

Small Area Prediction of the Mean of a Binomial Random Variable

Andreea L. Erciulescu*

Wayne A. Fuller†

Abstract

Direct estimates for small areas or subpopulations may not be reliable because of small sample sizes for such objects. Procedures based on implicit or explicit models have been used to construct better estimates for given small areas, by exploiting auxiliary information. In this paper we consider binary responses, and investigate predictors for situations with different amounts of available information. We use generalized linear mixed models and present bias and mean squared error results for different prediction methods.

Key Words: Logistic, Mixed models, Auxiliary information, Binary responses

1. Introduction

Procedures based on models have been used to construct estimates for small areas, by exploiting auxiliary information. In this paper, we study nested models with a binary response and random area effects. These models form a subclass of generalized linear mixed models. We also consider stochastic covariates.

Survey data often contain auxiliary variables with good correlation with the variable of interest. However, area level auxiliary data may be incomplete. We consider three cases of auxiliary information, when the covariates have known mean, when the covariates have unknown distribution, and when the covariates have unknown random mean. For the last two cases, we describe estimation methods for the area mean of the auxiliary data. Because the response variable is binary and the auxiliary information is not fixed, estimation and prediction are not as straight forward as in linear mixed models.

Mixed models with unit level auxiliary data have been used for small area estimation by a number of authors. Battese, Harter, and Fuller (1988) use a linear mixed model to predict the area planted with corn and soybeans in Iowa counties. Datta and Ghosh (1991) introduce the hierarchical Bayes predictor for general mixed linear models. Larsen (2003) compared estimators for proportions based on two unit level models, a simple model with no area level covariates and a model using the area level information. Malec (2005) proposes Bayesian small area estimates for means of binary responses using a multivariate binomial/multinomial model. Jiang (2007) reviews the classical inferential approach for linear and generalized linear mixed models and discusses the prediction for a function of fixed and random effects. Ghosh et al (2009) consider a small area model where covariates have unknown distribution. They assume the sample has been selected so that weights ω_{ij} are available satisfying $\sum_{j=1}^{n_i} \omega_{ij} = 1$. They consider both hierarchical Bayes and EB estimators and suggest predictors for the small area proportions of the form $\sum_{j=1}^{n_i} \omega_{ij} \tilde{p}_{ij}(\mathbf{x}_{ij})$, where $\tilde{p}_{ij}(\mathbf{x}_{ij})$ is either the hierarchical Bayes or EB predictor. Ghosh and Sinha (2007) propose EB estimators for the small area means, where the covariates in the super-population are subject to measurement error. Datta, Rao, and Torabi (2010) study a nested error linear regression model with area level covariates subject to measurement error. They propose a pseudo-Bayes predictor and a corresponding pseudo-empirical Bayes predictor of a small area mean. Montanari, Ranalli, and Vicarelli (2010) consider unit level linear mixed models

*Iowa State University, 1215 Department of Statistics, Ames, IA 50010

†Iowa State University, 1214 Department of Statistics, Ames, IA 50010

and logistic mixed models, for binary response variable and fully known auxiliary information. Vizcaino, Cortina, Morales Gonzalez (2011) derive small area estimators for labor force indicators in Galicia using a multinomial logit mixed model.

2. Models

Consider a binomial response variable \mathbf{y} , with realizations y_{ij} for m different areas and n_i different units within each area. That is $y_{ij}|b_i$ are independent, following a binomial distribution, with mean p_{ij} , where b_i are the random area effects. Let \mathbf{x}_i be independent and identically distributed stochastic vectors of auxiliary information, following a distribution $F_{\mathbf{x}_i}$, and let b_i be independent and identically distributed, with a density f_b with mean 0 and variance σ_b^2 .

Then our unit level model is

$$y_{ij} = h(\eta_{ij}) + e_{ij}, \quad \eta_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + b_i, \quad h(\eta_{ij}) = \frac{\exp(\eta_{ij})}{1 + \exp(\eta_{ij})} \quad (1)$$

for $\mathbf{x}_{ij} = (1, x_{ij})$, $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, n_i$, where i is the index for area, and j is the index for unit within area. We assume that b_i and x_{ij} are mutually independent. Note that the mean of y_{ij} given (\mathbf{x}_{ij}, b_i) is $h(\eta_{ij}) := p_{ij}(\mathbf{x}_{ij}, b_i)$. Under the assumptions of model (1), the true small area mean of \mathbf{y} is

$$\theta_i = \int p_{ij}(\mathbf{x}_{ij}, b_i) dF_{\mathbf{x}_i}(\mathbf{x}), \quad (2)$$

where $F_{\mathbf{x}_i}(\mathbf{x})$ is the distribution of \mathbf{x} in area i . Our objective is to construct predictions for θ_i .

An example of (1) is the simple unit level mean model for \mathbf{y}

$$p_{\alpha, ij} = \frac{\exp(\alpha + b_i)}{1 + \exp(\alpha + b_i)}, \quad (3)$$

where α is a location parameter and b_i is the random area effect.

We will have use for an area level model for the vector of covariates $\mathbf{x}_{xi} = (1, x_{ij})$, and assume

$$\boldsymbol{\mu}_{xi} \sim NI(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{\delta\delta}), \quad \mathbf{x}_i | \boldsymbol{\mu}_{xi} \sim NI(\boldsymbol{\mu}_{xi}, \boldsymbol{\Sigma}_{\epsilon\epsilon}). \quad (4)$$

3. Estimation and Prediction

The models (1) and (3) are generalized linear mixed models (GLMMs) and estimates for $\boldsymbol{\beta}$, σ_b^2 , α and σ_{2b}^2 can be computed using R, by maximizing a Laplacian approximation to the likelihood. Note that the predicted random area effects and the estimated random effects variance for model (3) differ from the estimated values under model (1), hence we denote those for model (3) by $\hat{\mathbf{b}}_2$ and $\hat{\sigma}_{2b}^2$, respectively.

We consider two methods for constructing predictions for θ_i . In the first method, the minimum mean squared error (MMSE) prediction method, we use the conditional distribution $f(b_i|y_{ij})$ to compute the unit means of \mathbf{y} and then we integrate over the distribution of \mathbf{x} to compute the predictions for θ_i . In the second method, the 'plug-in' method, we directly substitute the predicted random area effects vector $\hat{\mathbf{b}}$ in p_{ij} . As with the first method, we integrate estimated p_{ij} over the estimated distribution of \mathbf{x} to compute the predictions for θ_i . We compare these two methods using a simulation study.

3.1 MMSE Prediction

If the parameters of the distributions are known, the MMSE predictor of b_i as

$$\hat{b}_i = \int \frac{\int b_i \prod_{t=1}^{n_i} f(y_{it}|b_i) f_b(b_i) db_i}{\int \prod_{t=1}^{n_i} f(y_{it}|b_i) f_b(b_i) db_i} dF_{\mathbf{x}_i}(\mathbf{x}). \quad (5)$$

Let μ_{x_i} be the area mean of \mathbf{x}_i . We present predictions for θ_i , for different cases of auxiliary information, when μ_{x_i} is known, when the distribution of \mathbf{x} is unknown, and when μ_{x_i} is unknown random. For the first case we assume \mathbf{x} is normally distributed with unknown variance. For the second case, we estimate the distribution of \mathbf{x}_i following Ghosh et al (2009). For the third case, we estimate the area mean of \mathbf{x}_i using an area level model for the vector of covariates $\mathbf{x}_{ij} = (1, x_{ij})$, given in (4).

3.1.1 Covariate Mean Known

Consider the case when the mean of \mathbf{x} is known for area i and the form of the distribution is specified. Then, the MMSE predictor of the small area mean of \mathbf{y} is

$$\hat{\theta}_i = \int \frac{\int p_{ij}(\mathbf{x}_{ij}, b_i) \prod_{t=1}^{n_i} f(y_{it}|b_i) f_b(b_i) db_i}{\int \prod_{t=1}^{n_i} f(y_{it}|b_i) f_b(b_i) db_i} dF_{\mathbf{x}_i}(\mathbf{x}). \quad (6)$$

In some finite population situations, the entire finite population of \mathbf{x} values may be known and the integral in (6) is the sum over the population. In practice it is often necessary to estimate the parameters of the distribution $F_{\mathbf{x}_i}$.

3.1.2 Unspecified distribution for \mathbf{x}

If μ_{x_i} is unknown and treated as fixed, we estimate the distribution of \mathbf{x} at point c using the sample cumulative distribution function (CDF), $\sum_{j=1}^{n_i} \omega_{ij} I(x_{ij}, c)$, where $I(x_{ij}, c)$ is the indicator function. For known parameters, the predicted small area mean of \mathbf{y} is

$$\bar{p}_i = \sum_{j=1}^{n_i} \omega_{ij} \frac{\int p_{ij}(\mathbf{x}_{ij}, b_i) \prod_{t=1}^{n_i} f(y_{it}|b_i) f_b(b_i) db_i}{\int \prod_{t=1}^{n_i} f(y_{it}|b_i) f_b(b_i) db_i}. \quad (7)$$

See Ghosh et al (2009) for an example of the approach.

3.1.3 No Auxiliary Information Used

Under model (3), for known parameters, the MMSE predictor of the small area mean of \mathbf{y} is

$$\hat{p}_i = \frac{\int p_{\alpha,ij} \prod_{t=1}^{n_i} f(y_{it}|b_i) f_b(b_i) db_i}{\int \prod_{t=1}^{n_i} f(y_{it}|b_i) f_b(b_i) db_i}, \quad (8)$$

where $p_{\alpha,ij}$ is defined in (3).

3.1.4 Unknown Random Covariate Mean

Consider the model (1) for \mathbf{y} and the linear mixed model for x_{ij} given in (4):

$$x_{ij} = \mu_x + \delta_i + \epsilon_{ij}, \quad \delta_i \sim N(0, \sigma_\delta^2), \quad \epsilon_{ij} | \delta_i \sim N(0, \sigma_\epsilon^2) \quad (9)$$

A small area predictor of the mean of \mathbf{x}_i is

$$\hat{\mu}_{x_i} = \hat{\mu}_x + \hat{\gamma}_{x_i}(\bar{x}_i - \hat{\mu}_x), \quad (10)$$

where

$$\hat{\mu}_x = \sum_{i=1}^m (\hat{\sigma}_\delta^2 + n_i^{-1} \hat{\sigma}_\epsilon^2)^{-1} \bar{x}_i, \quad \hat{\gamma}_{xi} = (\hat{\sigma}_\delta^2 + n_i^{-1} \hat{\sigma}_\epsilon^2)^{-1} \hat{\sigma}_\delta^2$$

and

$$\hat{\sigma}_\epsilon^2 = \left(\sum_{i=1}^m (n_i - 1) \right)^{-1} \sum_{i=1}^m \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2.$$

In (10), $\bar{x}_i = n_i^{-1} \sum_{j=1}^{n_i} x_{ij}$ denotes the sample area mean of \mathbf{x}_i , and the variance of the random area effects δ_i is estimated by $\hat{\sigma}_\delta^2$, the REML estimate constructed as described in Rao (2003, page 119).

Then a predictor of the small area mean of \mathbf{y} is

$$\tilde{\theta}_i = \int \frac{\int p_{ij}(\mathbf{x}_{ij}, b_i) \prod_{t=1}^{n_i} f(y_{it}|b_i) f_b(b_i) db_i}{\int \prod_{t=1}^{n_i} f(y_{it}|b_i) f_b(b_i) db_i} d\tilde{F}_{\mathbf{x}_i}(\mathbf{x}), \quad (11)$$

where $\tilde{F}_{\mathbf{x}}(\mathbf{x})$ is the estimator of $F_{\mathbf{x}}(\mathbf{x})$ with parameter μ_x predicted on the basis of model (4).

If $F_{\mathbf{x}}$ and f_b are continuous distributions, there are many ways to approximate the integrals in (2,5,6,7,8,11). Algorithms are available in *R* or one can create a finite discrete approximation. We consider the normal distribution and let $z_k, k = 1, 2, \dots, K$ be a set of numbers such that

$$\frac{1}{K} \sum_{k=1}^K (z_k, z_k^2) = (0, 1) \quad (12)$$

and the $\{z_k\}$ is an approximation for the normal distribution. For example, z_k might be $\xi(k - 0.5K^{-1}), k = 1, 2, \dots, K - 1$, with $z_K = \xi(k + 0.5K^{-1})$, where $\xi(a)$ is the a th percentile of the normal distribution. The z_k are standardized to have mean zero and variance one. Let $\mathbf{x}_{ik}^* = (1, x_{ik}^*)$ and

$$x_{ik}^* = \mu_{xi} + z_k \sigma_\epsilon \text{ and } b_k^* = \sigma_b * z_k. \quad (13)$$

Then, the approximated random area predictions \hat{b}_i are

$$\hat{b}_i = \frac{\sum_{k=1}^K b_k^* \prod_{t=1}^{n_i} f(y_{it}|b_k^*)}{\sum_{k=1}^K \prod_{t=1}^{n_i} f(y_{it}|b_k^*)}.$$

Approximations for the integral expressions in (2,6,7,8,11) are:

(i) true small area mean of \mathbf{y}

$$\theta_i = K^{-1} \sum_{j=1}^K p_{ij}(\mathbf{x}_{ij}^*, b_i); \quad (14)$$

(ii) predicted small area mean of \mathbf{y} with μ_{xi} known

$$\hat{\theta}_i = \frac{1}{K} \sum_{j=1}^K \frac{\sum_{k=1}^K p_{ik}(\mathbf{x}_{ij}^*, b_k^*) \prod_{t=1}^{n_i} f(y_{it}|b_k^*)}{\sum_{k=1}^K \prod_{t=1}^{n_i} f(y_{it}|b_k^*)}, \quad (15)$$

where

$$x_{ij}^* = \mu_{xi} + z_j \hat{\sigma}_\epsilon, b_k^* = \hat{\sigma}_b * z_k, f(y_{it}|b_k^*) = I[y_{it} = 1] p_{it}(\mathbf{x}_{it}, b_k^*) + I[y_{it} = 0] (1 - p_{it}(\mathbf{x}_{it}, b_k^*)),$$

and σ_ϵ^2 is estimated using the pooled within-area mean squared

$$\hat{\sigma}_\epsilon^2 = \left(\sum_{i=1}^m n_i \right)^{-1} \sum_{i=1}^m \sum_{j=1}^{n_i} (x_{ij} - \mu_{xi})^2;$$

(iii) predicted small area mean of y using area sample CDF for \mathbf{x}

$$\bar{p}_i = n_i^{-1} \sum_{j=1}^{n_i} \bar{p}_{ij} = n_i^{-1} \sum_{j=1}^{n_i} \frac{\sum_{k=1}^K p_{ik}(\mathbf{x}_{ij}, b_k^*) \prod_{t=1}^{n_i} f(y_{it}|b_k^*)}{\sum_{k=1}^K \prod_{t=1}^{n_i} f(y_{it}|b_k^*)}, \quad (16)$$

where

$$b_k^* = \hat{\sigma}_b * z_k, f(y_{ij}|b_k^*) = I[y_{ij} = 1]p_{ik}(\mathbf{x}_{ij}, b_k^*) + I[y_{ij} = 0](1 - p_{ik}(\mathbf{x}_{ij}, b_k^*));$$

(iv) predicted small area mean of y using simple mean model for \mathbf{y}

$$\hat{p}_i = \frac{\sum_{k=1}^K p_{\alpha,ik}(b_{2k}^*) \prod_{t=1}^{n_i} f(y_{it}|b_{2k}^*)}{\sum_{k=1}^K \prod_{t=1}^{n_i} f(y_{it}|b_{2k}^*)}, \quad (17)$$

where

$$b_{2k}^* = \hat{\sigma}_{2b} * z_k, f(y_{it}|b_{2k}^*) = I[y_{ij} = 1]p_{ik}(b_{2k}^*) + I[y_{ij} = 0](1 - p_{ik}(b_{2k}^*));$$

(v) predicted small area mean of y using predicted small area mean of \mathbf{x}

$$\tilde{\theta}_i = \frac{1}{K} \sum_{j=1}^K \frac{\sum_{k=1}^K p_{ik}(\mathbf{x}_{ij}^*, b_k^*) \prod_{t=1}^{n_i} f(y_{it}|b_k^*)}{\sum_{k=1}^K \prod_{t=1}^{n_i} f(y_{it}|b_k^*)}, \quad (18)$$

where

$$x_{ij}^* = \hat{\mu}_{xi} + z_j \hat{\sigma}_\epsilon^*, b_k^* = \hat{\sigma}_b * z_k, f(y_{it}|b_k^*) = I[y_{it} = 1]p_{it}(\mathbf{x}_{it}, b_k^*) + I[y_{it} = 0](1 - p_{it}(\mathbf{x}_{it}, b_k^*)),$$

and

$$\hat{\sigma}_\epsilon^{*2} = \left(\sum_{i=1}^m (n_i - 1) \right)^{-1} \sum_{i=1}^m \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2.$$

In application, the parameters must be estimated. That is, $p_{ij}(\mathbf{x}_{ij}, b_i)$ is replaced with

$$\tilde{p}_{ij}(\mathbf{x}_{ij}, b_i) = \frac{\exp(\mathbf{x}'_{ij} \hat{\boldsymbol{\beta}} + b_i)}{1 + \exp(\mathbf{x}'_{ij} \hat{\boldsymbol{\beta}} + b_i)},$$

$\hat{\sigma}_b^2$ is estimated, and $p_{\alpha,ij}(b_i)$ is replaced with

$$\tilde{p}_{\alpha,ij}(b_i) = \frac{\exp(\hat{\alpha} + b_i)}{1 + \exp(\hat{\alpha} + b_i)}.$$

3.2 Simulation Results, MMSE Method

We performed a simulation study for $m = 36$ areas in three groups of 12 areas, with sizes $n_i \in \{2, 10, 40\}$ and unit level observations \mathbf{x}_{ij} . Each sample, (\mathbf{y}, \mathbf{x}) , is generated using model (1) with $\sigma_b^2 = 0.25$, $\mu_x = 0$, $\sigma_\delta^2 = 0.16$, and $\sigma_\epsilon^2 = 0.36$. Thus there is a random set of b_i for each MC sample. The vector of coefficients for the fixed effects is $(\beta_0, \beta_1) = (-0.8, 1)$ and, for each unit, the probability that $y_{ij} = 1$ is

$$p_{ij} = \frac{\exp(-0.8 + x_{ij} + b_i)}{1 + \exp(-0.8 + x_{ij} + b_i)}. \quad (19)$$

One thousand MC samples were generated satisfying the model.

Let the estimation models be

- Model 1: Model (1)-(4), with known auxiliary mean μ_{xi}
- Model 2: Model (1), with unknown distribution for \mathbf{x}_{ij}
- Model 3: Model (3), simple mean model for \mathbf{y}
- Model 4: Model (1)-(4), with unknown random auxiliary mean μ_{xi} .

We fit the estimation models (1) and (3) as generalized linear mixed models (GLMMs), with the binomial conditional distributions for the response. The model (4) for the covariate x_{ij} is fit as a linear mixed model (LMM).

The true small area mean of \mathbf{y} is given by (14) and the predicted area means of \mathbf{y} in the simulations are given in (15-18), with (β_0, β_1) and σ_b^2 estimated using GLMM in R. The integrals were approximated with $K = 50$. The values x_{ik}^* in (15) are constructed using the known μ_{xi} and the estimated σ_ϵ^2 defined for (15). Similarly, the values x_{ik}^* in (18) are constructed using the predicted μ_{xi} and the estimated σ_ϵ^{*2} defined for (18).

We denote the sample mean of \mathbf{y} by $\bar{\mathbf{y}}$. We computed the bias and the mean squared error (MSE) for the predictors averaged over the 1000 samples, averaged over areas with the same sample size, for the three different sample sizes.

Table 1 contains the estimated bias in predicting the small area mean y_{ij} as a percent of the standard error of prediction, under the MMSE method. The results are organized in three rows, corresponding to the three different sample sizes considered in this study. The simulation standard errors are presented in parentheses below the bias values. The estimator of the bias in the predictor is the simulation mean of the difference between the model predictor and the true parameter θ .

The mean squared errors for the predictions of the mean of y_{ij} and predictions for the random area effects b_i are presented in Table 2. The MSEs are multiplied by one thousand and are organized in three rows, corresponding to the three different sample sizes considered in this study. The simulation MSE standard errors are presented in parentheses below the MSE values. The estimator of the MSE is the simulation mean of the squared difference between the model predictor and the true parameter.

Because the estimated biases are small, relative to the standard error of prediction, the variance of the prediction error is approximately equal to the MSE. The smallest MSE corresponds to the prediction error in predicting the mean of y_{ij} under Model 1, when the auxiliary mean is known. Using Model 1 we estimate the sample variance of the auxiliary variable, and use the known value for the covariate mean to construct the predicted area mean of y_{ij} . On the other hand, for the case when the auxiliary mean is unknown and we make predictions based on the simple mean model of \mathbf{y} , we use no covariate information in predicting b_i in (3).

Table 1: MC BIAS of prediction error as percent of the standard error of prediction, MC BIAS of $\bar{y}_{ij} - \theta_i$ as percent of the standard error, and MC BIAS of $\hat{b}_i - b_i$ and MC BIAS of $\hat{b}_{2i} - b_{2i}$ as percents of the standard errors of predictions

n	$\hat{\theta} - \theta^1$	$\bar{p} - \theta^2$	$\hat{p} - \theta^3$	$\tilde{\theta} - \theta^4$	$\bar{y} - \theta$	$\hat{b} - b$	$\hat{b}_2 - b_2$
2	1.44 (1.16)	1.46 (1.06)	0.86 (1.15)	-0.24 (1.14)	-0.17 (0.93)	1.56 (0.87)	2.18 (0.88)
10	-1.62 (1.11)	-1.82 (1.07)	-1.64 (1.05)	-2.48 (1.08)	-1.60 (0.89)	0.69 (0.96)	0.71 (0.95)
40	0.37 (0.96)	0.16 (0.95)	0.12 (0.94)	-0.02 (0.96)	0.50 (0.90)	1.77 (1.23)	2.28 (1.01)

1. Model 1, known μ_{xi}
2. Model 2, unknown distribution for \mathbf{x}_{ij}
3. Model 3, simple mean model for \mathbf{y}
4. Model 4, unknown random μ_{xi}

For the case when the auxiliary mean is unknown, the smallest MSE comes from using Model 4. Making predictions based on Model 4 involves making predictions for the unknown random covariate mean, using the estimated grand mean of \mathbf{x} and estimated variance of \mathbf{x} . Using Model 2 gives smaller MSE than that of the simple mean model for large sample sizes, but the simple mean model predictor is superior to that based on Model 4 for small sample sizes.

Table 2: MC MSE (x1000) of prediction errors for the mean of y_{ij} , MC MSE (x1000) of $\bar{y}_{ij} - \theta_i$, MC MSE (x1000) of $\hat{b}_i - b_i$ and MC MSE (x1000) of $\hat{b}_{2i} - b_{2i}$

n	$\hat{\theta} - \theta^1$	$\bar{p} - \theta^2$	$\hat{p} - \theta^3$	$\tilde{\theta} - \theta^4$	$\bar{y} - \theta$	$\hat{b} - b$	$\hat{b}_2 - b_2$
2	9.31 (0.12)	16.17 (0.22)	14.21 (0.18)	12.46 (0.16)	101.91 (1.09)	228.88 (3.04)	236.13 (3.15)
10	7.24 (0.10)	8.63 (0.12)	9.83 (0.13)	8.37 (0.12)	20.66 (0.27)	184.79 (2.50)	210.32 (3.01)
40	3.54 (0.05)	3.93 (0.06)	4.15 (0.06)	3.90 (0.05)	5.17 (0.07)	105.09 (1.53)	176.08 (2.48)

1. Model 1, known μ_{xi}
2. Model 2, unknown distribution for \mathbf{x}_{ij}
3. Model 3, simple mean model for \mathbf{y}
4. Model 4, unknown random μ_{xi}

3.3 Plug-in Method for b_i

Because computer programs are available that give predictions of b_i , one may be tempted to ‘plug-in’ the predicted value of b_i into equation (14) to construct the predictor of θ_i . Let the estimated coefficients for the fixed effects be $\hat{\beta}$, $\hat{\alpha}$, and let the predicted values for the random area effects be \hat{b} , \hat{b}_2 , for models (1) and (3), respectively. We construct the plug-in small area mean prediction for the four methods by:

$$\hat{\theta}_{i,plugin} = K^{-1} \sum_{j=1}^K \tilde{p}_{ij}(\mathbf{x}_{ij}^*, \hat{b}_i), \text{ where } x_{ij}^* = \mu_{xi} + z_j \hat{\sigma}_\epsilon;$$

$$\bar{p}_{i,plugin} = n_i^{-1} \sum_{j=1}^{n_i} \tilde{p}_{ij}(\mathbf{x}_{ij}, \hat{b}_i);$$

$$\hat{p}_{i,plugin} = \frac{\exp(\hat{\alpha} + \hat{b}_{i2})}{1 + \exp(\hat{\alpha} + \hat{b}_{i2})};$$

and

$$\tilde{\theta}_{i,plugin} = K^{-1} \sum_{j=1}^K \tilde{p}_{ij}(x_{ij}^*, \hat{b}_i), \text{ where } x_{ij}^* = \hat{\mu}_{xi} + z_j \hat{\sigma}_\epsilon^*. \tag{20}$$

3.4 Simulation Results, Plug-in Method for b_i

We use the simulation setup of Section 3.2 and construct predictions of θ_i as defined in Section 3.3. Table 3 contains the estimated biases of the prediction error as percent of the standard error of prediction for the corresponding model. Some of the biases in the first four columns of Table 3 are significantly different from zero and arise because $p_{ij}(x_{ij}, b_i)$ of (19) is a nonlinear function of (x_{ij}, b_i) . The absolute values of the relative bias for the prediction errors for the mean of y_{ij} decrease with the increase in sample size, corresponding to a decrease in the variance of b_i . The smallest absolute values for the relative prediction bias are for estimation Model 1 and estimation Model 2. The absolute biases for Model 1 and Model 2 are comparable because the variance for Model 1 is smaller than the variance for Model 2. Model 1, Model 2 and Model 3 have the same variance of $\hat{\mathbf{b}} - \mathbf{b}$. The \hat{b}_2 associated with Model 3 estimation has a larger variance.

Table 3: MC BIAS of prediction error as percent of the standard error of prediction, ‘plug-in method’

n	$\hat{\theta}_{plugin} - \theta^1$	$\bar{P}_{plugin} - \theta^2$	$\hat{P}_{plugin} - \theta^3$	$\tilde{\theta}_{plugin} - \theta^4$
2	-3.49 (1.18)	-2.28 (1.06)	-5.88 (1.16)	-4.68 (1.15)
10	-4.69 (1.12)	-4.65 (1.08)	-5.30 (1.06)	-5.39 (1.09)
40	-1.02 (0.97)	-1.18 (0.96)	-1.24 (0.95)	-1.35 (0.96)

1. Model 1, known μ_{xi}
2. Model 2, unknown distribution for \mathbf{x}_{ij}
3. Model 3, simple mean model for \mathbf{y}
4. Model 4, unknown random μ_{xi}

The MC MSE of prediction errors for the mean of y_{ij} constructed using the ‘plug-in’ method are slightly larger than, but very close to, the values presented in Table 2. The procedure using estimated conditional mean is less biased and slightly more efficient than the ‘plug-in method.’

Table 4: MC MSE ($\times 1000$) of prediction errors for the mean of y_{ij} , ‘plug-in method’

n	$\hat{\theta}_{plugin} - \theta^1$	$\bar{P}_{plugin} - \theta^2$	$\hat{P}_{plugin} - \theta^3$	$\tilde{\theta}_{plugin} - \theta^4$
2	9.38 (0.13)	16.60 (0.22)	14.36 (0.19)	12.56 (0.17)
10	7.29 (0.10)	8.72 (0.12)	9.89 (0.14)	8.43 (0.12)
40	3.54 (0.05)	3.94 (0.06)	4.15 (0.06)	3.91 (0.05)

1. Model 1, known μ_{xi}
2. Model 2, unknown distribution for \mathbf{x}_{ij}
3. Model 3, simple mean model for \mathbf{y}
4. Model 4, unknown random μ_{xi}

4. Conclusions

This work was motivated by real survey situations, in particular those where there is incomplete auxiliary information. In this paper we presented a unit level model for binomial response variables, a specific case of a generalized linear mixed model, and constructed predictors for the area means for different cases of auxiliary information. We showed that using the ‘plug-in’ method can lead to the sizeable bias in predictions.

We presented results for a simulation study, generating data from the unit level model. The bias in the prediction errors was small, relative to the standard errors of the predictions for the mean of y_{ij} . The results indicate that, generally, it is better to include auxiliary information in the model and estimate the distribution, rather than to ignore the auxiliary information.

REFERENCES

- Battese, G.E., Harter, R.M., 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.
- Datta, G.S. and Ghosh, M. (1991), “Bayesian Prediction in Linear Models: Applications to Small Area Estimation,” *Annals of Statistics*, 19, 1748-1770.
- Datta, G.S., Rao, J.N.K., Torabi, M. (2010), “Pseudo-empirical Bayes estimation of small area means under a nested error linear regression model with functional measurements errors,” *Journal of Statistical Planning and Inference*, 140, 2952-2962.
- Ghosh, M., Kim, D., Sinha, K., Maiti, T., Katzoff, M., and Parsons, V.L. (2009), “Hierarchical and Empirical Bayes small domain estimation and proportion of persons without health insurance for minority subpopulations,” *Survey Methodology*, 35, 53–66.
- Jiang, J. (2007), *Linear and Generalized Linear Mixed Models and their Applications*, Springer, New York.
- Larsen, M. D. (2003), “Estimation of small-area proportions using covariates and survey data,” *Journal of Statistical Planning and Inference*, 112, 89-98.
- Lopez Vizcaino, E., Lombardia Cortina, M.J., and Morales Gonzalez, D. (2011), “Multinomial-based small area estimation of labour force indicators in Galicia,” *X Congreso Galego de Estatística e Investigación de Operacions Pontevedra*.
- Malec, D. (2005), ‘Small Area Estimation from the American Community Survey Using a Hierarchical Logistic Model of Persons and Housing Units,’ *Journal of Official Statistics*, 21, 3, 411-432.
- Montanari, G.E., Ranalli, G.M., and Vicarelli, C. (2010), “A comparison of small area estimators of counts aligned with direct higher level estimates,” *Scientific Meeting of the Italian Statistical Society* <http://homes.stat.unipd.it/mgri/SIS2010/Program/contributedpaper/678-1393-1-DR.pdf>
- Rao, J.N.K. (2003), *Small Area Estimation*, Wiley.