compared to the CDM HB and GDM HB intervals.

We also observe that the credible intervals by the DG HB model are on average larger than those developed from the other two models, except for the t-distributed e_{ij} scenario where the DG HB and CDM HB credible intervals have similar lengths. Though the DG HB credible intervals most often capture the true value $\theta_i^{(s)}$ and have low non-coverage probabilities, they are longer than the GDM HB credible intervals, which closely attain the target coverage probability. While CDM HB and GDM HB intervals have similar non-coverage probabilities and nearly achieve the target when e_{ij} is generated from a t-distribution, the ratios of average lengths (CDM/GDM) are consistently higher than one when greater levels of contamination are introduced into the population, indicating that the narrower GDM HB credible intervals are as successful as the CDM HB credible intervals in capturing the true values.

At 3% contamination of e_{ij} from the secondary distribution ($p_e = 0.97$), the non-coverage probabilities of the GDM HB and CDM HB credible intervals remain approximately equal, but the 90% and 95% intervals produced by the CDM HB model are up to 5% greater in length than their respective GDM HB measures. When 10% of e_{ij} come from a secondary distribution ($p_e = 0.90$), the non-coverage probabilities of the credible intervals found from the CDM HB approach are slightly lower than those found from the GDM HB model, but the CDM HB credible intervals are also about 10% greater in length than their respective GDM HB measures. When e_{ij} comes from a primary distribution with probability $p_e = 0.60$, the CDM HB model credible intervals are about 40% longer than those given by the GDM HB model but continue to have a slightly lower non-coverage probability. We note that the non-coverage probabilities of the GDM HB credible intervals seem to be consistent across all simulation setups. In contrast, the non-coverage probabilities of the CDM HB credible intervals appear to decrease when the concentration of e_{ij} from a secondary distribution increases, but the CDM HB credible intervals become wider relative to the GDM HB credible intervals in higher contamination cases.

6   Conclusion

Since Battese et al. (1988) introduced the NER model it has been the basis for many important developments in small area estimation for unit-level data. Datta and Ghosh (1991) applied the HB method to the NER model to develop Bayesian inference for small area means. This approach, however, is not robust in the presence of outliers or under non-normality of unit-level errors. The HB method proposed by Chakraborty et al. (2019), which also relied on an HB approach to the NER model, built upon the work

of Datta and Ghosh (1991) to accommodate populations contaminated with outliers due to unit-level errors.

The CDM model is robust in the presence of outliers, but not under circumstances where the proportion of unit-level errors from the secondary distribution is fairly large. In this paper, we propose an alternate HB approach to extend the NER model for more general cases where unit-level errors come from a mixture of two different normal distributions. Based on simulation studies, we find that the proposed model provides HB estimates with lower empirical MSEs, posterior variances and narrower credible intervals than the DG and CDM HB models. The consistent superior performance of the proposed model to the DG and CDM HB models regardless of the presence of mixture in the unit-level error indicates that there is no loss to applying it to all data sets.

References

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. Am. Stat. Assoc.* **83**, 28–36.

CHAKRABORTY, A., DATTA, G. S. and MANDAL, A. (2019). Robust Hierarchical Bayes Small Area Estimation for Nested Error Regression Model. *Int. Stat. Rev.* **87**, 158–176.

CHAMBERS, R. L. (1986). Outlier robust finite population estimation. *J. Am. Stat. Assoc.* **81**, 1063–1069.

CHAMBERS, R. L., CHANDRA, H., SALVATI, N. and TZAVIDIS, N. (2014). Outlier robust smallarea estimation. *Journal of the Royal Statistical Society Series B* **76**, 47–69.

CHAMBERS, R. L., CHANDRA, H. and TZAVIDIS, N. (2011). On bias-robust mean squared error estimation for pseudo-linear small area estimators. *Survey Methodology* **37**, 153–170.

DATTA, G. and GHOSH, M. (1991). Bayesian prediction in linear models: Applications to small area estimation. *Annals Stat.* **19**, 1748–1770.

DATTA, G. S. and LAHIRI, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Stat. Sin.* **10**, 613–627.

EFRON, B. and MORRIS, C. (1973). Stein's Estimation Rule and Its Competitors − An Empirical Bayes Approach. *J. Am. Stat. Assoc.* **68**, 117–130.

FAY, R. E. and HERRIOT, R. A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *J. Am. Stat. Assoc.* **74**, 269–277.

PRASAD, N. G. N. and RAO, J. N. K. (1990). On the estimation of mean square error of small area predictors. *J. Am. Stat. Assoc.* **85**, 163–171.

SINHA, S. K. and RAO, J. N. K. (2009). Robust small area estimation. *Can. J. Stat.* **37**, 381–399.

STEIN, C. (1955). Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution, Proccedings of the Third Berkeley Symposium. *University of California Press* **1**, 197–206.

Appendix A

A Integrability of Joint Posterior Probability Density Function

Chakraborty et al. (2019) showed that the joint posterior density function of $\boldsymbol{\beta}, \sigma_1^2, \sigma_2^2, p_e,$ and σ_v^2 is proper. In particular, they showed that the function

$$L(\boldsymbol{\beta}, \sigma_1^2, \sigma_2^2, p_e, \sigma_v^2) \frac{I_{\{\sigma_1^2 < \sigma_2^2\}}}{(\sigma_2^2)^2} \tag{A.1}$$

is integrable with respect to $\boldsymbol{\beta}, \sigma_1^2, \sigma_2^2, p_e,$ and σ_v^2, where $L(\boldsymbol{\beta}, \sigma_1^2, \sigma_2^2, p_e, \sigma_v^2)$ is the likelihood function based on the distribution $y_{ij}, j = 1, \ldots, n_i, i = 1, \ldots, m$ obtained as the marginal distribution from $(I)-(III)$ in Section 2. Similar arguments show that $L(\boldsymbol{\beta}, \sigma_1^2, \sigma_2^2, p_e, \sigma_v^2) \frac{I_{\{\sigma_1^2 \geq \sigma_2^2\}}}{(\sigma_1^2)^2}$ is also integrable with respect to the same variables. Now we note that

$$
\begin{aligned}
\frac{I_{\{2^{-1} < p_e < 1\}}}{(\sigma_1^2 + \sigma_2^2)^2} &\leq \frac{1}{(\sigma_1^2 + \sigma_2^2)^2} = \frac{I_{\{\sigma_1^2 < \sigma_2^2\}} + I_{\{\sigma_1^2 \geq \sigma_2^2\}}}{(\sigma_1^2 + \sigma_2^2)^2} \\
&= \frac{1}{(\sigma_2^2)^2} \left(\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \right)^2 I_{\{\sigma_1^2 < \sigma_2^2\}} + \frac{1}{(\sigma_1^2)^2} \left(\frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \right)^2 I_{\{\sigma_1^2 \geq \sigma_2^2\}} \\
&< \frac{I_{\{\sigma_1^2 < \sigma_2^2\}}}{(\sigma_2^2)^2} + \frac{I_{\{\sigma_1^2 \geq \sigma_2^2\}}}{(\sigma_1^2)^2}.
\end{aligned}
$$

This implies,

$$L(\boldsymbol{\beta}, \sigma_1^2, \sigma_2^2, p_e, \sigma_v^2) \frac{I_{\{2^{-1} < p_e < 1\}}}{(\sigma_1^2 + \sigma_2^2)^2} < L(\boldsymbol{\beta}, \sigma_1^2, \sigma_2^2, p_e, \sigma_v^2) \left(\frac{I_{\{\sigma_1^2 < \sigma_2^2\}}}{(\sigma_2^2)^2} + \frac{I_{(\sigma_1^2 \geq \sigma_2^2)}}{(\sigma_1^2)^2} \right). \tag{A.2}$$

The LHS of Eq. A.2 is bounded above by two integrable functions, hence it is also integrable.

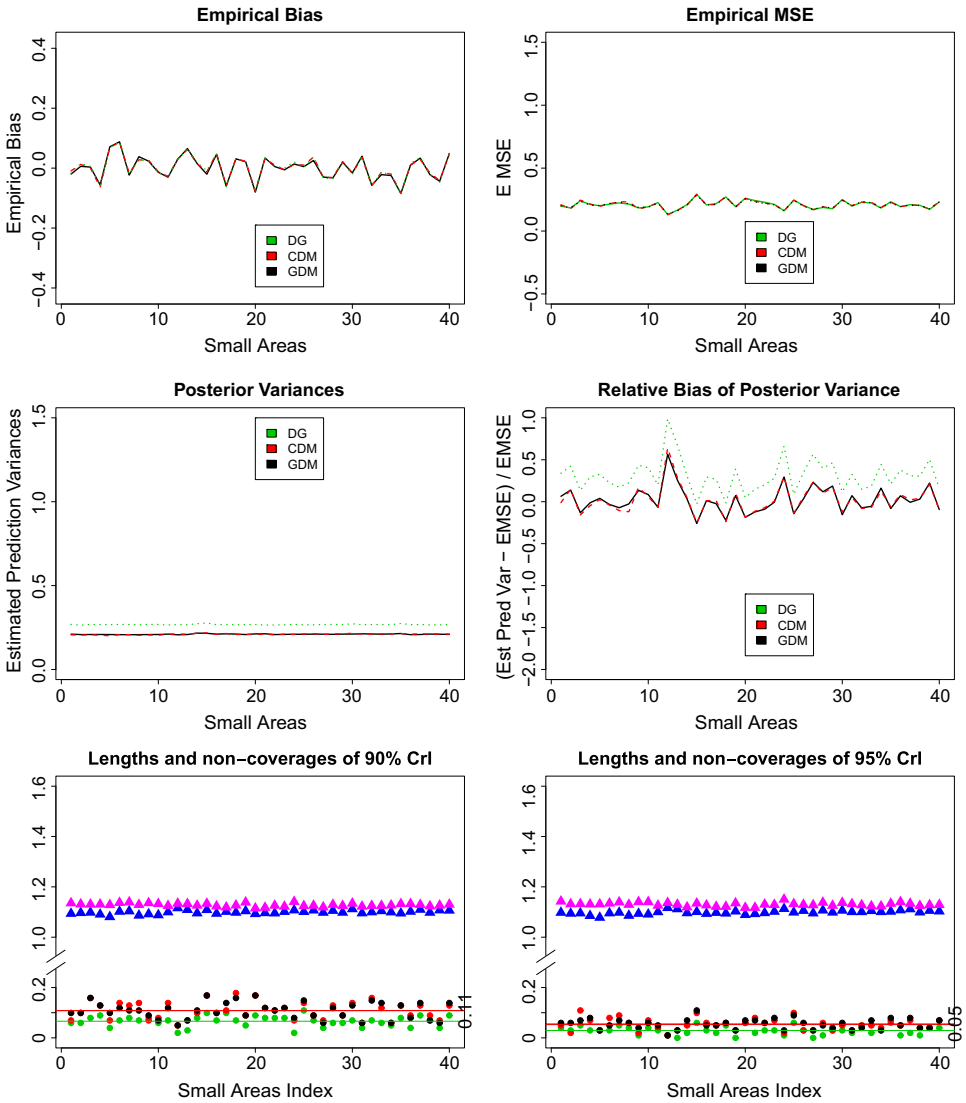### A.2 Simulation Results with no Contamination of Unit-Level Errors

Figure 6: Plots of various measures of $\hat{\theta}$s when no unit-level contamination is present

Shuchi Goyal
Department of Statistics,
University of California
at Los Angeles
Los Angeles, CA, 90095, USA

Gauri Sankar Datta and
Abhyuday Mandal
Department of Statistics, University
of Georgia, Athens, GA, 30602 USA
E-mail: amandal@stat.uga.edu