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

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

Gauri Sankar Datta
Department of Statistics
University of Georgia
Athens, GA 30602

Abhyuday Mandal
Department of Statistics
University of Georgia
Athens, GA 30602

April 1, 2019

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 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. (2018) 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 (1991) and the Chakraborty et al. (2018) HB methods. 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 prediction mean squared error, coverage probabilities and lengths of credible intervals for the small area means.

Key words: 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 geographic domains (e.g., states, counties) and/or demographic domains (e.g., age, race, gender). However, the sample available from many of these domains 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 appropriate 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.

Battese, Harter, and Fuller (1988) proposed the popular nested-error regression (NER) model, a unit-level model, to develop small area estimates based on data available on the individual sampled units. The NER model is developed under the normality assumption of small area random effects and unit-level random errors, which are also assumed to be

independently distributed with zero means and suitable variances. The NER model has been the basis for both frequentist and Bayesian approaches for unit-level data to develop reliable small area estimates. 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 (HB) estimates of finite population small area means.

While the HB estimates proposed by Datta and Ghosh (1991) are effective for populations in which the unit-level random errors come from a single normal distribution, they are less effective when the distribution of this error is a mixture of normal distributions. This scenario can be formulated by a two-component mixture model for the unit-level errors. The model accommodates two normal distributions underlying the unit-level error term 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 mean zero, but a larger variance than that of the non-outliers.

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 in the data, Chakraborty, Datta, and Mandal (2018) 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. (2018) 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").

It should be noted that the model proposed by Chakraborty et al. (2018) 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.

Sinha and Rao (2009) proposed a robust empirical best linear unbiased prediction (REBLUP) approach to address the presence of outliers in the unit-level error and/or in the area-level random effects. Chakraborty et al. (2018) showed in their simulations that their Bayesian method and the REBLUP method perform very similarly. For clarity in the presentation

of our simulations, we excluded the Sinha-Rao frequentist method and we focus only on Bayesian methods.

This paper is organized as follows: in Section 2, we review the relevant existing HB approaches to small area estimation for unit-level data. Next, we propose our new model in Section 3. We apply our method 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 twelve county-level crop areas cultivated for corn in some 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). 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 proposed method. Concluding comments are provided in Section 6, and relevant proofs and in-depth details are relegated to the Appendix and Supplementary Information sections.

2 Current Unit-Level HB Models for Small Area Estimation

At the base of many popular unit-level models for small area estimation is the nested-error regression (NER) model proposed by Battese et al. (1988). This model supposes that given a population with m small areas and N_i units in the ith small area,

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

where Y_{ij} is the response variable for the jth unit of the ith small area and $\boldsymbol{x_{ij}} = (x_{ij1}, \ldots, x_{ijq})^T$ is a $q \times 1$ vector of values for predictor variables for that observation, with $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_q)^T$ denoting the vector of regression coefficients. The variables v_i and e_{ij} account for area- and unit-level random errors, respectively, and are assumed to be independent of each other. We further assume that v_i's are i.i.d. $N(0, \sigma_v^2)$. Under appropriate distributional assumptions for the e_{ij}'s, the goal is to predict small area finite populations means $\bar{Y}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} Y_{ij}, i = 1, \ldots, m$, based on a sample of size $n = \sum_{i=1}^{m} n_i$, where n_i is the number of units sampled from the ith small area.

One Bayesian approach to extending the basic NER model requires assigning noninformative priors to parameters. Posterior means and variances for $\bar{Y}_i$'s for the HB method are calculated using MCMC methods such as Gibbs sampling. The method developed by Datta and Ghosh

(1991), referred to hereafter as DG, uses the above model based on independent normal distributions for e_{ij}'s, i.i.d. $N(0, \sigma_e^2)$, to derive hierarchical Bayesian prediction of $\bar{Y}_i$'s, $i = 1, \ldots, m$. For our data analysis and simulations in the next sections, we complete the DG hierarchical Bayes formulation using the following noninformative prior:

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

The DG model, which is built under normality assumptions, lacks robustness in the presence of unit-level error outliers. As a part of the effort to address this issue of robustness, Chakraborty et al. (2018) proposed a two-component normal mixture for the unit-level error distribution. The latter 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 $\boldsymbol{\beta} = (\beta_1, ..., \beta_q)^T, v_i, z_{ij}, p_e, \sigma_1^2, \sigma_2^2$, and σ_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 $j = 1, ..., N_i, i = 1, ..., m$.

(II). The indicator variables z_{ij} are i.i.d. with $P(z_{ij} = 1|p_e) = p_e$ and $P(z_{ij} = 0|p_e) = 1 - p_e$ 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 σ_v^2, $v_i \overset{\text{iid}}{\sim} N(0, \sigma_v^2)$ for all i.

A key component 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 occur in the various small areas with equal probability $(1 - p_e)$. The 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 $\boldsymbol{\beta} \sim Uniform(R^q), \sigma_v^2 \sim Uniform(R^+)$, $\pi(\sigma_1^2, \sigma_2^2) \propto \frac{1}{(\sigma_2^2)^2} I(\sigma_1^2 < \sigma_2^2)$, and $p_e \sim Uniform(0, 1)$.

Chakraborty et al. (2018) show via simulations that the CDM model performs as well as the DG 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). The REBLUP is a frequentist approach to develop robust predictors to handle outliers. We now propose a new HB model which is more appropriate for dealing with mixture populations comprised of two subgroups, differentiated by their variances.

3 An HB Normal Mixture Model for Unit-Level Error

The proposed model is a mixture extension of the nested-error regression model which accounts for unit-level error terms coming from two different normal distributions. While the models discussed in the previous section accommodate the presence of outliers, our proposed model further generalizes the mixture model for the case in which the proportion of observations coming from the subpopulation with a larger variance is large enough that these data points may no longer be considered outliers in the traditional sense.

We first consider the general form of the nested-error regression model for unit-level data, given in Equation (1). 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.

We also assign the same independent noninformative uniform priors to $\boldsymbol{\beta}$ and σ_v^2. An important difference in the priors assumed in our model and those used for CDM is that we assign $\pi(\sigma_1^2, \sigma_2^2) \propto \dfrac{1}{(\sigma_1^2 + \sigma_2^2)^2}$ and $p_e \sim Uniform(\frac{1}{2}, 1)$. In other words, as a part of our mixture formulation, we do not assume $\sigma_1^2 < \sigma_2^2$ but instead constrain $P(z_{ij} = 1) > P(z_{ij} = 0)$, which ensures that the parameters in the mixture model are identifiable.

To proceed with Bayesian inference aided by an improper prior, the propriety of the posterior distribution must be shown. This propriety is demonstrated in the Appendix A.1. Details on the procedure for Gibbs sampling for fitting the model can be found in the Supplementary Material included at the end of this paper.

4 Data Analysis

4.1 Prediction of County Crop Areas

Battese et al. (1988) illustrated the NER model using data on corn and soybean production from 12 Iowa counties. Based on the NER model and a frequentist approach, they considered EBLUP prediction of mean hectares of corn grown in each county based on two auxiliary variables provided by LANDSAT satellite data from the U.S. Department of Agriculture. Of 37 units observed from the 12 counties, one observation from Hardin County was suspected of being an outlier. Battese et al. (1988) suggested removing the suspected outlier 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. Chakraborty et al. (2018) demonstrated that when the full set of sampled observations is used, their HB prediction (CDM HB) of corn hectares in Hardin County is closer to estimates produced using other outlier-accommodating models, such as the robust EBLUP approach proposed by 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 as well as 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, which is appropriate for mixture of normal populations, 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, these estimates are substantially different from those from the non-robust DG HB method.

The GDM model identifies the second Hardin County observation as a likely member of a secondary subpopulation when analyzing the full data, as shown in the left panel of Figure 1. The posterior probability that this observation may belong to the secondary population

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

SA	n_i	DG		CDM		GDM	
		Mean	SD	Mean	SD	Mean	SD
1	1	123.8	11.7	123.4	9.8	123.6	11.3
2	1	124.9	11.4	126.6	10.3	125.8	10.2
3	1	110.0	12.3	108.0	11.3	107.7	11.7
4	2	114.2	10.7	112.3	10.2	112.0	10.7
5	3	140.3	10.8	142.1	8.1	142.4	8.4
6	3	110.0	9.6	111.4	7.6	111.6	7.3
7	3	116.0	9.7	114.3	7.6	113.7	7.9
8	3	123.2	9.5	122.7	7.9	122.3	7.7
9	4	112.6	9.9	113.9	6.9	114.3	6.8
10	5	124.4	8.9	123.5	6.1	123.6	6.1
11	5	111.3	8.9	108.2	6.8	108.1	6.9
12	6	130.7	8.3	135.3	7.5	136.5	7.4

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
1	1	122.0	11.6	121.7	9.7	121.9	10.2
2	1	126.4	10.9	127.2	9.7	126.3	9.8
3	1	107.6	12.4	105.6	10.1	105.3	10.9
4	2	108.9	10.5	108.2	8.7	108.0	9.3
5	3	143.6	9.7	144.1	7.0	144.3	7.0
6	3	112.3	9.7	112.5	6.5	112.3	6.7
7	3	113.4	9.1	112.5	6.8	111.5	7.4
8	3	121.9	8.8	121.9	6.6	121.8	6.7
9	4	115.5	9.2	115.7	5.7	115.8	6.1
10	5	124.8	8.4	124.4	5.4	124.6	5.5
11	5	107.7	8.5	106.3	5.7	106.0	5.6
12	5	142.6	9.0	143.5	5.9	143.6	5.6

is 0.62, which is about 2.5 times the corresponding prior probability 0.25. For the other observations, most of these posterior probabilities are near 0.25, not much different from their prior values. We note here that in the right panel of Figure 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 standard error given by each method for the reduced data is given in Table 2. For the first 11 counties, the estimates produced by each model change only slightly when compared to 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 most substantial in the DG estimate. This makes sense, as the estimates from the mixture models should be less sensitive to the presence of outliers. The corn hectare estimates in County 12 given by the three models are also much closer in value to each other relative to the full data set.

For the reduced data set, a comparison of posterior standard deviations of the HB estimates show that the standard deviations of the estimates by the mixture models are consistently lower than those given in by the DG 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 model seem to change only moderately.

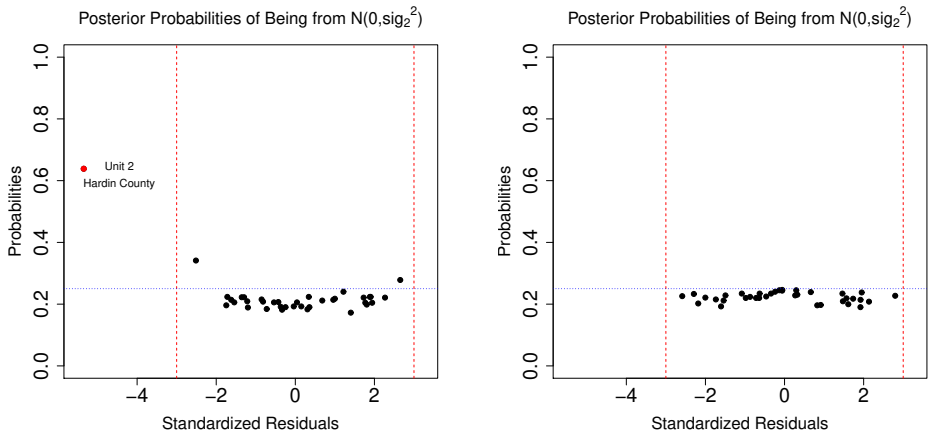

Figure 1: Posterior probabilities of observations coming from subpopulation 2 in *full* and *reduced* corn data

Tables 3 and 4 show posterior means 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 $\frac{1}{2}$ 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 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 estimates 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 $p_e > \frac{1}{2}$ and has the smaller variance σ_1^2.

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

Estimates	Datta-Ghosh HB				Chakraborty et al. HB				Proposed Mixture HB			
	Full Data		Reduced Data		Full Data		Reduced Data		Full Data		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
$\hat{\sigma}_v^2$	175.08	193.42	228.88	178.87	203.53	178.55	257.32	183.64	205.71	176.27	266.17	202.47
$\hat{\sigma}_1^2$	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	Datta-Ghosh HB				Chakraborty et al. HB				Proposed Mixture HB			
	Full Data		Reduced Data		Full Data		Reduced Data		Full Data		Reduced Data	
	Median	IQR	Median	IQR	Median	IQR	Median	IQR	Median	IQR	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

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*	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*	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

* suspected outlier

4.2 AAGIS Farm Data Analysis

Chambers et al. (2011) considered a 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 the farm based on the farm area covariate. We acknowledge Chambers et al. (2011) for the dataset. For our illustration purpose 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 x, the farm area.

Following Equation (1), we fit a model $Y^*_{ij} = \beta_0 + \beta_1 x^*_{ij} + v_i + e_{ij}$ to predict the $m = 27$ small area finite populations geometric means (GM) $\bar{Y}_{iG} = \{\prod_{j=1}^{N_i} Y_{ij}\}^{1/N_i}$, for $i = 1, \ldots, 27$, where

$x^*_{ij} = \log(x_{ij})$ and $Y^*_{ij} = \log(Y_{ij})$. The use of the GM has two advantages over the arithmetic mean $\bar{Y}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} Y_{ij}$. First, in computing prediction of $\bar{Y}_{iG}$, knowledge of unit-level x^*_{ij} will not be needed, unlike that for $\bar{Y}_i$. It is enough to know the population means $\bar{x}^*_i$. Also, the influence of an extreme (very large) observation on the GM is less than its influence on the arithmetic mean. (For n positive numbers $w_1, \cdots, w_n$, $\partial \log(\bar{w})/\partial w_i = 1/(n\bar{w}) < 1/(nw_i) = \partial \log(\bar{w}_G)/\partial w_i$ for a w_i larger than $\bar{w}$.)

We approximate inference of $\bar{Y}_{iG}$ by that of $\theta_i = \exp\{\beta_0 + \beta_1 \bar{x}^*_i + v_i\}$. The predictors $\hat{\theta}_i$'s are calculated for DG, CDM and GDM methods, and the true values of θ_i is given by the GM's of the Y_{ij}'s in the ith small area. With G Gibbs sampled values, for each competing methods, we calculate $\hat{\theta}_i$ by $\frac{1}{G}\sum_{k=1}^{G} \hat{\theta}_{i,k}$ where $\log(\hat{\theta}_{i,k})$ is given by $\beta_{0,k} + \beta_{1,k}\bar{x}^*_i + v_{i,k}$. Since the finite population of Y_{ij} values are known, we know the true $\bar{Y}_{iG}$ values for each small area. To evaluate the effectiveness of an estimator $\hat{\theta}$, 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):

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

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

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

Table 6: Performance of competing methods

	AAD	ASD	AARD	ASRD
DG	63754.99	6488672193	0.5228	0.6491
CDM	62575.70	6158475672	0.5136	0.6278
GDM	41379.93	3077805866	0.2528	0.1183

We also calculated the credible intervals for θ_i under the DG, CDM and GDM methods, and reported the ratio of their lengths in Table 7. In Figure 2 we plotted the posterior probabilities of each sampled observing coming from a subpopulation 2. We notice that the GDM method identifies the suspected outliers correctly belonging to the subpopulation 2.

Table 7: Summary results of CCT data analysis under GDM, CDM and GD methods

Small				Estimates			Standard Deviations			Ratio of CI	
Area$_i$	N_i	$\bar{\boldsymbol{x}}_i^*$	θ_i	$\hat{\theta}_{i,GDM}$	$\hat{\theta}_{i,CDM}$	$\hat{\theta}_{i,DG}$	GDM	CDM	DG	$\frac{CDM}{GDM}$	$\frac{DG}{GDM}$
111	30	9.89	201370	296635	288148	286318	123534	217026	239487	1.49	1.57
121	95	7.55	185680	232937	371268	396077	105410	284740	304627	2.38	2.75
122	103	7.06	129304	138383	169237	170695	35847	78543	87881	2.09	2.27
123	108	7.02	161197	137833	163745	164617	37045	77838	91257	1.99	2.16
131	81	6.83	122631	164568	188203	188424	55414	103254	119452	1.73	1.83
132	34	5.93	43616	93194	151337	153407	38906	92115	106497	2.34	2.43
221	55	6.93	108188	114051	159463	156218	40999	80615	95054	1.95	2.05
222	60	6.74	100614	84069	132039	136731	28750	64212	79009	2.14	2.44
223	73	6.23	80062	84780	276393	276656	46066	198431	234573	4.52	4.85
231	77	6.13	87463	115350	156562	155549	35867	82737	93934	2.18	2.36
312	46	11.43	327596	442005	361870	359552	212693	297268	345169	1.23	1.33
313	30	10.02	218926	209548	204906	189261	155131	133916	130460	0.91	0.85
314	40	9.94	255903	304915	275962	275211	129423	197411	220901	1.29	1.41
321	79	6.81	103095	90498	145805	142656	30983	78992	77490	2.26	2.38
322	117	8.18	198718	189878	172930	177376	59449	74691	89379	1.22	1.37
331	51	7.31	87874	120960	171823	173882	40977	92970	97898	2.25	2.44
332	19	8.61	178834	170278	215274	217694	61022	130002	166648	1.89	2.04
411	36	10.81	206493	240076	246761	240358	146572	180646	189762	1.40	1.40
421	51	7.64	135527	125931	166466	163524	37105	83112	87272	2.19	2.13
422	80	6.87	95703	97545	137330	140267	28776	66128	74603	2.23	2.42
431	74	6.81	120257	102240	147234	148054	31447	74554	83519	2.28	2.42
512	31	12.29	256312	394080	359739	359229	253326	316639	375523	1.34	1.51
521	83	7.60	215403	142684	168224	168090	37312	77969	90081	1.98	2.14
522	47	8.19	245206	169260	205402	199279	64568	135174	133997	1.80	1.77
531	60	6.52	124681	155370	171356	172361	67740	92367	108688	1.30	1.43
631	62	6.62	133561	213976	215842	215939	94033	131931	160268	1.23	1.28
711	30	12.46	503157	403845	366298	373478	302952	309189	443391	1.35	1.46

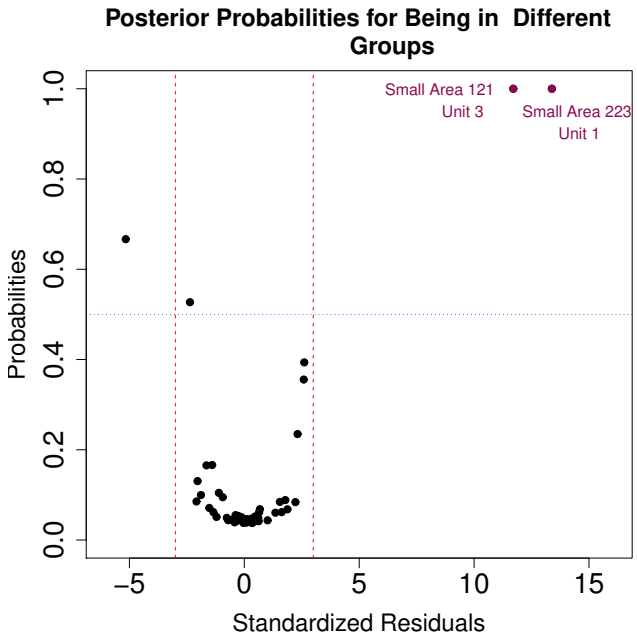

Figure 2: Posterior probabilities of observations coming from subpopulation 2

5 Simulation Study

Sinha and Rao (2009) and Chakraborty et al. (2018) employ a simulation study to evaluate and compare the 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 $N(0, 1)$. Each small area has a population of size $N_i = 200$. 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 portion 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 Equation (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 $\bar{Y}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} Y_{ij}, i = 1, \cdots, m$ for finite populations with large N_i and small ratio n_i/N_i. 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 model. These predictors are denoted as DG, CDM, and GDM, respectively, in subsequent data visualizations included in this paper. Because the three HB methods perform equally well in all observed HB methods 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 Figures 3 to 5.

For each simulation setup, we simulate $S = 100$ populations. For the s^{th} 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 $V_i^{(s)}$. In Figure 3, we provide plots of empirical biases and empirical mean squared errors (MSEs) for HB predictors considered. 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 mean squared errors as $eM_i = \frac{1}{S}\sum_{s=1}^{S}\left(\hat{\theta}_i^{(s)} - \theta_i^{(s)}\right)^2$. None of the HB predictors shows signs of systematic bias. However, in the simulation setup where $p_e = 0.6$, 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.

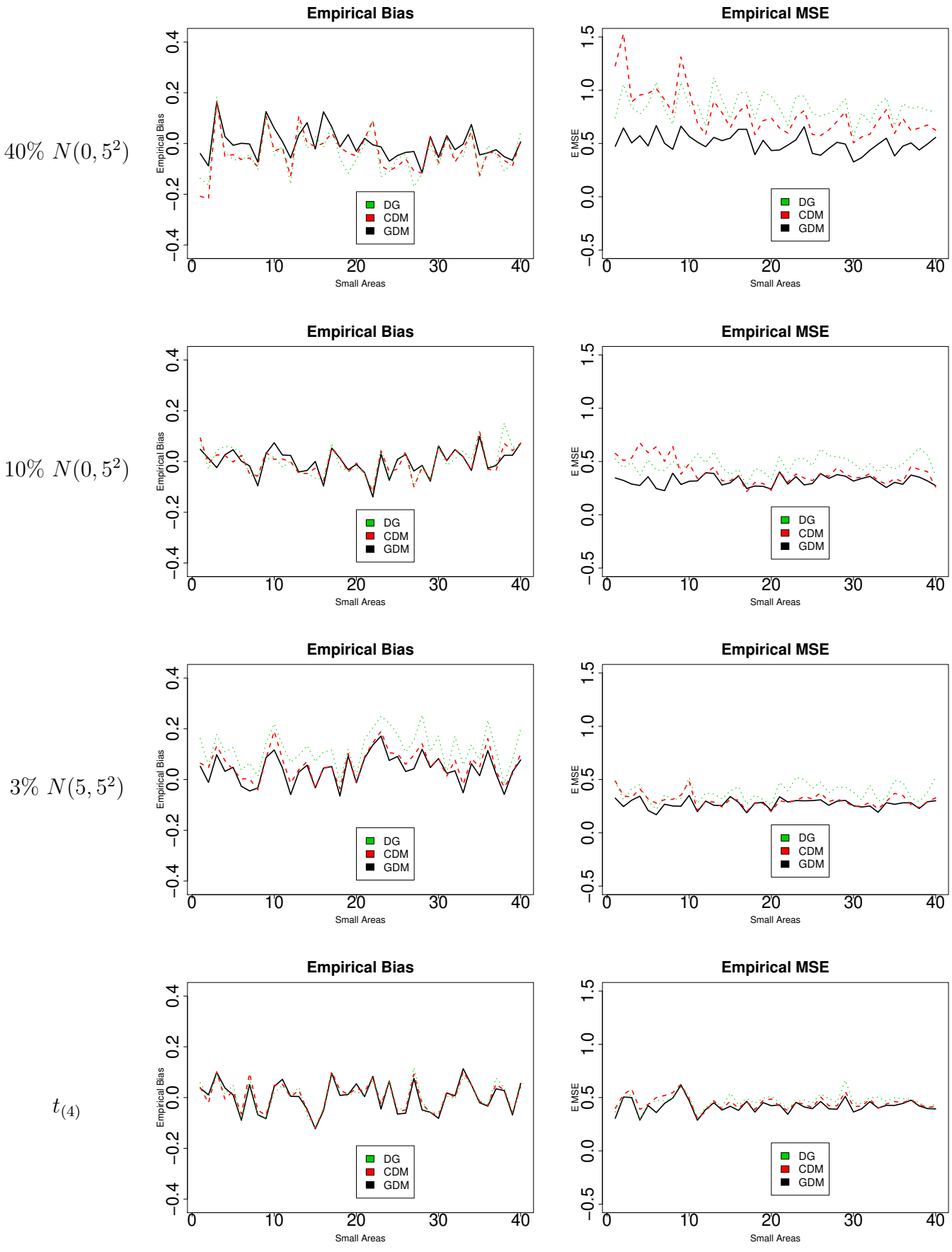

Figure 3: Plot of empirical biases and empirical MSEs of $\hat{\theta}$s

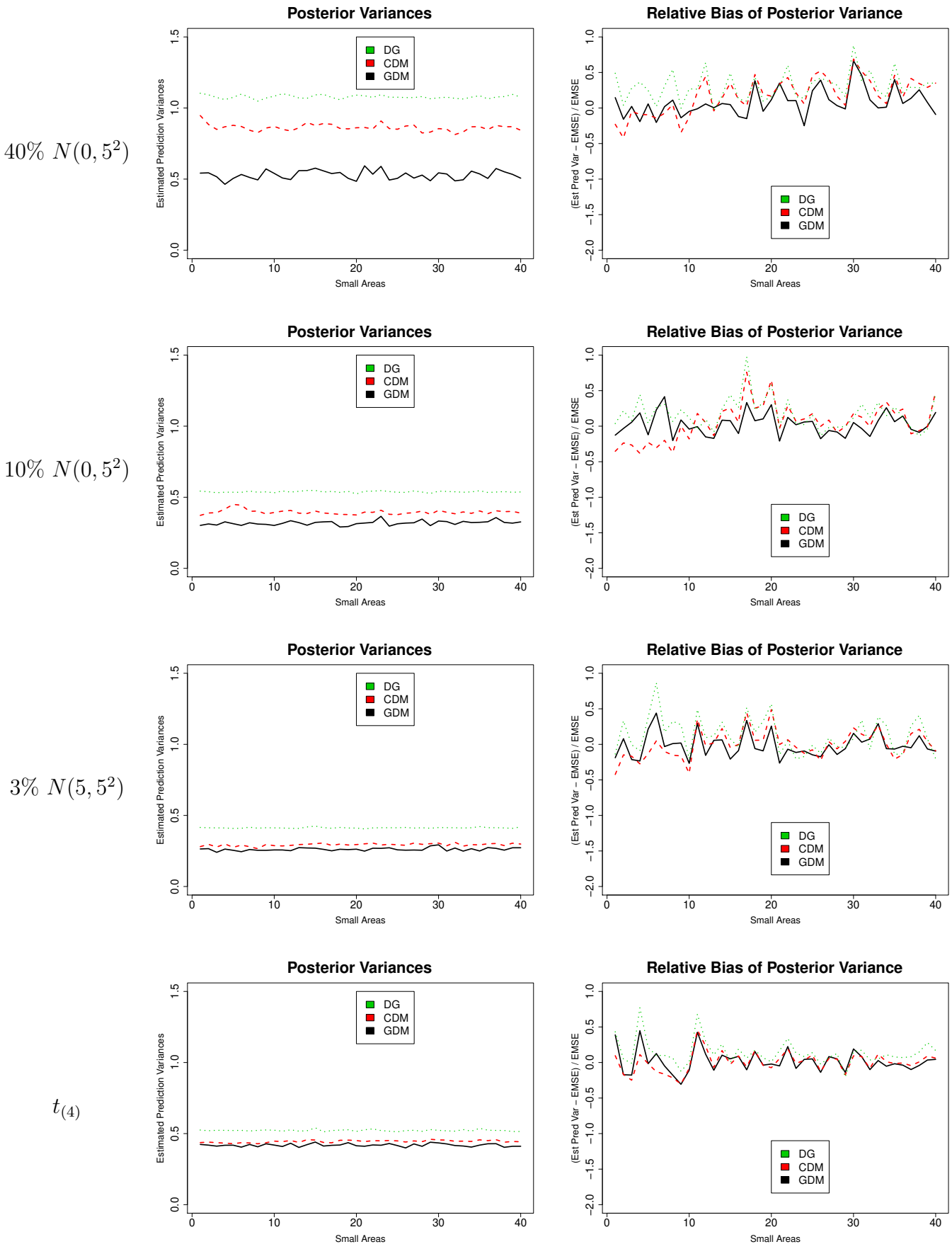

Figure 4: Plot of posterior variances and their empirical relative biases

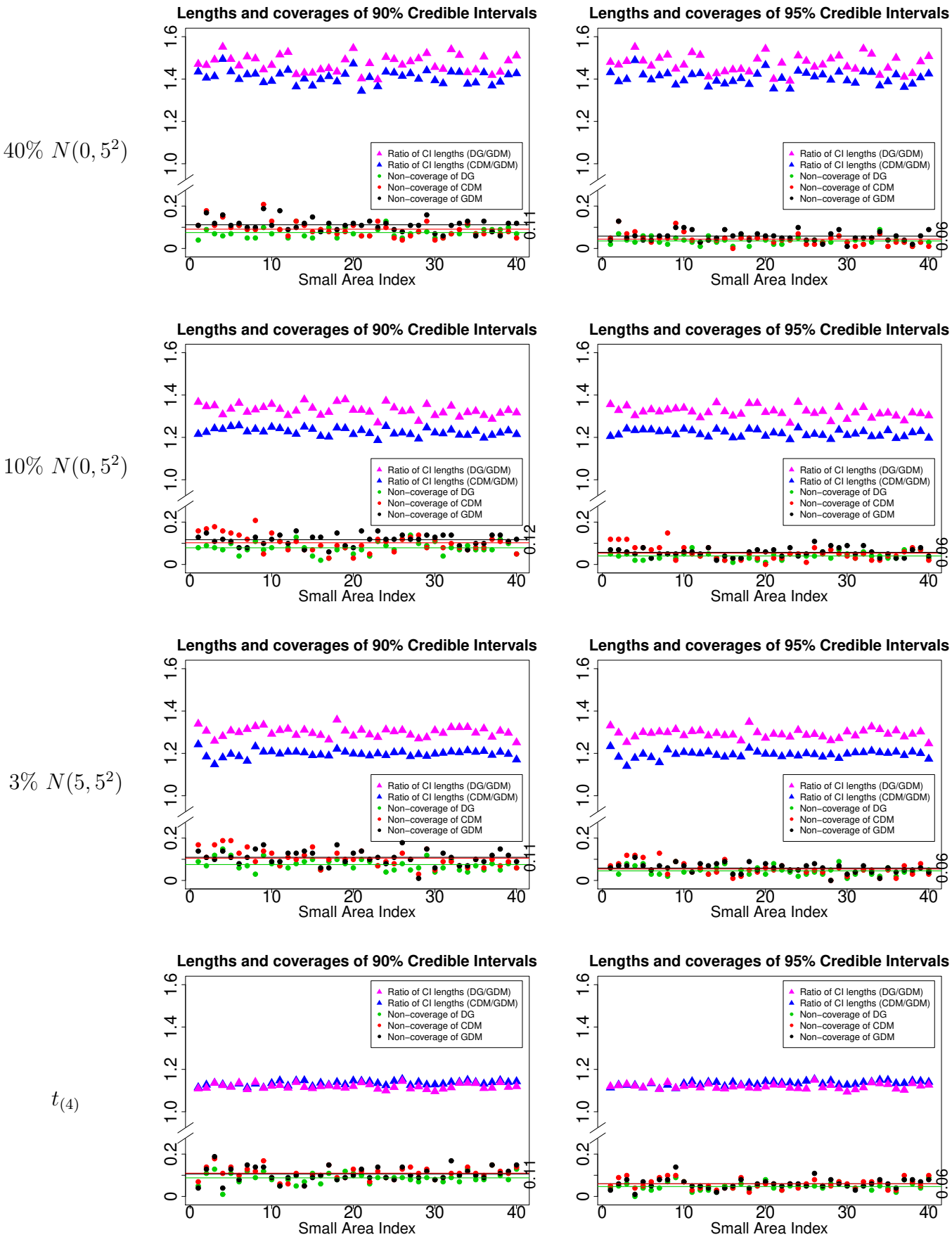

Figure 5: Plot of lengths and coverages of credible intervals

Next, Figure 4 shows posterior variances $V_i^{(s)}$ for each of 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 model 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 to each other.

Figure 5 shows the empirical non-coverage probabilities of 90% and 95% credible intervals of small area means θ_i. 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 plots also show two ratios which compare the lengths of the DG and CDM credible intervals to those of the GDM 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 $\bar{L}_{i,90} = \frac{1}{S}\sum_{i=1}^{S} L_{i,90}^{(s)}$. Again, this calculation is repeated for the 95% credible intervals. We see the credible intervals produced by the DG method consistently have the lowest non-coverage probabilities for each simulation, compared to the CDM and GDM 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 and CDM credible intervals have similar lengths. Though the DG credible intervals most often capture the true value $\theta_i^{(s)}$ and have low non-coverage probabilities, and thus conservative, they are longer than the GDM credible intervals, which closely attain the target coverage probability. While CDM and GDM 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 credible intervals are as successful as the CDM 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 and CDM credible intervals remain approximately equal, but the 90% and 95% intervals produced by the CDM model are up to 5% greater in length than their

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

6 Conclusion

Since its introduction by Battese et al. (1988), the NER model has been the basis for many important developments in small area estimation for unit-level data. Datta and Ghosh (1991) applied HB methods to the NER model to develop point estimators of small area means. This approach, however, is not robust in the presence of outliers. The HB method proposed by Chakraborty et al. (2018), 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 and posterior variances and narrower credible intervals than previous HB models discussed. The consistent superior performance of the proposed model as compared to the DG and CDM 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.

Acknowledgment

The authors are thankful to Dr. Ray Chambers for providing the dataset used in Section 4.2.

References

[1] 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, *Journal of the American Statistical Association*, **83**, 28–36

[2] Chakraborty, A., Datta, G. S. and Mandal, A. (2018) Robust Hierarchical Bayes Small Area Estimation for Nested Error Regression Model, *International Statistical Review*, under review

[3] Chambers, R. L. (1986) Outlier robust finite population estimation, *Journal of the American Statistical Association*, **81**, 1063–1069

[4] 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

[5] 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

[6] Datta, G. and Ghosh, M. (1991) Bayesian prediction in linear models: Applications to small area estimation, *Annals of Statistics*, **19**, 1748–1770

[7] Sinha, S. K. and Rao, J. N. K. (2009) Robust small area estimation, *The Canadian Journal of Statistics*, **37**, 381–399

A Appendix

A.1 Integrability of joint posterior probability density function

Chakraborty et al. (2018) show 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 show that the function

$$L(\boldsymbol{\beta}, \sigma_1^2, \sigma_2^2, p_e, \sigma_v^2) \frac{I_{(\sigma_1^2 < \sigma_2^2)}}{(\sigma_2^2)^2} \tag{3}$$

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 $y_{ij}, j = 1, \ldots, n_1, i = 1, \ldots, m$ 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. Note that

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

This implies,

$$L(\boldsymbol{\beta}, \sigma_1^2, \sigma_2^2, p_e, \sigma_v^2) \frac{I_{\frac{1}{2} < p_e < 1}}{(\sigma_1^2 + \sigma_2^2)^2} \quad < L(\boldsymbol{\beta}, \sigma_1^2, \sigma_2^2, p_e, \sigma_v^2) \left(\frac{I_{(\sigma_1^2 < \sigma_2^2)}}{(\sigma_2^2)^2} + \frac{I_{(\sigma_1^2 \geq \sigma_2^2)}}{(\sigma_1^2)^2} \right) \tag{4}$$

Since LHS of (4) is bounded above by two integrable functions, it is also integrable.

A.2 Simulation results with no contamination of unit-level errors

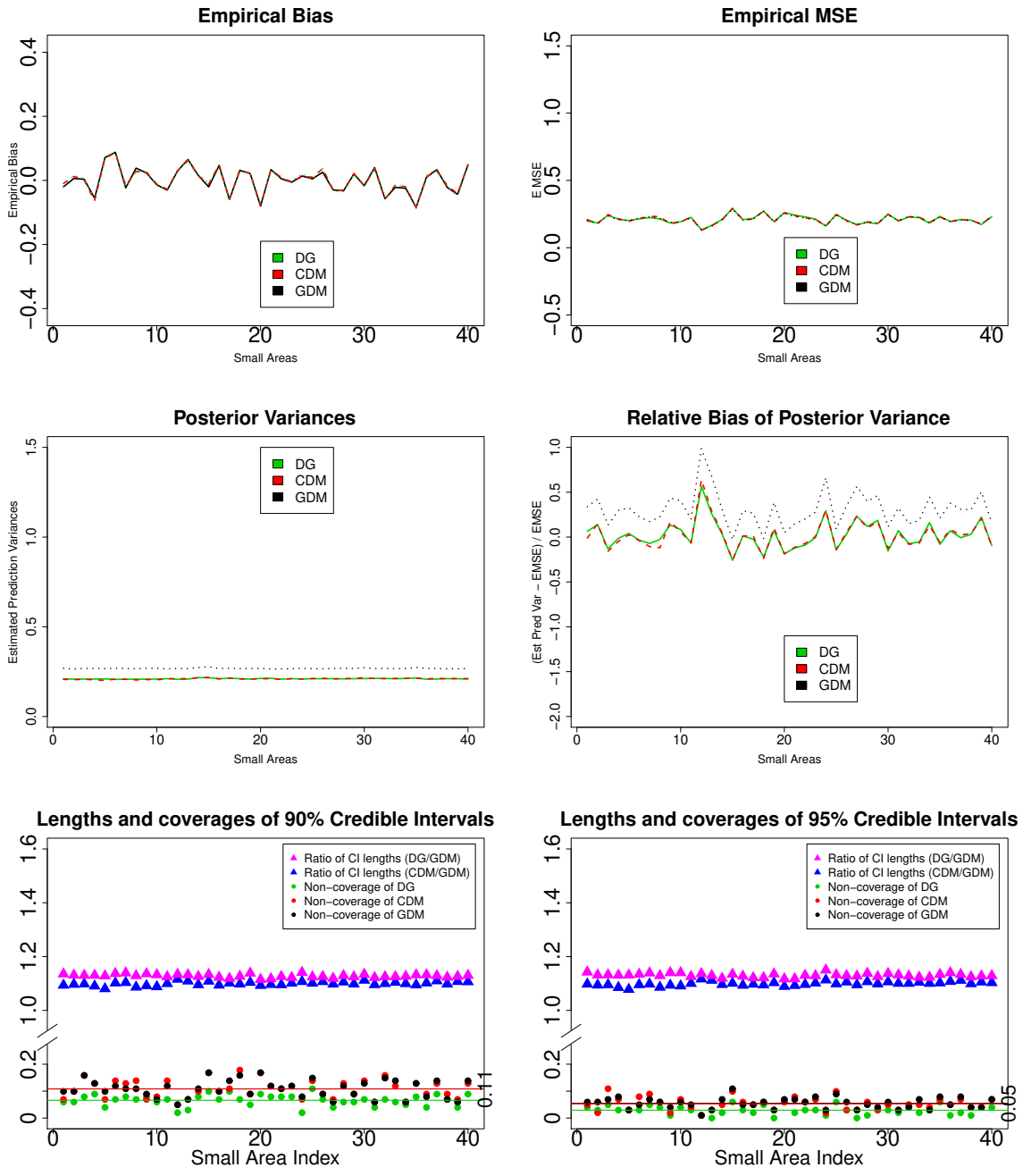

Figure 6: Plots of various measures of $\hat{\theta}$s when no unit-level contamination is present

Supplementary Materials

A priori probabilities

The a priori probability that a unit is from the secondary population is

$$P(z_{ij}=0)=\int P(z_{ij}=0|p_e)\pi(p_e)dp_e=\int_{\frac{1}{2}}^{1}(1-p_e)2dp_e=-(1-p_e)\Big|_{\frac{1}{2}}^{1}=\frac{1}{4}$$

and that a unit is from the primary population is

$$P(z_{ij}=1) \;\; = \;\; 1-P(z_{ij}=0)=\frac{3}{4}.$$

Joint posterior distribution

Implementation of the model through Gibbs sampling requires the conditional distributions to be derived from the full joint density. The joint density of $\boldsymbol{y}=\{y_{ij};j=1,...,n_i,i=1,...,m\}$, $\boldsymbol{z}=\{z_{ij};j=1,...,n_i,i=1,...,m\}$, and $\boldsymbol{v}=\{v_1,...v_m\}$ is written as:

$$f(\boldsymbol{y},\boldsymbol{v},\boldsymbol{\beta},\boldsymbol{z},\sigma_1^2,\sigma_2^2,\sigma_v^2,p_e) \propto \frac{\exp\left[-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(\frac{(y_{ij}-\boldsymbol{x_{ij}^T}\boldsymbol{\beta}-v_i)^2}{\sigma_1^2}z_{ij}+\frac{(y_{ij}-\boldsymbol{x_{ij}^T}\boldsymbol{\beta}-v_i)^2}{\sigma_2^2}(1-z_{ij})\right)\right]}{(\sigma_1^2)^{\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}z_{ij}}(\sigma_2^2)^{\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(1-z_{ij})}}$$

$$\times\frac{\exp\left(-\frac{1}{2}\sum_{i=1}^{m}\frac{v_i^2}{\sigma_v^2}\right)}{(\sigma_v^2)^{m/2}}\times p_e^{\sum_{i=1}^{m}\sum_{j=1}^{n_i}z_{ij}}(1-p_e)^{\sum_{i=1}^{m}\sum_{j=1}^{n_i}(1-z_{ij})}\times I_{\{\frac{1}{2}<p_e<1\}}\times\frac{1}{(\sigma_1^2+\sigma_2^2)^2}$$

To facilitate Gibbs sampling, we re-parametrize $\sigma_2^2=\eta\sigma_1^2$. The joint posterior distribution

of the reparametrized model is:

$$f(\boldsymbol{y},\boldsymbol{v},\boldsymbol{\beta},\boldsymbol{z},\sigma_1^2,\eta,\sigma_v^2,p_e) \propto \frac{\exp\left[-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(\frac{(y_{ij}-\boldsymbol{x_{ij}^T}\boldsymbol{\beta}-v_i)^2}{\sigma_1^2}z_{ij}+\frac{(y_{ij}-\boldsymbol{x_{ij}^T}\boldsymbol{\beta}-v_i)^2}{\eta\sigma_1^2}(1-z_{ij})\right)\right]}{(\sigma_1^2)^{\frac{n}{2}+1}(\eta)^{\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(1-z_{ij})}}$$

$$\times\frac{\exp\left(-\frac{1}{2}\sum_{i=1}^{m}\frac{v_i^2}{\sigma_v^2}\right)}{(\sigma_v^2)^{m/2}}\times p_e^{\sum_{i=1}^{m}\sum_{j=1}^{n_i}z_{ij}}(1-p_e)^{\sum_{i=1}^{m}\sum_{j=1}^{n_i}(1-z_{ij})}\times\frac{I_{\{\frac{1}{2}<p_e<1\}}}{(1+\eta)^2}$$

Conditional distributions for Gibbs sampling

We calculate the conditional distribution of each parameter using the joint pdf:

$$f(\boldsymbol{y},\boldsymbol{v},\boldsymbol{\beta},\boldsymbol{z},\sigma_1^2,\eta,\sigma_v^2,p_e) \propto \frac{\exp\left[-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(\frac{(y_{ij}-\boldsymbol{x_{ij}^T}\boldsymbol{\beta}-v_i)^2}{\sigma_1^2}z_{ij}+\frac{(y_{ij}-\boldsymbol{x_{ij}^T}\boldsymbol{\beta}-v_i)^2}{\eta\sigma_1^2}(1-z_{ij})\right)\right]}{(\sigma_1^2)^{\frac{n}{2}+1}(\eta)^{\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(1-z_{ij})}}$$

$$\times\frac{\exp\left(-\frac{1}{2}\sum_{i=1}^{m}\frac{v_i^2}{\sigma_v^2}\right)}{(\sigma_v^2)^{m/2}}\times p_e^{\sum_{i=1}^{m}\sum_{j=1}^{n_i}z_{ij}}(1-p_e)^{\sum_{i=1}^{m}\sum_{j=1}^{n_i}(1-z_{ij})}\times\frac{I_{\{\frac{1}{2}<p_e<1\}}}{(1+\eta)^2}$$

(I). $\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{v},\boldsymbol{z},\boldsymbol{\beta},\sigma_1^2,\eta,\sigma_v^2,p_e \sim N_p\left(S_\beta\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}-v_i)\left(\frac{z_{ij}}{\sigma_1^2}+\frac{1-z_{ij}}{\eta\sigma_1^2}\right)x_{ij},S_\beta\right)$

$$\text{where } S_\beta = \left[\sum_{i=1}^{m}\sum_{j=1}^{n_i}x_{ij}x_{ij}^T\left(\frac{z_{ij}}{\sigma_1^2}+\frac{1-z_{ij}}{\eta\sigma_1^2}\right)\right]^{-1}$$

(II). $v_i|\boldsymbol{y},\boldsymbol{\beta},\boldsymbol{z},\sigma_1^2,\eta,\sigma_v^2,p_e \sim N\left(\varphi_i\sum_{j=1}^{n_i}(y_{ij}-x_{ij}^T\boldsymbol{\beta})\left(\frac{z_{ij}}{\sigma_1^2}+\frac{1-z_{ij}}{\eta\sigma_1^2}\right),\varphi_i\right)$

$$\text{where } \varphi_i = \left(\frac{1}{\sigma_v^2}+\sum_{j=1}^{n_i}\left(\frac{z_{ij}}{\sigma_1^2}+\frac{1-z_{ij}}{\eta\sigma_1^2}\right)\right)^{-1}, i=1,...,m$$

(III). $z_{ij}|\boldsymbol{y},\boldsymbol{v},\boldsymbol{\beta},\sigma_1^2,\eta,\sigma_v^2,p_e \sim Bernoulli(p_{ij}^*), j=1,...,n, i=1,...,m$

$$\text{where } p_{ij}^* = \frac{p_e \times \exp\left(-\frac{(y_{ij}-\boldsymbol{x_{ij}^T}\boldsymbol{\beta}-v_i)^2}{2\sigma_1^2}\right)}{p_e \times \exp\left(-\frac{(y_{ij}-\boldsymbol{x_{ij}^T}\boldsymbol{\beta}-v_i)^2}{2\sigma_1^2}\right) + \frac{1-p_e}{\sqrt{\eta}}\exp\left(-\frac{(y_{ij}-\boldsymbol{x_{ij}^T}\boldsymbol{\beta}-v_i)^2}{2\eta\sigma_1^2}\right)}$$

(IV). $p_e|\boldsymbol{y},\boldsymbol{v},\boldsymbol{z},\boldsymbol{\beta},\sigma_1^2,\eta,\sigma_v^2 \sim Beta\left(\sum_{i=1}^{m}\sum_{j=1}^{n_i} z_{ij}+1, \sum_{i=1}^{m}\sum_{j=1}^{n_i}(1-z_{ij})+1\right) \times I_{\{\frac{1}{2}<p_e<1\}}$ In other words, we draw p_e from a truncated Beta distribution with the first shape parameter $\sum_{i=1}^{m}\sum_{j=1}^{n_i} z_{ij}+1$, the second shape parameter $\sum_{i=1}^{m}\sum_{j=1}^{n_i}(1-z_{ij})+1$, and lower truncation point $\frac{1}{2}$.

(V). $\frac{1}{\sigma_v^2}|\boldsymbol{y},\boldsymbol{v},\boldsymbol{z},\boldsymbol{\beta},\sigma_1^2,\eta,p_e \sim Gamma\left(\frac{m}{2}-1, \frac{1}{2}\sum_{i=1}^{m} v_i^2\right)$

(VI). $\frac{1}{\sigma_1^2}|\boldsymbol{y},\boldsymbol{v},\boldsymbol{z},\boldsymbol{\beta},\sigma_v^2,\eta,p_e \sim \pi(\frac{1}{\sigma_1^2}|\boldsymbol{y},\boldsymbol{v},\boldsymbol{z},\boldsymbol{\beta},\sigma_v^2,\eta,p_e)$

$$\propto Gamma\left(\frac{n}{2}, \frac{\sum_{i=1}^{m}\sum_{j=1}^{n_i}[(y_{ij}-\boldsymbol{x_{ij}^T}\boldsymbol{\beta}-v_i)^2(1+\eta z_{ij}-z_{ij})]}{2\eta}\right)$$

(VII). $\eta|\boldsymbol{y},\boldsymbol{v},\boldsymbol{z},\boldsymbol{\beta},\sigma_v^2,\sigma_1^2,p_e \sim \pi(\eta|\boldsymbol{y},\boldsymbol{v},\boldsymbol{z},\boldsymbol{\beta},\sigma_v^2,\sigma_1^2,p_e)$

$$= \underset{(\eta)}{\frac{\exp\left[-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(\frac{(y_{ij}-\boldsymbol{x_{ij}^T}\boldsymbol{\beta}-v_i)^2}{\eta\sigma_1^2}(1-z_{ij})\right)\right]}{\left(\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(1-z_{ij})+1\right)}} \times \frac{\eta}{(1+\eta)^2}$$