

2015

# Small area prediction based on unit level models when the covariate mean is measured with error

Andreea Luisa Erciulescu  
*Iowa State University*

Follow this and additional works at: <https://lib.dr.iastate.edu/etd>

 Part of the [Statistics and Probability Commons](#)

---

## Recommended Citation

Erciulescu, Andreea Luisa, "Small area prediction based on unit level models when the covariate mean is measured with error" (2015).  
*Graduate Theses and Dissertations*. 14674.  
<https://lib.dr.iastate.edu/etd/14674>

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact [digirep@iastate.edu](mailto:digirep@iastate.edu).

**Small area prediction based on unit level models when the covariate mean is  
measured with error**

by

**Andreea Luisa Erciulescu**

A dissertation submitted to the graduate faculty  
in partial fulfillment of the requirements for the degree of  
DOCTOR OF PHILOSOPHY

Major: Statistics

Program of Study Committee:

Wayne A. Fuller, Major Professor

Emily J. Berg

Song X. Chen

Daniel Nettleton

Jarad Niemi

Iowa State University

Ames, Iowa

2015

Copyright © Andreea Luisa Erciulescu, 2015. All rights reserved.

## DEDICATION

This dissertation is dedicated to the memory of my father, Marin.

## TABLE OF CONTENTS

<b>LIST OF TABLES</b> . . . . .	iv
<b>LIST OF FIGURES</b> . . . . .	v
<b>ACKNOWLEDGEMENTS</b> . . . . .	vi
<b>ABSTRACT</b> . . . . .	vii
<b>CHAPTER 1. INTRODUCTION</b> . . . . .	1
1.1 Bootstrap Properties . . . . .	8
1.2 Parametric Double Bootstrap Results . . . . .	9
<b>CHAPTER 2. SMALL AREA PREDICTION OF THE MEAN OF A BI- NOMIAL RANDOM VARIABLE</b> . . . . .	17
2.1 Introduction . . . . .	17
2.2 Models . . . . .	19
2.3 Estimation and Prediction . . . . .	19
2.3.1 MMSE Prediction . . . . .	20
2.3.2 Simulation Results, MMSE Method . . . . .	24
2.3.3 Plug-in Method for $b_i$ . . . . .	27
2.3.4 Simulation Results, Plug-in Method for $b_i$ . . . . .	28
2.4 Conclusions . . . . .	29
<b>CHAPTER 3. SMALL AREA PREDICTION UNDER ALTERNATIVE MODEL SPECIFICATIONS</b> . . . . .	32
3.1 Introduction . . . . .	32
3.2 Unit Level Nonlinear Models . . . . .	33

3.2.1	Introduction . . . . .	33
3.2.2	Predictors of $\theta_i$ . . . . .	34
3.2.3	Estimation . . . . .	37
3.2.4	Bootstrap MSE Estimation . . . . .	38
3.3	Simulations . . . . .	40
3.3.1	Refinement of Prediction MSE Estimators . . . . .	41
3.3.2	MSE for Different Types of Auxiliary Information . . . . .	43
3.3.3	Monte Carlo Properties of Prediction MSE Estimators . . . . .	44
3.3.4	Efficiency of Bootstrap Procedures . . . . .	45
3.4	Summary . . . . .	52
3.5	Appendix A. Optimal Number of Bootstrap Samples at Level Two . . . . .	54
<b>CHAPTER 4. BOOTSTRAP CONFIDENCE INTERVALS FOR SMALL</b>		
<b>AREA MEANS . . . . .</b>		<b>58</b>
4.1	Introduction . . . . .	58
4.2	Models . . . . .	61
4.2.1	Small Area Mean Prediction for the Unit Level Linear Mixed Model . . . . .	62
4.2.2	Estimation . . . . .	69
4.3	Bootstrap MSE Estimators . . . . .	71
4.4	Linear Approximation of the ULGLMM . . . . .	73
4.5	Symmetric two-sided $(1 - \alpha)$ -level CI . . . . .	78
4.5.1	Wald-type Symmetric $(1 - \alpha)$ -level CI for the Small Area Mean . . . . .	78
4.5.2	Level one Bootstrap Symmetric $(1 - \alpha)$ -level CI for the Small Area Mean . . . . .	79
4.5.3	Double Bootstrap Symmetric $(1 - \alpha)$ -level CI for the Small Area Mean . . . . .	81
4.5.4	General Purpose Statistic for CI . . . . .	84
4.6	Simulations . . . . .	86
4.6.1	Refinement of Estimators . . . . .	88
4.6.2	Results . . . . .	89
4.6.3	Alternative Methods and Simulation Results . . . . .	106
4.7	Summary . . . . .	108

4.8	Appendix A. Estimated Variance of the Variance of Random Area Effects . . .	110
4.8.1	Model for $\tilde{x}$ . . . . .	110
4.8.2	Model for $y$ . . . . .	110
4.9	Appendix B. Alternative Double Bootstrap Methods . . . . .	112
4.10	Appendix C. Two-sided $(1 - \alpha)$ -level CIs for other Model Parameters . . . . .	114
4.10.1	Random Effects Variance Component . . . . .	114
4.10.2	Fixed Effects Coefficients . . . . .	114
4.10.3	Simulation Results: Confidence Intervals for the Model Parameters . . .	115
4.11	Appendix D. Numerical Integration . . . . .	116
4.12	Appendix E. Additional Simulation Results . . . . .	118
4.12.1	Results for Model 1, Different Variance of Random Effects . . . . .	128
4.13	Appendix F. Plots for the Simulation Results for Model 2 . . . . .	129
	<b>CHAPTER 5. SUMMARY . . . . .</b>	<b>134</b>

## LIST OF TABLES

Table 2.1	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 . . . . .	26
Table 2.2	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 $\hat{b}_{2i} - b_{2i}$ . . . . .	27
Table 2.3	MC BIAS of Prediction Error as Percent of the Standard Error of Prediction, ‘plug-in method’ . . . . .	28
Table 2.4	MC MSE (x1000) of Prediction Errors for the Mean of $y_{ij}$ , ‘plug-in method’ . . . . .	29
Table 3.1	MSE for Different Types Auxiliary Information (Entries Multiplied by $10^3$ ) . . . . .	44
Table 3.2	Monte Carlo Properties of Prediction MSE Estimators . . . . .	45
Table 3.3	Estimated Variance Components for Variance of Estimated Prediction MSE  (Within is for $(B_1 = 100, B_2 = 1)$ . All Variances Multiplied by $10^6$ ) . . . . .	48
Table 3.4	Estimated Within Bootstrap Variance and Covariance of the Bootstrap Prediction MSE ( $B_1 = 1, B_2 = 1$ , Entries Multiplied by $10^6$ ) . . . . .	50
Table 3.5	Estimated Within Bootstrap Variance of the Bootstrap Prediction MSE. Entries Multiplied by $10^6$ . . . . .	51
Table 3.6	Equal Efficiency Bootstrap Procedures . . . . .	52

Table 4.1	Monte Carlo Properties of the Estimated Parameters, $\mu_{xi}$ known . . .	90
Table 4.2	Monte Carlo Properties of the Estimated Parameters, $\mu_{xi}$ unknown, $\tilde{\mu}_{xi}$ observed, $N = 400$ . . . . .	91
Table 4.3	Monte Carlo Properties of Intermediate Statistics, $\mu_{xi}$ known (Entries Multiplied by $10^3$ ) . . . . .	93
Table 4.4	Monte Carlo Properties of Intermediate Statistics, $\mu_{xi}$ unknown, $\tilde{\mu}_{xi}$ observed (400 MC Samples, Entries Multiplied by $10^3$ )	94
Table 4.5	Monte Carlo Properties of Prediction MSE Estimators, $\mu_{xi}$ unknown and fixed, $\tilde{\mu}_{xi}$ observed ( $B_1 = 400, B_2 = 1, N = 400$ , Variances Multiplied by $10^3$ ) . . . . .	98
Table 4.6	Empirical Coverages for 95% Wald-type CIs, Bootstrap Estimated Prediction MSE ( $B_1 = 400, B_2 = 1$ ) . . . . .	99
Table 4.7	Empirical Coverages ( $B_1 = 400, B_2 = 1$ ) . . . . .	102
Table 4.8	Empirical Coverages, $\mu_{xi}$ known ( $B_1 = 400, B_2 = 1, N = 400$ ) . . . . .	103
Table 4.9	Properties of Estimated Length of CIs, . . . . .	103
Table 4.10	Properties of Estimated Correction in Double Bootstrap CIs, $\mu_{xi}$ unknown and fixed, $\tilde{\mu}_{xi}$ observed ( $B_1 = 400, B_2 = 1, N = 400$ , Entries Multiplied by $10^3$ ) . . . . .	104
Table 4.11	Empirical Coverages for Bootstrap CIs, Estimated Distribution of $ T^* $ ( $B_1 = 400, B_2 = 1$ ) . . . . .	105
Table 4.12	Monte Carlo Properties of Estimated Parameters for the Student-t Dis- tribution, $\mu_{xi}$ unknown and fixed, $\tilde{\mu}_{xi}$ observed ( $B_1 = 400, B_2 = 1, N = 400$ ) . .	106
Table 4.13	Empirical Coverages for Bootstrap CIs ( $\mu_{xi}$ known, $B_1 = 400$ and 800 MC Samples) . . . . .	115
Table 4.14	Monte Carlo Properties of Prediction MSE Estimators, $\mu_{xi}$ known ( $B_1 = 400, B_2 = 1, N = 800$ , Variances Multiplied by $10^3$ ) .	118
Table 4.15	Monte Carlo Properties of Prediction MSE Estimators, $\mu_{xi}$ known ( $B_1 = 400, B_2 = 1, N = 400$ , Variances Multiplied by $10^3$ ) .	119



Table 4.16	Monte Carlo Properties of Prediction MSE Estimators, $\mu_{xi}$ unknown, random, $\tilde{\mu}_{xi}$ observed ( $B_1 = 400, B_2 = 1, N = 400$ , Variances Multiplied by $10^3$ ) . . . . .	120
Table 4.17	Properties of Estimated Length of 95% Wald-type CIs Bootstrap Estimated Prediction MSE ( $B_1 = 400, B_2 = 1$ , Entries Multiplied by $10^3$ ) . . . . .	121
Table 4.18	Properties of Estimated Length of 95% Wald-type CIs, $\mu_{xi}$ known, Bootstrap Estimated Prediction MSE ( $B_1 = 400, B_2 = 1, N = 400$ MC Samples, Entries Multiplied by $10^3$ ) . . . . .	121
Table 4.19	Properties of Estimated Length of CIs, $\mu_{xi}$ known ( $B_1 = 400, B_2 = 1, N = 800$ , Entries Multiplied by $10^3$ ) . . . . .	122
Table 4.20	Properties of Estimated Length of CIs, $\mu_{xi}$ known ( $B_1 = 400, B_2 = 1, N = 400$ , Entries Multiplied by $10^3$ ) . . . . .	123
Table 4.21	Properties of Estimated Length of CIs, $\mu_{xi}$ unknown, random, $\tilde{\mu}_{xi}$ observed ( $B_1 = 400, B_2 = 1, N = 400$ , En- tries Multiplied by $10^3$ ) . . . . .	124
Table 4.22	Properties of Estimated Correction in Double Bootstrap CIs, $\mu_{xi}$ known ( $B_1 = 400, B_2 = 1, N = 800$ , Entries Multiplied by $10^3$ ) . . . . .	124
Table 4.23	Properties of Estimated Correction in Double Bootstrap CIs, $\mu_{xi}$ known ( $B_1 = 400, B_2 = 1, N = 400$ , Entries Multiplied by $10^3$ ) . . . . .	125
Table 4.24	Properties of Estimated Correction in Double Bootstrap CIs, $\mu_{xi}$ unknown, random, $\tilde{\mu}_{xi}$ observed ( $B_1 = 400, B_2 = 1, N = 400$ , En- tries Multiplied by $10^3$ ) . . . . .	125
Table 4.25	Properties of the Length of Bootstrap CIs, Estimated Distribution of $ T^* $ , $\mu_{xi}$ known ( $B_1 = 400, B_2 = 1, N = 800$ ) . . . . .	125
Table 4.26	Properties of the Length of Bootstrap CIs, Estimated Distribution of $ T^* $ , $\mu_{xi}$ known ( $B_1 = 400, B_2 = 1, N = 400$ ) . . . . .	126

Table 4.27	Properties of the Length of Bootstrap CIs, Estimated Distribution of $ T^* $ , $\mu_{xi}$ unknown, $\tilde{\mu}_{xi}$ observed ( $B_1 = 400, B_2 = 1, N = 400$ ) . . . . .	126
Table 4.28	Monte Carlo Properties of Estimated Parameters for the Student-t Distribution, $\mu_{xi}$ known ( $B_1 = 400, B_2 = 1, N = 800$ ) . . . . .	126
Table 4.29	Monte Carlo Properties of Estimated Parameters for the Student-t Distribution, $\mu_{xi}$ known ( $B_1 = 400, B_2 = 1, N = 400$ ) . . . . .	127
Table 4.30	Monte Carlo Properties of Estimated Parameters for the Student-t Distribution, $\mu_{xi}$ unknown, random, $\tilde{\mu}_{xi}$ observed ( $B_1 = 400, B_2 = 1, N = 400$ ) . . . . .	127
Table 4.31	Empirical Coverages for Level one Bootstrap CIs, Estimated Distribution of $ T^* $ ( $\mu_{xi}$ known, 400 MC Samples, $B_1 = 400, B_2 = 1$ ) . . . . .	128

**LIST OF FIGURES**

Figure 4.1	CI's for the Small Area Means: Empirical Coverage versus Estimated Length . . . . .	129
Figure 4.2	General CI's for the Small Area Means: Estimated Parameters for the Student-t Distribution . . . . .	130

## ACKNOWLEDGEMENTS

I would like to express the deepest appreciation to my major professor, Dr. Wayne A. Fuller, for guiding me in various aspect of conducting research and the writing of scientific papers. I would like to thank Dr. Fuller for his availability in advising, for his enthusiasm for my work and for his patience.

Throughout my graduate career, I have benefited immensely from the instruction I received at Iowa State University (ISU). I would like to acknowledge the faculty in the Statistics Department and the Center for Survey Statistics and Methodology (CSSM) at ISU. I am especially grateful for the opportunity to have worked as a research assistant for CSSM.

Many people have constantly offered me love, support and encouragement. I would especially like to thank my family in Romania, my family in Colorado, and my friends all over the world.

## ABSTRACT

Agencies and policy makers are interested in constructing reliable estimates for areas with small sample sizes, where areas often refer to geographic areas and demographic groups. The estimation for such areas is known as small area estimation. Procedures based on models have been used to construct estimates for the small area means, by exploiting auxiliary information. Mixed models are suitable small area models because they combine different sources of information and contain different sources of error. The models studied in this dissertation are unit level generalized linear mixed models in situations where the mean of an auxiliary variable is subject to estimation error. Different cases of auxiliary information are considered. Prediction methods for the small area mean, estimation of the prediction mean squared error (MSE) and confidence intervals (CIs) for the small area means are presented for the case when the response variable is nonnormal. In the simulation studies, the response variable is binary.

In the first study, two methods for constructing small area mean predictions are considered. The first method is based on the conditional distribution of the random area effects given the response variables. The second method, called the 'plug-in method' is based on the direct substitution of the predicted random area effects into the small area mean expression. Using a simulation study, we show that the 'plug-in' predictor for the small area mean can have sizeable bias.

The estimation of prediction MSE for small area models is complicated, particularly in a nonlinear model setting. In the second study, the efficiency gains associated with the random specification for the auxiliary variable measured with error are demonstrated. The prediction MSE is smaller when additional auxiliary information is available and included in the estimation. The effect of including auxiliary information, if available, in the estimation is smaller for

the random mean model than for the fixed mean model for the covariates. A parametric fast double bootstrap procedure is proposed for the estimation of the MSE of the predictor. The proposed procedure has smaller bootstrap error than a classical fast double bootstrap procedure with the same number of samples. We call the proposed procedure telescoping fast double bootstrap.

Most small area studies, including the first two studies in this dissertation, focus on constructing predictors for the area means and on estimating the variance of the prediction errors. The ultimate goal of this dissertation is to construct CIs for the small area means. The most common CI is based on the estimated prediction MSE and approximates the distribution of parameter estimates with a normal distribution. The coverage error for such an interval can be large when the distribution of the parameter estimate is skewed and when the standard error is poorly estimated. We present two sided CIs for the small area means of a binary response variable. The estimation of the prediction error variance and the estimation of the cutoff points are key components in the construction of confidence intervals for the small area means. A linear approximation of the model is considered and a Taylor variance approximation is presented for the prediction error variance. We compare the normal approximation method, the percentile bootstrap method and the pivot-like bootstrap method for estimating the cutoff points using a simulation study. Level one bootstrap and telescoping fast double bootstrap methods are used to construct CIs for the small area means. Pivot-like bootstrap CIs perform better than the percentile bootstrap CIs, with respect to the coverage errors. Double bootstrap CIs perform well, but do not improve the coverage accuracy compared to the level one bootstrap CIs. A method for constructing bootstrap CIs for a general level is proposed. The user is given a degrees of freedom for the Student-t distribution and a standard error of the small area mean prediction. The CI for the small area mean can be constructed in the common form  $(\hat{\theta}_i \pm \zeta_{1-\alpha/2, i, df_i} se(\hat{\theta}_i))$ , where  $i$  denotes the area,  $1 - \alpha$  is the desired level,  $\hat{\theta}_i$  is the predicted small area mean,  $\zeta_{1-\alpha/2, i, df_i}$  is the  $100(1 - \alpha/2)^{th}$  quantile of the Student-t distribution with given degrees of freedom  $df_i$ , and  $se(\hat{\theta}_i)$  is the given standard error of  $\hat{\theta}_i$ . The coverage of the general bootstrap CI is comparable to the coverage of the level specific bootstrap CI.

## CHAPTER 1. INTRODUCTION

Procedures based on models have been used to construct estimates for the means of small areas, by exploiting auxiliary information. We study nested models with a binary response, stochastic covariates and random area effects. In the first paper we investigate predictors for situations with different amounts of available information. We present bias and mean squared error results for different prediction methods.

Statistical models containing fixed effects and random effects are called mixed models. In the small area models, the area specific random effects explain the between area variations in the data not explained by the fixed effects part of the model. 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 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  $\omega_{ij}$  are available satisfying  $\sum_{j=1}^{n_i} \omega_{ij} = 1$ . They consider both hierarchical Bayes and empirical Bayes (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

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 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, Spain, using a multinomial logit mixed model.

Jiang and Lahiri (2001) and Pfeffermann and Correa (2012) consider a unit level logistic model and construct estimates for the small area proportions using the conditional distribution of the random area effects given the response variables. We consider a unit level mixed logistic model and study two methods for constructing small area mean predictions. The first method is based on the conditional distribution of the random area effects given the response variables. The second method, called the 'plug-in method' is based on the direct substitution of the predicted random area effects into the small area mean expression. We show that the 'plug-in' predictor for the small area mean can have sizeable bias.

The estimation of prediction mean squared error (MSE) for small area models is complicated, particularly in a nonlinear model setting. In the second paper we study unit level generalized linear mixed models under situations where the mean of an auxiliary variable is subject to estimation error. The efficiency gains associated with the random specification for the auxiliary variable measured with error are demonstrated. A parametric bootstrap procedure is proposed for the mean squared error of the predictor based on a logit model. The proposed procedure has smaller bootstrap error than a classical double bootstrap procedure with the same number of samples.

Taylor methods have been shown to give good estimates of the prediction mean squared error, for predicted small area means; see p. 103 in Rao (2003) for area level models and p. 139 in Rao (2003) for unit level models. When the direct estimates of small area means are



nonlinear functions of the auxiliary information and the random area effects, the prediction of the small area mean is no longer a linear function of the observations. Bootstrap methods have been used for MSE estimators in this case. Double bootstrap methods reduce the order of the bias in the bootstrap prediction MSE estimators. There are many studies reporting point estimates for the small area means, as well as prediction MSE estimates. Ghosh, Sinha and Kim (2006) consider an area level linear model with random auxiliary variable mean, estimated jointly with the small area mean. Ybarra and Lohr (2008) consider an area level linear model with auxiliary mean estimated with error. Datta, Rao and Torabi (2010), following Ghosh and Sinha (2007), studied a nested error linear regression model with area level covariate subject to measurement error.

Hall and Maiti (2006) consider linear area level models and a unit level binomial model with fixed known covariates. They construct small area predictions for the logit of the small area means and nonnegative, bias-corrected MSE estimates using a double bootstrap procedure. Pfeffermann and Correa (2012) study a unit level binomial model with fixed known covariates and suggest a bootstrap procedure in which the bias in the estimator is estimated as a function of parameters and of a bootstrap estimator of bias.

Agencies and policy makers are often interested in confidence intervals for the small area estimates. Most studies that report confidence intervals (CIs) for the small area means, report for special cases of the Fay-Herriot model; see Hall and Maiti (2006), Chatterjee et al (2008), Dass et al (2012), Diao et al (2014) and Yoshimori and Lahiri (2014). In the third paper, we consider different procedures to estimate the small area mean prediction mean squared error and to construct confidence intervals for the small area means.

Let  $\theta_i$  be the area specific parameter of interest, where  $i$  denotes the area. Let a two-sided  $\alpha$  level confidence interval for  $\theta_i$  be the interval  $I = (\theta_{Li}, \theta_{Ui})$ , with desired coverage  $1 - \alpha = P(\theta_i \in I)$ . When  $P(\theta_i < \theta_{Li}) = P(\theta_i > \theta_{Ui}) = \alpha/2$ ,  $I$  is called equal tailed  $\alpha$  level confidence interval for  $\theta_i$ . Let  $\hat{\theta}_i$  be the sample estimate of  $\theta_i$ . If  $\theta_{Li} = \hat{\theta}_i - c$ ,  $\theta_{Ui} = \hat{\theta}_i + c$  and

$P(|\hat{\theta}_i - \theta_i| > c) = \alpha$ ,  $I$  is called symmetric  $\alpha$  level confidence interval for  $\theta_i$ . The difference between the actual coverage rate of a confidence interval and the claimed value  $1 - \alpha$  is the coverage error.

The most common confidence interval is based on the estimated MSE and approximating the distribution of parameter estimates with a normal distribution. This confidence interval is called a Wald-type confidence interval. The standard form of the  $\alpha$  level Wald-type CI for  $\theta_i$  is  $(\hat{\theta}_i \pm \zeta_{\alpha,i} se(\hat{\theta}_i))$ , where  $\zeta_{\alpha,i}$  is the normal percentile, for the desired  $\alpha$  level, and  $se(\hat{\theta}_i)$  is the standard error of the estimate  $\hat{\theta}_i$ , where  $se(\hat{\theta}_i)$  can be computed using Taylor methods or using bootstrap methods. The coverage error can be large when the distribution of the parameter estimate is skewed and when the standard error is poorly estimated.

Because Wald-type confidence intervals (CIs) often have poor coverage, correction procedures have been proposed. Correction of the coverage probability can be performed using asymptotic expansions of the correction and of the interval endpoints, in an iterative method. Diao et al (2014) construct second order correct confidence intervals for small area means based on the Fay-Herriot model. They calibrate the cutoff points using asymptotic expansions in a direct way, not involving resampling methods, where calibration refers to the bias adjustment in the point or interval estimates.

Bootstrap methods were introduced to construct confidence intervals in an algorithmic fashion, using fewer assumptions than those based on the normal approximation. Also, bootstrap confidence intervals can be constructed for complicated models and data structures. Efron (1983) introduced the bootstrap as a nonparametric tool for estimating standard errors and biases. Confidence intervals require more effort than parametric estimation. Methods of improvements have been developed since Efron (1983), such as the bootstrap accelerated method, bootstrap-t, iterated bootstrap and calibration, see Martin (1990), Hall (1992) and Shi (1992).

Hall (1986) proposed pivot-like statistics to reduce the two-sided bootstrap CI coverage error from  $O(m^{-1})$  to  $O(m^{-3/2})$ , where  $m$  is the number of areas. Chatterjee et al (2008) construct parametric bootstrap confidence intervals based on a pivot-like statistic and centered around the small area predictions, for a generalization of the Fay-Herriot model. The area specific confidence intervals constructed by the authors have error of order  $O(d^3 m^{-3/2})$ , where  $d$  is the number of model parameters and  $m$  is the number of small areas. Chatterjee et al. (2008) state that if calibrated, their intervals would be  $O(d^5 m^{-5/2})$  order correct. Liu and Diallo (2013) apply the method in Chatterjee et al(2008) and construct percentile parametric bootstrap confidence intervals for survey-weighted small area proportions based on the Fay-Herriot model.

The double bootstrap is a procedure designed to estimate the coverage error of a CI and adjust the interval, based on the error estimate, in such a way that the coverage probability improves. In the level one bootstrap, a large number of bootstrap samples are drawn. In the double bootstrap, a large number of bootstrap samples are drawn for every level one bootstrap iteration. Further nested levels of bootstrap samples can be drawn at every previous bootstrap level in the iteration procedure. Hall (1986), Beran (1987), Loh (1987), Martin (1990), Shi (1992), Davidson and Hinkley (1997), McCullough and Vidon (1998) and Nankervis (2005) summarize the theory of bootstrap iteration for confidence intervals. The authors consider basic, studentized, percentile and percentile-t confidence intervals, using pivot and pivot-free methods. McCullough and Vidon (1998) and Nankervis (2005) use a pivot-like statistic and correct the coverage rate of the two-sided equal-tailed and symmetric bootstrap confidence intervals by estimating the cutoff points in the second bootstrap level. A uniform random variable is constructed with realizations that are the proportions of the level two pivot values less than the level one pivot values and the quantiles of this uniform random variable are used to correct for the final cutoff points. In a Monte Carlo study, Nankervis (2005) compares the empirical coverage rates for percentile and for studentized bootstrap confidence intervals, at levels  $\alpha = 0.1, 0.05$ , for the mean from samples from a normal and lognormal populations, and for the cumulative impulse response in a second order autoregressive model (AR(2)). For the

Gaussian example, the empirical coverage rates of the level one bootstrap CIs for the mean and of the double bootstrap CIs for the mean are close to the nominal coverages, while for the lognormal example, the bootstrap confidence intervals result in undercoverage, with better coverage rates for the double bootstrap CIs for the mean. For the stationary Gaussian AR(2) model, the errors in the coverage rates of the studentized bootstrap CIs are smaller than those of the percentile intervals, and the double bootstrap improves the coverage rates for the level one bootstrap CIs. Shi (1992) studied the method described in Nankervis (2005), with no pivot-like statistic. Using Edgeworth expansions of the distributions and Cornish-Fisher inversion of quantiles, Shi proves that the difference between the bootstrap CI endpoint limit and the theoretical CI endpoint limit is  $O_p(m^{-3/2})$ , and that the difference in probability coverage between the bootstrap CI and the theoretical CI is  $O(m^{-1})$ . Martin (1990) constructs double bootstrap confidence intervals, using a coverage correction method based on the interpolation between the estimated true coverages at several nominal levels close to the desired level. He shows that the expected asymptotic length of the final intervals changes by an amount proportional to the coverage error of the original interval. The author discusses the advantages of bootstrap coverage-correction, such as the transformation invariance property of the percentile CIs, the simplicity of implementing the percentile CIs since no variance estimator is needed, and the asymptotic high-order coverage accuracy in CIs, and the disadvantages of bootstrap coverage-correction, such as the computational expense.

Hall and Maiti (2006) construct two sided, equal-tailed, double bootstrap calibrated CIs. The authors outline an algorithm for calibrating the CI coverage and constructing percentile confidence intervals for the parameter of interest. Linear and nonlinear models are considered, but the parameter of interest is always a linear function of the model parameters. For example, they consider a binary response variable, with area mean  $\psi(\theta_i)$ , where  $\psi$  is the inverse logit function and  $\theta_i = f_i(\boldsymbol{\beta}) + b_i = \mathbf{x}'_i\boldsymbol{\beta} + b_i$ . The index  $i$  denotes the area and  $f_i(\boldsymbol{\beta})$  is an area specific known smooth function of the covariates  $\mathbf{x}_i$  and the vector of parameters  $\boldsymbol{\beta}$ . The random effects  $b_i$  are independent and identically distributed with mean zero. The authors assume that  $\theta_i$  follows a normal distribution with mean  $\mathbf{x}'_i\boldsymbol{\beta}$  and variance  $\zeta$ . The parameter

of interest is  $\theta_i = \text{logit}(\psi(\mathbf{x}'_i\boldsymbol{\beta} + b_i)) = \mathbf{x}'_i\boldsymbol{\beta} + b_i$ , the logit of the small area mean  $\psi(\mathbf{x}'_i\boldsymbol{\beta} + b_i)$ . Hence, the parameter is a linear function of the area covariate  $\mathbf{x}_i$  and the area random effect  $b_i$ . The bootstrap CIs for  $\theta_i$  are constructed using the estimated distribution of the bootstrap predictions of  $\theta_i$ . The authors state that the coverage error of the level one bootstrap CI is of order  $O(m^{-2})$  and that the coverage error of the level two calibrated bootstrap CI is of order  $O(m^{-3})$ . A simulation study is conducted for the binary model, for  $m = 15$  areas with sample sizes in the range 48 to 287, vector of parameters  $\boldsymbol{\beta} = (0, 1)$ , and variance of random effects  $\zeta = 1$ . Bootstrap confidence intervals with nominal coverages  $\alpha = 0.80, 0.90, 0.95$  are constructed for  $\theta_i$ , for different models. The simulation results show undercoverage for the normal approximated CIs, good performance for the level one bootstrap CIs for  $\alpha = 0.20, 0.10$  and undercoverage for the level one bootstrap CIs for  $\alpha = 0.05$ . When the performance for the level one bootstrap CI for  $\theta_i$  was good, the calibrated double bootstrap CI either had no effect or it produced overcoverage. When the level one bootstrap CI for  $\theta_i$  undercovered, the calibrated double bootstrap CI for  $\theta_i$  corrected the coverage error, but it resulted in overcoverage.

Nested levels of bootstrap samples quickly become very costly. Several authors have considered analytical approximation to bootstrap distribution functions, to replace the inner levels of resampling in iterated bootstrap procedures. Davison and Hinkley (1988) and Diccio, Martin and Young (1992) proposed the use of saddlepoint approximations for constructing approximate iterated bootstrap CIs. Lee and Yong (1995) construct asymptotic iterated bootstrap confidence intervals, replacing the need for a second bootstrap level. They construct two-sided percentile CI with calibrated nominal coverage. The authors use Edgeworth expansions of the coverage and the endpoints of the CIs.

Davidson and MacKinnon (2007) introduced a fast double bootstrap procedure for bootstrap testing. Only one sample is drawn at the second bootstrap level. Giacomini et al. (2013) provide key properties for fast double bootstrap methods, under regularity conditions. The authors discuss applications of fast double bootstrap methods to assess the performance of bootstrap estimators, test statistics and confidence intervals. Erciulescu and Fuller (2014)

study a unit level binomial model and construct prediction mean squared error estimators for the small area means using fast double bootstrap procedures. Chang and Hall (2014) study the fast double bootstrap method described in Giacomini et al (2013) to produce third-order accurate confidence intervals. The authors show that the performance of the fast double bootstrap in reducing the order of magnitude of bias is comparable to the classic double bootstrap method. The authors describe a calibration method for the confidence interval cutoff points and show that the fast double bootstrap does not improve level one order of magnitude of coverage error of the CIs.

### 1.1 Bootstrap Properties

Suppose that  $X_1, X_2, \dots, X_n$  is a random sample from a distribution  $F$  and that  $\theta$  is the parameter of interest. Let  $\hat{\theta}$  be a sample estimator of  $\theta$ . The idea of bootstrap (Efron, 1979) is to treat the sample  $X_1, X_2, \dots, X_n$  as the population and to draw samples of size  $n$ , with replacement, from  $X_1, X_2, \dots, X_n$ , denoted by  $X_1^*, X_2^*, \dots, X_n^*$ . The bootstrap estimate of  $\hat{\theta}$  is  $\hat{\theta}^*$ , a function of the bootstrap sample  $X_1^*, X_2^*, \dots, X_n^*$ . The procedure of drawing a sample of size  $n$  from the original sample  $X_1, X_2, \dots, X_n$ , treated as the population, is called the nonparametric bootstrap, proposed by Efron (1979).

Assume that the form of distribution  $F$  is known and is determined by an unknown parameter  $\psi$ . The parametric bootstrap is the procedure of generating a sample of size  $n$  from a distribution  $\hat{F}$ , defined by  $\hat{\psi}$ , estimated from the original sample.

The bootstrap principle (p. 45 in Efron and Tibshirani, 1993) is that information on the relationship between the true parameter  $\theta$  and its estimator  $\hat{\theta}$  can be obtained by treating  $\hat{\theta}$  as the true parameter value in question and looking at the relationship between  $\hat{\theta}$  and  $\hat{\theta}^*$ , where  $\hat{\theta}^*$  is its bootstrap estimate. Based on the bootstrap principle, the distribution of  $\hat{\theta}^* - \hat{\theta}$  can be used to approximate the sampling distribution of  $\hat{\theta} - \theta$ , and hence to construct CIs.

**Bootstrap theoretical basis.** Our discussion follows Chapter 52 in Horowitz (2001). Suppose that  $X_1, X_2, \dots, X_n$  is a random sample from  $F$ . Let  $G = g(X_1, X_2, \dots, X_n; F)$  be the parameter of interest, with distribution

$$G_{F,n}(g) = P(G(X_1, X_2, \dots, X_n; F) \leq g | F),$$

and let the bootstrap estimate of  $G_{F,n}(g)$  be

$$G_{\hat{F},n}(g) = P(G(X_1^*, X_2^*, \dots, X_n^*; \hat{F}) \leq g | \hat{F}),$$

where  $X_1^*, X_2^*, \dots, X_n^*$  is a bootstrap sample and  $\hat{F}$  is the empirical distribution of  $G$ .

Suppose that, as  $n \rightarrow \infty$ ,  $\hat{F}$  falls into a neighbourhood,  $\mathcal{N}$ , of  $F$ , with probability one. The bootstrap is consistent if for any  $g$  and  $\epsilon > 0$ ,  $P(\sup_g |G_{\hat{F},n}(g) - G_{F,\infty}(g)| > \epsilon) \rightarrow 0$  as  $n \rightarrow \infty$ . The consistency holds under the following conditions:

- for any distribution  $\mathcal{A} \in \mathcal{N}$ ,  $G_{\mathcal{A},n}$  must converge weakly to a limit  $G_{\mathcal{A},\infty}$
- the convergence of  $G_{\mathcal{A},n}$  must be uniform on  $\mathcal{N}$
- the function mapping  $\mathcal{A}$  to  $G_{\mathcal{A},\infty}$  must be continuous.

The validity of the bootstrap for a given parameter  $G$  is given by the existence of an Edgeworth expansion for the statistic of interest. For example, a differentiable function of the sample moments is a smooth and stable statistic. The parameter of interest in the bootstrap methods should be a smooth function, details on the conditions are described in Bhattacharya and Ghosh (1978).

## 1.2 Parametric Double Bootstrap Results

Consider the parametric bootstrap procedure. The parameter  $\psi$  is estimated by  $\hat{\psi}$ , using the original sample. Suppose  $B_1$  samples are generated from the a distribution  $\hat{F}$ , determined by  $\hat{\psi}$ , and let  $\alpha_k^*$ ,  $k = 1, \dots, B_1$  be the parameter of interest for the  $k^{th}$  bootstrap sample. Given

$\hat{\psi}$ , the  $\alpha_k^*$  are identically distributed. The parameter  $\psi$  is estimated by  $\psi_k^*$ , using the bootstrap sample  $k$ , for  $k = 1, \dots, B_1$ . In the double bootstrap, the level one procedure is repeated for each level one bootstrap sample. For each level one bootstrap sample  $k$ , a new set of  $B_2$  samples are generated, using  $\psi_k^*$ . Let  $\alpha_{kt}^{**}$ ,  $k = 1, \dots, B_1$ ,  $t = 1, \dots, B_2$  be the parameter of interest for the  $k^{th}$  level one bootstrap sample,  $t^{th}$  level two bootstrap sample. Given  $\psi_k^*$ ,  $\alpha_{kt}^{**}$  are identically distributed, for  $k=1, \dots, B_1$ ,  $t = 1, \dots, B_2$ . In the classic double bootstrap, both  $B_1$  and  $B_2$  are large. In the classic fast double bootstrap,  $B_2 = 1$ .

Let level one bootstrap estimator of  $\alpha$  be

$$\hat{\alpha}^* = B_1^{-1} \sum_{k=1}^{B_1} \alpha_k^* = \bar{\alpha}^*. \quad (1.1)$$

Then the double bootstrap bias adjusted estimator of  $\alpha$  is

$$\hat{\alpha}^{**} = B_1^{-1} \sum_{k=1}^{B_1} \left( 2\alpha_k^* - B_2^{-1} \sum_{t=1}^{B_2} \alpha_{k,t}^{**} \right), \quad (1.2)$$

and a fast double bootstrap bias adjusted estimator of  $\alpha$  is

$$\hat{\alpha}^{**} = B_1^{-1} \sum_{k=1}^{B_1} (2\alpha_k^* - \alpha_k^{**}) = 2\bar{\alpha}^* - \bar{\alpha}^{**}. \quad (1.3)$$

We present two results for the parametric double bootstrap procedure. In the first result we show that, under suitable conditions, the expected value of the bias corrected double bootstrap estimator does not change with a change in the number of level two samples. Chang and Hall (2014) give a different proof for this result, showing that the order of magnitude of bias reduction is comparable for the fast double bootstrap and for the classic double bootstrap. In the second result we give an expression for the optimal number of level two bootstrap samples needed to minimize the variance of the bootstrap estimator, given a fixed number of bootstrap samples.

**Result 1.** Consider the parametric bootstrap procedure and let  $\alpha$  be parameter of interest, where  $\alpha$  is a smooth function of an unknown parameter  $\psi$ . Let the distribution  $F$  be a known smooth function of  $\psi$  and let the estimated distribution  $\hat{F}$  be a smooth function of  $\hat{\psi}$ .



Assume that the conditions given in Section 1.1 hold for  $G = \alpha$ . Then, the expected value of the additive bias corrected double bootstrap estimator does not change with a change in the number of level two samples,  $B_2, B_2 \geq 1$ .

*Proof.* The bootstrap estimated bias in the estimator of the parameter of interest,  $\alpha$ , is

$$\Delta_\alpha = B_1^{-1} \sum_{k=1}^{B_1} (B_2^{-1} \sum_{t=1}^{B_2} \alpha_{kt}^{**} - \alpha_k^*),$$

and its expected value is

$$\begin{aligned} E(\Delta_\alpha) &= E\left(B_1^{-1} \sum_{k=1}^{B_1} \left(B_2^{-1} \sum_{t=1}^{B_2} \alpha_{kt}^{**} - \alpha_k^*\right)\right) \\ &= E\left(E\left(B_1^{-1} \sum_{k=1}^{B_1} B_2^{-1} \sum_{t=1}^{B_2} \alpha_{kt}^{**} | \psi_k^*\right)\right) - E\left(B_1^{-1} \sum_{k=1}^{B_1} \alpha_k^*\right). \end{aligned}$$

Since  $\alpha_{kt}^{**}$  are identically distributed, given  $\psi_k^*$ ,

$$\begin{aligned} E(\Delta_\alpha) &= E\left(E\left(B_1^{-1} \sum_{k=1}^{B_1} \alpha_{k1}^{**} | \psi_k^*\right)\right) - E\left(B_1^{-1} \sum_{k=1}^{B_1} \alpha_k^*\right) \\ &= E(B_1^{-1} \sum_{k=1}^{B_1} \alpha_{k1}^{**} - \alpha_k^*), \end{aligned}$$

for any  $B_2$ . ■

The variance of the bootstrap estimator of  $\alpha$  has two components. The first, that we call *between*, is the variance one would obtain if one used an infinite number of bootstrap samples. The second, that we call *within*, is the variability due to the fact that our set of bootstrap samples is a sample of samples.

**Result 2.** Consider the parametric bootstrap procedure and let  $\alpha$  be the parameter of interest, where  $\alpha$  is a smooth function of  $\psi$ . Let the distribution  $F$  be a known smooth function of  $\psi$  and let the estimated distribution  $\hat{F}$  be a smooth function of  $\hat{\psi}$ . Assume that the conditions given in Section 1.1 hold for  $G = \alpha$ . Then, given a fixed number of bootstrap

samples,  $B$ , the within bootstrap variance of the bias corrected double bootstrap estimator is minimized when the number of level two bootstrap samples,

$$B_2 = \sqrt{\frac{E\left(V(\alpha^{**}|\psi^*)|\hat{\psi}\right)}{4V_w(\alpha^*) - 4C_w(\alpha^*, \alpha^{**}) + E\left\{V\left(E(\alpha^{**}|\psi^*)|\hat{\psi}\right)\right\}}}, \quad (1.4)$$

where

$$V_w(\alpha^*) = E\left(V(\alpha^*|\hat{\psi})\right),$$

$$V_w(\alpha^{**}) = E\left\{V\left(E(\alpha^{**}|\psi^*)|\hat{\psi}\right)\right\} + E\left(V(\alpha^{**}|\psi^*)|\hat{\psi}\right),$$

$$C_w(\alpha^*, \alpha^{**}) = E\{C(\alpha^*, E(\alpha^{**}|\psi^*))\}.$$

*Proof.* The double bootstrap, bias corrected, estimator of  $\alpha$  is

$$\hat{\alpha}^{**} = B_1^{-1} \sum_{k=1}^{B_1} \alpha_k^* - \Delta_\alpha = B_1^{-1} \sum_{k=1}^{B_1} 2\alpha_k^* - B_1^{-1} B_2^{-1} \sum_{k=1}^{B_1} \sum_{t=1}^{B_2} \alpha_{kt}^{**} =: 2\bar{\alpha}^* - \bar{\alpha}^{**},$$

and the within bootstrap variance of  $\hat{\alpha}^{**}$  is

$$\begin{aligned} V_w(\hat{\alpha}^{**}) &= V_w(2\bar{\alpha}^* - \bar{\alpha}^{**}) \\ &= V_w(2\bar{\alpha}^*) + V_w(\bar{\alpha}^{**}) - 2C_w(2\bar{\alpha}^*, \bar{\alpha}^{**}) \\ &= E\left(V(2B_1^{-1} \sum_{k=1}^{B_1} \alpha_k^*|\hat{\psi})\right) + E\left\{V\left(E(B_1^{-1} \sum_{k=1}^{B_1} \bar{\alpha}_k^{**}|\psi_k^*)|\hat{\psi}\right)\right\} \\ &\quad + E\left(V(B_1^{-1} B_2^{-1} \sum_{k=1}^{B_1} \sum_{t=1}^{B_2} \alpha_{kt}^{**}|\psi_k^*)|\hat{\psi}\right) - 4E\left\{C\left(\bar{\alpha}^*, E(B_1^{-1} \sum_{k=1}^{B_1} \bar{\alpha}_k^{**}|\psi_k^*)\right)\right\}, \end{aligned}$$

where  $\bar{\alpha}_k^{**} = B_2^{-1} \sum_{t=1}^{B_2} \alpha_{kt}^{**}$ . Since  $\alpha_{kt}^{**}$  are identically distributed, given  $\psi_k^*$  and  $\alpha_k^*$  are identically distributed, given  $\hat{\psi}$ ,

$$\begin{aligned} V_w(\hat{\alpha}^{**}) &= 4B_1^{-1} E\left(V(\alpha^*|\hat{\psi})\right) + B_1^{-1} E\left\{V\left(E(\alpha_1^{**}|\psi^*)|\hat{\psi}\right)\right\} \\ &\quad + B_1^{-1} B_2^{-1} E\left(V(\alpha^{**}|\psi^*)|\hat{\psi}\right) - 4B_1^{-1} E\{C(\alpha^*, E(\alpha^{**}|\psi^*))\} \\ &= 4B_1^{-1} V_w(\alpha^*) + B_1^{-1} E\left\{V\left(E(\alpha^{**}|\psi^*)|\hat{\psi}\right)\right\} + B_1^{-1} B_2^{-1} E\left(V(\alpha^{**}|\psi^*)|\hat{\psi}\right) - 4B_1^{-1} C_w(\alpha^*, \alpha^{**}). \end{aligned}$$

We would like to minimize  $V_w(\hat{\alpha}^{**})$  with respect to the restriction  $B_1 B_2 + B_1 = B$ , where  $B$  is a constant representing the total number of bootstrap samples.

To consider this problem as a Lagrangian multiplier problem, let

$$\begin{aligned} L(B_1, B_2, \lambda) = & 4B_1^{-1}V_w(\alpha^*) - 4B_1^{-1}C_w(\alpha^*, \alpha^{**}) + B_1^{-1}E \left\{ V \left( E(\alpha^{**}|\psi^*)|\hat{\psi} \right) \right\} \\ & + B_1^{-1}B_2^{-1}E \left( V(\alpha^{**}|\psi^*)|\hat{\psi} \right) + \lambda(B_1 B_2 + B_1 - B) \end{aligned}$$

where  $\lambda$  the Lagrangian multiplier. The resulting system of three equations is:

$$0 = -4B_1^{-2}V_w(\alpha^*) + 4B_1^{-2}C_w(\alpha^*, \alpha^{**}) - B_1^{-2}E \left\{ V \left( E(\alpha^{**}|\psi^*)|\hat{\psi} \right) \right\}$$

$$-B_1^{-2}B_2^{-1}E \left( V(\alpha^{**}|\psi^*)|\hat{\psi} \right) + \lambda(B_2 + 1)$$

$$0 = -B_1^{-1}B_2^{-2}E \left( V(\alpha^{**}|\psi^*)|\hat{\psi} \right) + \lambda B_1$$

$$0 = B_1 B_2 + B_1 - B.$$

The solution for  $B_2$  is (1.4) and

$$B_1 = B \left( \sqrt[4]{\frac{E \left( V(\alpha^{**}|\psi^*)|\hat{\psi} \right)}{4V_w(\alpha^*) - 4C_w(\alpha^*, \alpha^{**}) + E \left\{ V \left( E(\alpha^{**}|\psi^*)|\hat{\psi} \right) \right\}}} + 1 \right)^{-1}.$$

■

## Bibliography

- [1] Chang, J. and Hall, P. (2014), “Double-bootstrap methods that use a single double-bootstrap simulation,” *Biometrika*, submitted.
- [2] Chatterjee, S., Lahiri, P., Li, H. (2008), “Parametric Bootstrap Approximation to the Distribution of EBLUP and Related Prediction Intervals in Linear Mixed Models,” *The Annals of Statistics*, 36, 3, 1221-1245.
- [3] Dass, S., Maiti, T., Ren, H., Sinha, S. (2001), “Confidence interval estimation of small area parameters shrinking both means and variances,” *Survey Methodology*, 38, 2, 173-187.
- [4] Datta, G.S. and Lahiri, P. (2000), “A unified measure of uncertainty of best linear unbiased predictors in small area estimation problems,” *Statistica Sinica*, 10, 613-627.
- [5] Datta, G.S., Rao, J.N.K. and Smith, D. (2005) “On measuring the variability of small area estimators under a basic area level model. ” *Biometrika*, 92, 183-196.
- [6] Datta, G.S., Rao, J.N.K. and Smith, D. (2012) “Amendments and Corrections: On measuring the variability of small area estimators under a basic area level model, ” *Biometrika*, 99, 2, 509.
- [7] 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.
- [8] Davidson, R. and MacKinnon, J.G. (2007), “Improving the reliability of bootstrap tests with the fast double bootstrap,” *Computational Statistics and Data Analysis*, 51, 3259-3281.

- [9] Diao, L., Smith, D., Datta, G.S., Maiti, T., Opsomer, J. (2014), “ Accurate Confidence Interval Estimation of Small Area Parameters Under the Fay-Herriot Model,” *Scandinavian Journal of Statistics*, 41, 497-515.
- [10] Erciulescu, A.L. and Fuller, W.A. (2013), “Small Area Prediction of the Mean of a Binomial Random Variable,” *Survey Research Methods Section, JSM Proceedings*, 855-863.
- [11] Erciulescu, A.L. and Fuller, W.A. (2014), “Small Area Prediction under Alternative Model Specifications,” *Statistics in Transition and Survey Methodology*. Manuscript submitted for publication.
- [12] Fay, R.E. and Herriot, R.A. (1979), “Estimates of income for small places: an application of James-Stein procedures to census data,” *Journal of the American Statistical Association*, 74, 269-277.
- [13] Ghosh, M. and Sinha, K. (2007), “Empirical Bayes estimation in finite population sampling under functional measurement error models,” *Journal of Statistical Planning and Inference*, 137, 2759-2773.
- [14] Ghosh, M., Sinha, K. and Kim, D. (2006), “Empirical and Hierarchical Bayesian Estimation in Finite Population Sampling under Structural Measurement Error Models,” *Scandinavian Journal of Statistics*, 33, 3, 591-608.
- [15] Giacomini, R., Politis, D. and White H. (2013), ”A Warp-Speed Method for Conducting Monte Carlo Experiments Involving Bootstrap Estimators,” *Econometric Theory*, 29, 567-589.
- [16] Hall P.(1986), “On the bootstrap and confidence intervals,” *Annals of Statistics*, 14, 4, 1431-1452.
- [17] Hall P. and Maiti T. (2006), “On parametric bootstrap methods for small area prediction,” *J.R. Statist. Soc. B*, 68, 2, 221-238.
- [18] Jiang, J. and Lahiri, P. (2001), “Empirical best prediction for small area inference with binary data,” *Annals of the Institute of Statistical Mathematics*, 53, 2, 217-243..

- [19] Liu, B. and Diallo, M. (2013), "Parametric Bootstrap Confidence Intervals for Survey-Weighted Small Area Proportions," *JSM 2013- SRMS Proceedings*, 109-121.
- [20] Pfeiffermann, D. and Correa, S. (2012), "Empirical bootstrap bias correction and estimation of prediction mean square error in small area estimation." *Biometrika*, 99, 457-472.
- [21] Rao J.N.K. (2003), "Small Area Estimation," *Wiley Series in Survey Methodology*.
- [22] Yoshimori, M. and Lahiri, P. (2014), "A second-order efficient empirical Bayes confidence interval," *Annals of Statistics*, 42, 4, 1233-1261.
- [23] Ybarra, L.M.R. and Lohr, S.L. (2008), " Small area estimation when auxiliary information is measured with error," *Biometrika*, 95, 919-931.

## CHAPTER 2. SMALL AREA PREDICTION OF THE MEAN OF A BINOMIAL RANDOM VARIABLE

A paper in *Survey Research Methods Section, JSM Proceedings*

Andreea L. Erciulescu and 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.

### **2.1 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  $\omega_{ij}$  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 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.



## 2.2 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  $\sigma_b^2$ .

Then our unit level model is

$$y_{ij} = h(\eta_{ij}) + e_{ij}, \quad \eta_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + b_i, \quad h(\eta_{ij}) = \frac{\exp(\eta_{ij})}{1 + \exp(\eta_{ij})} \quad (2.1)$$

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 (2.1), the true small area mean of  $\mathbf{y}$  is

$$\theta_i = \int p_{ij}(\mathbf{x}_{ij}, b_i) dF_{\mathbf{x}_i}(\mathbf{x}), \quad (2.2)$$

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

An example of (2.1) is the simple unit level mean model for  $\mathbf{y}$

$$p_{\alpha, ij} = \frac{\exp(\alpha + b_i)}{1 + \exp(\alpha + b_i)}, \quad (2.3)$$

where  $\alpha$  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  $\mathbf{x}_{ij} = (1, x_{ij})$ , and assume

$$\boldsymbol{\mu}_{xi} \sim NI(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{\delta\delta}), \quad \mathbf{x}_i | \boldsymbol{\mu}_{xi} \sim NI(\boldsymbol{\mu}_{xi}, \boldsymbol{\Sigma}_{\epsilon\epsilon}). \quad (2.4)$$

## 2.3 Estimation and Prediction

The models (2.1) and (2.3) are generalized linear mixed models (GLMMs) and estimates for  $\boldsymbol{\beta}$ ,  $\sigma_b^2$ ,  $\alpha$  and  $\sigma_{2b}^2$  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 (2.3) differ from the estimated values under model (2.1), hence we denote those for model (2.3) by  $\hat{\mathbf{b}}_2$  and  $\hat{\sigma}_{2b}^2$ , respectively.

We consider two methods for constructing predictions for  $\theta_i$ . 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  $\theta_i$ . 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  $\theta_i$ . We compare these two methods using a simulation study.

### 2.3.1 MMSE Prediction

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

$$\hat{b}_i = \int \frac{\int b_i \prod_{t=1}^{n_i} f(y_{it}|b_i) f_b(b_i) db_i}{\int \prod_{t=1}^{n_i} f(y_{it}|b_i) f_b(b_i) db_i} dF_{\mathbf{x}_i}(\mathbf{x}). \quad (2.5)$$

Let  $\boldsymbol{\mu}_{x_i}$  be the area mean of  $\mathbf{x}_i$ . We present predictions for  $\theta_i$ , for different cases of auxiliary information, when  $\boldsymbol{\mu}_{x_i}$  is known, when the distribution of  $\mathbf{x}$  is unknown, and when  $\boldsymbol{\mu}_{x_i}$  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 (2.4).

### 2.3.1.1 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

$$\hat{\theta}_i = \int \frac{\int p_{ij}(\mathbf{x}_{ij}, b_i) \prod_{t=1}^{n_i} f(y_{it}|b_i) f_b(b_i) db_i}{\int \prod_{t=1}^{n_i} f(y_{it}|b_i) f_b(b_i) db_i} dF_{\mathbf{x}_i}(\mathbf{x}). \quad (2.6)$$

In some finite population situations, the entire finite population of  $\mathbf{x}$  values may be known and the integral in (2.6) is the sum over the population. In practice it is often necessary to estimate the parameters of the distribution  $F_{\mathbf{x}_i}$ .

### 2.3.1.2 Unspecified distribution for $\mathbf{x}$

If  $\mu_{x_i}$  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

$$\bar{p}_i = \sum_{j=1}^{n_i} \omega_{ij} \frac{\int p_{ij}(\mathbf{x}_{ij}, b_i) \prod_{t=1}^{n_i} f(y_{it}|b_i) f_b(b_i) db_i}{\int \prod_{t=1}^{n_i} f(y_{it}|b_i) f_b(b_i) db_i}. \quad (2.7)$$

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

### 2.3.1.3 No Auxiliary Information Used

Under model (2.3), for known parameters, the MMSE predictor of the small area mean of  $\mathbf{y}$  is

$$\hat{p}_i = \frac{\int p_{\alpha,ij} \prod_{t=1}^{n_i} f(y_{it}|b_i) f_b(b_i) db_i}{\int \prod_{t=1}^{n_i} f(y_{it}|b_i) f_b(b_i) db_i}, \quad (2.8)$$

where  $p_{\alpha,ij}$  is defined in (2.3).

### 2.3.1.4 Unknown Random Covariate Mean

Consider the model (2.1) for  $\mathbf{y}$  and the linear mixed model for  $x_{ij}$  given in (2.4):

$$x_{ij} = \mu_x + \delta_i + \epsilon_{ij}, \quad \delta_i \sim N(0, \sigma_\delta^2), \quad \epsilon_{ij} | \delta_i \sim N(0, \sigma_\epsilon^2) \quad (2.9)$$

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

$$\hat{\mu}_{xi} = \hat{\mu}_x + \hat{\gamma}_{xi}(\bar{x}_i - \hat{\mu}_x), \quad (2.10)$$

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 (2.10),  $\bar{x}_i = n_i^{-1} \sum_{j=1}^{n_i} x_{ij}$  denotes the sample area mean of  $\mathbf{x}_i$ , and the variance of the random area effects  $\delta_i$  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

$$\tilde{\theta}_i = \int \frac{\int p_{ij}(\mathbf{x}_{ij}, b_i) \prod_{t=1}^{n_i} f(y_{it} | b_i) f_b(b_i) db_i}{\int \prod_{t=1}^{n_i} f(y_{it} | b_i) f_b(b_i) db_i} d\tilde{F}_{\mathbf{x}_i}(\mathbf{x}), \quad (2.11)$$

where  $\tilde{F}_{\mathbf{x}}(\mathbf{x})$  is the estimator of  $F_{\mathbf{x}}(\mathbf{x})$  with parameter  $\boldsymbol{\mu}_x$  predicted on the basis of model (2.4).

If  $F_{\mathbf{x}}$  and  $f_b$  are continuous distributions, there are many ways to approximate the integrals in (2.2, 2.5, 2.6, 2.7, 2.8, 2.11). 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

$$\frac{1}{K} \sum_{k=1}^K (z_k, z_k^2) = (0, 1) \quad (2.12)$$

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

$$x_{ik}^* = \mu_{xi} + z_k \sigma_\epsilon \text{ and } b_k^* = \sigma_b * z_k. \quad (2.13)$$

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 (2.2,2.6,2.7,2.8,2.11) are:

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

$$\theta_i = K^{-1} \sum_{j=1}^K p_{ij}(\mathbf{x}_{ij}^*, b_i); \quad (2.14)$$

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

$$\hat{\theta}_i = \frac{1}{K} \sum_{j=1}^K \frac{\sum_{k=1}^K p_{ik}(\mathbf{x}_{ij}^*, b_k^*) \prod_{t=1}^{n_i} f(y_{it}|b_k^*)}{\sum_{k=1}^K \prod_{t=1}^{n_i} f(y_{it}|b_k^*)}, \quad (2.15)$$

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  $\sigma_\epsilon^2$  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  $\mathbf{y}$  using area sample CDF for  $\mathbf{x}$

$$\bar{p}_i = n_i^{-1} \sum_{j=1}^{n_i} \bar{p}_{ij} = n_i^{-1} \sum_{j=1}^{n_i} \frac{\sum_{k=1}^K p_{ik}(\mathbf{x}_{ij}, b_k^*) \prod_{t=1}^{n_i} f(y_{it}|b_k^*)}{\sum_{k=1}^K \prod_{t=1}^{n_i} f(y_{it}|b_k^*)}, \quad (2.16)$$

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  $\mathbf{y}$  using simple mean model for  $\mathbf{y}$

$$\hat{p}_i = \frac{\sum_{k=1}^K p_{\alpha,ik}(b_{2k}^*) \prod_{t=1}^{n_i} f(y_{it}|b_{2k}^*)}{\sum_{k=1}^K \prod_{t=1}^{n_i} f(y_{it}|b_{2k}^*)}, \quad (2.17)$$

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  $\mathbf{y}$  using predicted small area mean of  $\mathbf{x}$

$$\tilde{\theta}_i = \frac{1}{K} \sum_{j=1}^K \frac{\sum_{k=1}^K p_{ik}(\mathbf{x}_{ij}^*, b_k^*) \prod_{t=1}^{n_i} f(y_{it}|b_k^*)}{\sum_{k=1}^K \prod_{t=1}^{n_i} f(y_{it}|b_k^*)}, \quad (2.18)$$

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

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

### 2.3.2 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

$$p_{ij} = \frac{\exp(-0.8 + x_{ij} + b_i)}{1 + \exp(-0.8 + x_{ij} + b_i)}. \quad (2.19)$$

One thousand MC samples were generated satisfying the model. Let the estimation models be

- Model 1: Model (2.1)-(2.4), with known auxiliary mean  $\mu_{xi}$
- Model 2: Model (2.1), with unknown distribution for  $\mathbf{x}_{ij}$
- Model 3: Model (2.3), simple mean model for  $\mathbf{y}$
- Model 4: Model (2.1)-(2.4), with unknown random auxiliary mean  $\mu_{xi}$ .

We fit the estimation models (2.1) and (2.3) as generalized linear mixed models (GLMMs), with the binomial conditional distributions for the response. The model (2.4) for the covariate  $x_{ij}$  is fit as a linear mixed model (LMM).

The true small area mean of  $\mathbf{y}$  is given by (2.14) and the predicted area means of  $\mathbf{y}$  in the simulations are given in (2.15-2.18), with  $(\beta_0, \beta_1)$  and  $\sigma_b^2$  estimated using GLMM in R. The integrals were approximated with  $K = 50$ . The values  $x_{ik}^*$  in (2.15) are constructed using the known  $\mu_{xi}$  and the estimated  $\sigma_\epsilon^2$  defined for (2.15). Similarly, the values  $x_{ik}^*$  in (2.18) are constructed using the predicted  $\mu_{xi}$  and the estimated  $\sigma_\epsilon^{*2}$  defined for (2.18).

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 2.1 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  $\boldsymbol{\theta}$ .

Table 2.1: 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	$\hat{\theta} - \theta^1$	$\bar{p} - \theta^2$	$\hat{p} - \theta^3$	$\tilde{\theta} - \theta^4$	$\bar{y} - \theta$	$\hat{\mathbf{b}} - \mathbf{b}$	$\hat{\mathbf{b}}_2 - \mathbf{b}_2$
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  $\mu_{xi}$
2. Model 2, unknown distribution for  $\mathbf{x}_{ij}$
3. Model 3, simple mean model for  $\mathbf{y}$
4. Model 4, unknown random  $\mu_{xi}$

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 2.2. 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 (2.3).

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 2.2: 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	$\hat{\theta} - \theta^1$	$\bar{\mathbf{p}} - \theta^2$	$\hat{\mathbf{p}} - \theta^3$	$\tilde{\theta} - \theta^4$	$\bar{\mathbf{y}} - \theta$	$\hat{\mathbf{b}} - \mathbf{b}$	$\hat{\mathbf{b}}_2 - \mathbf{b}_2$
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  $\mu_{xi}$
2. Model 2, unknown distribution for  $\mathbf{x}_{ij}$
3. Model 3, simple mean model for  $\mathbf{y}$
4. Model 4, unknown random  $\mu_{xi}$

### 2.3.3 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 (2.14) to construct the predictor of  $\theta_i$ . Let the estimated coefficients for the fixed effects be  $\hat{\beta}, \hat{\alpha}$ , and let the predicted values for the random area effects be  $\hat{\mathbf{b}}, \hat{\mathbf{b}}_2$ , for models (2.1) and (2.3), 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

$$\tilde{\theta}_{i,plugin} = K^{-1} \sum_{j=1}^K \tilde{p}_{ij}(x_{ij}^*, \hat{b}_i), \text{ where } x_{ij}^* = \hat{\mu}_{xi} + z_j \hat{\sigma}_\epsilon^*. \quad (2.20)$$

### 2.3.4 Simulation Results, Plug-in Method for $b_i$

We use the simulation setup of Section 2.3.2 and construct predictions of  $\theta_i$  as defined in Section 2.3.3. Table 2.3 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 2.3 are significantly different from zero and arise because  $p_{ij}(x_{ij}, b_i)$  of (2.19) 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 2.3: MC BIAS of Prediction Error as Percent of the Standard Error of Prediction, ‘plug-in method’

n	$\hat{\theta}_{plugin} - \theta^1$	$\bar{\mathbf{p}}_{plugin} - \theta^2$	$\hat{\mathbf{p}}_{plugin} - \theta^3$	$\tilde{\theta}_{plugin} - \theta^4$
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  $\mu_{xi}$
2. Model 2, unknown distribution for  $\mathbf{x}_{ij}$
3. Model 3, simple mean model for  $\mathbf{y}$
4. Model 4, unknown random  $\mu_{xi}$

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 2.2. The procedure

using estimated conditional mean is less biased and slightly more efficient than the ‘plug-in method.’

Table 2.4: MC MSE (x1000) 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  $\mu_{xi}$
2. Model 2, unknown distribution for  $\mathbf{x}_{ij}$
3. Model 3, simple mean model for  $\mathbf{y}$
4. Model 4, unknown random  $\mu_{xi}$

## 2.4 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.

## Bibliography

- [1] 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.
- [2] Datta, G.S. and Ghosh, M. (1991), “Bayesian Prediction in Linear Models: Applications to Small Area Estimation,” *Annals of Statistics*, 19, 1748-1770.
- [3] 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.
- [4] Diao, L., Smith, D., Datta, G.S., Maiti, T., Opsomer, J. (2014), “Accurate Confidence Interval Estimation of Small Area Parameters Under the Fay-Herriot Model,” *Scandinavian Journal of Statistics*, 41, 497-515.
- [5] 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.
- [6] Ghosh, M. and Sinha, K. (2007), “Empirical Bayes estimation in finite population sampling under functional measurement error models,” *Journal of Statistical Planning and Inference*, 137, 2759-2773.
- [7] Ghosh, M., Sinha, K. and Kim, D. (2006), “Empirical and Hierarchical Bayesian Estimation in Finite Population Sampling under Structural Measurement Error Models,” *Scandinavian Journal of Statistics*, 33, 3, 591-608.

- [8] Giacomini, R., Politis, D. and White H. (2013), "A Warp-Speed Method for Conducting Monte Carlo Experiments Involving Bootstrap Estimators," *Econometric Theory*, 29, 567-589.
- [9] Hall P.(1986), "On the bootstrap and confidence intervals," *Annals of Statistics*, 14, 4, 1431-1452.
- [10] Hall P. and Maiti T. (2006), "On parametric bootstrap methods for small area prediction," *J.R. Statist. Soc. B*, 68, 2, 221-238.
- [11] Jiang, J. (2007), *Linear and Generalized Linear Mixed Models and their Applications*, Springer, New York.
- [12] Jiang, J. and Lahiri, P. (2001), "Empirical best prediction for small area inference with binary data," *Annals of the Institute of Statistical Mathematics*, 53, 2, 217-243..
- [13] Larsen, M. D. (2003), "Estimation of small-area proportions using covariates and survey data," *Journal of Statistical Planning and Inference*, 112, 89-98.
- [14] 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 Operacións Pontevedra*.
- [15] 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.
- [16] 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>
- [17] Rao J.N.K. (2003), "Small Area Estimation," *Wiley Series in Survey Methodology*.

## CHAPTER 3. SMALL AREA PREDICTION UNDER ALTERNATIVE MODEL SPECIFICATIONS

A modified paper submitted to *Statistics in Transition and Survey Methodology*

Andreea L. Erciulescu and Wayne A. Fuller

### Abstract

Construction of small area predictors and estimation of the prediction mean squared error, given different types of auxiliary information are illustrated for a unit level model. Of interest are situations where the mean and variance of an auxiliary variable are subject to estimation error. Fixed and random specifications for the auxiliary variables are considered. The efficiency gains associated with the random specification for the auxiliary variable measured with error are demonstrated. A parametric bootstrap procedure is proposed for the mean squared error of the predictor based on a logit model. The proposed procedure has smaller bootstrap error than a classical double bootstrap procedure with the same number of samples.

### 3.1 Introduction

Small area estimation can be more efficient than direct estimation in two ways. First, the assumption that the area-to-area differences are random permits the use of prediction (shrinkage estimators). Second, introducing auxiliary information through models provides the possibility for efficiency gains. Typical auxiliary information is in the form of observations on variables associated with the variable to be predicted. We are interested in situations in which the auxiliary variables are measured with error. The effect on estimation of measurement error in the auxiliary variables depends on the nature of the statistical model.

A number of papers consider sampling variability in the auxiliary variables. Ghosh, Sinha and Kim (2006) and Torabi, Datta and Rao (2009) consider an area level linear model with random auxiliary variable mean, estimated with error. The Ghosh, Sinha and Kim (2006) random mean model is a special case of the multivariate model studied by Fuller and Harter (1987). Ybarra and Lohr (2008) consider an area level linear model with fixed auxiliary mean estimated with error. Datta, Rao and Torabi (2010), following Ghosh and Sinha (2007), studied a nested error linear regression model with fixed area level covariate subject to measurement error.

We study unit level generalized linear mixed models under situations where the mean of an auxiliary variable is subject to estimation error. We propose a parametric bootstrap procedure for prediction mean squared error estimation and compare the proposed procedure to a classical double bootstrap procedure using a simulation study. Estimation with different types of auxiliary information is illustrated.

## 3.2 Unit Level Nonlinear Models

### 3.2.1 Introduction

Consider the unit level generalized linear mixed model

$$y_{ij} = g(\mathbf{x}_{ij}, \boldsymbol{\beta}, b_i) + e_{ij}, \quad (3.1)$$

$$\mathbf{x}_{ij} = \boldsymbol{\mu}_x + \boldsymbol{\delta}_i + \boldsymbol{\epsilon}_{ij} =: \boldsymbol{\mu}_{xi} + \boldsymbol{\epsilon}_{ij}, \quad (3.2)$$

$$\tilde{\mathbf{x}}_{ij'} = \boldsymbol{\mu}_{xi} + \boldsymbol{\epsilon}_{ij'}, \quad (3.3)$$

$i = 1, \dots, m$ , where  $m$  is the number of areas and  $j = 1, \dots, n_i$ , where  $n_i$  is the number of units within area  $i$ . The vector  $(y_{ij}, \mathbf{x}_{ij})$  is observed. In addition to  $\mathbf{x}_{ij}$ , a vector of auxiliary information,  $\tilde{\mathbf{x}}_{ij'}$ , may be available, where  $j' = 1, \dots, n'_i$ ,  $n'_i$  is the number of additional observations

in area  $i$ . The vector of random variables  $(b_i, \delta_i, e_{ij}, \epsilon_{ij})$  is unobserved, and  $\boldsymbol{\beta}$  is the vector of coefficients. Of interest is the small area mean of  $\mathbf{y}$

$$\theta_i = \int g(\mathbf{x}_{ij}, \boldsymbol{\beta}, b_i) dF_{\mathbf{x}_i}(\mathbf{x}), \quad (3.4)$$

where  $F_{\mathbf{x}_i}(\mathbf{x})$  is the distribution of  $\mathbf{x}$  in area  $i$ . Also of interest is the prediction mean squared error

$$\alpha_i = E(\hat{\theta}_i - \theta_i)^2, \quad (3.5)$$

where  $\hat{\theta}_i$  is the predictor. The nature of the estimation-prediction problem is determined by the distributional properties of the vector  $(b_i, \delta_i, e_{ij}, \epsilon_{ij})$ . The nonlinear model is more complicated than the linear model for several reasons. First, parameter estimation is more difficult because no closed form estimator exists. Likewise, closed form estimators of the mean squared error do not exist. Lastly, the small area mean of the auxiliary variable is not sufficient for the estimation of  $\theta_i$ .

As an example of model (3.1), consider a Bernoulli response variable  $\mathbf{y}$ , with realizations  $y_{ij}$  for  $m$  different areas and  $n_i$  different units within each area. To simplify the presentation, we consider scalar  $x_{ij}$  for the remainder of our discussion. Let  $x_{ij}$  be independent and identically distributed, following a distribution  $F_{\mathbf{x}_i}$ . Let  $b_i$  be independent and identically distributed, with a density  $f_b$  with mean 0 and variance  $\sigma_b^2$  and let  $\delta_i$  be independent and identically distributed, with a density  $f_\delta$  with mean 0 and variance  $\sigma_\delta^2$ . The mean of  $y$  given  $(\mathbf{x}_{ij}, b_i)$  is

$$g(\mathbf{x}_{ij}, \boldsymbol{\beta}, b_i) = \frac{\exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + b_i)}{1 + \exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + b_i)}, \quad (3.6)$$

where  $\mathbf{x}_{ij} = (1, x_{ij})$ ,  $\tilde{\mathbf{x}}_{ij} = (1, \tilde{x}_{ij})$  and  $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ . We assume that  $b_i \sim NI(0, \sigma_b^2)$  and that the elements of  $(b_i, \delta_i, e_{ij}, \epsilon_{ij})$  are mutually independent.

### 3.2.2 Predictors of $\theta_i$

We present predictors of  $\theta_i$  for model (3.6), under alternative specifications for  $\mathbf{x}_{ij}$  and for different levels of auxiliary information, given known parameters  $(\sigma_b^2, \sigma_\epsilon^2, \sigma_\delta^2, \boldsymbol{\beta}, \mu_x)$ .



### 3.2.2.1 Known Covariate Distribution

Let the distribution of  $x_{ij}$  be known. Then, given known parameters, the minimum mean squared error (MMSE) predictor of the small area mean of  $y$  is

$$\hat{\theta}_i = E \left[ \hat{\theta}(b) | (\mathbf{x}_i, \mathbf{y}_i) \right],$$

where  $\mathbf{x}_i = (x_{i,1}, x_{i,2}, \dots, x_{i,n_i})$ ,  $\mathbf{y}_i = (y_{i,1}, y_{i,2}, \dots, y_{i,n_i})$ , and

$$\hat{\theta}(b) = \int_{\mathbf{x}} g(x, \boldsymbol{\beta}, b) dF_{\mathbf{x}}(x)$$

and

$$\hat{\theta}_i = \frac{\int_b \hat{\theta}(b) \prod_{t=1}^{n_i} f(y_{it} | x_{it}, b_i) f(x_{it} | \mu_{xi}) dF_{b_i}(b)}{\int_b \prod_{t=1}^{n_i} f(y_{it} | x_{it}, b_i) f(x_{it} | \mu_{xi}) dF_{b_i}(b)}. \quad (3.7)$$

Since  $f(x_{it} | \mu_{xi})$  is free of  $b$ , the predictor given in (3.7) simplifies to

$$\hat{\theta}_i = \frac{\int_b \hat{\theta}(b) \prod_{t=1}^{n_i} f(y_{it} | x_{it}, b_i) dF_{b_i}(b)}{\int_b \prod_{t=1}^{n_i} f(y_{it} | x_{it}, b_i) dF_{b_i}(b)}. \quad (3.8)$$

In some finite population situations, the entire finite population of  $x$  values may be known and the integral expression for  $\hat{\theta}(b)$  in (3.8) is the sum over the population. In the simulations of Section 3.4 for this model we assume  $x_{ij} \sim NI(\mu_{xi}, \sigma_c^2)$  with  $\mu_{xi}$  known and  $\sigma_c^2$  known.

### 3.2.2.2 Unknown, Unspecified Covariate Distribution

If the distribution of  $x$  is unknown, an estimate of the distribution of  $x$  at point  $c$  is given by the sample cumulative distribution function (CDF),  $\sum_{j=1}^{n_i} w_{ij} I(x_{ij}, c)$ , where  $I(x_{ij}, c)$  is the indicator function and  $w_{ij}$  are sampling weights. Then, given known  $(\sigma_b^2, \boldsymbol{\beta})$ , the predictor of the small area mean of  $y$  is

$$\hat{\theta}_i = E \left[ \hat{\theta}_i(b) | (\mathbf{x}_i, \mathbf{y}_i) \right],$$

where

$$\hat{\theta}_i(b) = \sum_{j=1}^{n_i} w_{ij} g(x_{ij}, \boldsymbol{\beta}, b_i)$$

and

$$\hat{\theta}_i = \frac{\int_b \hat{\theta}_i(b) \prod_{t=1}^{n_i} f(y_{it} | x_{it}, b_i) dF_{b_i}(b)}{\int_b \prod_{t=1}^{n_i} f(y_{it} | x_{it}, b_i) dF_{b_i}(b)}. \quad (3.9)$$

See Ghosh et al (2009) for an example using the sample CDF.

### 3.2.2.3 Unknown, Unspecified Covariate Distribution, Additional Information

$\tilde{\mathbf{x}}_i$

Let a vector of additional information  $\tilde{\mathbf{x}}_i$  be available, where it is assumed  $\tilde{\mathbf{x}}_i$  and  $\mathbf{x}_i$  are probability samples from the population of area  $i$ . Let  $w_{ij}$  be the sampling weights for the combined sample. As in Section 3.2.2.2, an estimate of the distribution of  $x$  at point  $c$  is given by the sample cumulative distribution function (CDF),  $\sum_{j=1}^{n_i+n'_i} w_{ij} I(x_{ij}, c)$ . Then, given known  $(\sigma_b^2, \boldsymbol{\beta})$ , the predictor of the small area mean of  $y$  is

$$\hat{\theta}_i = E \left[ \hat{\theta}_i(b) | (\mathbf{x}_i, \mathbf{y}_i, \tilde{\mathbf{x}}_i) \right],$$

where

$$\hat{\theta}_i(b) = \sum_{j=1}^{n_i+n'_i} w_{ij} g(x_{ij}, \boldsymbol{\beta}, b_i)$$

and

$$\hat{\theta}_i = \frac{\int_b \hat{\theta}_i(b) \prod_{t=1}^{n_i} f(y_{it} | x_{it}, b_i) dF_{b_i}(b)}{\int_b \prod_{t=1}^{n_i} f(y_{it} | x_{it}, b_i) dF_{b_i}(b)}. \quad (3.10)$$

### 3.2.2.4 Unknown Random Covariate Mean

Assume the distribution of  $\mathbf{x}$  for area  $i$  is  $F_{\mathbf{x}_i}$ , with unknown parameters  $(\mu_{xi}, \sigma_\epsilon^2)$ . Assume  $\mu_{xi} \sim NI(\mu_x, \sigma_\delta^2)$ . Then, given known  $(\sigma_b^2, \sigma_\epsilon^2, \sigma_\delta^2, \boldsymbol{\beta}, \mu_x)$ , the MMSE predictor of the small area mean of  $y$  is

$$\hat{\theta}_i = E \left[ \hat{\theta}(b, \delta) | (\mathbf{x}_i, \mathbf{y}_i) \right],$$

where

$$\hat{\theta}(b, \delta) = \int g(\mu_x + \delta + \epsilon, \boldsymbol{\beta}, b) dF_\epsilon(\epsilon)$$

and

$$\hat{\theta}_i = \frac{\int_b \int_\delta \hat{\theta}(b, \delta) \prod_{t=1}^{n_i} f(y_{it} | x_{it}, b_i) f(x_{it} | \delta_i) dF_{\delta_i}(\delta) dF_{b_i}(b)}{\int_b \int_\delta \prod_{t=1}^{n_i} f(y_{it} | x_{it}, b_i) f(x_{it} | \delta_i) dF_{\delta_i}(\delta) dF_{b_i}(b)}. \quad (3.11)$$

In the simulations of Section 3.4 we assume  $x_{ij} \sim NI(\mu_{xi}, \sigma_\epsilon^2)$ .

### 3.2.2.5 Unknown Random Covariate Mean, Additional Information $\tilde{\mathbf{x}}_i$

Let the assumptions of Section 3.2.2.4 hold. Let  $\tilde{\mathbf{x}}_i$  information as described in Section 3.2.2.3 be available. Then, given known  $(\sigma_b^2, \sigma_\epsilon^2, \sigma_\delta^2, \boldsymbol{\beta}, \mu_x)$ , the MMSE predictor of the small

area mean of  $y$  is

$$\hat{\theta}_i = E \left[ \hat{\theta}(b, \delta) | (\mathbf{x}_i, \mathbf{y}_i, \tilde{\mathbf{x}}_i) \right],$$

where

$$\begin{aligned} \hat{\theta}(b, \delta) &= \int g(\mu_x + \delta + \epsilon, \boldsymbol{\beta}, b) dF_{\boldsymbol{\epsilon}}(\epsilon), \\ \hat{\theta}_i &= \frac{\int_b \int_{\delta} \hat{\theta}(b, \delta) \prod_{t=1}^{n_i} f(y_{it} | x_{it}, b_i) f(x_{it} | \delta_i) \prod_{t'=1}^{n'_i} f(\tilde{x}_{it'} | \delta_i) dF_{\delta_i}(\delta) dF_{b_i}(b)}{\int_b \int_{\delta} \prod_{t=1}^{n_i} f(y_{it} | x_{it}, b_i) f(x_{it} | \delta_i) \prod_{t'=1}^{n'_i} f(\tilde{x}_{it'} | \delta_i) dF_{\delta_i}(\delta) dF_{b_i}(b)}. \end{aligned}$$

In the simulations of Section 3.4 we assume  $\tilde{x}_{ij'} \sim NI(\mu_{xi}, \sigma_{\epsilon}^2)$ , so  $\tilde{\mu}_{xi} = (n'_i)^{-1} \sum_{j'=1}^{n'_i} \tilde{x}_{ij'}$  is a sufficient statistic for  $\mu_{xi}$  and the predictor simplifies to

$$\hat{\theta}_i = \frac{\int_b \int_{\delta} \hat{\theta}(b, \delta) \prod_{t=1}^{n_i} f(y_{it} | x_{it}, b_i) f(x_{it} | \delta_i) f(\tilde{\mu}_{xi} | \delta_i) dF_{\delta_i}(\delta) dF_{b_i}(b)}{\int_b \int_{\delta} \prod_{t=1}^{n_i} f(y_{it} | x_{it}, b_i) f(x_{it} | \delta_i) f(\tilde{\mu}_{xi} | \delta_i) dF_{\delta_i}(\delta) dF_{b_i}(b)}. \quad (3.12)$$

### 3.2.3 Estimation

In practice, the vector of parameters  $\psi = (\sigma_b^2, \sigma_{\epsilon}^2, \sigma_{\delta}^2, \boldsymbol{\beta}, \mu_x)$  is not known and needs to be estimated. Consider the model specified by (3.1), (3.2), (3.3), (3.6) and described in Section 3.2.2.5. The likelihood is

$$L(\sigma_b^2, \sigma_{\epsilon}^2, \sigma_{\delta}^2, \boldsymbol{\beta}, \mu_x | \mathbf{y}, \mathbf{x}, \tilde{\mathbf{x}}) = \prod_{i=1}^m L_i,$$

where

$$\begin{aligned} L_i &= \int_{b_i} \int_{\delta_i} \prod_{j=1, j'=1}^{n_i, n'_i} f(y_{ij}, x_{ij}, \tilde{x}_{ij'} | b_i, \delta_i, \psi) f(b_i | \psi) f(\delta_i | \psi) d\delta_i db_i \\ &= \int_{b_i} \prod_{j=1}^{n_i} f(y_{ij} | b_i, x_{ij}, \boldsymbol{\beta}) f(b_i | \sigma_b^2) db_i \int_{\delta_i} \prod_{j=1}^{n_i + n'_i} f(x_{ij}^* | \delta_i, \mu_x, \sigma_{\epsilon}^2) f(\delta_i | \sigma_{\delta}^2) d\delta_i, \end{aligned}$$

and  $\mathbf{x}^* = (\mathbf{x}, \tilde{\mathbf{x}})$  is the vector of all available auxiliary information.

Notice that the likelihood  $L(\sigma_b^2, \sigma_{\epsilon}^2, \sigma_{\delta}^2, \boldsymbol{\beta}, \mu_x | \mathbf{y}, \mathbf{x}, \tilde{\mathbf{x}})$  factors into  $L(\sigma_b^2, \boldsymbol{\beta} | \mathbf{y})$  and  $L(\sigma_{\epsilon}^2, \sigma_{\delta}^2, \mu_x | \mathbf{x}, \tilde{\mathbf{x}})$ . Hence, the parameters  $(\sigma_{\epsilon}^2, \sigma_{\delta}^2, \mu_x)$  can be estimated separately from the parameters  $(\sigma_b^2, \boldsymbol{\beta})$ . The estimation of  $(\sigma_{\epsilon}^2, \sigma_{\delta}^2, \mu_x)$  is based on maximizing the likelihood for the linear mixed model specified in (3.2) and (3.3). In the simulation study, we construct the estimated generalized least squares estimator for  $\mu_x$  and REML estimates for  $\sigma_{\epsilon}^2$  and  $\sigma_{\delta}^2$ , using the *lmer* function, in R.

The estimation of  $(\sigma_b^2, \boldsymbol{\beta})$  is based on maximizing the likelihood for the generalized linear mixed model specified in (3.1) and (3.6). There is no closed form expression for the likelihood and no exact form of the estimating equations. Two commonly used methods of estimation are based on pseudo-likelihood and on integral approximation. In the simulation study, we use the integral approximation method in the *glmer* function, in R, that uses the Laplace approximation to the likelihood.

### 3.2.4 Bootstrap MSE Estimation

In this section we consider estimation of the MSE of  $\hat{\theta}_i$  as a predictor of  $\theta_i$ . Let  $\boldsymbol{\psi}$  be the parameter that defines the distribution of the sample observations, and let  $\hat{\boldsymbol{\psi}}$  be an estimator of  $\boldsymbol{\psi}$ . Let  $\boldsymbol{\alpha}$  be a vector of parameters of interest and let  $\boldsymbol{\alpha}^*$  be a parametric bootstrap (simulation) estimator of  $\boldsymbol{\alpha}$ . For the models considered in Section 3.2.2, let  $\alpha_i$  be the MSE of the prediction error for area  $i$ , as defined in (3.5). For the nonlinear small area model with known distribution for  $x_{ij}$ , the vector of parameters is  $\boldsymbol{\psi} = (\sigma_b^2, \boldsymbol{\beta})$ . For the nonlinear small area models with unknown random  $\mu_{xi}$ , the vector of parameters is  $\boldsymbol{\psi} = (\sigma_b^2, \boldsymbol{\beta}, \sigma_\epsilon^2, \mu_x, \sigma_\delta^2)$ . Because there is no closed form expression for the prediction MSE given in (3.5), we consider bootstrap MSE estimation. Hall and Maiti (2006) constructed nonnegative, bias-corrected MSE estimates using a double bootstrap procedure. They considered area level models and a unit level binomial model with fixed known covariates. Pfeffermann and Correa (2012) suggested a bootstrap procedure in which the bias in the estimator is estimated as a function of parameters and of a bootstrap estimator of bias.

A sample generated with  $\boldsymbol{\psi}$  and random number seed  $r$  is said to be created with data generator  $(\boldsymbol{\psi}, r)$ , denoted  $DG(\boldsymbol{\psi}, r)$ . Let  $B_1$  bootstrap samples be generated using random number seeds  $r_{1,1}, r_{1,2}, \dots, r_{1,B_1}$ . Let  $\boldsymbol{\psi}_k^*$  be the estimator of  $\boldsymbol{\psi}$  from the  $k$ th bootstrap sample

generated using  $DG(\hat{\boldsymbol{\psi}}, r_{1,k})$ . The bootstrap estimator of prediction MSE for area  $i$  is

$$\hat{\alpha}_i^* = B_1^{-1} \sum_{k=1}^{B_1} (\hat{\theta}_{i,k}^* - \theta_{i,k}^*)^2 =: B_1^{-1} \sum_{k=1}^{B_1} \alpha_{i,k}^* = \bar{\alpha}_i^*, \quad (3.13)$$

where  $\theta_{i,k}^*$  is the true small area mean generated for the  $k$ th bootstrap sample,  $\hat{\theta}_{i,k}^*$  is the sample predictor of  $\theta_{i,k}^*$  and  $\alpha_{i,k}^*$  is the prediction squared error for the  $k$ th bootstrap sample. The estimator (3.13) is called the level-one bootstrap estimator.

In the double bootstrap, a sample estimator, denoted by  $\alpha_i^{**}$ , is generated using  $\boldsymbol{\psi}^*$  from the level-one generated sample. Typically a large number of  $\alpha_i^{**}$  is generated for each  $\alpha_i^*$  and the bias adjusted estimator is

$$\tilde{\alpha}_i^{**} = B_1^{-1} \sum_{k=1}^{B_1} 2\alpha_{i,k}^* - B_1^{-1} B_2^{-1} \sum_{k=1}^{B_1} \sum_{t=1}^{B_2} \alpha_{i,k,t}^{**}. \quad (3.14)$$

where  $\alpha_{i,k,t}^{**}$  is generated using  $DG(\boldsymbol{\psi}_k^*, r_{2,k,t})$ ,  $B_1$  is the number of level-one bootstrap samples,  $B_2$  is the number of level-two bootstrap samples per level-one sample, and the  $r_{2,k,t}$ ,  $k = 1, 2, \dots, B_1$ ,  $t = 1, 2, \dots, B_2$ , are independent random numbers, independent of  $r_{1,k}$ .

We use a double bootstrap estimator based on the work of Davidson and MacKinnon (2007) who give a fast double bootstrap procedure for bootstrap testing. See also Giacomini, Politis and White (2013). In the fast double bootstrap, a single  $\alpha_i^{**}$  is generated for each  $\alpha_i^*$ . Let  $r_{2,1}, r_{2,2}, \dots, r_{2,B_1}$  be a second independent sequence of random numbers. Given the sequence of random numbers, define  $\alpha_{i,k}^{**}$  to be calculated from data generated with  $DG(\boldsymbol{\psi}_k^*, r_{2,k})$ . The (classic) double bootstrap estimator used in this study is

$$\tilde{\alpha}_{i,C}^{**} = B_1^{-1} \sum_{k=1}^{B_1} (2\alpha_{i,k}^* - \alpha_{i,k}^{**}) = 2\bar{\alpha}_i^* - \bar{\alpha}_i^{**}. \quad (3.15)$$

To construct an even more efficient bootstrap estimator, define  $\alpha_{i,k,2}^*$  to be calculated from data generated with  $DG(\hat{\boldsymbol{\psi}}, r_{2,k})$ . Then a bias adjusted (double bootstrap) estimator is

$$\hat{\alpha}_i^{**} = B_1^{-1} \sum_{k=1}^{B_1} (\alpha_{i,k}^* + \alpha_{i,k,2}^* - \alpha_{i,k}^{**}), \quad (3.16)$$

where the quantity  $\alpha_{i,k}^{**} - \alpha_{i,k}^*$  is a one-degree-of-freedom estimator of the bias. If one uses  $r_{2,1}$  as  $r_{1,2}$ ,  $r_{2,2}$  as  $r_{1,3}$ , etc., a form of (3.16) becomes

$$\tilde{\alpha}_{i,T}^{**} = B_1^{-1} \sum_{k=1}^{B_1} (\alpha_{i,k}^* + \alpha_{i,k+1}^* - \alpha_{i,k}^{**}), \quad (3.17)$$

where  $\alpha_{i,k+1}^*$  is generated with  $DG(\hat{\psi}, r_{1,k+1})$  and  $\alpha_{i,k}^{**}$  is generated with  $DG(\psi_k^*, r_{1,k+1})$ . We call the estimator (3.17) a telescoping bootstrap because it is of the form (3.16) using lagged values of  $\alpha_{i,k}^*$ . If the use of  $r_{2,k}$  in place of an independent random number results in positive correlation between  $\alpha_{i,k}^*$  and  $\alpha_{i,k-1}^{**}$ , then  $\tilde{\alpha}_{i,T}^{**}$  will have smaller simulation variance than  $\tilde{\alpha}_{i,C}^{**}$  of (3.15).

### 3.3 Simulations

In the simulation study we consider  $m = 36$  areas with unit level observations  $x_{ij}$  in three groups of 12 areas, with sizes  $n_i \in \{2, 10, 40\}$ . The number of additional unit level observations is  $n'_i = 10$ , for each area  $i$ . Each sample,  $(\mathbf{y}, \mathbf{x}, \tilde{\mathbf{x}})$ , is generated using model (3.1 - 3.3) with  $\sigma_b^2 = 0.25$ ,  $\mu_x = 0$ ,  $\sigma_\delta^2 = 0.16$ , and  $\sigma_\epsilon^2 = 0.36$ . 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

$$g(x_{ij}, \boldsymbol{\beta}, b_i) = \frac{\exp(-0.8 + x_{ij} + b_i)}{1 + \exp(-0.8 + x_{ij} + b_i)}. \quad (3.18)$$

The population mean of  $g(x_{ij}, \boldsymbol{\beta}, b_i)$  is 0.334 with variance 0.029. An area with  $\mu_{x_i} = 0.4$  has mean 0.412 with variance 0.028. Four hundred Monte Carlo samples were generated satisfying the model.

The estimation models are:

- Model 1: Specified by (3.1) and (3.6) and described in Section 3.2.2.1. Known normal distribution for  $x_{ij}$ . The distribution of  $y_{it}$  is

$$f(y_{it}|x_{it}, b_i) = I(y_{it}, 1)g(x_{it}, \boldsymbol{\beta}, b_i) + I(y_{it}, 0)(1 - g(x_{it}, \boldsymbol{\beta}, b_i)),$$

where  $I(y_{it}, \cdot)$  is the indicator function defined in Section 3.2.2.2, and  $g(x_{it}, \boldsymbol{\beta}, b_i)$  is defined in (3.18). The distribution of  $b$  is  $N(0, 0.25)$ .

- Model 2: Specified by (3.1) and (3.6) and described in Section 3.2.2.2. Unknown, unspecified distribution of  $\mathbf{x}$ .
- Model 2\*: Specified by (3.1) and (3.6) and described in Section 3.2.2.3. Unknown, unspecified distribution of  $\mathbf{x}$ , observed  $\tilde{\mathbf{x}} = (\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_m)$ .
- Model 3: Specified by (3.1), (3.2), (3.6) and described in Section 3.2.2.4. Unknown random auxiliary mean  $\mu_{xi}$ . Distributions of  $\mathbf{y}$  and  $b$  are same as those for Model 1. The distribution of  $\mathbf{x}$  is given in Section 3.2.2.4.
- Model 4: Specified by (3.1), (3.2), (3.3), (3.6) and described in Section 3.2.2.5. Unknown random auxiliary mean  $\mu_{xi}$ , observed  $\tilde{\mathbf{x}} = (\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_m)$ .

The models are fitted as generalized linear mixed models (GLMMs), using the *lmer*, *glmer* functions in the *lme4* package in R. The true small area mean of  $\mathbf{y}$  is given by (3.4) and the predicted area means of  $\mathbf{y}$  are given in (3.8 - 3.12), with estimated  $(\mu_x, \beta_0, \beta_1, \sigma_b^2, \sigma_\delta^2, \sigma_\epsilon^2)$ . The integrals in (3.4, 3.8 - 3.12) were approximated using a 26-point approximation to the normal distribution.

### 3.3.1 Refinement of Prediction MSE Estimators

We bound the estimator of  $\sigma_\delta^2$  by

$$K_{\delta,s} = 0.5 \left[ \hat{V}(\hat{\sigma}_\delta^2 | \sigma_\delta^2 = 0) \right]^{0.5},$$

where  $\hat{V}(\hat{\sigma}_\delta^2 | \sigma_\delta^2 = 0)$  is the estimated variance of  $\hat{\sigma}_\delta^2$ , given  $\sigma_\delta^2 = 0$ . The bound was suggested in Wang and Fuller (2003). Because of the large degrees of freedom for  $\hat{\sigma}_\epsilon^2$ , we set  $K_{\delta,s}$  equal to the true value of 0.008 in the simulations,

$$K_{\delta,s} = 0.5 \left[ 2m(m-1)^{-1} \left( \sum_{i=1}^m ((n_i + n'_i)^{-1} \sigma_\epsilon^2)^{-2} \right)^{-1} \right]^{0.5} = 0.008.$$

Similarly, we bound the estimator of  $\sigma_b^2$  by

$$K_{b,s} = 0.5 \left[ V(\hat{\sigma}_b^2 | \sigma_b^2 = 0) \right]^{0.5} = 0.006.$$

The proportion of sample estimators  $\hat{\sigma}_b^2$  that hit the bound is 0.025, the proportion of level one estimators of  $\hat{\sigma}_b^{2*}$  that hit the bound is 0.111. If  $\hat{\sigma}_{b,k}^2 = 0.006$  we set  $\alpha_{i,k}^{**}$  equal to  $\alpha_{i,k}^*$ . That is, the estimated bias is zero for such samples.

Using (3.15), one can obtain an unacceptable double bootstrap prediction MSE estimator, where the estimated bias for a sample is greater than the estimate. In practice, one would increase the number of bootstrap samples. Rather than build such a procedure into our Monte Carlo algorithm, we defined bounds for the estimator.

Thus, the final estimator is

$$\hat{\alpha}_{i,C}^{**} = \begin{cases} 1.60\bar{\alpha}_i^*, & \text{if } \bar{\alpha}_i^{*-1}\bar{\alpha}_i^{**} > 1.60 \\ 0.83\bar{\alpha}_i^*, & \text{if } \bar{\alpha}_i^{*-1}\bar{\alpha}_i^{**} < 0.83 \\ \tilde{\alpha}_{i,C}^{**}, & \text{otherwise,} \end{cases} \quad (3.19)$$

where 0.83 and 1.60 are the 0.025 and 0.975 points of the chi-square distribution with 199 ( $B_1 - 1$ ) degrees of freedom, and  $\tilde{\alpha}_{i,C}^{**}$  is defined in (3.15). The analogous definition holds for the telescoping estimator of (3.16). See Hall and Maiti (2006) for an alternative definition of the direct double bootstrap estimates.

The proportions of sample estimators of  $\hat{\alpha}_{i,T}^{**}$  that hit the lower bound defined in (3.19) are 0.016, 0.016 and 0.013, for the areas of sizes 2, 10 and 40, respectively. The proportions of sample estimators of  $\hat{\alpha}_{i,T}^{**}$  that hit the upper bound defined in (3.19) are 0.026, 0.069 and 0.084, for the areas of sizes 2, 10 and 40, respectively. Due to larger variability in the classic double bootstrap estimators, the proportions of sample estimators of  $\hat{\alpha}_{i,C}^{**}$  that hit the lower bound defined in (3.19) are 0.058, 0.048 and 0.041, for the areas of sizes 2, 10 and 40, respectively, and the proportions of sample estimators of  $\hat{\alpha}_{i,C}^{**}$  that hit the upper bound defined in (3.19) are 0.155, 0.201 and 0.183, for the areas of sizes 2, 10 and 40, respectively.



### 3.3.2 MSE for Different Types of Auxiliary Information

The coefficient of variation for  $\hat{\sigma}_b^2$  calculated for the 400 Monte Carlo samples is about 0.64, approximately the CV of a Chi-square with five degrees of freedom. The Monte Carlo relative bias of the estimator of  $\hat{\sigma}_b^2$  is about  $-0.12$ , which is approximately equal to eighteen Monte Carlo standard errors.

Table 3.1 contains estimates of  $\alpha = MSE$  for fixed and random models with different amounts of auxiliary information. The simulation MSE standard errors are presented in parentheses below the MSE values. The smallest MSE is for Model 1, where the covariate distribution is known. The next smallest MSE is for Model 4, where the form of the covariate distribution is known, the covariate mean is random and the auxiliary information is available. The largest MSE is for Model 2, where the covariate distribution is not specified. The small area mean predictor for Model 3 is the conditional expected value formula given in (3.11). Notice that in the construction of the small area predictor for Model 4, given in (3.12), the conditioning is also on the additional source of information,  $\tilde{\mathbf{x}}$ , available for the areas. By including the ten additional unit level observations, the estimated MSE is closer to the MSE of the known distribution case than to the MSE for the case with no additional information.

The extra observations on  $x_{ij}$  represent additional information available about the distribution of  $\mathbf{x}$  for the area. Hence, the large gain in efficiency associated with  $\tilde{\mathbf{x}}$  for sample size two (compare 10.94 for Model 2\* to 17.29 for Model 2). Model 3 differs from Model 2 in that the distribution of  $x_{ij}$  is assumed to be normal and the area mean is also assumed to be normally distributed. Adding these distributional assumptions changes the MSE from 17.29 to 13.22 for sample size two. The effect of added information is much smaller for the random  $\mu_{xi}$  models (Models 2\* and 4) than for the fixed  $\mu_{xi}$  models (Models 2 and 3).

Table 3.1: MSE for Different Types Auxiliary Information (Entries Multiplied by  $10^3$ )

Size	$\bar{y}$	Model 1	Model 2	Model 2*A	Model 3	Model 4
2	102.14	9.88	17.29	10.94	13.22	10.72
	(6.13)	(0.71)	(1.24)	(0.79)	(0.92)	(0.76)
10	20.15	7.15	8.56	7.87	8.26	7.76
	(1.40)	(0.52)	(0.63)	(0.57)	(0.60)	(0.56)
40	5.14	3.46	3.81	3.74	3.78	3.72
	(0.37)	(0.25)	(0.27)	(0.27)	(0.27)	(0.27)

Model 1: known distribution for  $x_{ij}$ , Model 2: unknown distribution for  $x_{ij}$ , with no  $\tilde{\mathbf{x}}$ , Model 2\*A: unknown distribution for  $x_{ij}$ , with observed  $\tilde{\mathbf{x}}$ , Model 3: random  $\mu_{xi}$ , with no  $\tilde{\mathbf{x}}$ , Model 4: random  $\mu_{xi}$ , with observed  $\tilde{\mathbf{x}}$

### 3.3.3 Monte Carlo Properties of Prediction MSE Estimators

The relative performances of bootstrap prediction MSE estimators under the different types of auxiliary information are similar. Therefore, we only present properties of prediction MSE estimators for Model 4, where the area mean  $\mu_{xi}$  is random and auxiliary information  $\tilde{\mathbf{x}}$  is available.

Table 3.2 contains results for  $(\hat{\alpha}^*, \hat{\alpha}_T^{**}, \hat{\alpha}_C^{**})$  for the three area sample sizes, in groups of five lines. Each line is the average of the results for the 12 areas with the same sample size. The first line is the Monte Carlo estimates of the prediction MSE,  $\hat{\alpha}$ . The next four lines are of the bias relative to the mean, the coefficient of variation, the bias relative to the standard deviation and the bias relative to the standard error. The definitions are

$$\begin{aligned}
RelBias &= \sum_{is=1}^{12} (\hat{\alpha}_{.,is}^{EST} - \hat{\alpha}_{.,is}) / \sum_{is=1}^{12} \hat{\alpha}_{.,is}, \\
CV &= \sum_{is=1}^{12} \sqrt{(400-1)^{-1} \sum_{\zeta=1}^{400} (\hat{\alpha}_{\zeta,is}^{EST} - \hat{\alpha}_{.,is}^{EST})^2} / \sum_{is=1}^{12} \hat{\alpha}_{.,is}, \\
Bias/sd &= \sum_{is=1}^{12} (\hat{\alpha}_{.,is}^{EST} - \hat{\alpha}_{.,is}) / \sum_{is=1}^{12} \sqrt{(400-1)^{-1} \sum_{\zeta=1}^{400} (\hat{\alpha}_{\zeta,is}^{EST} - \hat{\alpha}_{.,is}^{EST})^2}, \\
Bias/se &= Bias/(20sd),
\end{aligned}$$

where  $\zeta$  indexes the Monte Carlo samples,  $i$  denotes an area from a group of areas of sample size  $s$ ,  $\hat{\alpha}_{.,is} = (400)^{-1} \sum_{\zeta=1}^{400} \hat{\alpha}_{\zeta,is}$  is the average of the Monte Carlo prediction error estimators,  $\hat{\alpha}_{.,is}^{EST} = (400)^{-1} \sum_{\zeta=1}^{400} \hat{\alpha}_{\zeta,is}^{EST}$  is the average of the bootstrap prediction MSE estimators, and  $\hat{\alpha}^{EST} \in \{\hat{\alpha}^*, \hat{\alpha}_T^{**}, \hat{\alpha}_C^{**}\}$  is the bootstrap estimator for an area.

Table 3.2: Monte Carlo Properties of Prediction MSE Estimators  
 ( $B_1 = 200, B_2 = 1$  and 400 MC Samples, Variances Multiplied by  $10^3$ )

size		$\hat{\alpha}^*$	$\hat{\alpha}_T^{**}$	$\hat{\alpha}_C^{**}$
2	$V(\hat{\theta} - \theta)$	10.723	10.723	10.723
	RelBias	-0.143	-0.058	-0.062
	$CV(\hat{\alpha})$	0.403	0.456	0.477
	Bias/sd	-0.355	-0.127	-0.130
	Bias/se	-7.097	-2.537	-2.609
10	$V(\hat{\theta} - \theta)$	7.758	7.758	7.758
	RelBias	-0.133	-0.032	-0.039
	$CV(\hat{\alpha})$	0.318	0.365	0.385
	Bias/sd	-0.417	-0.087	-0.102
	Bias/se	-8.336	-1.738	-2.034
40	$V(\hat{\theta} - \theta)$	3.721	3.721	3.721
	RelBias	-0.082	0.016	0.009
	$CV(\hat{\alpha})$	0.222	0.260	0.286
	Bias/sd	-0.372	0.062	0.032
	Bias/se	-7.430	1.249	0.636

The estimated prediction MSEs have CV's of about 40%, 32% and 22% for 200 bootstrap samples for sample sizes 2, 10, and 40, respectively. In all cases the telescoping double bootstrap, denoted with a subscript T, has lower MSE than the classic double bootstrap, denoted with a subscript C. The estimators  $\hat{\alpha}_T^{**}$  and  $\hat{\alpha}_C^{**}$  have the same bias if the bound (3.19) is not used. The double bootstrap reduces the absolute value of the bias for all the sample sizes. However, the absolute bias of the double bootstrap is about 6% of the true value for sample size 2.

### 3.3.4 Efficiency of Bootstrap Procedures

The variance of the bootstrap estimator of  $\theta$  has two components. The first, that we call *between*, is the variance one would obtain if one used an infinite number of bootstrap samples. The second, that we call *within*, is the variability due to the fact that our set of bootstrap

samples is a sample of samples. In this section, we estimate the *between* and the *within* variance components of the variance of a bootstrap estimator of the prediction MSE. Consider the parametric bootstrap method, with the difference bias correction method. The performance of the classic double bootstrap, the classic fast double bootstrap and telescoping fast double bootstrap methods are compared.

We estimate the components for Model 4, where the area mean  $\mu_{xi}$  is random and auxiliary information  $\tilde{\mathbf{x}}$  is available. For each Monte Carlo sample, two independent sets of  $B_1 = 100$  level one bootstrap samples are generated using the estimated vector of parameters  $\hat{\psi}$  based on the Monte Carlo sample and two independent sequences of random seeds  $r_{1,k}$ ,  $k = 1, \dots, B_1$ , for the first set, and  $r'_{1,k}$ ,  $k = 1, \dots, B_1$ , for the second set. For each level one bootstrap sample, three independent sets of  $B_2$  double bootstrap samples are generated using the estimated vector of parameters  $\psi^*$  based on the level one bootstrap sample:

- For the classic double bootstrap,  $B_2 = 2$  and the two sequences of random seeds  $r_{21,k}$  and  $r_{22,k}$ ,  $k = 1, \dots, B_1$ , for the first set are independent of the two sequences of random seeds  $r'_{21,k}$  and  $r'_{22,k}$ ,  $k = 1, \dots, B_1$ , for the second set.
- For the classic fast double bootstrap,  $B_2 = 1$  and the sequence of random seeds  $r_{2,k}$ ,  $k = 1, \dots, B_1$ , for the first set is independent of the sequence of random seeds  $r'_{2,k}$ ,  $k = 1, \dots, B_1$ , for the second set.
- For the telescoping fast double bootstrap,  $B_2 = 1$  and the sequences of random seeds are  $r_{1,k}$  and  $r'_{1,k}$ ,  $k = 2, \dots, B_1$ , used in the level one bootstrap.

The sets of random seeds  $r_{1,k}, r'_{1,k}, r_{2,k}, r'_{2,k}, r_{21,k}, r'_{21,k}, r_{22,k}, r'_{22,k}$  are independent. Define

$$V_w(\alpha^{EST}) = E(V(\alpha^{EST}|\psi^{EST})),$$

$$V_b(\alpha^{EST}) = V(E(\alpha^{EST}|\psi^{EST})),$$

$$V_T(\alpha^{EST}) = V_w(\alpha^{EST}) + V_b(\alpha^{EST}),$$

where  $\alpha^{EST} \in \{\hat{\alpha}^*, \hat{\alpha}_T^{**}, \hat{\alpha}_C^{**}\}$  and  $\psi^{EST} \in \{\hat{\psi}, \psi^*\}$ .

Let  $\zeta$  index the Monte Carlo samples and let  $i$  denote an area within a group of areas of sample size  $s$ . Let  $(\alpha_{\zeta,k,is}^*, \alpha_{T,\zeta,k,is}^{**}, \alpha_{C,\zeta,k,is}^{**})$  be the level one bootstrap and fast double bootstrap prediction MSE values for the first set of bootstrap samples and let  $(\alpha_{\zeta,k,is}^{*2}, \alpha_{T,\zeta,k,is}^{**2}, \alpha_{C,\zeta,k,is}^{**2})$  be the level one bootstrap and fast double bootstrap prediction MSE values for the second set of bootstrap samples. The estimated total variance of a bootstrap prediction MSE estimator, averaged over the areas of sample size  $s$ , is

$$\hat{V}_T^{EST} = (400)(12)^{-1} \sum_{is=1}^{12} \left( (399)^{-1} \sum_{\zeta=1}^{400} (\bar{\alpha}_{\zeta,;is}^{EST} - \bar{\bar{\alpha}}_{;is}^{EST})^2 \right), \quad (3.20)$$

where  $\alpha^{EST} \in \{\alpha^*, \alpha_T^{**}, \alpha_C^{**}\}$ ,

$$\bar{\alpha}_{\zeta,;is}^{EST} = B_1^{-1} \sum_{k=1}^{B_1} \alpha_{\zeta,k,is}^{EST}$$

and

$$\bar{\bar{\alpha}}_{;is}^{EST} = 400^{-1} \sum_{\zeta=1}^{400} B_1^{-1} \sum_{k=1}^{B_1} \alpha_{\zeta,k,is}^{EST}.$$

The within variance component for  $B_1 = 100$  is estimated by half of the mean of squared differences between the two prediction MSE values,

$$\hat{V}_w^{EST} = (12)^{-1} \sum_{is=1}^{12} \left( (400)^{-1} \sum_{\zeta=1}^{400} (\alpha_{\zeta,is}^{EST} - \alpha_{2,\zeta,is}^{EST})^2 \right) / 2, \quad (3.21)$$

where  $\alpha^{EST} \in \{\alpha^*, \alpha_T^{**}, \alpha_C^{**}\}$  and  $\alpha_2^{EST} \in \{\alpha_2^*, \alpha_T^{**2}, \alpha_C^{**2}\}$ .

The variance components for the bootstrap prediction MSE estimators  $(\hat{\alpha}^*, \hat{\alpha}_T^{**}, \hat{\alpha}_C^{**})$  are given in Table 3.3 for  $(B_1 = 100, B_2 = 1)$ . The estimated between variance component is the difference between the estimated total variance and the estimated within variance component.

Using the entries in Table 3.3, we compare the performance of the level one bootstrap, the performance of the fast classic double bootstrap and the performance of the fast telescoping double bootstrap. The between component for the level one bootstrap is about 80% to 86% of the between component for the double bootstrap procedures. This is not surprising as bias reduction procedures often increase the variance. The bootstrap sampling variance, the within component, for the classic double bootstrap is about four times that of the level one bootstrap.

Table 3.3: Estimated Variance Components for Variance of Estimated Prediction MSE  
(Within is for  $(B_1 = 100, B_2 = 1)$ . All Variances Multiplied by  $10^6$ )

Source of Variation	Size	$\hat{\alpha}^*$	$\hat{\alpha}_T^{**}$	$\hat{\alpha}_C^{**}$
Between	2	17.8860	22.2159	22.2159
Within		2.0985	3.9029	10.9261
Total		19.9845	26.1188	33.1420
Between	10	5.5618	6.8598	6.8598
Within		1.0991	2.3235	5.5663
Total		6.6609	9.1833	12.4260
Between	40	0.5440	0.6303	0.6303
Within		0.2642	0.6133	1.3417
Total		0.8082	1.2437	1.9720

The telescoping bootstrap is 1.8 to 2.3 times as efficient as the classic double bootstrap.

The performance of the classic double bootstrap estimator changes when the number of level two bootstrap samples is greater than one. For  $B_2 \geq 1$ , the within variance of  $\alpha_C^{**}$  has two components corresponding to the variance of the conditional mean of  $\alpha_C^{**}$  given  $\psi^*$  and to the level two bootstrap sampling variance.

Define  $\alpha^*$  and  $\alpha^{**}$  to be the level one bootstrap and the classic double bootstrap sample random variables with realizations the values of the bootstrap prediction MSE, for a Monte Carlo sample and for a specific area. The within variance for the classic double bootstrap MSE estimator is a function of the variance of  $\alpha^*$ , the variance of  $\alpha^{**}$  and the covariance between  $\alpha^*$  and  $\alpha^{**}$ . Define

$$\begin{aligned}
 V_w(\alpha^*) &= E \left( V(\alpha^* | \hat{\psi}) \right), \\
 V_w(\alpha^{**}) &= E \left\{ V \left( E(\alpha^{**} | \psi^*) | \hat{\psi} \right) \right\} + E \left( V(\alpha^{**} | \psi^*) | \hat{\psi} \right), \\
 C_w(\alpha^*, \alpha^{**}) &= E \{ C(\alpha^*, E(\alpha^{**} | \psi^*)) \}.
 \end{aligned} \tag{3.22}$$

The within variance components for  $(\alpha^*, \alpha^{**})$  and the corresponding within covariance component are given in Table 3.4 for  $(B_1 = 1, B_2 = 1)$ . Using the entries in this table, one can

calculate the variance of classic double bootstrap prediction MSE estimator for  $B_1$  level one samples combined with  $B_2$  level two samples for each level one sample,

$$\begin{aligned} V_w(\hat{\alpha}^{**C}) &= 4B_1^{-1}V_w(\alpha^*) + B_1^{-1}E \left\{ V \left( E(\alpha^{**}|\psi^*)|\hat{\psi} \right) \right\} \\ &\quad + B_1^{-1}B_2^{-1}E \left( V(\alpha^{**}|\psi^*)|\hat{\psi} \right) - 4B_1^{-1}C_w(\alpha^*, \alpha^{**}). \end{aligned}$$

Let  $(\alpha^{*1}, \alpha_1^{**1}, \alpha_2^{**1})$  be the realizations of  $(\alpha^*, \alpha^{**})$  for the level one bootstrap and classic double bootstrap, for the first set of bootstrap samples, and let  $(\alpha^{*2}, \alpha_1^{**2}, \alpha_2^{**2})$  be the realizations  $(\alpha^*, \alpha^{**})$  for the level one bootstrap and fast double bootstrap, for the second set of bootstrap samples. The within variance of  $\alpha^*$  for  $B_1 = 1$  is estimated by

$$V_w(\widehat{\alpha^*}) = 100(12)^{-1} \sum_{is=1}^{12} \left( (400)^{-1} \sum_{\zeta=1}^{400} (\alpha_{\zeta, is}^* - \alpha_{2, \zeta, is}^*)^2 \right) / 2.$$

Define  $a_{11}, a_{21}, a_{31}$  for the first set of bootstrap samples to be

$$\begin{aligned} a_{11} &= \alpha^{*1}, \\ a_{21} &= 0.5(\alpha_1^{**1} + \alpha_2^{**1}), \\ a_{31} &= \alpha_1^{**1} - \alpha_2^{**1}, \end{aligned}$$

and define  $a_{12}, a_{22}, a_{32}$  for the second set of bootstrap samples to be

$$\begin{aligned} a_{12} &= \alpha^{*2}, \\ a_{22} &= 0.5(\alpha_1^{**2} + \alpha_2^{**2}), \\ a_{32} &= \alpha_1^{**2} - \alpha_2^{**2}. \end{aligned}$$

Let the average of  $(a_{11}, a_{21}, a_{31}, a_{12}, a_{22}, a_{32})$  over the bootstrap samples be  $(\bar{a}_{11}, \bar{a}_{21}, \bar{a}_{31}, \bar{a}_{12}, \bar{a}_{22}, \bar{a}_{32})$ . The within covariance for  $B_1 = 100$  is estimated by half of the mean of the product of the differences between the two average values of the prediction MSE at the two bootstrap levels,

$$\widehat{Cov}_{w,100}(\alpha^*, \alpha^{**}) = (12)^{-1} \sum_{is=1}^{12} \left( (400)^{-1} \sum_{\zeta=1}^{400} (\bar{a}_{11, is} - \bar{a}_{12, is})(\bar{a}_{21, is} - \bar{a}_{22, is}) \right) / 2.$$

The within covariance for  $B_1 = 1$  is estimated by

$$\widehat{Cov}_w(\alpha^*, \alpha^{**}) = 100\widehat{Cov}_{w,100}(\alpha^*, \alpha^{**}).$$

The within variance of  $\alpha^{**}$  for  $B_1 = 100$  is estimated by one half of the mean of squared differences between the two statistics  $\bar{a}_{21}$  and  $\bar{a}_{22}$ ,

$$\hat{V}_{w,100}(\alpha^{**}) = (12)^{-1} \sum_{is=1}^{12} \left( (400)^{-1} \sum_{\zeta=1}^{400} (\bar{a}_{21,is} - \bar{a}_{22,is})^2 \right) / 2.$$

The variance component of the bootstrap variance of  $\alpha^{**}$  due to level two bootstrap sampling, for  $B_1 = 100$ , is estimated by one fourth of the mean of squared differences between the two statistics  $\bar{a}_{31}$  and  $\bar{a}_{32}$ ,

$$E_{100} \left( V(\widehat{\alpha^{**}|\psi^*})|\hat{\psi} \right) = (12)^{-1} \sum_{is=1}^{12} \left( (400)^{-1} \sum_{\zeta=1}^{400} (\bar{a}_{31,is} - \bar{a}_{32,is})^2 \right) / 4,$$

and for  $B_1 = 1$  is estimated by

$$E \left( V(\widehat{\alpha^{**}|\psi^*})|\hat{\psi} \right) = 100E_{100} \left( V(\widehat{\alpha^{**}|\psi^*})|\hat{\psi} \right).$$

The variance component of the bootstrap variance of  $\alpha^{**}$  corresponding to the variance of the conditional mean of  $\alpha^{**}$  given  $\psi^*$ , for  $B_1 = 100$ , is estimated by the difference between the estimated total level two within variance of  $\alpha^{**}$  and one half of the estimated within variance in  $\alpha^{**}$  due to bootstrap sampling,

$$E_{100} \left\{ V \left( \widehat{E(\alpha^{**}|\psi^*)}|\hat{\psi} \right) \right\} = \hat{V}_w(\alpha^{**}) - E \left( V(\widehat{\alpha^{**}|\psi^*})|\hat{\psi} \right) / 2,$$

and for  $B_1 = 1$  is estimated by

$$E \left\{ V \left( \widehat{E(\alpha^{**}|\psi^*)}|\hat{\psi} \right) \right\} = 100E_{100} \left\{ V \left( \widehat{E(\alpha^{**}|\psi^*)}|\hat{\psi} \right) \right\}.$$

Table 3.4: Estimated Within Bootstrap Variance and Covariance of the Bootstrap Prediction MSE ( $B_1 = 1, B_2 = 1$ , Entries Multiplied by  $10^6$ )

	$V_w(\widehat{\alpha^*})$	$E \left( V(\widehat{\alpha^{**} \psi^*}) \hat{\psi} \right)$	$E \left\{ V \left( \widehat{E(\alpha^{**} \psi^*)} \hat{\psi} \right) \right\}$	$C_w(\widehat{\alpha^*}, \alpha^{**})$
2	209.2073	192.4625	71.2836	1.9917
10	109.1077	93.4928	28.5335	0.4580
40	26.1292	22.0451	6.3739	-0.3074



Table 3.5 contains estimates of the within variance components for the classic double bootstrap prediction MSE estimator  $\hat{\alpha}_C^{**}$  and for the telescoping double bootstrap prediction MSE estimator  $\hat{\alpha}_T^{**}$ , for different double bootstrap designs, that is for different combinations of  $B_1$  level one samples combined with  $B_2$  level two samples for each level one sample. The choice of  $(B_1, B_2)$  pair has an effect on the estimated within variance of the bootstrap prediction MSE estimator. For the parameters in the simulation study, the optimal number of bootstrap samples is  $B_2 = 0.45$ ; see Appendix A. That is, for a given number of bootstrap samples  $B$ , the fast double bootstrap with  $B_2 = 1$  minimizes the variance of the classic double bootstrap prediction MSE estimator. In practice it is better to use a large number of bootstrap samples at level one and one bootstrap sample at level two, than to use a large number of bootstrap samples at level one and a large number of bootstrap samples at level two.

Table 3.5: Estimated Within Bootstrap Variance of the Bootstrap Prediction MSE. Entries Multiplied by  $10^6$

$n_i$	Classic				Telescoping		
	$V_{100,1}$	$V_{1000,1}$	$V_{5000,1}$	$V_{100,50}$	$V_{100,1}$	$V_{1000,1}$	$V_{5000,1}$
2	10.9261	1.0926	0.2185	9.0400	3.9029	0.3903	0.0781
10	5.5663	0.5566	0.1113	4.6500	2.3235	0.2324	0.0465
40	1.3417	0.1342	0.0268	1.1256	0.6133	0.0613	0.0123

Consider the prediction MSE estimators for the areas of size  $n_i = 2$ . Using the results in Tables 3.5 and 3.3, we conclude that increasing the number of bootstrap samples to  $B_1 = 1000$  reduces the within variance component to about 3.0% of the total variance for the classic method and to about 1.4% of the total variance for the telescoping method. For the classic bootstrap method based on a total of 200 samples, the estimated within variance component is about 33% of the total variance for the design  $(B_1 = 100, B_2 = 1)$ . For the classic bootstrap method based on a total of 10000 samples, the estimated within variance component is about 0.7% of the total variance for the design  $(B_1 = 5000, B_2 = 1)$  and about 27.3% of the total variance for the design  $(B_1 = 100, B_2 = 50)$ .

### 3.3.4.1 Equal Efficiency Bootstrap Designs

We give bootstrap sample sizes such that the bootstrap variance of the estimated prediction MSE is the same under different bootstrap sampling procedures. Table 3.6 contains the number of level one bootstrap samples needed in the classic bootstrap method in order to produce prediction MSE estimates as efficient as the prediction MSE estimates produced using the telescoping bootstrap method with  $(B_1 = 100, B_2 = 1)$ . The last column in Table 3.6 contains the total number of bootstrap samples for each procedure, for each design.

Table 3.6: Equal Efficiency Bootstrap Procedures

Bootstrap Method/Design	$n_i$	$B_1$	Total
Telescoping (100, 1)	2	100	200
Classic ( $B_1, 1$ )		280	560
Classic ( $B_1, 50$ )		232	11832
Telescoping (100, 1)	10	100	200
Classic ( $B_1, 1$ )		240	480
Classic ( $B_1, 50$ )		200	10200
Telescoping (100, 1)	40	100	200
Classic ( $B_1, 1$ )		219	438
Classic ( $B_1, 50$ )		184	9384

## 3.4 Summary

We used a simulation study of a unit level logistic model to compare the impact of different levels of auxiliary information. The minimum mean squared error (MMSE) predictors for the small area means were obtained by conditioning on the information available for an area. That information is the unit level response realizations, the unit level covariate observations, and the, sometimes available, additional unit level auxiliary information. We considered fixed and random mean models for the covariates, as well as known and unknown distribution for the covariates. The prediction MSE is smaller when the covariate distribution is specified. Also, the

prediction MSE is smaller when additional auxiliary information is available and included in the estimation. The effect of including auxiliary information in the estimation is much smaller for the random mean model than for the fixed mean model for the covariates.

We presented a parametric double bootstrap procedure for the prediction MSE for unit level logistic models. The fast double bootstrap procedure, where the number of level-two bootstrap samples is  $B_2 = 1$ , has superior bootstrap efficiency relative to classic double bootstrap procedure with  $B_2 > 1$ . The double bootstrap reduces the prediction MSE estimation bias to 60 to 80% of that of the level one bootstrap. The double bootstrap increases the standard error of the prediction MSE estimator by 13 to 17% relative to that of the level one bootstrap.

### 3.5 Appendix A. Optimal Number of Bootstrap Samples at Level Two

Given a fixed number of bootstrap samples,  $B$ , the within bootstrap variance of the bias corrected double bootstrap estimator is minimized when the number of level two bootstrap samples,

$$B_2 = \sqrt{\frac{E\left(V(\alpha^{**}|\psi^*)|\hat{\psi}\right)}{4V_w(\alpha^*) - 4C_w(\alpha^*, \alpha^{**}) + E\left\{V\left(E(\alpha^{**}|\psi^*)|\hat{\psi}\right)\right\}}}, \quad (3.23)$$

where

$$V_w(\alpha^*) = E\left(V(\alpha^*|\hat{\psi})\right),$$

$$V_w(\alpha^{**}) = E\left\{V\left(E(\alpha^{**}|\psi^*)|\hat{\psi}\right)\right\} + E\left(V(\alpha^{**}|\psi^*)|\hat{\psi}\right),$$

$$C_w(\alpha^*, \alpha^{**}) = E\{C(\alpha^*, E(\alpha^{**}|\psi^*))\}.$$

*Proof.* The double bootstrap, bias corrected, estimator of  $\alpha$  is

$$\hat{\alpha}^{**} = B_1^{-1} \sum_{k=1}^{B_1} \alpha_k^* - \Delta_\alpha = B_1^{-1} \sum_{k=1}^{B_1} 2\alpha_k^* - B_1^{-1} B_2^{-1} \sum_{k=1}^{B_1} \sum_{t=1}^{B_2} \alpha_{kt}^{**} =: 2\bar{\alpha}^* - \bar{\alpha}^{**},$$

and the within bootstrap variance of  $\hat{\alpha}^{**}$  is

$$\begin{aligned} V_w(\hat{\alpha}^{**}) &= V_w(2\bar{\alpha}^* - \bar{\alpha}^{**}) \\ &= V_w(2\bar{\alpha}^*) + V_w(\bar{\alpha}^{**}) - 2C_w(2\bar{\alpha}^*, \bar{\alpha}^{**}) \\ &= E\left(V(2B_1^{-1} \sum_{k=1}^{B_1} \alpha_k^*|\hat{\psi})\right) + E\left\{V\left(E(B_1^{-1} \sum_{k=1}^{B_1} \bar{\alpha}_k^{**}|\psi_k^*)|\hat{\psi}\right)\right\} \\ &\quad + E\left(V(B_1^{-1} B_2^{-1} \sum_{k=1}^{B_1} \sum_{t=1}^{B_2} \alpha_{kt}^{**}|\psi_k^*)|\hat{\psi}\right) - 4E\left\{C\left(\bar{\alpha}^*, E(B_1^{-1} \sum_{k=1}^{B_1} \bar{\alpha}_k^{**}|\psi_k^*)\right)\right\}, \end{aligned}$$

where  $\bar{\alpha}_k^{**} = B_2^{-1} \sum_{t=1}^{B_2} \alpha_{kt}^{**}$ . Since  $\alpha_{kt}^{**}$  are identically distributed, given  $\psi_k^*$  and  $\alpha_k^*$  are identi-

cally distributed, given  $\hat{\psi}$ ,

$$\begin{aligned}
V_w(\hat{\alpha}^{**}) &= 4B_1^{-1}E\left(V(\alpha^*|\hat{\psi})\right) + B_1^{-1}E\left\{V\left(E(\alpha_1^{**}|\psi^*)|\hat{\psi}\right)\right\} \\
&\quad + B_1^{-1}B_2^{-1}E\left(V(\alpha^{**}|\psi^*)|\hat{\psi}\right) - 4B_1^{-1}E\left\{C\left(\alpha^*, E(\alpha^{**}|\psi^*)\right)\right\} \\
&= 4B_1^{-1}V_w(\alpha^*) + B_1^{-1}E\left\{V\left(E(\alpha^{**}|\psi^*)|\hat{\psi}\right)\right\} + B_1^{-1}B_2^{-1}E\left(V(\alpha^{**}|\psi^*)|\hat{\psi}\right) - 4B_1^{-1}C_w(\alpha^*, \alpha^{**}).
\end{aligned}$$

We would like to minimize  $V_w(\hat{\alpha}^{**})$  with respect to the restriction  $B_1B_2 + B_1 = B$ , where  $B$  is a constant representing the total number of bootstrap samples.

To consider this problem as a Lagrangian multiplier problem, let

$$\begin{aligned}
L(B_1, B_2, \lambda) &= 4B_1^{-1}V_w(\alpha^*) - 4B_1^{-1}C_w(\alpha^*, \alpha^{**}) + B_1^{-1}E\left\{V\left(E(\alpha^{**}|\psi^*)|\hat{\psi}\right)\right\} \\
&\quad + B_1^{-1}B_2^{-1}E\left(V(\alpha^{**}|\psi^*)|\hat{\psi}\right) + \lambda(B_1B_2 + B_1 - B)
\end{aligned}$$

where  $\lambda$  the Lagrangian multiplier. The resulting system of three equations is:

$$\begin{aligned}
0 &= -4B_1^{-2}V_w(\alpha^*) + 4B_1^{-2}C_w(\alpha^*, \alpha^{**}) - B_1^{-2}E\left\{V\left(E(\alpha^{**}|\psi^*)|\hat{\psi}\right)\right\} \\
&\quad - B_1^{-2}B_2^{-1}E\left(V(\alpha^{**}|\psi^*)|\hat{\psi}\right) + \lambda(B_2 + 1) \\
0 &= -B_1^{-1}B_2^{-2}E\left(V(\alpha^{**}|\psi^*)|\hat{\psi}\right) + \lambda B_1 \\
0 &= B_1B_2 + B_1 - B.
\end{aligned}$$

The solution for  $B_2$  is (3.23) and

$$B_1 = B \left( \sqrt{\frac{E\left(V(\alpha^{**}|\psi^*)|\hat{\psi}\right)}{4V_w(\alpha^*) - 4C_w(\alpha^*, \alpha^{**}) + E\left\{V\left(E(\alpha^{**}|\psi^*)|\hat{\psi}\right)\right\}} + 1} \right)^{-1}.$$

■

## Bibliography

- [1] Datta, G.S., Rao, J.N.K. and Smith, D. (2005) “On measuring the variability of small area estimators under a basic area level model. ” *Biometrika*, 92, 183-196.
- [2] Datta, G.S., Rao, J.N.K. and Smith, D. (2012) “Amendments and Corrections: On measuring the variability of small area estimators under a basic area level model, ” *Biometrika*, 99, 2, 509.
- [3] 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.
- [4] Davidson, R. and MacKinnon, J.G. (2007), “Improving the reliability of bootstrap tests with the fast double bootstrap,” *Computational Statistics and Data Analysis*, 51, 3259-3281.
- [5] Diao, L., Smith, D., Datta, G.S., Maiti, T., Opsomer, J. (2014), “ Accurate Confidence Interval Estimation of Small Area Parameters Under the Fay-Herriot Model, ” *Scandinavian Journal of Statistics*, 41, 497-515.
- [6] Fuller, W.A. and Harter, R.M. (1987), “The multivariate components of variance model for small area estimation,” *Small Area Statistics: An International Symposium* , Platek, R., Rao, J.N.K., Sarndal, C.E. and Singh, M.P. (Eds.), John Wiley , New York, 103-123.
- [7] 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.

- [8] Ghosh, M. and Sinha, K. (2007), "Empirical Bayes estimation in finite population sampling under functional measurement error models," *Journal of Statistical Planning and Inference*, 137, 2759-2773.
- [9] Hall P. and Maiti T. (2006), "On parametric bootstrap methods for small area prediction," *J.R. Statist. Soc. B*, 68, 2, 221-238.
- [10] Pfeffermann, D. and Correa, S. (2012), "Empirical bootstrap bias correction and estimation of prediction mean square error in small area estimation." *Biometrika*, 99, 457-472.
- [11] Rao J.N.K. (2003), "Small Area Estimation," *Wiley Series in Survey Methodology*.
- [12] Torabi, M., Datta G. and Rao, J.N.K. (2009), "Empirical Bayes Estimation of Small Area Means under a Nested Error Linear Regression Model with Measurement Errors in the Covariates," *Scandinavian Journal of Statistics*, 36, 355-368.
- [13] Wang, J. and Fuller, W.A. (2003), "The mean squared error of small area predictors constructed with estimated area variances," *Journal of the American Statistical Association*, 98, 716-723.
- [14] Ybarra, L.M.R. and Lohr, S.L. (2008), "Small area estimation when auxiliary information is measured with error," *Biometrika*, 95, 919-931.

## CHAPTER 4. BOOTSTRAP CONFIDENCE INTERVALS FOR SMALL AREA MEANS

A paper to be submitted for publication

Andreea L. Erciulescu and Wayne A. Fuller

### **Abstract**

Most small area studies focus on constructing predictors for the area means and on estimating the variance of the prediction errors. However, agencies and policy makers are often interested in confidence intervals for the small area predictors. We present two sided confidence intervals for the small area means of a binary response variable. We consider unit level data and stochastic covariates. The estimation of the prediction error variance and the estimation of the cutoff points are key components in the construction of confidence intervals for the small area means. A linear approximation of the model is considered and a Taylor variance approximation is presented for the prediction error variance. We compare different bootstrap estimation methods for the cutoff points using a simulation study.

### **4.1 Introduction**

Procedures based on models have been used to construct estimates for the means of small areas, by exploiting auxiliary information. Agencies and policy makers are often interested in confidence intervals for the small area estimates. In this paper, we study nested models with a binary response, stochastic covariates and random area effects. We consider different procedures to estimate the small area mean prediction mean squared error and to construct confidence intervals for the small area means.



Taylor methods have been shown to give good estimates of the prediction mean squared error (MSE), for predicted small area means based on linear models; see p. 103 in Rao (2003) for area level models and p. 139 in Rao (2003) for unit level models. There are many studies reporting point estimates for the small area means, as well as prediction MSE estimates. Ghosh, Sinha and Kim (2006) consider an area level linear model with random auxiliary variable mean, estimated jointly with the small area mean. Ybarra and Lohr (2008) consider an area level linear model with auxiliary mean estimated with error. Datta, Rao and Torabi (2010), following Ghosh and Sinha (2007), studied a nested error linear regression model with area level covariate subject to measurement error. When the direct estimates of small area means are nonlinear functions of the auxiliary information and of the random area effects, the prediction of the small area mean is no longer a linear function of the observations. Bootstrap methods have been used for MSE estimators in this case. Erciulescu and Fuller (2014) consider a nonlinear unit level model where the mean of the auxiliary variable is measured with error. The authors construct small area predictions for the small area means of a binary response variable and present double bootstrap prediction MSE estimates.

Hall and Maiti (2006) consider a linear area level model and a unit level binomial model with fixed known covariates. They construct small area predictions for the logit of the small area means and nonnegative, bias-corrected MSE estimates using a double bootstrap procedure. Pfeiffermann and Correa (2012) study a unit level binomial model with fixed known covariates and suggest a bootstrap procedure in which the bias in the estimator is estimated as a function of parameters and of a bootstrap estimator of bias.

Most studies that report confidence intervals (CIs) for the small area means, report for special cases of the Fay-Herriot model; see Hall and Maiti (2006), Chatterjee et al. (2008), Dass et al. (2012), Diao et al. (2014) and Yoshimori and Lahiri (2014). Hall (1986) proposed pivot-like statistics to reduce the two-sided bootstrap CI coverage error relative to the normal approximated CI coverage error, from  $O(m^{-1})$  to  $O(m^{-3/2})$ , where  $m$  is the number of areas. Chatterjee et al. (2008) construct parametric bootstrap confidence intervals based on a pivot-

like statistic and centered around the small area predictions. The authors consider a general linear mixed model with unknown random effects variance and unknown sampling variance. The area specific parametric bootstrap prediction intervals constructed by Chatterjee et al. (2008) have error of order  $O(d^3m^{-3/2})$ , where  $d$  is the number of model parameters and  $m$  is the number of small areas. Chatterjee et al. (2008) state that if calibrated, their intervals would be  $O(d^5m^{-5/2})$  order correct, where calibration refers to the bootstrap bias adjustment in the point or interval estimates. Liu and Diallo (2013) apply the method in Chatterjee et al. (2008) and construct percentile parametric bootstrap confidence intervals for survey-weighted small area proportions based on the Fay-Herriot model.

Hall and Maiti (2006) construct two sided, equal-tailed, double bootstrap calibrated CIs. The authors outline an algorithm for calibrating the CI coverage and constructing percentile confidence intervals for the parameter of interest. Linear and nonlinear models are considered, but the parameter of interest,  $\theta_i$ , is always a linear function of the model parameters. The bootstrap CIs for  $\theta_i$  are constructed using the estimated distribution of the bootstrap predictions of  $\theta_i$ . The authors state that the coverage error of the level one bootstrap CI is of order  $O(m^{-2})$  and that the coverage error of the level two calibrated bootstrap CI is of order  $O(m^{-3})$ . A simulation study is conducted for the binary model and bootstrap confidence intervals with nominal coverages  $1 - \alpha = 0.80, 0.90, 0.95$  are constructed for  $\theta_i$ , for different models.

Davidson and MacKinnon (2007) introduced a fast double bootstrap procedure for bootstrap hypothesis testing. Only one sample is drawn at the second bootstrap level. Giacomini et al. (2013) provide key properties for the fast double bootstrap methods, under regularity conditions. The authors discuss applications of fast double bootstrap methods to assess the performance of bootstrap estimators, test statistics and confidence intervals. Erciulescu and Fuller (2014) study a unit level binomial model and construct prediction mean squared error estimators for the small area means using fast double bootstrap procedures. Chang and Hall (2014) study the fast double bootstrap method described in Giacomini et al (2013) to produce third-order accurate confidence intervals. Chang and Hall (2014) show that the performance

of the fast double bootstrap in reducing the order of magnitude of bias is comparable to the classic double bootstrap method. The authors describe a calibration method for the confidence interval cutoff points and show that the fast double bootstrap does not improve level one order of magnitude of coverage error of the CIs.

In this paper we present bootstrap confidence intervals for the small area mean constructed using parametric single and double bootstrap methods. The model of primary interest is the unit level logistic model. Stochastic covariates and different cases of auxiliary information are considered. Different estimators for the prediction error variance, including a Taylor approximation and bootstrap estimators, and different estimators for the cutoff points of the CIs for the small area means (proportions) are compared in a simulation study.

## 4.2 Models

Define the unit level generalized linear mixed model (ULGLMM) by

$$\begin{aligned} y_{ij} &= g(x_{ij}, \boldsymbol{\beta}, b_i) + e_{ij}, \\ x_{ij} &= \mu_x + \delta_i + \epsilon_{ij} =: \mu_{xi} + \epsilon_{ij}, \\ \tilde{\mu}_{xi} &= \mu_{xi} + u_i, \end{aligned} \tag{4.1}$$

for  $i = 1, \dots, m$ , where  $m$  denotes the number of areas and  $j = 1, \dots, n_i$ , where  $n_i$  denotes the number of units within area  $i$ . Assume that the random area effects  $b_i$  are independent and identically distributed, with a density  $f_b$  with mean and variance  $(0, \sigma_b^2)$ , respectively. Assume that the sampling errors  $e_{ij}$  are independent  $(0, \sigma_{e_{ij}}^2)$  random variables, independent of  $b_k$ , for all  $i, j$  and  $k$ . Assume  $\sigma_{e_{ij}}^2 = \sigma_e^2 k_{ij}^2$ , for known constants  $k_{ij}$  and assume the vector  $(b_i, e_{ij})$  is independent of  $x_{kt}$  for all  $i, j, k$ , and  $t$ .

The relationship between the response vector  $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})$  and the explanatory vector  $\mathbf{x}_i = (x_{i1}, \dots, x_{in_i})$  and the random area effects  $b_i$  need not be a linear function. We assume that  $g(x_{ij}, \boldsymbol{\beta}, b_i)$  is a continuous, with continuous partial derivatives with respect to the vector of fixed effects coefficients  $\boldsymbol{\beta}$  and with respect to the random area effects  $b_i$ .

The small area mean of  $\mathbf{x}$ ,  $\mu_{xi}$  can be fixed or random, known or unknown. If  $\mu_{xi}$  is random, assume that the area effects  $\delta_i$  are independent and identically distributed, with a density  $f_\delta$  with mean and variance  $(0, \sigma_\delta^2)$ , respectively. Assume that the sampling errors  $\epsilon_{ij}$  are independent and identically distributed, with mean and variance  $(0, \sigma_\epsilon^2)$ , respectively, independent of  $\delta_k$ , for all  $i, j$  and  $k$ .

Auxiliary information about  $\mu_{xi}$  is denoted by  $\tilde{\mu}_{xi}$ . Assume that the sampling errors  $u_i$  are independent and identically distributed, with mean and variance  $(0, \sigma_u^2)$ , respectively, and assume that the auxiliary information  $\tilde{\boldsymbol{\mu}}_x = (\tilde{\mu}_{x1}, \dots, \tilde{\mu}_{xm})$  is independent of  $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m)$ , where  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{in_i})$  is the covariate information for area  $i$ .

Let  $\theta_i$  be the small area mean of  $y$ ,

$$\theta_i = \int g(x, \boldsymbol{\beta}, b_i) dF_{\mathbf{x}_i}(x), \quad (4.2)$$

where  $F_{\mathbf{x}_i}(x)$  is the distribution of  $x$  in area  $i$ .

#### 4.2.1 Small Area Mean Prediction for the Unit Level Linear Mixed Model

Consider first the case when  $g(x_{ij}, \boldsymbol{\beta}, b_i)$  is a linear function of the parameters. Assume  $\mu_{xi}$  is fixed and known. Then the unit level linear mixed model (ULLMM) is

$$y_{ij} = \mathbf{x}_{ij}\boldsymbol{\beta} + b_i + e_{ij}, \quad (4.3)$$

where  $\mathbf{x}_{ij} = (1, x_{ij})$ ,  $b_i \sim NI(0, \sigma_b^2)$ ,  $e_{ij} \sim NI(0, \sigma_{eij}^2)$  and  $\sigma_{eij}^2 = \sigma_e^2 k_{ij}^2$ , for known constants  $k_{ij}$ . The small area mean of  $\mathbf{y}$  is

$$\theta_i = \bar{\mathbf{x}}_{Ni}\boldsymbol{\beta} + b_i, \quad (4.4)$$

where  $\bar{\mathbf{x}}_{Ni}$  is the population mean of  $\mathbf{x}$  for area  $i$ . Let

$$\hat{\boldsymbol{\beta}} = \left( \sum_{i=1}^m \mathbf{x}'_i V_i^{-1} \mathbf{x}_i \right)^{-1} \left( \sum_{i=1}^m \mathbf{x}'_i V_i^{-1} \mathbf{y}_i \right) \quad (4.5)$$

be the best linear unbiased estimator (BLUE) of  $\boldsymbol{\beta}$ , for  $V_i = \sigma_b^2 \mathbf{1}_{n_i} \mathbf{1}'_{n_i} + \sigma_e^2 \text{Diag}(k_{ij}^2)$ , where  $\mathbf{1}_{n_i}$  denotes the vector of length  $n_i$  with entries equal to one and  $\text{Diag}(k_{ij}^2)$  denotes the  $n_i \times n_i$  diagonal matrix with diagonal entries  $k_{ij}^2$ . The variance of the  $\hat{\boldsymbol{\beta}}$  is

$$V(\hat{\boldsymbol{\beta}}) = \left( \sum_{i=1}^m \mathbf{x}'_i V_i^{-1} \mathbf{x}_i \right)^{-1}. \quad (4.6)$$

In practice, the vector of parameters  $(\sigma_b^2, \sigma_e^2)$  needs to be estimated. Let parameters  $(\sigma_b^2, \sigma_e^2)$  be estimated by  $(\hat{\sigma}_b^2, \hat{\sigma}_e^2)$ , where, for example,  $(\hat{\sigma}_b^2, \hat{\sigma}_e^2)$  are REML estimators. Then the empirical best linear unbiased predictor (EBLUP) of  $\theta_i$  is

$$\hat{\theta}_i^{EBLUP} = \bar{\mathbf{x}}_{N_i} \hat{\boldsymbol{\beta}} + \hat{b}_i = \bar{\mathbf{x}}_{N_i} \hat{\boldsymbol{\beta}} + \hat{\gamma}_i (\bar{y}_i - \bar{\mathbf{x}}_i \hat{\boldsymbol{\beta}}), \quad (4.7)$$

where

$$\hat{\gamma}_i = \hat{\sigma}_b^2 \left( \hat{\sigma}_b^2 + \left( \sum_{j=1}^m k_{ij}^{-2} \right)^{-1} \hat{\sigma}_e^2 \right)^{-1}$$

and

$$(\bar{y}_i, \bar{x}_i) = \left( \sum_{j=1}^m k_{ij}^{-2} \right)^{-1} \sum_{j=1}^m k_{ij}^{-2} (y_{ij}, x_{ij})$$

is the sample weighted mean of  $(\mathbf{y}, \mathbf{x})$ , for area  $i$ .

The Prasad and Rao (1990) expression for the prediction MSE is

$$MSE(\hat{\theta}_i^{EBLUP} - \theta_i) = g_{1i} + g_{2i} + g_{3i}, \quad (4.8)$$

where  $g_{1i}$  is the prediction MSE when the parameters  $(\boldsymbol{\beta}, \sigma_b^2, \sigma_e^2)$  are known,  $g_{2i}$  is due to the estimation of the vector  $\boldsymbol{\beta}$ , and  $g_{3i}$  is due to the estimation of the variance components,  $\sigma_b^2$  and  $\sigma_e^2$ .

Prasad and Rao (1990) derive an estimator of the prediction MSE, with bias of order  $o(m^{-1})$ ,

$$\hat{MSE}(\hat{\theta}_i^{EBLUP} - \theta_i) = \hat{g}_{1i} + \hat{g}_{2i} + 2\hat{g}_{3i}, \quad (4.9)$$

where  $\hat{g}_{1i}, \hat{g}_{2i}, \hat{g}_{3i}$  are functions of the estimated variance components,  $\hat{\sigma}_b^2, \hat{\sigma}_e^2$ . Extension of the result to  $k_{ij} \neq 1$  gives

$$\begin{aligned}\hat{g}_{1i} &= \hat{\gamma}_i \left( \sum_{j=1}^m k_{ij}^{-2} \right)^{-1} \hat{\sigma}_e^2, \\ \hat{g}_{2i} &= (1 - \hat{\gamma}_i)^2 \bar{\mathbf{x}}_{Ni} \hat{V}(\hat{\boldsymbol{\beta}}) \bar{\mathbf{x}}'_{Ni}, \\ \hat{g}_{3i} &= \left( \hat{\sigma}_b^2 + \left( \sum_{j=1}^m k_{ij}^{-2} \right)^{-1} \hat{\sigma}_e^2 \right) \hat{V}(\hat{\gamma}_i),\end{aligned}$$

where

$$\hat{V}(\hat{\gamma}_i) = \left( \hat{\sigma}_b^2 + \left( \sum_{j=1}^m k_{ij}^{-2} \right)^{-1} \hat{\sigma}_e^2 \right)^{-4} \left( \hat{\sigma}_b^4 \left( \sum_{j=1}^m k_{ij}^{-2} \right)^{-2} \hat{V}(\hat{\sigma}_e^2) + \left( \sum_{j=1}^m k_{ij}^{-2} \right)^{-2} \hat{\sigma}_e^4 \hat{V}(\hat{\sigma}_b^2) \right)$$

and  $\hat{V}(\hat{\boldsymbol{\beta}})$  is defined in (4.6). The estimation of  $\gamma_i \left( \sum_{j=1}^m k_{ij}^{-2} \right)^{-1} \sigma_e^2$  requires care because the bias in the typical estimator of  $\gamma_i$  is  $O(m^{-1})$ . The third term,  $\hat{g}_{3i}$  appears twice in the expression for  $M\hat{S}E(\hat{\theta}_i^{EBLUP} - \theta_i)$  due to a bias correction in the first term  $g_{1i}$ .

Using the fact that  $\bar{u}_{yi}^2 := (\bar{y}_i - \bar{\mathbf{x}}_i \hat{\boldsymbol{\beta}})^2$  is an approximately unbiased estimator for  $\sigma_b^2 + \left( \sum_{j=1}^m k_{ij}^{-2} \right)^{-1} \sigma_e^2$ , an alternative estimator for  $g_{3i}$  is

$$\hat{g}_{3i} = \left( \hat{\sigma}_b^2 + \left( \sum_{j=1}^m k_{ij}^{-2} \right)^{-1} \hat{\sigma}_e^2 \right)^{-4} \left( \hat{\sigma}_b^4 \left( \sum_{j=1}^m k_{ij}^{-2} \right)^{-2} \hat{V}(\hat{\sigma}_e^2) + \left( \sum_{j=1}^m k_{ij}^{-2} \right)^{-2} \hat{\sigma}_e^4 \hat{V}(\hat{\sigma}_b^2) \right) \bar{u}_{yi}^2, \quad (4.10)$$

see Fuller (1990). The estimator (4.9) does not depend directly on the area-specific data  $y_i$ . However, the estimator (4.9) with (4.10) depends on the area-specific data  $y_i$ .

Lahiri and Rao (1995) show that the estimator (4.9) is robust to departures from normality of the random area effects, but not to normality of the sampling errors. Alternative expressions for the estimated MSE and extensions are given by Fuller and Harter (1987), Fuller (1990), Datta and Lahiri (2000), and Wang and Fuller (2003). Datta, Rao and Smith (2005) give

results of an extensive study of mean squared error estimators for different estimators of  $\sigma_b^2$  and different distributions of the random effects.

#### 4.2.1.1 Small Area Mean Prediction for $\mu_{xi}$ in ULGLMM

Let a model for  $x_{ij}$  be of the form given in (4.1)

$$\begin{aligned} x_{ij} &= \mu_x + \delta_i + \epsilon_{ij}, \\ \tilde{\mu}_{xi} &= \mu_{xi} + u_i. \end{aligned} \quad (4.11)$$

The observations are  $\tilde{\mathbf{x}}_i = (\mathbf{x}_i, \tilde{\mu}_{xi})$ , with sampling errors  $\tilde{\boldsymbol{\epsilon}}_i := (\boldsymbol{\epsilon}_i, u_i)$ . The number of observations in area  $i$  is  $n_i + 1$ . The fixed parameter is  $\mu_x$ . The constants  $k_{ij} := a_{ijx}^{-2}$  equal 1 for  $j = 1, \dots, n_i$  and  $k$  for  $j = n_i + 1$ , where  $k = \sigma_\epsilon^2 / \sigma_u^2$  is assumed to be known. Hence, the estimation and prediction expressions in (4.5, 4.6, 4.7, 4.9, 4.10) are appropriate for the ULLMM for  $\tilde{\mathbf{x}}$ .

Let parameters  $(\sigma_\delta^2, \sigma_\epsilon^2)$  be estimated by  $(\hat{\sigma}_\delta^2, \hat{\sigma}_\epsilon^2)$ . Let  $\hat{\mu}_x$  be the empirical BLUE of  $\mu_x$  defined in (4.5)

$$\hat{\mu}_x = \left( \sum_{i=1}^m \mathbf{1}' \hat{V}_{ix}^{-1} \mathbf{1} \right)^{-1} \left( \sum_{i=1}^m \mathbf{1}' \hat{V}_{ix}^{-1} \tilde{\mathbf{x}}_i \right), \quad (4.12)$$

for  $\hat{V}_{ix} = \hat{\sigma}_\delta^2 \mathbf{1}_{n_i} \mathbf{1}'_{n_i} + \hat{\sigma}_\epsilon^2 \text{Diag}(k_{ij}^2)$ . The variance of  $\hat{\mu}_x$  given in (4.6) is estimated by  $\hat{V}(\hat{\mu}_x)$ ,

$$\hat{V}(\hat{\mu}_x) = \left( \sum_{i=1}^m \mathbf{1}' \hat{V}_{ix}^{-1} \mathbf{1} \right)^{-1}. \quad (4.13)$$

The EBLUP of  $\mu_{xi}$ , the small area mean of the auxiliary variable, is  $\hat{\mu}_{xi}$  defined in (4.7),

$$\hat{\mu}_{xi} = \hat{\mu}_x + \hat{\delta}_i = \hat{\mu}_x + \hat{\gamma}_{ix}(\tilde{x}_i - \hat{\mu}_x), \quad (4.14)$$

where  $\hat{\gamma}_{ix} = \hat{\sigma}_\delta^2 (\hat{\sigma}_\delta^2 + a_{i,x}^{-1} \hat{\sigma}_\epsilon^2)^{-1}$ ,  $\tilde{x}_i = \sum_{j=1}^{n_i} a_{ijx} a_{i,x}^{-1} \tilde{x}_{ij}$  and  $a_{i,x} = \sum_{j=1}^{n_i} a_{ijx}$ .

The estimated variance of  $\hat{\mu}_{xi}$  is  $\hat{V}(\hat{\mu}_{xi})$  and is derived in the form of (4.9) with (4.10),

$$\hat{V}(\hat{\mu}_{xi}) = \hat{g}_{1i} + \hat{g}_{2i} + 2\hat{g}_{3i}, \quad (4.15)$$

where

$$\hat{g}_{1i} = \hat{\gamma}_{ix} a_{i,x}^{-1} \hat{\sigma}_\epsilon^2,$$

$$\hat{g}_{2i} = (1 - \hat{\gamma}_{ix})^2 \hat{V}(\hat{\mu}_x),$$

$$\hat{g}_{3i} = (\hat{\sigma}_\delta^2 + a_{i,x}^{-1} \hat{\sigma}_\epsilon^2)^{-4} \left( \hat{\sigma}_\delta^4 a_{i,x}^{-2} \hat{V}(\hat{\sigma}_\epsilon^2) + a_{i,x}^{-2} \hat{\sigma}_\epsilon^4 \hat{V}(\hat{\sigma}_\delta^2) \right) \bar{u}_{ix}^2,$$

where  $\bar{u}_{ix}^2 := (\bar{x}_i - \hat{\mu}_x)^2$ .

#### 4.2.1.2 Small Area Mean Prediction for $y$ in ULGLMM

Small area mean predictors for the unit level model in (4.1), with nonlinear  $g(x_{ij}, \boldsymbol{\beta}, b_i)$  function, with fully known auxiliary information have been studied in Jiang and Lahiri (2001), Montanari, Ranalli and Vicarelli (2010), Lopez-Vizcaino, Lombardia-Cortina, Morales-Gonzalez (2011), and Pfeffermann and Correa (2012). The model with different amounts of available information was studied in Erciulescu and Fuller (2013).

Jiang and Lahiri (2001) and Pfeffermann and Correa (2012) consider the inverse logit  $g(x_{ij}, \boldsymbol{\beta}, b_i)$  function and construct estimates for the small area proportions using the conditional distribution of the random area effects given the response variables. Erciulescu and Fuller (2013) consider the inverse logit  $g(x_{ij}, \boldsymbol{\beta}, b_i)$  function and study two methods for constructing small area mean (proportion) predictions. The first method is based on the conditional distribution of the random area effects given the response variables. The second method, called the 'plug-in method' is based on the direct substitution of the predicted random area effects into the small area mean expression. They showed that the 'plug-in' predictor for the small area mean can have sizeable bias.

We present predictors of  $\theta_i$  for model (4.1), under alternative specifications for  $\mathbf{x}_{ij}$  and for different levels of auxiliary information.



*Known  $\mu_{xi}$* 

Let the small area mean of  $x_{ij}$  be fixed, known. The unit level information  $(\mathbf{y}_{ij}, \mathbf{x}_{ij})$  is observed, for all  $i = 1, \dots, m$ ,  $j = 1, \dots, n_i$ . Then, given known parameters for model (4.1),  $(\sigma_b^2, \sigma_\epsilon^2, \boldsymbol{\beta}, \mu_{xi})$ , and a sample  $(\mathbf{x}, \mathbf{y})$ , the minimum mean squared error (MMSE) predictor of the small area mean of  $y$  is

$$\hat{\theta}_i = E \left[ \hat{\theta}(b) | (\mathbf{x}_i, \mathbf{y}_i) \right],$$

where  $\mathbf{x}_i = (x_{i,1}, x_{i,2}, \dots, x_{i,n_i})$ ,  $\mathbf{y}_i = (y_{i,1}, y_{i,2}, \dots, y_{i,n_i})$ , and

$$\hat{\theta}(b) = \int_{\boldsymbol{\epsilon}} g(\mu_{xi} + \epsilon, \boldsymbol{\beta}, b) dF_{\boldsymbol{\epsilon}}(\epsilon)$$

and

$$\hat{\theta}_i = \frac{\int_b \hat{\theta}(b) \prod_{t=1}^{n_i} f(y_{it}|x_{it}, b) f(x_{it}|\mu_{xi}) dF_{b_i}(b)}{\int_b \prod_{t=1}^{n_i} f(y_{it}|x_{it}, b) f(x_{it}|\mu_{xi}) dF_{b_i}(b)}. \quad (4.16)$$

Since  $f(x_{it}|\mu_{xi})$  is free of  $b$ , the predictor given in (4.16) simplifies to

$$\hat{\theta}_i = \frac{\int_b \hat{\theta}(b) \prod_{t=1}^{n_i} f(y_{it}|x_{it}, b) dF_{b_i}(b)}{\int_b \prod_{t=1}^{n_i} f(y_{it}|x_{it}, b) dF_{b_i}(b)}. \quad (4.17)$$

*Unknown Fixed  $\mu_{xi}$ , Additional Information  $\tilde{\mu}_{xi}$* 

Let the small area mean of  $x_{ij}$  be fixed, unknown. The area level information  $(\mathbf{y}_i, \mathbf{x}_i, \tilde{\mu}_{xi})$  is observed, for all  $i = 1, \dots, m$ . Then, given known parameters for model (4.1),  $(\sigma_b^2, \sigma_\epsilon^2, \boldsymbol{\beta}, \mu_{xi})$ , and a sample  $(\tilde{\mathbf{x}} = (\mathbf{x}, \tilde{\boldsymbol{\mu}}_x), \mathbf{y})$ , the minimum mean squared error (MMSE) predictor of the small area mean of  $y$  is

$$\hat{\theta}_i = E \left[ \hat{\theta}(b) | (\mathbf{x}_i, \mathbf{y}_i, \tilde{\mu}_{xi}) \right],$$

where  $\mathbf{x}_i = (x_{i,1}, x_{i,2}, \dots, x_{i,n_i})$ ,  $\mathbf{y}_i = (y_{i,1}, y_{i,2}, \dots, y_{i,n_i})$ , and

$$\hat{\theta}(b) = \int_{\boldsymbol{\epsilon}} g(\mu_{xi} + \epsilon, \boldsymbol{\beta}, b) dF_{\boldsymbol{\epsilon}}(\epsilon)$$

and

$$\hat{\theta}_i = \frac{\int_b \hat{\theta}(b) \left( \prod_{t=1}^{n_i} f(y_{it}|x_{it}, b) f(x_{it}|\mu_{xi}) \right) f(\tilde{\mu}_{xi}|\mu_{xi}) dF_{b_i}(b)}{\int_b \left( \prod_{t=1}^{n_i} f(y_{it}|x_{it}, b) f(x_{it}|\mu_{xi}) \right) f(\tilde{\mu}_{xi}|\mu_{xi}) dF_{b_i}(b)}. \quad (4.18)$$

Since  $f(x_{it}|\mu_{xi})$  and  $f(\tilde{\mu}_{xi}|\mu_{xi})$  are free of  $b$ , the predictor given in (4.18) simplifies to

$$\hat{\theta}_i = \frac{\int_b \hat{\theta}(b) \prod_{t=1}^{n_i} f(y_{it}|x_{it}, b) dF_{b_i}(b)}{\int_b \prod_{t=1}^{n_i} f(y_{it}|x_{it}, b) dF_{b_i}(b)}. \quad (4.19)$$

For model (4.3), when  $g(x_{ij}, \boldsymbol{\beta}, b_i) = (1, x_{ij})' \boldsymbol{\beta} + b_i$ ,  $\hat{\theta}(b)$  defined for (4.16) and (4.18) is

$$\hat{\theta}(b) = \int_{\boldsymbol{\epsilon}} ((1, \mu_{xi} + \epsilon)' \boldsymbol{\beta} + b) dF_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}) = (1, \mu_{xi})' \boldsymbol{\beta} + b$$

and

$$\hat{\theta}_i = \frac{\int_b ((1, \mu_{xi})' \boldsymbol{\beta} + b) \prod_{t=1}^{n_i} f(y_{it}|x_{it}, b) dF_{b_i}(b)}{\int_b \prod_{t=1}^{n_i} f(y_{it}|x_{it}, b) dF_{b_i}(b)} = (1, \mu_{xi})' \boldsymbol{\beta} + E(b_i | \mathbf{y}_i, \mathbf{x}_i). \quad (4.20)$$

#### *Unknown Random $\mu_{xi}$ , Additional Information $\tilde{\mu}_{xi}$*

Consider model (4.1) with unknown random  $\mu_{xi}$ . The area level information  $(\mathbf{y}_i, \mathbf{x}_i, \tilde{\mu}_{xi})$  is observed, for all  $i = 1, \dots, m$ . Then, given known parameters for model (4.1),  $(\sigma_b^2, \sigma_\epsilon^2, \sigma_\delta^2, \boldsymbol{\beta}, \mu_x)$ , and a sample  $(\tilde{\mathbf{x}} = (\mathbf{x}, \tilde{\boldsymbol{\mu}}_x), \mathbf{y})$ , the minimum mean squared error (MMSE) predictor of the small area mean of  $y$  is

$$\hat{\theta}_i = E \left[ \hat{\theta}(b, \delta) | (\mathbf{x}_i, \mathbf{y}_i, \tilde{\mu}_{xi}) \right],$$

where  $\mathbf{x}_i = (x_{i,1}, x_{i,2}, \dots, x_{i,n_i})$ ,  $\mathbf{y}_i = (y_{i,1}, y_{i,2}, \dots, y_{i,n_i})$ , and

$$\hat{\theta}(b, \delta) = \int_{\boldsymbol{\epsilon}} g(\mu_x + \delta + \epsilon, \boldsymbol{\beta}, b) dF_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon})$$

and

$$\hat{\theta}_i = \frac{\int_b \int_\delta \hat{\theta}(b, \delta) \left( \prod_{t=1}^{n_i} f(y_{it}|x_{it}, b) f(x_{it}|\delta_i) \right) f(\tilde{\mu}_{xi}|\delta_i) dF_{\delta_i}(\delta) dF_{b_i}(b)}{\int_b \int_\delta \left( \prod_{t=1}^{n_i} f(y_{it}|x_{it}, b) f(x_{it}|\delta_i) \right) f(\tilde{\mu}_{xi}|\delta_i) dF_{\delta_i}(\delta) dF_{b_i}(b)}. \quad (4.21)$$

For model (4.3), when  $g(x_{ij}, \boldsymbol{\beta}, b_i) = (1, x_{ij})' \boldsymbol{\beta} + b_i$ ,  $\hat{\theta}(b, \delta)$  defined for (4.21) is

$$\hat{\theta}(b, \delta) = \int_{\boldsymbol{\epsilon}} ((1, \mu_x + \delta + \epsilon)' \boldsymbol{\beta} + b) dF_{\boldsymbol{\epsilon}}(\epsilon) = (1, \mu_x + \delta)' \boldsymbol{\beta} + b$$

and

$$\begin{aligned} \hat{\theta}_i &= \frac{\int_b \int_{\delta} ((1, \mu_x + \delta)' \boldsymbol{\beta} + b) (\prod_{t=1}^{n_i} f(y_{it}|x_{it}, b) f(x_{it}|\delta_i)) f(\tilde{\mu}_{xi}|\delta_i) dF_{\delta_i}(\delta) dF_{b_i}(b)}{\int_b \int_{\delta} (\prod_{t=1}^{n_i} f(y_{it}|x_{it}, b) f(x_{it}|\delta_i)) f(\tilde{\mu}_{xi}|\delta_i) dF_{\delta_i}(\delta) dF_{b_i}(b)} \\ &= \frac{\int_{\delta} (1, \mu_x + \delta)' \boldsymbol{\beta} (\prod_{t=1}^{n_i} f(y_{it}|x_{it}, b) f(x_{it}|\delta_i)) f(\tilde{\mu}_{xi}|\delta_i) dF_{\delta_i}(\delta)}{\int_b \int_{\delta} (\prod_{t=1}^{n_i} f(y_{it}|x_{it}, b) f(x_{it}|\delta_i)) f(\tilde{\mu}_{xi}|\delta_i) dF_{\delta_i}(\delta) dF_{b_i}(b)} \\ &\quad + \frac{\int_b b (\prod_{t=1}^{n_i} f(y_{it}|x_{it}, b) f(x_{it}|\delta_i)) f(\tilde{\mu}_{xi}|\delta_i) dF_{b_i}(b)}{\int_b \int_{\delta} (\prod_{t=1}^{n_i} f(y_{it}|x_{it}, b) f(x_{it}|\delta_i)) f(\tilde{\mu}_{xi}|\delta_i) dF_{\delta_i}(\delta) dF_{b_i}(b)} \\ &= (1, \mu_x + E(\delta_i | (\mathbf{x}_i, \tilde{\mu}_{xi})))' \boldsymbol{\beta} + E(b_i | \mathbf{y}_i, \mathbf{x}_i). \end{aligned} \tag{4.22}$$

#### 4.2.2 Estimation

In practice, the parameters  $\sigma_b^2, \boldsymbol{\beta}, \sigma_{\epsilon}^2, \mu_{xi}$  are not known and need to be estimated. Consider the model specified by (4.1) and (4.11), under the different cases of auxiliary information presented in Section 4.2.1. The parameters  $(\sigma_{\epsilon}^2, \mu_{xi})$  are estimated separately from the parameters  $(\sigma_b^2, \boldsymbol{\beta})$ , see Erciulescu and Fuller (2014).

The estimation of  $(\sigma_b^2, \boldsymbol{\beta})$  is based on maximizing the likelihood for the generalized linear mixed model specified in (4.1). There is no closed form expression for the likelihood and no exact form of the estimating equations. Two commonly used methods of estimation are based on pseudo-likelihood and on integral approximation. In the simulation study, we use the integral approximation method in the *glmer* function, in R, that uses the Laplace approximation to the likelihood.

#### 4.2.2.1 Known $\mu_{xi}$

For the case when  $\mu_{xi}$  is known, the sampling variance  $\sigma_\epsilon^2$  is estimated by the pooled within-area mean squared

$$\hat{\sigma}_\epsilon^2 = \left( \sum_{i=1}^m n_i \right)^{-1} \sum_{i=1}^m \sum_{j=1}^{n_i} (x_{ij} - \mu_{xi})^2. \quad (4.23)$$

#### 4.2.2.2 Unknown Fixed $\mu_{xi}$ , Additional Information $\tilde{\mu}_{xi}$

For the case when  $\mu_{xi}$  is unknown fixed and additional information  $\tilde{\mu}_{xi}$  is observed, the estimated small area mean of  $\mathbf{x}$  is

$$\hat{\mu}_{xi} = \frac{\sigma_u^2 \bar{x}_i + n_i^{-1} \sigma_\epsilon^2 \tilde{\mu}_{xi}}{\sigma_u^2 + n_i^{-1} \sigma_\epsilon^2} = \frac{k^{-1} \bar{x}_i + n_i^{-1} \tilde{\mu}_{xi}}{k^{-1} + n_i^{-1}}, \quad (4.24)$$

and the estimated sampling variance is

$$\hat{\sigma}_\epsilon^2 = \left( \sum_{i=1}^m n_i \right)^{-1} \left( \sum_{i=1}^m \sum_{j=1}^{n_i} (x_{ij} - \hat{\mu}_{xi})^2 + k \sum_{i=1}^m (\tilde{\mu}_{xi} - \hat{\mu}_{xi})^2 \right). \quad (4.25)$$

The estimated variance of  $\hat{\mu}_{xi}$  is

$$\hat{V}(\hat{\mu}_{xi}) = \frac{\hat{\sigma}_u^2 n_i^{-1} \hat{\sigma}_\epsilon^2}{\hat{\sigma}_u^2 + n_i^{-1} \hat{\sigma}_\epsilon^2} = \hat{\sigma}_\epsilon^2 (n_i + k)^{-1}. \quad (4.26)$$

#### 4.2.2.3 Unknown Random $\mu_{xi}$ , Additional Information $\tilde{\mu}_{xi}$

For the case when  $\mu_{xi}$  is unknown random and additional information  $\tilde{\mu}_{xi}$  is observed, the estimation of  $(\sigma_\epsilon^2, \sigma_\delta^2, \mu_x)$  is based on maximizing the likelihood for the linear mixed model specified in (4.11). In the simulation study, we construct the estimated weighted least squares estimator for  $\mu_x$  and REML estimates for  $\sigma_\epsilon^2$  and  $\sigma_\delta^2$ , using the *lmer* function, in R.

### 4.3 Bootstrap MSE Estimators

In this section we consider bootstrap estimation of the MSE of  $\hat{\theta}_i$  as a predictor of  $\theta_i$ . Let

$$\alpha_i = E(\hat{\theta}_i - \theta_i)^2 \quad (4.27)$$

be the prediction MSE for  $\hat{\theta}_i$  and let  $\hat{\alpha}_i$  be an estimator of  $\alpha_i$ . For model (4.1) and known  $\mu_{xi}$ , the prediction MSE for  $\hat{\theta}_i$  is a function of the parameters  $\boldsymbol{\psi} = (\sigma_b^2, \boldsymbol{\beta}, \sigma_\epsilon^2)$ . For model (4.1) and unknown fixed  $\mu_{xi}$ , the prediction MSE for  $\hat{\theta}_i$  is a function of the parameters  $\boldsymbol{\psi} = (\sigma_b^2, \boldsymbol{\beta}, \sigma_\epsilon^2, \mu_{xi})$ . For model (4.1) and unknown random  $\mu_{xi}$ , the prediction MSE for  $\hat{\theta}_i$  is a function of the parameters  $\boldsymbol{\psi} = (\sigma_b^2, \boldsymbol{\beta}, \sigma_\epsilon^2, \mu_x, \sigma_\delta^2)$ . For nonlinear  $g(x_{ij}, \boldsymbol{\beta}, b_i)$ , there is no closed form expression of the type (4.27) for  $\alpha_i$ . In this section we consider bootstrap estimation of  $\alpha_i$ . Hall and Maiti (2006) showed that

$$E(\hat{\alpha}_i) = \alpha_i + m^{-1}B_i(\boldsymbol{\psi}) + O(m^{-2}), \quad (4.28)$$

where  $B_i$  is a smooth function of its arguments. The bootstrap estimate of the distribution of  $\boldsymbol{\psi}$  is consistent for the distribution of  $\boldsymbol{\psi}$ . Therefore, the bootstrap estimate of the distribution of  $\alpha_i$  is consistent for the distribution of  $\alpha_i$  and a smooth function of  $\hat{\boldsymbol{\psi}}$ , and the conditions for use of the bootstrap hold for  $\alpha_i$ .

Define  $DG(\boldsymbol{\psi}, r)$ , for the vector of parameters  $\boldsymbol{\psi}$  and the random number seed  $r$ , to be the generator for a sample from the model of interest. In the level one bootstrap,  $B_1$  samples are generated using the vector of random number seeds  $\mathbf{r}_1 = \{r_{1,1}, r_{1,2}, \dots, r_{1,B_1}\}$ . Let  $\hat{\boldsymbol{\psi}}$  be an estimator of  $\boldsymbol{\psi}$  for the original sample and let  $\boldsymbol{\psi}_k^*$  be the estimator of  $\boldsymbol{\psi}$  from the  $k$ th bootstrap sample generated using  $DG(\hat{\boldsymbol{\psi}}, r_{1,k})$ . Let  $\theta_{i,k}^*$  be the true small area mean for the  $k$ th bootstrap sample and let  $\hat{\theta}_{i,k}^*$  be the sample predictor of  $\theta_{i,k}^*$ . Then, the level one bootstrap estimator of the prediction MSE for area  $i$  is

$$\hat{\alpha}_i^* = B_1^{-1} \sum_{k=1}^{B_1} (\hat{\theta}_{i,k}^* - \theta_{i,k}^*)^2 =: B_1^{-1} \sum_{k=1}^{B_1} \alpha_{i,k}^* = \bar{\alpha}_i^*, \quad (4.29)$$

where  $\alpha_{i,k}^*$  is the prediction squared error for the  $k$ th bootstrap sample.

In the double bootstrap, a sample estimator  $\alpha_{i,k,t}^{**}$  is generated using  $\psi_k^*$  from the level one generated sample  $k = 1, 2, \dots, B_1$ , for  $t = 1, 2, \dots, B_2$ , where  $B_2$  is the number of level two bootstrap samples per level one sample. Typically a large number of  $\alpha_{i,k,t}^{**}$  is generated for each  $\alpha_{i,k}^*$  and the bias adjusted estimator is

$$\tilde{\alpha}_i^{**} = B_1^{-1} \sum_{k=1}^{B_1} (2\alpha_{i,k}^* - B_2^{-1} \sum_{t=1}^{B_2} \alpha_{i,k,t}^{**}). \quad (4.30)$$

where  $\alpha_{i,k,t}^{**}$  is generated using  $DG(\psi_k^*, r_{2,k,t})$  and the  $r_{2,k,t}$ ,  $k = 1, 2, \dots, B_1$ ,  $t = 1, 2, \dots, B_2$ , are independent random numbers, independent of  $r_{1,k}$ . The estimator (4.30) is called the classic double bootstrap estimator of  $\alpha_i$  and the bootstrap bias correction method is called the difference bootstrap bias correction method.

Based on the work of Davidson and MacKinnon (2007) and Giacomini, Politis and White (2013), we construct a classic fast double bootstrap estimator of  $\alpha_i$ , with  $B_2 = 1$ . Let  $r_{2,k}$  be a second sequence of independent random numbers, independent of  $r_{1,k}$ , for  $k = 1, 2, \dots, B_1$  and define  $\alpha_{i,k}^{**}$  to be calculated from data generated with  $DG(\psi_k^*, r_{2,k})$ . Notice that a single  $\alpha_{i,k}^{**}$  is generated for each  $\psi_{i,k}^*$ , for  $k = 1, 2, \dots, B_1$ . The classic fast double bootstrap estimator of the prediction MSE for area  $i$  is

$$\tilde{\alpha}_{i,C}^{**} = B_1^{-1} \sum_{k=1}^{B_1} (2\alpha_{i,k}^* - \alpha_{i,k}^{**}) = 2\bar{\alpha}_i^* - \bar{\alpha}_i^{**}. \quad (4.31)$$

Erciulescu and Fuller (2014) proposed a more efficient double bootstrap estimator,

$$\tilde{\alpha}_{i,T}^{**} = B_1^{-1} \sum_{k=1}^{B_1} (\alpha_{i,k}^* + \alpha_{i,k+1}^* - \alpha_{i,k}^{**}), \quad (4.32)$$

where  $\alpha_{i,k+1}^*$  is generated with  $DG(\hat{\psi}, r_{1,k+1})$  and  $\alpha_{i,k}^{**}$  is generated with  $DG(\psi_k^*, r_{1,k+1})$ . They called the estimator (4.32) a telescoping fast double bootstrap estimator.

Let  $\hat{\alpha}_i^{Tay}$  be an alternative estimator of  $\alpha_i$ , typically based on Taylor expansions. For example, for a linear function  $g(x_{ij}, \beta, b_i)$ ,  $\hat{\alpha}_i^{Tay}$  could be the Taylor estimator given in (4.9). Let  $\alpha_{i,k}^{Tay*}$ , for  $k = 1, 2, \dots, B_1$ , be the level one bootstrap values for  $\hat{\alpha}_i^{Tay}$ . Then, the level one

bootstrap estimated bias in  $\hat{\alpha}_i^{Tay}$  is

$$\hat{\Delta}_{\alpha^{Tay},i}^* = B_1^{-1} \sum_{k=1}^{B_1} (\alpha_{i,k}^{Tay*} - \hat{\alpha}_i^{Tay}), \quad (4.33)$$

and a bias corrected level one bootstrap estimator of the prediction MSE for area  $i$  is obtained by subtracting the estimated bias in  $\hat{\alpha}_i^{Tay}$  from the bootstrap estimator of the prediction MSE for area  $i$  to give

$$\hat{\alpha}_{i,\Delta}^* = \bar{\alpha}_i^* - \hat{\Delta}_{\alpha^{Tay},i}^*. \quad (4.34)$$

Let  $\alpha_{i,k}^{Tay**,C}$  and  $\alpha_{i,k}^{Tay**,T}$ , for  $k = 1, 2, \dots, B_1$ , be the level two classic and telescoping bootstrap values for  $\hat{\alpha}_i^{Tay}$ , respectively. The level two bootstrap estimated bias in  $\hat{\alpha}_i^{Tay}$ , for the classic fast double bootstrap procedure, is

$$\hat{\Delta}_{\alpha^{Tay},i}^{**,C} = B_1^{-1} \sum_{k=1}^{B_1} (3\alpha_{i,k}^{Tay*} - \alpha_{i,k}^{Tay**,C} - 2\hat{\alpha}_i^{Tay}), \quad (4.35)$$

and a classic fast double bootstrap estimator of the prediction MSE for area  $i$  is

$$\hat{\alpha}_{i,\Delta}^{**,C} = \bar{\alpha}_i^* - \hat{\Delta}_{\alpha^{Tay},i}^{**,C}. \quad (4.36)$$

Similarly, the level two bootstrap estimated bias in  $\hat{\alpha}_i^{Tay}$ , for the fast telescoping double bootstrap procedure, is

$$\hat{\Delta}_{\alpha^T,i}^{**,Tay} = B_1^{-1} \sum_{k=1}^{B_1} (3\alpha_{i,k}^{Tay*} - \alpha_{i,k}^{Tay**,T} - 2\hat{\alpha}_i^{Tay}), \quad (4.37)$$

and a fast telescoping double bootstrap estimator of the prediction MSE for area  $i$  is

$$\hat{\alpha}_{i,\Delta}^{**,T} = \bar{\alpha}_i^* - \hat{\Delta}_{\alpha^T,i}^{**,Tay}. \quad (4.38)$$

#### 4.4 Linear Approximation of the ULGLMM

In this section, we construct a linear approximation of the model (4.1) and use the approximation to approximate the predicted small area mean of  $\mathbf{y}$ . The method is illustrated for the general case when  $\mu_{xi}$  is unknown and random, and additional information  $\tilde{\mu}_{xi}$  is observed.

A Taylor approximation for the function  $g(x_{ij}, \boldsymbol{\beta}, b)$  expanded about  $(\hat{\boldsymbol{\beta}}, 0)$  is

$$y_{ij} \approx g(x_{ij}, \hat{\boldsymbol{\beta}}, 0) + \mathbf{h}_{\boldsymbol{\beta}, ij}(x_{ij}, \hat{\boldsymbol{\beta}}, 0)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + h_{b_i, ij}(x_{ij}, \hat{\boldsymbol{\beta}}, 0)b_i + e_{ij}, \quad (4.39)$$

where  $(\hat{\boldsymbol{\beta}}', \hat{\sigma}_b^2)$  are estimates of  $(\boldsymbol{\beta}', \sigma_b^2)$  based on model (4.1),  $\mathbf{h}_{\boldsymbol{\beta}, ij}(x_{ij}, \hat{\boldsymbol{\beta}}, 0)$  is the partial derivative of  $g(x_{ij}, \boldsymbol{\beta}, b_i)$  with respect to  $\boldsymbol{\beta}$ , evaluated at  $(x_{ij}, \hat{\boldsymbol{\beta}}, 0)$  and  $h_{b_i, ij}(x_{ij}, \hat{\boldsymbol{\beta}}, 0)$  is the partial derivative of  $g(x_{ij}, \boldsymbol{\beta}, b_i)$  with respect to  $b_i$ , evaluated at  $(x_{ij}, \hat{\boldsymbol{\beta}}, 0)$ .

Suppose that the distribution  $F_{\mathbf{x}_i}(x)$  is defined by the first two moments of  $\mathbf{x}$ . Suppose that the integral in (4.2) can be approximated numerically, using a set of numbers  $z_k, k = 1, 2, \dots, K$  and a set of weights  $w_k, k = 1, 2, \dots, K$ , by

$$\theta_i \approx \sum_{k=1}^K w_k g(\mu_{xi} + z_k \sigma_\epsilon, \boldsymbol{\beta}, b_i). \quad (4.40)$$

For  $F_{\mathbf{x}_i}(x)$  the normal distribution, the numerical approximation is described in Appendix D.

To simplify the approximation (4.39) and later approximations, we replace  $\mathbf{h}_{\boldsymbol{\beta}, ij}(x_{ij}, \hat{\boldsymbol{\beta}}, 0)$  and  $h_{b_i, ij}(x_{ij}, \hat{\boldsymbol{\beta}}, 0)$  by averages,

$$y_{ij} \approx g(x_{ij}, \hat{\boldsymbol{\beta}}, 0) + \bar{h}_{\boldsymbol{\beta}, i}(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \bar{h}_{b_i, i}(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0)b_i + e_{ij}, \quad (4.41)$$

where  $(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon)$  is an estimator of  $(\mu_{xi}, \sigma_\epsilon)$  and

$$\begin{aligned} \bar{h}_{\boldsymbol{\beta}, i}(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0) &:= \sum_{k=1}^K w_k \frac{\partial g(\mu_{xi} + z_k \sigma_\epsilon, \boldsymbol{\beta}, b_i)}{\partial \boldsymbol{\beta}} \Big|_{(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0)} \\ \bar{h}_{b_i, i}(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0) &:= \sum_{k=1}^K w_k \frac{\partial g(\mu_{xi} + z_k \sigma_\epsilon, \boldsymbol{\beta}, b_i)}{\partial b_i} \Big|_{(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0)}. \end{aligned} \quad (4.42)$$

Using expression (4.41), the predicted random area effect for area  $i$  is

$$\bar{h}_{b_i, i}(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0)\hat{b}_i = \tilde{\gamma}_i \bar{u}_{yi}, \quad (4.43)$$



where

$$\tilde{\gamma}_i = \left( \bar{h}_{b,i}^2(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0) \hat{\sigma}_b^2 + n_i^{-1} \hat{\sigma}_{ei}^2 \right)^{-1} \bar{h}_{b,i}^2(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0) \hat{\sigma}_b^2,$$

$$u_{y,ij} = y_{ij} - g(x_{ij}, \hat{\boldsymbol{\beta}}, 0),$$

$\bar{u}_{yi} = n_i^{-1} \sum_{j=1}^{n_i} u_{y,ij}$  and  $\hat{\sigma}_{ei}^2$  is the estimated model sampling variance of  $\mathbf{y}$  based on the linear approximation in (4.41).

Using a Taylor approximation for the function  $g(x_{ij}, \boldsymbol{\beta}, b)$ , with respect to the parameters  $(\boldsymbol{\beta}, \sigma_\epsilon, b_i, \mu_{xi})$ , about  $(\hat{\boldsymbol{\beta}}, \hat{\sigma}_\epsilon, 0, \hat{\mu}_{xi})$ , the small area mean of  $\mathbf{y}$  in (4.40) is approximated by

$$\begin{aligned} \theta_i \approx & \sum_{k=1}^K w_k g(\hat{\mu}_{xi} + z_k \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0) + \bar{h}_{\mu_{xi}}(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0) (\mu_{xi} - \hat{\mu}_{xi}) \\ & + \bar{h}_{\sigma_\epsilon, i}(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0) (\sigma_\epsilon - \hat{\sigma}_\epsilon) + \bar{h}_{\boldsymbol{\beta}, i}(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0) (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \bar{h}_{b, i}(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0) b_i, \end{aligned} \quad (4.44)$$

where

$$\begin{aligned} \bar{h}_{\mu_{xi}}(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0) &:= \sum_{k=1}^K w_k \frac{\partial g(\mu_{xi} + z_k \sigma_\epsilon, \boldsymbol{\beta}, b_i)}{\partial \mu_{xi}} \Big|_{(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0)}, \\ \bar{h}_{\sigma_\epsilon, i}(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0) &:= \sum_{k=1}^K w_k \frac{\partial g(\mu_{xi} + z_k \sigma_\epsilon, \boldsymbol{\beta}, b_i)}{\partial \sigma_\epsilon} \Big|_{(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0)}, \end{aligned} \quad (4.45)$$

and  $\bar{h}_{\boldsymbol{\beta}, i}(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0)$  and  $\bar{h}_{b, i}(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0)$  are defined in (4.42).

The small area mean prediction

$$\hat{\theta}_i = \sum_{k=1}^K w_k g(\hat{\mu}_{xi} + z_k \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, \hat{b}_i), \quad (4.46)$$

$$\approx \sum_{k=1}^K w_k g(\hat{\mu}_{xi} + z_k \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0) + \bar{h}_{b, i}(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0) \hat{b}_i,$$

for  $\hat{b}_i$  defined in (4.43).

In (4.44), the mean of the dependent variable  $y_i$  is approximated by a linear function of  $(\sigma_\epsilon, \mu_{xi}, \boldsymbol{\beta}, b_i)$ . Therefore, expressions analogous to (4.8) and (4.9) can be derived for the prediction mean squared error, and for the estimated prediction mean squared error, respectively.

Using the approximations in (4.44) and (4.46), the prediction error  $\hat{\theta}_i - \theta_i$  is approximated by

$$\begin{aligned} \hat{\theta}_i - \theta_i &= \bar{h}_{\mu_{xi}}(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0)(\hat{\mu}_{xi} - \mu_{xi}) + \bar{h}_{\sigma_{\epsilon,i}}(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0)(\hat{\sigma}_\epsilon - \sigma_\epsilon) \\ &\quad + \bar{\mathbf{h}}_{\boldsymbol{\beta},i}(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + \bar{h}_{b,i}(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0)(\hat{b}_i - b_i). \end{aligned} \quad (4.47)$$

The variance of the approximation (4.47) depends on the model and the parameter estimates. The vector  $(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}})$  is constructed as described in Section 4.2.2. The  $\bar{u}_{yi}^2$ , defined in (4.43), is an approximate estimator for  $\bar{h}_{b,i}^2(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0)\sigma_b^2 + n_i^{-1}\sigma_{\epsilon_i}^2$ . A Taylor MSE estimator for the prediction error (4.47) of type (4.9) with  $k_{ij} = 1$  is

$$M\hat{S}E(\hat{\theta}_i - \theta_i) = \hat{g}_{1i}(\hat{\sigma}_b^2, \tilde{\sigma}_{\epsilon_i}^2) + \hat{g}_{2i}(\hat{\sigma}_b^2, \tilde{\sigma}_{\epsilon_i}^2) + 2\hat{g}_{3i}(\hat{\sigma}_b^2, \tilde{\sigma}_{\epsilon_i}^2), \quad (4.48)$$

where

$$\hat{g}_{1i}(\hat{\sigma}_b^2, \tilde{\sigma}_{\epsilon_i}^2) = \tilde{\gamma}_i n_i^{-1} \tilde{\sigma}_{\epsilon_i}^2,$$

$$\begin{aligned} \hat{g}_{2i}(\hat{\sigma}_b^2, \tilde{\sigma}_{\epsilon_i}^2) &= (1 - \tilde{\gamma}_i)^2 \bar{\mathbf{h}}'_{\boldsymbol{\beta},i}(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0) \hat{V}(\hat{\boldsymbol{\beta}}) \bar{\mathbf{h}}_{\boldsymbol{\beta},i}(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0) + \bar{h}_{\sigma_{\epsilon,i}}^2(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0) \hat{V}(\hat{\sigma}_\epsilon) \\ &\quad + \bar{h}_{\mu_{xi}}^2(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0) \hat{V}(\hat{\mu}_{xi}), \end{aligned}$$

$$\hat{g}_{3i}(\hat{\sigma}_b^2, \tilde{\sigma}_{\epsilon_i}^2) = \frac{\left( \bar{h}_{b,i}^4(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0) \hat{\sigma}_b^4 n_i^{-2} \hat{V}(\hat{\sigma}_{\epsilon_i}^2) + n_i^{-2} \hat{\sigma}_{\epsilon_i}^4 \hat{V}(\bar{h}_{b,i}^2(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0) \hat{\sigma}_b^2) \right) \bar{u}_{yi}^2}{\left( \bar{h}_{b,i}^2(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0) \hat{\sigma}_b^2 + n_i^{-1} \tilde{\sigma}_{\epsilon_i}^2 \right)^4},$$

$$\hat{V}(\bar{h}_{b,i}^2(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0) \hat{\sigma}_b^2) = \bar{h}_{b,i}^4(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0) \hat{V}(\hat{\sigma}_b^2),$$

$\tilde{\gamma}_i$  and  $\tilde{\sigma}_{\epsilon_i}^2$  are defined for (4.43),  $\bar{\mathbf{h}}_{\boldsymbol{\beta},i}(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0)$  and  $\bar{h}_{b,i}(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0)$  are defined in (4.42),  $\bar{h}_{\mu_{xi}}(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0)$  and  $\bar{h}_{\sigma_{\epsilon,i}}(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0)$  are defined in (4.45),  $\hat{V}(\hat{\sigma}_\epsilon)$  is the estimated variance of  $\hat{\sigma}_\epsilon$ ,  $\hat{V}(\hat{\boldsymbol{\beta}})$  is the estimated covariance matrix of  $\hat{\boldsymbol{\beta}}$ ,  $\hat{V}(\hat{\sigma}_{\epsilon_i}^2)$  is the estimated variance of the  $\hat{\sigma}_{\epsilon_i}^2$ ,  $\hat{V}(\hat{\mu}_{xi})$  is the estimated variance of  $\hat{\mu}_{xi}$  and  $\hat{V}(\hat{\sigma}_b^2)$  is the estimated variance of  $\hat{\sigma}_b^2$ .

For model (4.1) with unknown  $\mu_{xi}$ , the area mean of  $\mathbf{x}$ ,  $\mu_{xi}$ , is estimated/predicted based on the linear mixed model for  $\tilde{\mathbf{x}}$  given in (4.11), as described in Section 4.2.2. Given the independence assumptions for the random area effects  $\mathbf{b}$ , the sampling errors  $\boldsymbol{\epsilon}$  and the observations  $\mathbf{x}$ ,

described for model (4.1), the variance covariance matrix for the vector of area level information  $(\bar{u}_{y,i}, \bar{x}_i)$  has zero off-diagonal elements, where  $(\bar{u}_{y,i}, \bar{x}_i) = (n_i^{-1} \sum_{j=1}^{n_i} u_{y,ij}, n_i^{-1} \sum_{j=1}^{n_i} \tilde{x}_{ij})$ , for  $u_{y,ij}$  defined for (4.43) and  $\tilde{x}_{ij}$  defined for model (4.11). Moreover, the estimated vector of parameters  $(\hat{\beta}, \hat{\sigma}_b^2)$  is independent from the estimated  $\sigma_\epsilon^2$ , as described in Section 4.2.2. Therefore, the covariance terms associated with the estimation of  $(\mu_{xi}, \sigma_\epsilon, \beta, b_i)$  are zero in expression (4.48).

If we ignore the estimation error in  $\hat{\sigma}_\epsilon^2$  and the estimation error in  $\tilde{\sigma}_{ei}^2$  in constructing the Taylor estimators (4.15, 4.48), the terms  $\hat{g}_{2i}(\hat{\sigma}_b^2, \tilde{\sigma}_{ei}^2), \hat{g}_{3i}(\hat{\sigma}_b^2, \tilde{\sigma}_{ei}^2)$  in (4.48) become

$$\hat{g}_{2i}(\hat{\sigma}_b^2, \tilde{\sigma}_{ei}^2) = (1 - \tilde{\gamma}_i)^2 \bar{\mathbf{h}}_{\beta,i}'(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\beta}, 0) \hat{V}(\hat{\beta}) \bar{\mathbf{h}}_{\beta,i}(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\beta}, 0) + \bar{h}_{\mu_{xi}}^2(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\beta}, 0) \hat{V}(\hat{\mu}_{xi}) \quad (4.49)$$

with

$$\hat{V}(\hat{\mu}_{xi}) = \hat{\gamma}_{ix} a_{i,x}^{-1} \hat{\sigma}_\epsilon^2 + (1 - \hat{\gamma}_{ix})^2 V(\hat{\mu}_x) + 2(\hat{\sigma}_\delta^2 + a_{i,x}^{-1} \hat{\sigma}_\epsilon^2)^{-4} a_{i,x}^{-2} \hat{\sigma}_\epsilon^4 V(\hat{\sigma}_\delta^2) \bar{u}_{ix}^2$$

and

$$\hat{g}_{3i}(\hat{\sigma}_b^2, \tilde{\sigma}_{ei}^2) = \frac{(n_i^{-1} \tilde{\sigma}_{ei}^2)^2 \bar{h}_{b,i}^4(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\beta}, 0) \hat{V}(\hat{\sigma}_b^2) \bar{u}_{yi}^2}{\left( \bar{h}_{b,i}^2(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\beta}, 0) \hat{\sigma}_b^2 + n_i^{-1} \tilde{\sigma}_{ei}^2 \right)^4}, \quad (4.50)$$

respectively. The approximate estimated variance of the ML (or REML) estimated variance components  $\hat{\sigma}_\delta^2$  and  $\hat{\sigma}_b^2$  are

$$\hat{V}(\hat{\sigma}_\delta^2) = 2(m-1)^{-1} m \left[ \sum_{i=1}^m (\hat{\sigma}_\delta^2 + a_{i,x}^{-1} \hat{\sigma}_\epsilon^2)^{-2} \right]^{-1}, \quad (4.51)$$

and

$$\hat{V}(\hat{\sigma}_b^2) = 2(m-1)^{-1} m \left[ \sum_{i=1}^m \bar{h}_{b,i}^4(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\beta}, 0) \left( \bar{h}_{b,i}^2(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\beta}, 0) \hat{\sigma}_b^2 + n_i^{-1} \tilde{\sigma}_{ei}^2 \right)^{-2} \right]^{-1}, \quad (4.52)$$

where  $a_{i,x}$  is defined for (4.14). The estimators are based on the expression (21) given in Wang and Fuller (2003). The steps are illustrated in Appendix A.

The approximation (4.39) is not a typical bivariate Taylor expansion of the  $g(x_{ij}, \beta, b_i)$  function because it has an approximation error of  $O(1)$ . However,  $\bar{\mathbf{h}}_{\mu_{xi}}(\mu_{xi}, \beta, \sigma_\epsilon, 0)$ ,

$\bar{\mathbf{h}}_{\boldsymbol{\beta},i}(\mu_{xi}, \boldsymbol{\beta}, \sigma_\epsilon, 0)$ ,  $\bar{\mathbf{h}}_{b_i,i}(\mu_{xi}, \boldsymbol{\beta}, \sigma_\epsilon, 0)$  and  $\bar{\mathbf{h}}_{\sigma_\epsilon,i}(\mu_{xi}, \boldsymbol{\beta}, \sigma_\epsilon, 0)$  are continuous differentiable functions of the parameters  $(\mu_{xi}, \boldsymbol{\beta}, \sigma_\epsilon)$ . Therefore, given root- $m$  consistent estimators

$$(\hat{\mu}_{xi}, \hat{\boldsymbol{\beta}}, \hat{\sigma}_\epsilon, \hat{\sigma}_b, \hat{\sigma}_e) - (\mu_{xi}, \boldsymbol{\beta}, \sigma_\epsilon, \sigma_b, \sigma_e) = O_p(m^{-0.5}),$$

the expression for  $M\hat{S}E(\hat{\theta}_i - \theta_i)$  of (4.48) converges to a limiting quantity. Furthermore, the difference between  $M\hat{S}E(\hat{\theta}_i - \theta_i)$  defined by (4.48) and its limiting value is  $O_p(m^{-0.5})$ . Hence, the conditions required to use the bootstrap to estimate the quantiles of the distribution of (4.54) below with  $\hat{V}(\hat{\theta}_i - \theta_i) = M\hat{S}E(\hat{\theta}_i - \theta_i)$  defined by (4.48) hold.

#### 4.5 Symmetric two-sided $(1 - \alpha)$ -level CI

Let model (4.1) hold, let  $\theta_i$  be the small area mean defined in (4.2) and let  $\hat{\theta}_i$  be the predicted small area mean defined in (4.17,4.19,4.21) for different cases of auxiliary information. In this section we consider the general model with  $\mu_{xi}$  unknown random and construct a symmetric two-sided  $(1 - \alpha)$ -level CI for the area mean  $\theta_i$ . The confidence level  $(1 - \alpha)$  represents the percentage of the hypothetically observed CIs that would hold the true value of the area mean  $\theta_i$ . We consider the percentile method and the pivot-like method to construct CIs. The percentile CIs are based on the bootstrap distributions of the bootstrap prediction errors  $\hat{\theta}_{i,k}^* - \theta_{i,k}^*$ , and  $\hat{\theta}_{i,k}^{**} - \theta_{i,k}^{**}$ . The pivot-type, two-sided, symmetric, CI of nominal level  $1 - \alpha$  is of the form

$$\left( \hat{\theta}_i \pm \hat{\zeta}_{i,\alpha} \sqrt{\hat{V}(\hat{\theta}_i - \theta_i)} \right), \quad (4.53)$$

where the notation defines the interval  $\left( \hat{\theta}_i - \hat{\zeta}_{i,\alpha} \sqrt{\hat{V}(\hat{\theta}_i - \theta_i)}, \hat{\theta}_i + \hat{\zeta}_{i,\alpha} \sqrt{\hat{V}(\hat{\theta}_i - \theta_i)} \right)$ ,  $\hat{\zeta}_{i,\alpha}$  is determined by the procedure and  $\hat{V}(\hat{\theta}_i - \theta_i)$  is an estimator of the variance of  $\hat{\theta}_i - \theta_i$ . We consider bootstrap estimators and Taylor estimators of  $\hat{V}(\hat{\theta}_i - \theta_i)$ . The cutoff point  $\zeta_{i,\alpha}$  is specific to the area  $i$  and to  $\alpha$ . Different methods of estimating the cutoff points  $\zeta_{i,\alpha}$  are presented next.

### 4.5.1 Wald-type Symmetric $(1 - \alpha)$ -level CI for the Small Area Mean

The pivot-type CIs are constructed using the statistic

$$\hat{T}_i := \frac{\hat{\theta}_i - \theta_i}{\sqrt{\hat{V}(\hat{\theta}_i - \theta_i)}}. \quad (4.54)$$

Given  $\hat{V}(\hat{\theta}_i - \theta_i)$ , one can approximate the distribution of  $\hat{T}_i$  by the standard normal distribution and estimate the cutoff point  $\zeta_{i,\alpha}$  by the  $100(1 - \alpha)\%$  standard normal distribution quantile,  $\hat{\zeta}_{i,\alpha,z}$ . Let

$$I_W = \left( \hat{\theta}_i \pm \hat{\zeta}_{i,\alpha,z} \sqrt{\hat{V}(\hat{\theta}_i - \theta_i)} \right) \quad (4.55)$$

denote the basic Wald-type confidence interval for the area mean  $\theta_i$ .

In the next subsections, bootstrap methods to estimate  $\zeta_{i,\alpha}$  are described.

### 4.5.2 Level one Bootstrap Symmetric $(1 - \alpha)$ -level CI for the Small Area Mean

In the level one bootstrap,  $B_1$  bootstrap samples are generated using the estimated parameters  $(\hat{\sigma}_b^2, \hat{\beta}, \hat{\mu}_x, \hat{\sigma}_\delta^2, \hat{\sigma}_\epsilon^2)$  and a set of random seeds  $\mathbf{r}_1$ . The bootstrap small area mean  $\theta_i^*$  is computed using (4.2). The parameter estimates of  $(\hat{\sigma}_b^2, \hat{\beta}, \hat{\mu}_{xi}, \hat{\sigma}_\epsilon^2)$  for a bootstrap sample  $k, k = 1, \dots, B_1$ , are denoted by  $(\sigma_{bk}^{2*}, \beta_k^*, \mu_{xik}^*, \sigma_{\epsilon k}^{2*})$ . The predicted small area mean for the bootstrap sample  $k$  is denoted by  $\hat{\theta}_{ik}^*$ , and computed using (4.21).

#### 4.5.2.1 Percentile Method

The distribution of the prediction error  $\hat{\theta}_i - \theta_i$  is estimated by the distribution of the bootstrap prediction errors  $\hat{\theta}_{i,k}^* - \theta_{i,k}^*, k = 1, \dots, B_1$ . A percentile symmetric, two-sided, level one bootstrap CI with nominal coverage  $(1 - \alpha)$  is

$$I_{p1} := \left( \hat{\theta}_i \pm \hat{\zeta}_{p,i,\alpha} \right), \quad (4.56)$$

where the cutoff point is the  $100(1 - \alpha)^{th}$  quantile of the bootstrap estimated distribution of the absolute value of the level one bootstrap prediction error,

$$\hat{\zeta}_{p,i,\alpha} = |\hat{\theta}^* - \theta^*|_{i,([ (1-\alpha)B_1 ] + 1)},$$

where we denote the integer-part function by  $[.]$  and the ordered value by  $(.)$ .

#### 4.5.2.2 Pivot-like Method

For the  $k^{th}$  bootstrap sample, let the estimated prediction MSE for  $\hat{\theta}_{ik}^*$  be as in (4.48) and denoted by  $(\sigma_{i,Taylor}^*)_k^2$ , where

$$(\sigma_{i,Taylor}^*)_k^2 = \hat{g}_{1i}(\sigma_{bk}^{2*}, \sigma_{eik}^{2*}) + \hat{g}_{2i}(\sigma_{bk}^{2*}, \sigma_{eik}^{2*}) + 2\hat{g}_{3i}(\sigma_{bk}^{2*}, \sigma_{eik}^{2*}), \quad (4.57)$$

and

$$\begin{aligned} \hat{g}_{1i}(\sigma_{bk}^{2*}, \sigma_{eik}^{2*}) &= \gamma_{ik}^* n_i^{-1} \sigma_{eik}^{2*}, \\ \hat{g}_{2i}(\sigma_{bk}^{2*}, \sigma_{eik}^{2*}) &= (1 - \gamma_{ik}^*)^2 \bar{\mathbf{h}}'_{\beta,i}(\mu_{xik}^*, \sigma_{\epsilon k}^*, \beta_k^*, 0) \hat{V}(\beta_k^*) \bar{\mathbf{h}}_{\beta,i}(\mu_{xik}^*, \sigma_{\epsilon k}^*, \beta_k^*, 0) \\ &\quad + \bar{h}_{\sigma_{\epsilon,i}}^2(\mu_{xik}^*, \sigma_{\epsilon k}^*, \beta_k^*, 0) \hat{V}(\sigma_{\epsilon}^*) + \bar{h}_{\mu_{xi}}^2(\mu_{xi}^*, \sigma_{\epsilon}^*, \beta_k^*, 0) \hat{V}(\mu_{xik}^*), \\ \hat{g}_{3i}(\sigma_{bk}^{2*}, \sigma_{eik}^{2*}) &= \frac{(\bar{h}_{b,i}^4(\mu_{xik}^*, \sigma_{\epsilon k}^*, \beta_k^*, 0) \sigma_{bk}^{4*} n_i^{-2} \hat{V}(\sigma_{eik}^{2*})) \bar{u}_{yik}^{2*}}{(\bar{h}_{b,i}^2(\mu_{xik}^*, \sigma_{\epsilon k}^*, \beta_k^*, 0) \sigma_{bk}^{2*} + n_i^{-1} \sigma_{eik}^{2*})^4} \\ &\quad + \frac{(n_i^{-2} \sigma_{eik}^{4*} \hat{V}(\bar{h}_{b,i}^2(\mu_{xik}^*, \sigma_{\epsilon k}^*, \beta_k^*, 0) \sigma_b^{2k*})) \bar{u}_{yik}^{2*}}{(\bar{h}_{b,i}^2(\mu_{xik}^*, \sigma_{\epsilon k}^*, \beta_k^*, 0) \sigma_{bk}^{2*} + n_i^{-1} \sigma_{eik}^{2*})^4}, \end{aligned}$$

$$\hat{V}(\bar{h}_{b,i}^2(\mu_{xik}^*, \sigma_{\epsilon k}^*, \beta_k^*, 0) \sigma_b^{2k*}) = \bar{h}_{b,i}^4(\mu_{xik}^*, \sigma_{\epsilon k}^*, \beta_k^*, 0) \hat{V}(\sigma_b^{2k*}),$$

where  $\gamma_{ik}^*$  and  $\sigma_{eik}^{2*}$  are defined for (4.43),  $\bar{\mathbf{h}}_{\beta,i}(\mu_{xik}^*, \sigma_{\epsilon k}^*, \beta_k^*, 0)$  and  $\bar{h}_{b,i}(\mu_{xik}^*, \sigma_{\epsilon k}^*, \beta_k^*, 0)$  are defined in (4.42),  $\bar{h}_{\mu_{xi}}(\mu_{xi}^*, \sigma_{\epsilon k}^*, \beta_k^*, 0)$  and  $\bar{h}_{\sigma_{\epsilon,i}}(\mu_{xik}^*, \sigma_{\epsilon k}^*, \beta_k^*, 0)$  are defined in (4.45),  $\hat{V}(\sigma_{\epsilon k}^*)$  is the estimated variance of  $\sigma_{\epsilon k}^*$  based on (4.41), and  $\hat{V}(\beta_k^*)$  is the estimated covariance matrix of

the estimated parameters  $\beta_k^*$  based on (4.1),  $\hat{V}(\sigma_{\epsilon ik}^{2*})$  is the estimated variance of the  $\sigma_{\epsilon ik}^{2*}$  and  $\hat{V}(\mu_{xik}^*)$  is the estimated variance of  $\mu_{xik}^*$ ,  $\hat{V}(\sigma_b^{2k*})$  is the estimated variance of  $\sigma_b^{2k*}$ . Estimation of the parameters  $(\mu_{xik}^*, \sigma_{\epsilon k}^*, \beta_k^*)$  and of the variance of  $\mu_{xik}^*$  is specific to the model and described in Section 4.2.2.

Let

$$\hat{T}_{i,k}^* = \frac{\hat{\theta}_{i,k}^* - \theta_{i,k}^*}{\sqrt{\hat{V}(\hat{\theta}_{i,k}^* - \theta_{i,k}^*)}}, \quad (4.58)$$

where  $\hat{V}(\hat{\theta}_{i,k}^* - \theta_{i,k}^*) = (\sigma_{i,Taylor}^*)_k^2$  be the level one bootstrap pivot-like statistic.

The distribution of (4.54) is estimated by the distribution of (4.58) and the symmetric, two-sided, level one bootstrap CI, based on the pivot-like statistic  $T_{ik}^*$ , is

$$I_1 := \left( \hat{\theta}_i \pm \hat{\zeta}_{i,\alpha,B} \hat{\sigma}_{i,Taylor} \right), \quad (4.59)$$

where the cutoff point  $\zeta_{i,\alpha}$  is estimated by the  $100(1 - \alpha)^{th}$  quantile of bootstrap estimated distribution of the absolute value of the level one bootstrap statistic,

$$\hat{\zeta}_{i,\alpha,B} = |T^*|_{i,((1-\alpha)B_1+1)},$$

and  $\hat{\sigma}_{i,Taylor} = \sqrt{\hat{V}(\hat{\theta}_i - \theta_i)}$  is the square root of the Taylor MSE estimator defined in (4.48).

### 4.5.3 Double Bootstrap Symmetric $(1 - \alpha)$ -level CI for the Small Area Mean

In the fast double bootstrap, one bootstrap sample is generated for each set of estimated parameters  $(\sigma_{bk}^{2*}, \beta_k^*, \mu_{xk}^*, \sigma_{\delta k}^{2*}, \sigma_{\epsilon}^{2*})$ . The double bootstrap small area mean  $\theta_{i,k}^{**}$  is computed using (4.2). The parameter estimates of  $(\hat{\sigma}_b^2, \hat{\beta}, \hat{\mu}_{xi}, \hat{\sigma}_{\epsilon}^2)$  for the double bootstrap sample generated for the level one bootstrap sample  $k$  are denoted by  $(\sigma_{bk}^{2**}, \beta_k^{**}, \mu_{xik}^{**}, \sigma_{\epsilon k}^{2**})$ . The predicted small area mean for the double bootstrap sample generated for the level one bootstrap sample  $k$  is denoted by  $\hat{\theta}_{ik}^{**}$ , and computed using (4.21).

### 4.5.3.1 Percentile Method

For bootstrap sample  $k$ , let the double bootstrap prediction errors be  $\hat{\theta}_{i,k}^{**} - \theta_{i,k}^{**}$ ,  $k = 1, \dots, B_1$ . The double bootstrap symmetric  $(1 - \alpha)$ -level CI, constructed using the percentile method, is

$$I_{p2} := \left( \hat{\theta}_i \pm \hat{\zeta}_{p,i,\alpha} \right), \quad (4.60)$$

where the cutoff point is

$$\hat{\zeta}_{p,i,\alpha} = |\hat{\theta}^* - \theta^*|_{i, \left( [B_1^{-1} \sum_{k=1}^{B_1} I(|\hat{\theta}_{i,k}^{**} - \theta_{i,k}^{**}| < |\hat{\theta}^* - \theta^*|_{i, ((1-\alpha)B_1] + 1)}] B_1 \right) + 1}.$$

Recall that  $|\hat{\theta}^* - \theta^*|_{i, ((1-\alpha)B_1] + 1}$  is the  $100(1 - \alpha)^{th}$  quantile of the bootstrap estimated distribution of the absolute value of the level one bootstrap prediction error  $\hat{\theta}^* - \theta^*$ , and the cutoff point used for the level one bootstrap confidence interval using the percentile method. The double bootstrap adjustment in the cutoff point is based on the proportion of double bootstrap prediction errors  $\hat{\theta}_{i,k}^{**} - \theta_{i,k}^{**}$  that are smaller than  $|\hat{\theta}^* - \theta^*|_{i, ((1-\alpha)B_1] + 1}$ , denoted by  $B_1^{-1} \sum_{k=1}^{B_1} I(|\hat{\theta}_{i,k}^{**} - \theta_{i,k}^{**}| < |\hat{\theta}^* - \theta^*|_{i, ((1-\alpha)B_1] + 1})$ .

### 4.5.3.2 Pivot-like Method

For bootstrap sample  $k$ , the Taylor estimated prediction MSE for  $\hat{\theta}_{ik}^{**}$  is approximated as in (4.48) and denoted by  $(\sigma_{i,Taylor}^{**})_k^2$ , where

$$(\sigma_{i,Taylor}^{**})_k^2 = \hat{g}_{1i}(\sigma_{bk}^{2**}, \sigma_{eik}^{2**}) + \hat{g}_{2i}(\sigma_{bk}^{2**}, \sigma_{eik}^{2**}) + 2\hat{g}_{3i}(\sigma_{bk}^{2**}, \sigma_{eik}^{2**}), \quad (4.61)$$



and

$$\hat{g}_{1i}(\sigma_{bk}^{2**}, \sigma_{eik}^{2**}) = \gamma_{ik}^{**} n_i^{-1} \sigma_{eik}^{2**},$$

$$\begin{aligned} \hat{g}_{2i}(\sigma_{bk}^{2**}, \sigma_{eik}^{2**}) &= (1 - \gamma_{ik}^{**})^2 \bar{\mathbf{h}}'_{\beta,i}(\mu_{xik}^{**}, \sigma_{ek}^{**}, \beta_k^{**}, 0) \hat{V}(\beta_k^{**}) \bar{\mathbf{h}}_{\beta,i}(\mu_{xik}^{**}, \sigma_{ek}^{**}, \beta_k^{**}, 0) \\ &\quad + \bar{h}_{\sigma_{\epsilon,i}}^2(\mu_{xik}^{**}, \sigma_{ek}^{**}, \beta_k^{**}, 0) \hat{V}(\sigma_{ek}^{**}) + \bar{h}_{\mu_{xi}}^2(\mu_{xik}^{**}, \sigma_{ek}^{**}, \beta_k^{**}, 0) \hat{V}(\mu_{xik}^{**}), \end{aligned}$$

$$\hat{g}_{3i}(\sigma_{bk}^{2**}, \sigma_{eik}^{2**}) = \frac{\left( \bar{h}_{b,i}^4(\mu_{xik}^{**}, \sigma_{ek}^{**}, \beta_k^{**}, 0) \sigma_{bk}^{4**} n_i^{-2} \hat{V}(\sigma_{eik}^{2**}) + n_i^{-2} \sigma_{eik}^{4**} \hat{V}(\bar{h}_{b,i}^2(\mu_{xik}^{**}, \sigma_{ek}^{**}, \beta_k^{**}, 0) \sigma_{bk}^{2**}) \right) \bar{u}_{yik}^{2**}}{\left( \bar{h}_{b,i}^2(\mu_{xik}^{**}, \sigma_{ek}^{**}, \beta_k^{**}, 0) \sigma_{bk}^{2**} + n_i^{-1} \sigma_{eik}^{2**} \right)^4},$$

$$\hat{V}(\bar{h}_{b,i}^2(\mu_{xik}^{**}, \sigma_{ek}^{**}, \beta_k^{**}, 0) \sigma_b^{2k**}) = \bar{h}_{b,i}^4(\mu_{xik}^{**}, \sigma_{ek}^{**}, \beta_k^{**}, 0) \hat{V}(\sigma_b^{2k**}),$$

where  $\gamma_{ik}^{**}$  and  $\sigma_{eik}^{2**}$  are defined for (4.43),  $\bar{h}_{\beta,i}(\mu_{xik}^{**}, \sigma_{ek}^{**}, \beta_k^{**}, 0)$  and  $\bar{h}_{b,i}(\mu_{xik}^{**}, \sigma_{ek}^{**}, \beta_k^{**}, 0)$  are defined in (4.42),  $\bar{h}_{\mu_{xi}}(\mu_{xik}^{**}, \sigma_{ek}^{**}, \beta_k^{**}, 0)$  and  $\bar{h}_{\sigma_{\epsilon,i}}(\mu_{xik}^{**}, \sigma_{ek}^{**}, \beta_k^{**}, 0)$  are defined in (4.45),  $\hat{V}(\sigma_{ek}^{**})$  is the estimated variance of  $\sigma_{ek}^{**}$  based on (4.41), and  $\hat{V}(\beta_k^{**})$  is the estimated covariance matrix of the estimated parameters  $\beta_k^{**}$  based on (4.1),  $\hat{V}(\sigma_{eik}^{2**})$  is the estimated variance of the  $\sigma_{eik}^{2**}$ ,  $\hat{V}(\mu_{xik}^{**})$  is the estimated variance of  $\mu_{xik}^{**}$  and  $\hat{V}(\sigma_b^{2k**})$  is the estimated variance of  $\sigma_b^{2k**}$ . Estimation of the parameters  $(\mu_{xik}^{**}, \sigma_{ek}^{**}, \beta_k^{**})$  and of the variance of  $\mu_{xik}^{**}$  is specific to the model and described in Section 4.2.2.

Let

$$\hat{T}_{i,k}^{**} = \frac{\hat{\theta}_{i,k}^{**} - \theta_{i,k}^{**}}{\sqrt{\hat{V}(\hat{\theta}_{i,k}^{**} - \theta_{i,k}^{**})}}, \quad (4.62)$$

where  $\hat{V}(\hat{\theta}_{i,k}^{**} - \theta_{i,k}^{**}) = (\sigma_{i,Taylor}^{**})_k^2$  be the double bootstrap pivot-like statistic.

Following Chang, J. and Hall, P. (2014), a double bootstrap CI for  $\theta_i$  constructed using the pivot-like method is

$$I_2 := \left( \hat{\theta}_i \pm \hat{\zeta}_{i,\alpha,DB} \hat{\sigma}_{i,Taylor} \right), \quad (4.63)$$

where the cutoff point is estimated by the  $100B_1^{-1} \sum_{k=1}^{B_1} I(|T_{i,k}^{**}| < |T^*|_{i,((1-\alpha)B_1+1)})^{th}$  quantile of bootstrap estimated distribution of the absolute value of the level one bootstrap statistic,

$$\hat{\zeta}_{i,\alpha,DB} = |T^*|_{i,([B_1^{-1} \sum_{k=1}^{B_1} I(|T_{i,k}^{**}| < |T^*|_{i,((1-\alpha)B_1+1)})B_1]+1)}$$

and  $\hat{\sigma}_{i,Taylor}$  is defined for (4.59).

The level one bootstrap cutoff point is estimated by the  $100(1-\alpha)^{th}$  quantile of the bootstrap estimated distribution of the absolute value of the level one bootstrap statistic  $T_i^*$ . The double bootstrap cutoff point is estimated by the  $100B_1^{-1} \sum_{k=1}^{B_1} I(|T_{i,k}^{**}| < |T^*|_{i,((1-\alpha)B_1+1)})^{th}$  quantile of the bootstrap estimated distribution of the absolute value of the level one bootstrap statistic  $T_i^*$ . Hence, the double bootstrap bias correction in the level of the confidence interval is

$$d_{1,\alpha,i} := 1 - \alpha - B_1^{-1} \sum_{k=1}^{B_1} I(|T_{i,k}^{**}| < |T^*|_{i,((1-\alpha)B_1+1)}). \quad (4.64)$$

#### 4.5.4 General Purpose Statistic for CI

The CI endpoints depend on the level. Hence the confidence interval in, for example (4.55), needs to be constructed separately, for each  $\alpha$ . We now introduce a method of constructing CIs applicable for a set of  $\alpha$ 's. We propose constructing the CI cutoff points using the quantiles of a multiple of Student-t distribution. The multiple  $\tau_i$  and the degrees of freedom  $df_i$  are to be estimated.

##### 4.5.4.1 Level One Bootstrap Symmetric $(1-\alpha)$ -level CI

Let  $T_{i,k}^*$  be the statistic defined in (4.58), where  $i$  denotes the area and  $k$  denotes the bootstrap sample and let (4.59) be the level one bootstrap symmetric  $(1-\alpha)$ -level CI for  $\theta_i$ . We assume that the level one bootstrap CI cutoff point  $\zeta_{i,\alpha,B}$  is approximated by the  $100(1-\alpha/2)^{th}$  quantile of a Student-t distribution with degrees of freedom  $df_i$ ,  $\zeta_{i,\alpha,B} = q_{t,df_i,1-\alpha/2}$ . Let

$$\mathbf{q}_{i,B} = \hat{\zeta}_{i,\alpha,B} \quad (4.65)$$

be a vector of level one bootstrap CIs cutoff points defined for (4.59), where  $\boldsymbol{\alpha}$  is a vector of CI levels of interest. Let  $\mathbf{q}_{t_{df_i}}$  be the vector of quantiles of a Student-t distribution with degrees of freedom  $df_i$ ,

$$\mathbf{q}_{t_{df_i}} := F_{t_{df_i}}^{-1}([1 - \boldsymbol{\alpha}/2]), \quad (4.66)$$

where  $F_{t_{df_i}}$  is the Student-t distribution with  $df_i$  degrees of freedom.

Assume that

$$\mathbf{q}_{i,B} = \tau_i \mathbf{q}_{t_{df_i}} + e_{q,i,B},$$

where  $e_{q,i,B}$  is the error in  $\mathbf{q}_{i,B}$  due to estimation and approximation. Then, the parameters  $\tau_i$  and  $df_i$  are estimated by minimizing an objective function  $Q_i(\mathbf{q}_{i,B}, \tau_i, df_i)$  with respect to  $\tau_i$  and  $df_i$ . Let the objective function be

$$Q_i(\mathbf{q}_{i,B}, \tau_i, df_i) := (\mathbf{q}_{i,B} - \tau_i \mathbf{q}_{t_{df_i}}) V_{q_i}^{-1} (\mathbf{q}_{i,B} - \tau_i \mathbf{q}_{t_{df_i}})', \quad (4.67)$$

where  $V_{q_i}$  is the variance covariance matrix of  $\tau_i \mathbf{q}_{t_{df_i}}$ . Using Bahadur (1966) representation,  $V_{q_i}$  is the  $l \times l$  matrix with entries

$$V_{q_i}^{r,c} = \frac{(1 - \alpha_r/2) \wedge (1 - \alpha_c/2) - (1 - \alpha_r/2)(1 - \alpha_c/2)}{f_{t_{\tau_i}, df_i}(F_{t_{\tau_i}, df_i}^{-1}(1 - \alpha_r/2)) f_{t_{\tau_i}, df_i}(F_{t_{\tau_i}, df_i}^{-1}(1 - \alpha_c/2))}, r = 1, 2, \dots, l, c = 1, 2, \dots, l,$$

where  $l$  is the length of  $\boldsymbol{\alpha}$ ,  $\wedge$  denotes the *or* binary operator,  $f_{t_{\tau_i}, df_i}$  denotes the multiple of a Student-t density with parameters  $\tau_i$  and  $df_i$  and  $F_{t_{\tau_i}, df_i}$  denotes the multiple of a Student-t distribution with parameters  $\tau_i$  and  $df_i$ .

Let  $(\hat{\tau}_{i,B}, \hat{df}_{i,B})$  be the solution to the minimization problem. Then the level one bootstrap CI is

$$I_{1,General} := \left( \hat{\theta}_i \pm q_{t, \hat{df}_{i,B}, 1-\alpha/2} \hat{s}e_i \right), \quad (4.68)$$

where  $1 - \alpha$  is the level of interest,  $\hat{s}e_i = \hat{\tau}_{i,B} \hat{\sigma}_{i,Taylor}$ , and  $\hat{\theta}_i$  and  $\hat{\sigma}_{i,Taylor}$  are defined for (4.59).

#### 4.5.4.2 Double Bootstrap Symmetric $(1 - \alpha)$ -level CI

Let  $T_{i,k}^{**}$  be the statistic defined in (4.62), where  $i$  denotes the area and  $k$  denotes the bootstrap sample and let (4.63) be the double bootstrap symmetric  $(1 - \alpha)$ -level CI for  $\theta_i$ . We assume that the double bootstrap CI cutoff point  $\zeta_{i,\alpha,DB}$  is the  $100(1 - \alpha/2)^{th}$  quantile of a Student-t distribution with degrees of freedom  $df_i$ ,  $\zeta_{i,\alpha,DB} = q_{t,df_i,1-\alpha/2}$ . Let

$$\mathbf{q}_{i,DB} = \hat{\zeta}_{i,\alpha,DB} \quad (4.69)$$

be the vector of level one bootstrap CIs cutoff points, as defined for (4.63), where  $\alpha$  is a vector of CI levels of interest. Assume that

$$\mathbf{q}_{i,DB} = \tau_i \mathbf{q}_{t,df_i} + e_{q,i,DB},$$

where  $e_{q,i,DB}$  is the error in  $\mathbf{q}_{i,DB}$  due to estimation and approximation. Then, the parameters  $\tau_i$  and  $df_i$  are estimated by minimizing the objective function  $Q_i(\mathbf{q}_{i,DB}, \tau_i, df_i)$  defined in (4.67) with respect to  $\tau_i$  and  $df_i$ .

Let  $(\hat{\tau}_{i,DB}, \hat{df}_{DB})$  be the solution to the minimization problem. Then the double bootstrap CI is

$$I_{2,General} := \left( \hat{\theta}_i \pm q_{t,\hat{df}_{i,DB},1-\alpha/2} \hat{s}\hat{e}_i \right), \quad (4.70)$$

where  $1 - \alpha$  is the level of interest,  $\hat{s}\hat{e}_i = \hat{\tau}_{i,DB} \hat{\sigma}_{i,Taylor}$ , and  $\hat{\theta}_i$  and  $\hat{\sigma}_{i,Taylor}$  are defined for (4.59).

## 4.6 Simulations

In the simulation study there are  $m = 36$  areas with unit level observations  $x_{ij}$  in three size groups. Two sample configurations are used. In the first, there are 12 areas of size  $n_i = 2$ , 12 areas of size  $n_i = 10$  and 12 areas of size  $n_i = 40$ . In the second, there are 12 areas of size  $n_i = 10$ , 12 areas of size  $n_i = 12$  and 12 areas of size  $n_i = 40$ . Each sample,  $(\mathbf{y}, \mathbf{x}, \tilde{\boldsymbol{\mu}}_x) = ((y_{11}, y_{12}, \dots, y_{mn_m}), (x_{11}, x_{12}, \dots, x_{mn_m}), (\tilde{\mu}_{x1}, \tilde{\mu}_{x2}, \dots, \tilde{\mu}_{xm}))$ , is generated using model (4.1) with  $\sigma_b^2 = 0.25$ ,  $\sigma_\epsilon^2 = 0.36$ ,  $\sigma_u^2 \sigma_\epsilon^{-2} = k = 0.10$  and  $k_{ij} = 1$ , for all  $i = 1, \dots, 36, j = 1, \dots, n_i$ . For all the

cases of auxiliary information considered in the study, the area means of  $\mathbf{x}$  are generated from a normal distribution with mean  $\mu_x = 0$  and variance  $\sigma_\delta^2 = 0.16$ . The within-area distribution of  $x_{ij}$  is normal with mean  $\mu_{xi}$  and variance  $\sigma_\epsilon^2$ . The distribution of  $\tilde{\mu}_{xi}$  is normal with mean  $\mu_{xi}$  and variance  $\sigma_u^2$ . The distribution of  $b$  is normal with mean 0 and variance  $\sigma_b^2$ . 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

$$g(x_{ij}, \boldsymbol{\beta}, b_i) = \frac{\exp(-0.8 + x_{ij} + b_i)}{1 + \exp(-0.8 + x_{ij} + b_i)}. \quad (4.71)$$

The distribution of  $y_{ij}$  is

$$f(y_{ij}|x_{ij}, b_i) = I(y_{ij}, 1)g(x_{ij}, \boldsymbol{\beta}, b_i) + I(y_{ij}, 0)(1 - g(x_{ij}, \boldsymbol{\beta}, b_i)),$$

where  $I(y_{ij}, \cdot)$  is the indicator function, and  $g(x_{it}, \boldsymbol{\beta}, b_i)$  is defined in (4.71).

The population mean of  $g(x_{ij}, \boldsymbol{\beta}, b_i)$  is 0.334 with variance 0.029. An area with  $\mu_{xi} = 0.4$  has mean 0.412 with variance 0.028. Eight hundred Monte Carlo samples were generated for some statistics with four hundred Monte Carlo samples for most statistics. For each Monte Carlo sample, four hundred level one bootstrap samples and one double bootstrap sample per level one sample were generated.

The estimation models for different types of auxiliary information are:

- Model 1: Specified by (4.1) with (4.71) and normal distribution for  $\mathbf{x}$ . The small area mean of  $\mathbf{x}$ ,  $\mu_{xi}$  is known. The Taylor approximation of the  $g(x_{ij}, \boldsymbol{\beta}, b_i)$  function in (4.44) and (4.46) is about  $(\mu_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0)$ .
- Model 2: Specified by (4.1) with (4.71) and normal distribution for  $\mathbf{x}$ . The small area mean of  $\mathbf{x}$ ,  $\mu_{xi}$  is unknown and fixed. The Taylor approximation of the  $g(x_{ij}, \boldsymbol{\beta}, b_i)$  function in (4.44) and (4.46) is about  $(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0)$ .

- Model 3: Specified by (4.1) with (4.71) and normal distribution for  $\mathbf{x}$ . The small area mean of  $\mathbf{x}$ ,  $\mu_{xi}$  is unknown and random. The Taylor approximation of the  $g(x_{ij}, \boldsymbol{\beta}, b_i)$  function in (4.44) and (4.46) is about  $(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0)$ .

The Taylor approximation of the prediction MSE is computed using (4.48) with (4.49) and (4.50) with the expressions for the derivatives of the  $g(x_{ij}, \boldsymbol{\beta}, b_i)$  function of (4.71) being

$$\bar{\mathbf{h}}_{\boldsymbol{\beta}, i}(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0) = \sum_{k=1}^K w_k \left[ g(\hat{\mu}_{xi} + z_k \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0) (1 - g(\hat{\mu}_{xi} + z_k \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0)) (1, \hat{\mu}_{xi} + z_k \hat{\sigma}_\epsilon)' \right],$$

$$\bar{h}_{b, i}(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0) = \sum_{k=1}^K w_k \left[ g(\hat{\mu}_{xi} + z_k \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0) (1 - g(\hat{\mu}_{xi} + z_k \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0)) \right],$$

$$\bar{h}_{\mu_{xi}}(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0) = \sum_{k=1}^K w_k \left[ g(\hat{\mu}_{xi} + z_k \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0) (1 - g(\hat{\mu}_{xi} + z_k \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0)) \right] \hat{\boldsymbol{\beta}}_1,$$

and the expression for the sampling variance of  $\mathbf{y}$  being

$$\tilde{\sigma}_{ei}^2 = \left[ \sum_{k=1}^K w_k g(\hat{\mu}_{xi} + z_k \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0) \right] \left[ 1 - \sum_{k=1}^K w_k g(\hat{\mu}_{xi} + z_k \hat{\sigma}_\epsilon, \hat{\boldsymbol{\beta}}, 0) \right]. \quad (4.72)$$

The models are fitted as generalized linear mixed models (GLMMs), using the *lmer*, *glmer* functions in the *lme4* package in R by restricted maximum likelihood (REML) and Laplace approximation to the likelihood. The R output from fitting the models includes the estimates of  $(\beta_0, \beta_1, \sigma_b^2, \mu_x, \sigma_\delta^2, \sigma_\epsilon^2)$  and the estimated variances of  $\hat{\boldsymbol{\beta}}, \hat{\mu}_x$ . The true small area mean of  $\mathbf{y}$  is given by (4.2) and the predicted area means of  $\mathbf{y}$  are given by (4.17, 4.19, 4.21), with estimated  $(\beta_0, \beta_1, \sigma_b^2, \mu_x, \sigma_\delta^2, \sigma_\epsilon^2)$ . The integrals in (4.2, 4.17, 4.19, 4.21) are approximated using a 26-point approximation to the normal distribution and the approximation in (4.40) is based on the same 26-points.

The estimation procedure for  $d_{fi}$  and  $\tau_i$  described in Section 4.5.4 is implemented for a set of three  $\alpha$ 's,  $\boldsymbol{\alpha} = (0.1, 0.05, 0.01)$  with a working matrix

$$V_{qi}(\tau_i = 1, d_{fi} = 5) = \begin{pmatrix} 0.0074 & 0.0073 & 0.0072 \\ 0.0073 & 0.0368 & 0.0362 \\ 0.0072 & 0.0362 & 0.0729 \end{pmatrix}.$$

The Nelder-Mead (1965) method is used to minimize  $Q_i(\cdot, \tau_i, d_{fi})$  in (4.67). The coverage probability of the confidence intervals is estimated by

$$N^{-1} \sum_{t=1}^N I(\theta_{i,t} \in I_t),$$

where  $N$  is the number of Monte Carlo samples and  $I_t$  represents a symmetric confidence interval of the type presented in Section 4.5.

#### 4.6.1 Refinement of Estimators

The estimators of the variance components  $\sigma_b^2, \sigma_\delta^2$  may be zero. In the first simulation, where  $n_i \in \{2, 10, 40\}$ , the estimator of  $\sigma_b^2$  is bounded by  $K_{b,s} = 0.006$  and the estimator of  $\sigma_\delta^2$  is bounded by  $K_{\delta,s} = 0.008$ . If  $\hat{\sigma}_{b,k}^2 = 0.006$  we set  $\alpha_{i,k}^{**}$  equal to  $\alpha_{i,k}^*$ , for  $k = 1, \dots, B_1$ . In the second simulation, where  $n_i \in \{10, 20, 40\}$ , the estimator of  $\sigma_b^2$  is bounded by  $K_{b,s} = 0.005$  and the estimator of  $\sigma_\delta^2$  is bounded by  $K_{\delta,s} = 0.007$ . If  $\hat{\sigma}_{b,k}^2 = 0.005$  we set  $\alpha_{i,k}^{**}$  equal to  $\alpha_{i,k}^*$ , for  $k = 1, \dots, B_1$ . The bounds are somewhat larger than those suggested in Wang and Fuller (2003).

The estimator of the prediction MSE  $\alpha_i$  given in (4.32) may be nonpositive. In the simulation studies, the estimator of  $\alpha_i$  is bounded using the procedure in Erciulescu and Fuller (2014).

#### 4.6.2 Results

Monte Carlo (MC) properties of the estimated parameters  $(\hat{\sigma}_b^2, \hat{\boldsymbol{\beta}}, \hat{\mu}_x, \hat{\sigma}_\delta, \hat{\sigma}_\epsilon)$  are presented in Tables 4.1 and 4.2. The simulation values for  $(\sigma_b^2, \boldsymbol{\beta}, \mu_x, \sigma_\delta, \sigma_\epsilon)$  are given in parentheses. The

MC absolute bias in  $\hat{\beta}$  and the MC absolute bias in  $\hat{\sigma}_\epsilon^2$  are less than two MC standard errors. The MC absolute bias in  $\hat{\mu}_x$  is significant, being approximately equal to 2.3 MC standard errors, and the MC absolute bias in  $\hat{\sigma}_\delta^2$  is significant, being approximately equal to 3.0 standard errors. This is not surprising because the REML estimates are known to have smaller bias than the maximum likelihood estimators for the variance components, but are not necessarily unbiased estimates of the variance components; see Bates (2008). Also, there are no explicit expressions for the finite-sample properties for the empirical BLUE  $\hat{\mu}_x$ .

Consider simulation Model 1, when  $\mu_{xi}$  is known. In the simulation study with sample size configuration  $n_i \in \{2, 10, 40\}$ , the proportion of sample estimators  $\hat{\sigma}_b^2$  that hit the bound is 0.0275 and the proportion of level one estimators  $\hat{\sigma}_b^{2*}$  that hit the bound is 0.1154. The coefficient of variation for  $\hat{\sigma}_b^2$  calculated for 800 Monte Carlo samples is about 0.6, approximately the CV of a Chi-square with five degrees of freedom. The Monte Carlo absolute bias of the estimator of  $\hat{\sigma}_b^2$  is about 0.0168, which is approximately equal to three Monte Carlo standard errors.

In the simulation study with sample size configuration  $n_i \in \{10, 20, 40\}$ , the proportion of sample estimators  $\hat{\sigma}_b^2$  that hit the bound is 0.0150, the proportion of level one estimators  $\hat{\sigma}_b^{2*}$  that hit the bound is 0.0670. The coefficient of variation for  $\hat{\sigma}_b^2$  calculated for 400 Monte Carlo samples is about 0.57, approximately the CV of a Chi-square with six degrees of freedom. The Monte Carlo absolute bias of the estimator of  $\hat{\sigma}_b^2$  is about 0.012, which is less than 2 Monte Carlo standard errors.

Consider simulation Models 2 and 3, when  $\mu_{xi}$  is unknown and  $\tilde{\mu}_{xi}$  observed. The proportion of sample estimators  $\hat{\sigma}_b^2$  that hit the bound is 0.0275 and the proportion of level one estimators  $\hat{\sigma}_b^{2*}$  that hit the bound is 0.1097 for Model 2 and 0.1099 for Model 3. The difference in the proportion of level one estimators  $\hat{\sigma}_b^{2*}$  that hit the bound for Models 1, 2 and 3 is due to the different estimated model parameters used to generate the samples. For Model 1,  $(\sigma_\epsilon^2)$  is estimated as in (4.23), for Model 2,  $(\sigma_\epsilon^2)$  is estimated as in (4.25) and for Model 3,  $(\mu_x, \sigma_\delta, \sigma_\epsilon^2)$  is estimated using REML. The coefficient of variation for  $\hat{\sigma}_b^2$  calculated for 400 Monte Carlo



Table 4.1: Monte Carlo Properties of the Estimated Parameters,  $\mu_{xi}$  known

MC	$n_i$		$\hat{\sigma}_b^2(0.25)$	$\hat{\beta}_0(-0.80)$	$\hat{\beta}_1(1.00)$	$\hat{\sigma}_\epsilon(0.36)$
800 (Model 1)	{2, 10, 40}	median	0.2128	-0.7990	1.0032	0.3616
		mean	0.2332	-0.7973	1.0035	0.3614
		sd	0.1502	0.1389	0.1553	0.0209
		se	0.0053	0.0049	0.0055	0.0007
400 (Model 1)	{10, 20, 40}	median	0.2221	-0.8048	1.0041	0.3625
		mean	0.2380	-0.8050	1.0096	0.3612
		sd	0.1368	0.1306	0.1417	0.0195
		se	0.0068	0.0065	0.0071	0.0010

samples is about 0.65, approximately the CV of a Chi-square with five degrees of freedom. The Monte Carlo absolute bias of the estimator of  $\hat{\sigma}_b^2$  is about 0.0152, which is equal to two Monte Carlo standard errors.

The parameters  $(\sigma_b^2, \beta)$  are estimated by maximizing the Laplace approximation of the likelihood, and are estimated separately from the estimation of  $(\mu_x, \sigma_\delta^2, \sigma_\epsilon^2)$ , see Section 4.2.2. Therefore, the differences in the Monte Carlo properties of  $(\hat{\sigma}_b^2, \hat{\beta})$  reported in Tables 4.1 and 4.2 are due to the different Monte Carlo samples only. Pfeffermann and Correa (2012) study a unit level logit model with two known covariates,  $m = 30$  areas and  $n_i = 25$  observations in each area. The authors report simulation results for the parameters  $(\sigma_b^2, \beta)$ , estimated by maximizing the Laplace approximation of the likelihood. The results in Pfeffermann and Correa (2012) show small but significant bias in the model parameters  $(\sigma_b^2, \beta)$ .

Properties of  $\bar{h}_{bi}, \bar{h}_{\beta_{i,2}}, \bar{h}_{\mu_{xi}}, \hat{\gamma}_i, \hat{g}_{1i}$ , averaged by the sample size and over the Monte Carlo samples are presented in Tables 4.3 and 4.4. The vector  $\bar{h}_{\beta_i}$  has length two. The first element of  $\bar{h}_{\beta_i}$  is  $\bar{h}_{bi}$  and represents the multiplier for the estimated variance of the estimated intercept in model (4.1),  $\hat{\beta}_0$ , in expression (4.48). The second element of  $\bar{h}_{\beta_i}$  represents the multiplier for the estimated variance of the estimated  $\mathbf{x}$  coefficient in model (4.1),  $\hat{\beta}_1$ , in expression (4.48). We denote the second element of  $\bar{h}_{\beta_i}$  by  $\bar{h}_{\beta_{i,2}}$ .

Table 4.2: Monte Carlo Properties of the Estimated Parameters,  
 $\mu_{xi}$  unknown,  $\tilde{\mu}_{xi}$  observed,  $N = 400$

		$\hat{\sigma}_b^2(0.25)$	$\hat{\beta}_0(-0.80)$	$\hat{\beta}_1(1.00)$	$\hat{\mu}_x(0.00)$	$\hat{\sigma}_\delta(0.16)$	$\hat{\sigma}_\epsilon(0.36)$
$\mu_{xi}$ fixed (Model 2)	median	0.2087	-0.7915	0.9986			0.3603
	mean	0.2348	-0.7892	0.9988			0.3625
	sd	0.1521	0.1368	0.1585			0.0258
	se	0.0076	0.0068	0.0079			0.0013
$\mu_{xi}$ random (Model 3)	median	0.2087	-0.7915	0.9986	0.0092	0.1642	0.3590
	mean	0.2348	-0.7892	0.9988	0.0076	0.1679	0.3590
	sd	0.1521	0.1368	0.1585	0.0669	0.0520	0.0204
	se	0.0076	0.0068	0.0079	0.0033	0.0026	0.0010

Since the simulation population mean of  $g(x_{ij}, \beta, b_i)$  is approximately 0.3, the expected value of  $\tilde{\sigma}_{\epsilon i}^2$  is approximately 0.21. Notice that the expression for  $\bar{h}_{bi}$  in (4.72) is close to the expression for  $\tilde{\sigma}_{\epsilon i}^2$  in (4.72). Hence, we would expect that the expected value of  $\bar{h}_{bi}$  to be close to 0.21. The Monte Carlo mean of  $\bar{h}_{bi}$  is about 0.20. The average value of  $\bar{h}_{\beta_{i,2}}$  is about 0.03, for all the areas and for all the models considered in the simulation study. Hence, the greatest contribution to the  $g_{2i}$  term in (4.48) is the estimated variance of  $\hat{\beta}_0$ . For the models with unknown  $\mu_{xi}$ , the average value of  $\bar{h}_{\mu_{xi}}$  is about 0.20 for all the areas. This is because  $\bar{h}_{\mu_{xi}} = \hat{\beta}_1 \bar{h}_{bi}$  and  $\hat{\beta}_1$  is close to 1, see (4.72). The differences in the properties of the intermediate statistics  $\bar{h}_{bi}, \bar{h}_{\beta_{i,2}}, \bar{h}_{\mu_{xi}}, \hat{\gamma}_i$  are due to the different predictors of  $\mu_{xi}$  and the different estimators of  $\sigma_\epsilon^2$  for Models 1, 2, and 3.

In the fourth column in Tables 4.3 and 4.4 are properties of the estimated  $g_{1i}$  term, the largest of the three terms in the estimated prediction MSE of (4.48). The expression for  $\hat{g}_{1i}$  is a function of the estimated  $\gamma_i$ , the estimated  $\sigma_{\epsilon i}^2$  and the area sample size  $n_i$ , see (4.48). The  $\hat{g}_{1i}$  decreases with an increase in sample size. The estimated  $\gamma_i$  is a function of  $\bar{h}_{b,i}(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\beta}, 0)$ , given in (4.72), and the estimated  $\sigma_{\epsilon i}^2$ , given in (4.72), is a function of  $g(\hat{\mu}_{xi}, \hat{\sigma}_\epsilon, \hat{\beta}, 0)$  given in (4.71). Since  $\hat{\mu}_{xi}$  and  $\hat{\sigma}_\epsilon$  differ for Models 1, 2, and 3, the MC properties of  $\hat{g}_{1i}$  in Tables 4.3 and 4.4 differ for Models 1, 2, and 3, too.

For the models with unknown  $\mu_{xi}$ , the small area mean of  $\mathbf{x}$  is estimated/predicted to be  $\hat{\mu}_{xi}$  and the variance of  $\hat{\mu}_{xi}$  is estimated, see Section 4.2.1 and 4.2.2. Given known  $\sigma_\epsilon^2$  and unknown fixed  $\mu_{xi}$ ,  $V(\hat{\mu}_{xi})$  given in (4.25) is equal to (0.0300, 0.0180, 0.0072) for  $n_i = (2, 10, 40)$ . The results in Table 4.4 indicate a MC absolute bias in  $\hat{V}(\hat{\mu}_{xi})$ , for fixed  $\mu_{xi}$ , less than two MC standard errors, for  $n_i = (2, 10, 40)$ . Given known parameters  $\mu_x, \sigma_\delta^2, \sigma_\epsilon^2$  and unknown random  $\mu_{xi}$ ,  $V(\hat{\mu}_{xi})$  given in (4.15) is approximately equal to (0.0253, 0.0162, 0.0069) for  $n_i = (2, 10, 40)$ . The theoretical bias in  $\hat{V}(\hat{\mu}_{xi})$  is  $o(m^{-1})$ . The results in Table 4.4 indicate a MC absolute bias in  $\hat{V}(\hat{\mu}_{xi})$ , for random  $\mu_{xi}$ , approximately equal to (4.56, 2.87, 0.80) MC standard errors, for  $n_i = (2, 10, 40)$ . The estimated  $V(\hat{\mu}_{xi})$  is lower, with lower MC standard error, for the model with random  $\mu_{xi}$  than for the model with fixed  $\mu_{xi}$ . The component of the estimated prediction MSE of the area mean of  $\mathbf{y}$  for Models 2 and 3 that is not due to the estimation of the parameters  $\beta, \sigma_b^2$ , is  $\hat{g}_{1i} + \bar{h}_{\mu_{xi}}^2 \hat{V}(\hat{\mu}_{xi})$ . The MC mean of  $\hat{g}_{1i} + \bar{h}_{\mu_{xi}}^2 \hat{V}(\hat{\mu}_{xi})$  is lower for Model 3 than for Model 2, showing the efficiency gain associated with the random specification for mean of the auxiliary variable. The MC mean of  $\hat{g}_{1i} + \bar{h}_{\mu_{xi}}^2 \hat{V}(\hat{\mu}_{xi})$  is lower for Model 1, when  $\mu_{xi}$  is known, than for Model 3. These results are consistent with the results in Erciulescu and Fuller (2014).

Table 4.3: Monte Carlo Properties of Intermediate Statistics,  
 $\mu_{xi}$  known (Entries Multiplied by  $10^3$ )

	$n_i$		$\bar{h}_{bi}$	$\bar{h}_{\beta_{i,2}}$	$\hat{\gamma}_i$	$\hat{g}_{1i}$
800 MC Samples (Model 1)	2	mean	198.3756	28.6705	76.3877	8.2382
		sd	25.3333	69.5531	44.8755	5.1152
		se	0.8957	2.4591	1.5866	0.1808
	10	mean	198.1536	27.9012	275.0239	5.9159
		sd	25.3668	68.8024	128.7306	2.9829
		se	0.8969	2.4325	4.5513	0.1055
	40	mean	199.1922	31.2796	566.7435	3.0531
		sd	24.9419	70.0753	182.9751	1.0996
		se	0.8818	2.4775	6.4691	0.0389
400 MC Samples (Model 1)	10	mean	197.8696	29.0197	184.4065	7.2084
		sd	25.3205	69.2630	136.9015	4.0727
		se	1.2660	3.4631	6.8451	0.2036
	20	mean	197.6387	28.2339	357.4052	5.3032
		sd	25.3083	68.5454	155.8462	2.4743
		se	1.2654	3.4273	7.7923	0.1237
	40	mean	198.6496	31.7173	582.8685	3.1340
		sd	25.0471	69.7743	165.2565	1.0233
		se	1.2524	3.4887	8.2628	0.0512

Table 4.4: Monte Carlo Properties of Intermediate Statistics,  
 $\mu_{xi}$  unknown,  $\tilde{\mu}_{xi}$  observed (400 MC Samples, Entries Multiplied by  $10^3$ )

	$n_i$		$\bar{h}_{bi}$	$\bar{h}_{\beta_{i,2}}$	$\hat{\gamma}_i$	$\hat{g}_{1i}$	$\hat{V}_{\mu_{xi}}$	$\bar{h}_{\mu_{xi}}$
$\mu_{xi}$ fixed (Model 2)	2	mean	198.8480	32.0193	77.0395	8.3373	30.2064	197.6681
		sd	26.8505	75.8948	45.8172	5.2690	2.1527	37.1846
	10	se	1.3425	3.7947	2.2909	0.2635	0.1076	1.8592
		sd	198.4142	28.2986	276.1925	5.9448	18.1238	197.2436
	40	sd	26.1575	71.5878	130.3471	3.0174	1.2916	36.7650
		se	1.3079	3.5794	6.5174	0.1509	0.0646	1.8383
	40	sd	199.8571	31.9783	567.8212	3.0659	7.2495	198.7973
		sd	25.0977	71.4102	184.9909	1.1088	0.5167	36.5318
		se	1.2549	3.5705	9.2495	0.0554	0.0258	1.8266
$\mu_{xi}$ random (Model 3)	2	mean	201.8742	27.4370	78.1215	8.5050	25.6851	200.8600
		sd	19.2271	52.0755	45.6473	5.0780	1.8519	33.1381
	10	se	0.9614	2.6038	2.2824	0.2539	0.0926	1.6569
		sd	199.8413	27.5617	277.7639	6.0057	16.3248	198.6924
	40	sd	23.6765	63.5487	130.3923	2.9892	1.0100	35.2189
		se	1.1838	3.1774	6.5196	0.1495	0.0505	1.7609
	40	sd	200.2392	31.0950	568.4912	3.0724	6.9056	199.1907
		sd	24.3353	68.7389	184.8713	1.1020	0.3893	36.0782
		se	1.2168	3.4369	9.2436	0.0551	0.0195	1.8039

#### 4.6.2.1 Prediction MSE Properties

Let  $\hat{\alpha}^{Taylor}$  be the Taylor prediction MSE estimator given in (4.48). The two bootstrap bias correction methods presented in Section 4.3 are

- *Bias Correction 1.* In this method, the bootstrap bias is estimated by the difference in the level two sample estimator of  $\alpha_i$  and the level one sample estimator of  $\alpha_i$ . The bias adjusted bootstrap estimators of  $\alpha_i$  using this method are given in expressions (4.29 - 4.32), in Section 4.3. This method is called the difference bootstrap bias correction method.
- *Bias Correction 2.* In this method, the bias adjusted bootstrap estimators of  $\alpha_i$  are obtained by subtracting the estimated bootstrap bias in  $\hat{\alpha}_i^{Taylor}$ , computed using the difference bootstrap bias correction method at the corresponding bootstrap level, from the bootstrap sample estimate of  $\alpha_i$ . The bias adjusted bootstrap estimators of  $\alpha_i$  using this method are given in expressions (4.34 - 4.38), in Section 4.3.

Let  $(\hat{\alpha}^*, \hat{\alpha}_\Delta^*, \hat{\alpha}_T^{**}, \hat{\alpha}_{\Delta,T}^{**})$  be the bootstrap prediction MSE estimators in Section 4.3,

- $\hat{\alpha}^*$  is the level one bootstrap MSE estimator given in (4.29)
- $\hat{\alpha}_\Delta^*$  is the bias corrected level one bootstrap MSE estimator given in (4.34) that uses method *Bias Correction 2*
- $\hat{\alpha}_T^{**}$  is the telescoping bias corrected double bootstrap MSE estimator given in (4.32) that uses method *Bias Correction 1*
- $\hat{\alpha}_{\Delta,T}^{**}$  is the telescoping bias corrected double bootstrap MSE estimator given in (4.38) that uses method *Bias Correction 2*.

Table 4.5 contains Monte Carlo properties of  $(\hat{\alpha}^{Taylor}, \hat{\alpha}^*, \hat{\alpha}_\Delta^*, \hat{\alpha}_T^{**}, \hat{\alpha}_{\Delta,T}^{**})$ , organized by the three area sample sizes, in groups of five lines, for the simulation Model 2. The results presented in Table 4.5 are similar for all the models and for the two sample size configurations considered in the simulation study. Appendix E contains tables for other configurations. Each line in Table 4.5 is the average of the results for the 12 areas with the same sample size. The first line is the

Monte Carlo estimate of the prediction MSE,  $\hat{\alpha}$ . The next four lines are of the bias relative to the mean, the coefficient of variation, the bias relative to the standard deviation and the bias relative to the Monte Carlo standard error. The definitions are

$$\begin{aligned} RelBias &= \sum_{is=1}^{12} (\hat{\alpha}_{.,is}^{EST} - \hat{\alpha}_{.,is}) / \sum_{is=1}^{12} \hat{\alpha}_{.,is}, \\ CV &= \sum_{is=1}^{12} \sqrt{(N-1)^{-1} \sum_{\zeta=1}^N (\hat{\alpha}_{\zeta,is}^{EST} - \hat{\alpha}_{.,is}^{EST})^2} / \sum_{is=1}^{12} \hat{\alpha}_{.,is}, \\ Bias/sd &= \sum_{is=1}^{12} (\hat{\alpha}_{.,is}^{EST} - \hat{\alpha}_{.,is}) / \sum_{is=1}^{12} \sqrt{(N-1)^{-1} \sum_{\zeta=1}^N (\hat{\alpha}_{\zeta,is}^{EST} - \hat{\alpha}_{.,is}^{EST})^2}, \\ Bias/se &= Bias / (\sqrt{N} sd), \end{aligned}$$

where  $N$  is the number of Monte Carlo samples,  $\hat{\alpha}^{EST} \in \{\hat{\alpha}^{Taylor}, \hat{\alpha}^*, \hat{\alpha}_{\Delta}^*, \hat{\alpha}_T^{**}, \hat{\alpha}_{\Delta,T}^{**}\}$  is the prediction MSE estimator for an area,  $\zeta$  indexes the Monte Carlo samples,  $i$  denotes an area within a group of areas of sample size  $s$ ,  $\hat{\alpha}_{.,is} = (N)^{-1} \sum_{\zeta=1}^N \hat{\alpha}_{\zeta,is}$  is the average of the Monte Carlo prediction error estimators, and  $\hat{\alpha}_{.,is}^{EST} = (N)^{-1} \sum_{\zeta=1}^N \hat{\alpha}_{\zeta,is}^{EST}$  is the average of the prediction MSE estimators.

The Taylor estimated prediction MSE has CV of about 53%, 42% and 32% for sample sizes 2, 10, and 40, respectively. The level one bootstrap estimated prediction MSE has CV of about 41%, 32% and 20% for 400 bootstrap samples for sample sizes 2, 10, and 40, respectively. The level one bootstrap estimated prediction MSE using method *Bias correction 2* has similar CV to the CV of the Taylor estimated prediction MSE. In the simulation study in Pfeffermann and Correa (2012), for the logit model,  $B_1 = 100$ ,  $B_2 = 100$  and  $m = 30$  areas, each of size  $n_i = 25$ . The parametric bootstrap estimated prediction MSE in Pfeffermann and Correa (2012) has an average CV of about 30%, for the level one bootstrap, and an average CV of about 26 – 73%, for the different double bootstrap methods considered, where the CVs are averaged over the areas.

Recall that the Taylor estimated prediction MSE in (4.48) is constructed based on the approximation (4.39). Even if the approximation (4.39) has an approximation error of  $O(1)$ , the Taylor estimated prediction MSE in (4.48) performs well in the simulation studies. The CV of the Taylor estimated prediction MSE in (4.48) is not much greater than the CV of the boot-

strap estimated prediction MSE. The absolute relative bias of the Taylor estimated prediction MSE in (4.48) is smaller than the absolute relative bias of the level one bootstrap estimated prediction MSE using method *Bias correction 1* and close to the absolute relative bias of the level one bootstrap estimated prediction MSE using method *Bias correction 2*.

The double bootstrap reduces the absolute value of the bias for all sample sizes, relative to the level one bootstrap. The absolute relative bias of the double bootstrap using either method *Bias correction 1* or method *Bias correction 2* is not larger than 4%, among all the models and the sample sizes considered in the simulation. The absolute relative bias of the double bootstrap using method *Bias correction 2* is less than 2% for the sample sizes considered in the simulation. The results are comparable with the results in Pfeffermann and Correa (2012).

In the simulation in Pfeffermann and Correa (2012), the average percent of the absolute relative bias in the parametric double bootstrap estimator is in the range 1.1% to 4.6%, with the largest value being 15.6%, for the different bootstrap bias correction methods. Datta et al. (2005) propose second-order accurate approximations to the mean squared error of model-based small area estimators, for the Fay-Herriot model, without using resampling methods. In a simulation study, the authors consider a basic area level model without covariates and  $m = 15$  areas divided into five groups of unequal sample sizes. The average relative bias in the prediction MSE estimator based on the REML estimates, is about 1.8–2.6%, in absolute value.

Hall and Maiti (2006) (HM) conducted a simulation study for the binary model. They denoted the model by M4. In the simulation study, HM considered  $m = 15$  areas with sample sizes in the range 48 to 287, vector of parameters  $\boldsymbol{\beta} = (0, 1)$ , and variance of random effects  $\sigma_b^2 = 1$ . HM considered known  $\mu_{xi}$  and constructed parametric double bootstrap prediction MSE estimators for the area parameter  $\theta_i^C$ , where  $\theta_i^C$  is the inverse logit of the small area mean of  $\mathbf{y}$ , using different bootstrap bias correction methods. In the simulation study in HM,  $B_1 = 100$  and  $B_2 = 50$ . HM report a relative bias in the double bootstrap estimator of about 13%, which is approximately 70% larger than the absolute relative bias reported in Table 4.5,



and a CV of about 43%, larger than the CV reported in 4.5. Our results are not directly comparable with the results in HM, because our parameter of interest  $\theta_i$  is the small area mean of a binary response variable, while the HM parameter of interest  $\theta_i^C$  is the inverse logit of the small area mean of a binary response variable. The true simulation value for  $\sigma_b^2$  in HM is four times larger than the true simulation value for  $\sigma_b^2$  in our simulation study (EF). The area sample sizes in HM are larger than the area sample sizes in EF. Hence, the estimated CV of  $\hat{\sigma}_b^2$  in HM is about 0.37 and the estimated CV of  $\hat{\sigma}_b^2$  in EF is about 0.60. The large bias in the double bootstrap estimator in HM may be due to different estimators for the random effects variance. HM use maximum likelihood for M4 and EF use a Laplace approximation to the likelihood for the random effects variance in the model for  $\mathbf{y}$ , and restricted maximum likelihood for the random effects variance in the model for  $\mathbf{x}$ .

Table 4.5: Monte Carlo Properties of Prediction MSE Estimators,  
 $\mu_{xi}$  unknown and fixed,  $\tilde{\mu}_{xi}$  observed  
( $B_1 = 400, B_2 = 1, N = 400$ , Variances Multiplied by  $10^3$ )

$n_i$		Taylor	Level 1		Level 2	
		$\hat{\alpha}^T$	$\hat{\alpha}^*$	$\hat{\alpha}_{\Delta}^*$	$\hat{\alpha}_T^{**}$	$\hat{\alpha}_{\Delta,T}^{**}$
2	$V(\hat{\theta} - \theta)$	10.9678	10.9678	10.9678	10.9678	10.9678
	RelBias	-0.0341	-0.1052	-0.0146	-0.0207	0.0083
	$CV(\hat{\alpha})$	0.5263	0.4088	0.5211	0.4523	0.6777
	Bias/sd	-0.0647	-0.2573	-0.0281	-0.0458	0.0122
	Bias/se	-1.2945	-5.1464	-0.5618	-0.9169	0.2436
10	$V(\hat{\theta} - \theta)$	7.9540	7.9540	7.9540	7.9540	7.9540
	RelBias	-0.0111	-0.1182	-0.0101	-0.0232	0.0179
	$CV(\hat{\alpha})$	0.4235	0.3164	0.4258	0.3520	0.5819
	Bias/sd	-0.0263	-0.3734	-0.0237	-0.0658	0.0307
	Bias/se	-0.5261	-7.4679	-0.4740	-1.3166	0.6145
40	$V(\hat{\theta} - \theta)$	3.9343	3.9343	3.9343	3.9343	3.9343
	RelBias	0.0425	-0.1176	-0.0395	-0.0332	-0.0160
	$CV(\hat{\alpha})$	0.3152	0.2020	0.3482	0.2251	0.5914
	Bias/sd	0.1350	-0.5822	-0.1135	-0.1474	-0.0270
	Bias/se	2.6994	-11.6449	-2.2693	-2.9473	-0.5395

#### 4.6.2.2 Confidence Intervals for the Small Area Means

Let  $I_{B,\Delta}$ ,  $I_{DBT,\Delta}$ ,  $I_B$  and  $I_{DBT}$  be confidence intervals given in (4.55), using the bootstrap prediction MSE estimators  $\alpha_{\Delta}^*$ ,  $\alpha_{\Delta,T}^{**}$ ,  $\alpha^*$ ,  $\alpha_T^{**}$  given in (4.34, 4.32, 4.29, 4.32). Table 4.6 contains the average empirical coverages of the CIs  $I_{B,\Delta}$ ,  $I_{DBT,\Delta}$ ,  $I_B$ ,  $I_{DBT}$  for  $\alpha = 0.05$ , for each area group, by sample size. Properties of the estimated lengths of the CIs  $I_{B,\Delta}$ ,  $I_{DBT,\Delta}$ ,  $I_B$ ,  $I_{DBT}$  for  $\alpha = 0.05$ , for each area group, by sample size, are presented in Appendix E.

Table 4.6: Empirical Coverages for 95% Wald-type CIs,  
Bootstrap Estimated Prediction MSE ( $B_1 = 400, B_2 = 1$ )

	$n_i$	$I_{B,\Delta}$	$I_{DBT,\Delta}$	$I_B$	$I_{DBT}$
$\mu_{xi}$ known	2	88.6	88.7	88.1	86.7
	10	90.5	90.3	89.9	87.9
800 MC Samples	40	92.6	92.0	91.8	89.9
$\mu_{xi}$ fixed	2	90.2	90.2	89.6	88.2
	10	91.0	90.9	90.7	89.2
400 MC Samples	40	91.7	91.7	91.3	89.7
$\tilde{\mu}_{xi}$ observed	2	90.2	90.2	89.7	88.3
	10	91.0	90.9	90.3	89.1
400 MC Samples	40	91.6	91.6	91.4	89.7
$\mu_{xi}$ random	10	91.3	91.2	91.0	89.8
	20	92.4	92.4	92.2	90.7
400 MC Samples	40	93.3	93.2	93.1	91.7

The Wald-type CIs using the bootstrap prediction MSE estimators result in undercoverage. The empirical coverages increase with the increase in sample size, but do not reach the desired nominal level. The bias correction method *Bias Correction 2* in the level one bootstrap and in the double bootstrap MSE estimator improves the coverage of the bootstrap CIs, relative to the coverages of the bootstrap CIs computed using the bootstrap MSE estimators using *Bias Correction 1*. The coverages of the telescoping double bootstrap CIs are similar to the coverages of the level one bootstrap CIs.

For each area  $i$ , we constructed 90%, 95% and 99% confidence intervals using the different methods described in Section 4.5, based on Taylor estimators for the prediction MSE. Recall

that  $I_W$  is the Wald-type CI based on the pivot-like method,  $I_{p1}$  is the level one bootstrap CI based on the percentile method and  $I_1$  is the level one bootstrap CI based on the pivot-like method. Let  $I_{p2T}$  be the telescoping double bootstrap CI based on the percentile method and  $I_{2T}$  be the telescoping double bootstrap CI based on the pivot-like method. Tables 4.7 and 4.8 contain the average empirical coverages of the CIs  $I_W, I_{p1}, I_1, I_{p2T}, I_{2T}$  for each  $\alpha$  level and for each area group, by sample size, for all the models considered in the simulation. Monte Carlo properties of the estimated lengths of the CIs  $I_W, I_{p1}, I_1, I_{p2T}, I_{2T}$  for each  $\alpha$  level and for each area group, by sample size, for simulation Model 2 are presented in Table 4.9; see Appendix E for results on the estimated lengths of the CIs for the simulation Models 1 and 3, and for the different sample size configuration considered in the study.

The Wald-type CIs using the pivot-like method have undercoverage of about 2 – 7% for  $n_i = 2$ , with better coverage for the simulation configuration with  $n_i = 10, 20, 40$  and for the models with unknown  $\mu_{xi}$ . The empirical coverages increase with the increase in sample size, resulting in undercoverage of less 1.0% for  $n_i = 40$  and overcoverage of less than 0.3% for  $n_i = 40$ . The results agree with the results for model M4 in the simulations in Hall and Maiti (2006). The authors report undercoverages of about 1 – 8% for 80%, 90%, 95% Wald-type CIs for the studied parameter  $\theta_i^C$ , for model M4.

The level one bootstrap CIs and the double bootstrap CIs based on the percentile method result in undercoverage. The level one bootstrap pivot-type CIs have coverage close to the nominal coverage for  $\alpha = 0.10, 0.05$ . The coverages of the double bootstrap pivot-type CIs are similar to the coverages of the level one bootstrap pivot-type CIs.

Hall and Maiti (2006) construct percentile bootstrap confidence intervals with nominal coverages  $1 - \alpha = 0.80, 0.90, 0.95$ , for the parameter  $\theta_i^C$ , for model M4. The HM simulation results show good performance for the level one percentile bootstrap CIs for  $\alpha = 0.20, 0.10$  and undercoverage for the level one bootstrap CIs for  $\alpha = 0.05$ . This can be explained by the distribution for the parameter of interest and by the area sample size. HM assume that  $\theta_i^C$  is a

linear function of  $(x_i, b_i, \boldsymbol{\beta})$  and its simulation values are generated from a normal distribution. The parameter of interest in our study is  $\theta_i$  in (4.2) with (4.71), hence, a nonlinear function of  $(x_{ij}, b_i, \boldsymbol{\beta})$ . Therefore, the estimated bootstrap distribution of  $\theta_i^C$  is less skewed than the estimated bootstrap distribution of  $\theta_i$  in (4.2) with (4.71). Also, the HM simulation set-up is based on areas of sample size larger than 48, while the largest sample size for the areas in our simulation set-up is 40. The results in Tables 4.7 and 4.8 show that the coverage of the level one percentile bootstrap CI for  $\theta_i$  in (4.2) with (4.71) increases with the increase in sample size.

The results in Table 4.7 show that the 99% pivot-type CIs have undercoverage of about 0.7 – 1.6% for the configuration  $n_i = 2, 10, 40$  and the simulation Models 1 and 2, when  $\mu_{xi}$  is fixed. By increasing the area sample sizes to 10, 20, 40 in the second simulation configuration, the undercoverage of the 99% pivot-type CIs is reduced to 0.3% for  $n_i = 2$  and the overcoverage of the 99% pivot-type CIs is 0.2% for  $n_i = 40$ , when  $\mu_{xi}$  is random; see Table 4.8.

The estimated length and the estimated variance of the estimated length of the pivot-type bootstrap CIs are larger than the corresponding values for the Wald-type CIs, but also the coverage for the pivot-type bootstrap CIs is better than the coverage for the Wald-type CIs. For all the CIs  $I_W, I_{p1}, I_1, I_{p2T}, I_{2T}$ , the empirical coverage increases, with the increase in the estimated length. See Appendix F for graphical representations of empirical coverage versus estimated length. The coverage of the double bootstrap CIs is similar to the coverage of the level one bootstrap CIs, while the estimated variability of the estimated length of the double bootstrap CIs is increased.

Table 4.7: Empirical Coverages ( $B_1 = 400, B_2 = 1$ )

			Wald-type Pivot	Level 1 Percentile	Level 1 Pivot	Level 2 Percentile	Level 2 Pivot
	$100(1 - \alpha)\%$	$n_i$	$I_W$	$I_{p1}$	$I_1$	$I_{p2T}$	$I_{2T}$
$\mu_{xi}$ known 800 MC Samples	90%	2	83.4	82.2	91.0	83.2	91.7
		10	86.1	83.6	90.8	85.4	91.9
		40	89.4	86.0	89.8	87.5	91.0
	95%	2	88.8	88.0	95.2	87.8	94.6
		10	91.6	89.8	95.2	89.9	94.9
		40	94.1	91.8	94.6	92.2	94.9
	99%	2	94.3	94.0	97.4	92.2	96.5
		10	95.9	95.2	97.7	94.2	96.9
		40	98.0	97.0	98.4	96.5	98.0
$\mu_{xi}$ unknown fixed $\tilde{\mu}_{xi}$ observed 400 MC Samples	90%	2	84.8	84.0	89.9	85.2	90.9
		10	86.9	84.6	89.9	86.1	90.9
		40	89.1	86.0	88.9	87.3	90.4
	95%	2	90.3	89.6	94.7	90.0	94.5
		10	91.8	90.6	94.5	90.6	94.4
		40	93.8	91.2	94.0	92.0	94.4
	99%	2	96.0	95.5	97.9	94.3	97.0
		10	96.6	95.8	98.0	95.0	97.3
		40	98.3	97.4	98.4	96.8	97.9
$\mu_{xi}$ unknown random $\tilde{\mu}_{xi}$ observed 400 MC Samples	90%	2	85.3	83.2	90.0	84.4	90.5
		10	86.6	84.5	89.7	85.9	90.8
		40	89.2	85.9	89.0	87.1	90.2
	95%	2	90.6	89.4	94.8	89.3	94.4
		10	92.2	90.4	94.4	90.8	94.3
		40	94.1	91.2	94.0	91.8	94.3
	99%	2	96.0	95.3	97.8	94.3	97.1
		10	96.7	96.0	97.9	95.3	97.2
		40	98.3	97.3	98.3	96.7	97.8

Table 4.8: Empirical Coverages,  $\mu_{xi}$  known ( $B_1 = 400, B_2 = 1, N = 400$ )

$100(1 - \alpha)\%$	$n_i$	Wald-type	Level 1	Level 1	Level 2	Level 2
		Pivot	Percentile	Pivot	Percentile	Pivot
		$I_W$	$I_{p1}$	$I_1$	$I_{p2T}$	$I_{2T}$
90%	10	88.2	85.8	91.0	86.8	92.2
	20	88.2	86.4	90.1	87.8	91.6
	40	90.3	88.1	90.6	88.9	91.7
95%	10	92.6	90.9	95.7	91.2	95.8
	20	93.2	92.0	95.5	92.5	95.8
	40	95.0	92.9	95.3	93.2	95.7
99%	10	97.0	96.3	98.7	95.7	98.0
	20	97.9	97.2	98.9	96.9	98.3
	40	98.7	97.8	99.2	97.6	98.8

Table 4.9: Properties of Estimated Length of CIs,  $\mu_{xi}$  unknown and fixed,  $\tilde{\mu}_{xi}$  observed ( $B_1 = 400, B_2 = 1, N = 400$ , Entries Multiplied by  $10^3$ )

$100(1 - \alpha)\%$	$n_i$		Wald-type	Level 1	Level 1	Level 2	Level 2
			Pivot	Percentile	Pivot	Percentile	Pivot
			$I_W$	$I_{p1}$	$I_1$	$I_{p2T}$	$I_{2T}$
90%	2	mean	325.4	314.8	374.5	327.4	394.9
		sd	27.0	22.4	29.0	24.4	32.1
	10	mean	284.1	268.9	313.3	282.5	334.0
		sd	19.2	15.7	20.6	17.6	23.5
	40	mean	207.8	191.4	209.4	200.6	221.4
		sd	10.0	7.4	10.5	8.9	12.6
95%	2	mean	387.8	376.3	470.8	382.6	483.3
		sd	32.1	26.7	37.0	29.2	41.0
	10	mean	338.5	321.6	390.5	331.8	408.9
		sd	22.9	18.8	25.9	21.4	30.2
	40	mean	247.7	229.2	256.1	237.3	268.4
		sd	11.9	8.9	13.1	10.9	15.8
99%	2	mean	509.6	494.7	707.6	483.0	682.8
		sd	42.2	35.7	62.2	39.0	69.2
	10	mean	444.9	424.1	581.8	424.6	586.6
		sd	30.1	25.5	43.5	30.2	53.5
	40	mean	325.5	301.9	362.9	306.3	373.8
		sd	15.6	12.1	20.8	15.9	27.2

In Section 4.5 we define the difference between the single bootstrap distribution and the double bootstrap distribution of the pivot-like statistics  $T^*$ ,  $T^{**}$ , respectively, evaluated at the final cutoff point used to construct the level one bootstrap CI for the area mean; see  $d_1$  defined for (4.63). Properties of  $d_1$  for Model 2 are presented in Table 4.10. The results in Table (4.10) are similar to results on properties of  $d_1$  for Models 1 and 3, and for the different sample size configuration considered in the study; see Appendix E. The Monte Carlo average correction in the double bootstrap CI is not significantly different from zero; results consistent with the previous results where the coverages of the double bootstrap CIs were similar to the coverages of the level one bootstrap CIs. The average absolute values of  $d_1$  are higher for the 90% CIs than for the 95% and 99% CIs. The average values of  $d_1$  are negative for the 90% and 95% CIs and positive for the 99% CIs, suggesting heavier tail for the distribution of the pivot-like statistic  $T^*$ , than the tail for the distribution of  $T^{**}$ ; results consistent with the previous results where the coverage error for the 99% CIs was higher than the coverage error for the 90% and 95% CIs.

Table 4.10: Properties of Estimated Correction in Double Bootstrap CIs,  $\mu_{xi}$  unknown and fixed,  $\tilde{\mu}_{xi}$  observed ( $B_1 = 400, B_2 = 1, N = 400$ , Entries Multiplied by  $10^3$ )

$n_i$		90%	95%	99%
2	mean	-5.6984	-2.5943	0.7750
	sd	18.1342	13.0431	6.0635
	se	0.9067	0.6522	0.3032
10	mean	-6.9031	-2.8688	1.1198
	sd	20.4628	14.8211	7.0856
	se	1.0231	0.7411	0.3543
40	mean	-5.9479	-1.9255	1.5396
	sd	20.6586	14.8019	7.2238
	se	1.0329	0.7401	0.3612

#### 4.6.2.3 General Purpose Statistic for CI

We construct general  $(1 - \alpha)\%$  bootstrap CIs using an estimated multiple of a Student-t distribution, as described in Section 4.5.4. The empirical coverages of the general bootstrap CIs are presented in Table 4.11 for the simulation models considered in the study. See Appendix

E for results on the estimated lengths of the general bootstrap CIs. The results are similar to the results in Tables 4.7 and 4.8, for the bootstrap CIs constructed for a specific  $\alpha$ . The novelty of this method is that it enables the user to construct bootstrap CIs at a set of confidence levels at once, without implementing the computationally intensive bootstrap for every  $\alpha$ .

Table 4.11: Empirical Coverages for Bootstrap CIs, Estimated Distribution of  $|T^*|$   
( $B_1 = 400, B_2 = 1$ )

MC Samples		$n_i$	Level 1			Level 2		
			90%	95%	99%	90%	95%	99%
800	$\mu_{xi}$ known	2	91.2	95.0	97.5	91.6	94.6	96.8
		10	90.8	95.2	97.7	91.7	95.0	97.1
		40	89.7	94.7	98.5	90.8	95.0	98.2
400	$\mu_{xi}$ unknown fixed $\tilde{\mu}_{xi}$ observed	2	89.8	94.5	97.9	90.8	94.5	97.4
		10	89.7	94.5	98.0	90.6	94.5	97.6
		40	88.8	94.2	98.4	90.1	94.7	98.2
400	$\mu_{xi}$ unknown random $\tilde{\mu}_{xi}$ observed	2	90.0	94.8	97.8	90.4	94.5	97.4
		10	89.6	94.4	97.9	90.5	94.5	97.5
		40	89.0	94.1	98.4	89.9	94.5	98.2
400	$\mu_{xi}$ known	10	90.7	95.7	98.7	91.9	95.9	98.2
		20	90.1	95.5	98.9	91.1	96.0	98.5
		40	90.6	95.3	99.2	91.4	95.8	99.0

Let  $\hat{\tau}_{i,B}, \hat{d}f_{i,B}$  be the estimated parameters of the Student-t distribution defined for (4.68), for the level one bootstrap, and let  $\hat{\tau}_{i,DB,T}, \hat{d}f_{i,DB,T}$  be the estimated parameters of the Student-t distribution defined for (4.70), for the telescoping double bootstrap. The Monte Carlo properties of  $\hat{\tau}_{i,B}, \hat{d}f_{i,B}, \hat{\tau}_{i,DB,T}$  and  $\hat{d}f_{i,DB,T}$ , for the simulation Model 2 are presented in Table 4.12; see Appendix E for results on the Monte Carlo properties of  $\hat{\tau}_{i,B}, \hat{d}f_{i,B}, \hat{\tau}_{i,DB,T}$  and  $\hat{d}f_{i,DB,T}$  for simulation models 1 and 3, and for the different sample size configuration. Table 4.12 contains the MC 5<sup>th</sup> quantile and the MC 50<sup>th</sup> quantile of the estimated distribution of the estimated bootstrap parameters  $\hat{\tau}_{i,B}, \hat{d}f_{i,B}, \hat{\tau}_{i,DB,T}$  and  $\hat{d}f_{i,DB,T}$ . Also, the the percentage of MC estimated values of  $\hat{d}f_{i,B}$  and  $\hat{d}f_{i,DB,T}$ , greater than 500, are reported in Table 4.12. The estimated parameters  $\hat{\tau}_{i,B}, \hat{d}f_{i,B}, \hat{\tau}_{i,DB,T}$  and  $\hat{d}f_{i,DB,T}$  increase with an increase in sample size, with larger values for the double bootstrap than for the level one bootstrap. The MC median of



$(\hat{d}f_{i,B}, \hat{d}f_{i,DB,T})$  is approximately (7, 18), (7, 14) and (12, 23), for the sample sizes 2, 10 and 40. The MC median of  $(\hat{\tau}_{i,B}, \hat{\tau}_{i,DB,T})$  is approximately (0.84, 0.95), (0.85, 0.96) and (0.92, 0.99), for the sample sizes 2, 10 and 4. The estimated multiple parameter  $(\hat{\tau}_{i,B}, \hat{\tau}_{i,DB,T})$  increases with an increase in the estimated degrees of freedom  $(\hat{d}f_{i,B}, \hat{d}f_{i,DB,T})$ ; see Appendix F for graphical representations of the estimated parameters of the Student-t distribution.

Table 4.12: Monte Carlo Properties of Estimated Parameters for the Student-t Distribution,  $\mu_{xi}$  unknown and fixed,  $\tilde{\mu}_{xi}$  observed ( $B_1 = 400, B_2 = 1, N = 400$ )

		Level1	Level 2	Level1	Level 2
	$n_i$	$\hat{d}f_{i,B}$	$\hat{d}f_{i,DB,T}$	$\hat{\tau}_{i,B}$	$\hat{\tau}_{i,DB,T}$
5 <sup>th</sup> quantile	2	3.5310	3.7240	0.6882	0.7213
	10	3.6388	3.3597	0.7014	0.7086
	40	4.5278	3.6756	0.7615	0.7382
50 <sup>th</sup> quantile	2	6.9599	17.6029	0.8418	0.9488
	10	7.4205	14.0913	0.8544	0.9550
	40	11.9556	22.6981	0.9167	0.9894
% greater than 500	2	3.8125	30.0625		
	10	5.4375	27.8750		
	40	13.9375	33.4583		

### 4.6.3 Alternative Methods and Simulation Results

Alternative Taylor MSE estimators were constructed using the  $g_{3i}$  term in (4.9) and using a Beale type estimator for the  $g_{1i}$  term in (4.9). The alternative Taylor MSE estimators indicated less bias and larger variance for the areas of small sample size. Using the expression (4.10) relative to using the expression defined for (4.9), the correlation between the  $\hat{g}_{1i}$  and  $\hat{g}_{3i}$  is reduced and the correlation between the prediction error  $\hat{\theta}_i^{EBLUP} - \theta_i$  and the estimated prediction MSE  $M\hat{S}E(\hat{\theta}_i^{EBLUP} - \theta_i)$  is increased, leading to less extreme values of statistics of the form (4.54). Hence, the coverages of the CIs constructed with the proposed Taylor MSE estimator in (4.48) were better than the coverages of the CIs constructed with the alternative Taylor MSE estimators. The results agree with the results in Fuller (1990). The author proposes two prediction MSE estimators for the true small area value of a normal response variable. The first proposed estimator is the unconditional prediction MSE estimator, using a  $g_{3i}$  term of the form in (4.9).

The second proposed estimator is the conditional prediction MSE, a function of the area specific data, using a  $g_{3i}$  term of the form (4.10). In the simulation study, Fuller (1990) considers 1000 Monte Carlo samples and a range of error variance from 0.021 to 0.211. The coverages of the 95% CIs constructed using the conditional estimator for the prediction MSE are close to the nominal coverages for the simulation model with normal errors and greater variance of the true values, and greater than the nominal coverages for the simulation models with smaller variance of the true values. The CIs constructed using the unconditional prediction MSE estimator have greater coverage error than the CIs constructed using the conditional prediction MSE estimator.

Alternative double bootstrap  $(1 - \alpha)\%$  CIs were constructed. One double bootstrap method corrects for the bias in the coverage level of the level one bootstrap CI using the average of the differences in the level one bootstrap and level two bootstrap distributions of the pivot-like statistics evaluated at the cutoff points of the level one bootstrap CI and the double bootstrap CI. Another double bootstrap method is based on estimating the cutoff point using local fitting. The normal quantiles of the level one bootstrap distribution of the pivot-like statistic evaluated at five cutoff points are regressed on the level two bootstrap distribution of the pivot-like statistic evaluated at five cutoff points. The set of five cutoff points are equally spaced around and centered at the normal  $100(1 - \alpha)^{th}$  quantile. The simulation results indicated no improvement in the coverage of the double bootstrap CIs constructed with the alternative methods over the coverages of the CIs constructed with the method described for (4.63).

We conducted two simulation studies with different values for  $\sigma_b^2$ . For the case when  $\sigma_b^2 = 0.49$ , the number of samples with estimated  $\sigma_b^2$  equal to the bound is zero and the bootstrap CIs have overcoverage of about 0.2 to 1.9. For the case when  $\sigma_b^2 = 0.0064$ , the proportion of samples with estimated  $\sigma_b^2$  equal to the bound increases to about 63%, resulting in larger coverage errors for the 90% and 95% bootstrap CIs, than the coverage errors for the CIs constructed when  $\sigma_b^2 = 0.25$ .

In this paper we constructed bootstrap CIs for the small area predictions. Using similar methods to the ones described in Section 4.5, we constructed bootstrap CIs for the model parameters  $\beta$  and  $\sigma_b^2$ . For  $\beta_1$ , normal approximated CIs resulted in undercoverage of less than 3% and the level one pivot-type bootstrap CIs resulted in undercoverage of less than 2%. For  $\beta_2$ , normal approximated CIs resulted in undercoverage of less than 1.25% and the level one pivot-type bootstrap CIs resulted in overcoverage of less than 0.25%. For  $\sigma_b^2$ , chi-squared approximated CIs resulted in coverage error of less than 3% and the level one pivot-type bootstrap CIs resulted in undercoverage in the range 2.6% to 7.5%.

#### 4.7 Summary

We study a nonlinear unit level model, with binary response variable. Taylor prediction MSE estimators are presented, for a linear approximation of the model. We propose a new bootstrap bias correction method for the small area mean prediction MSE estimator that reduces the bias in the estimators, for areas of small sample size, at a cost of increasing the variance.

Basic CIs, based on the normal approximation of the parameter, and bootstrap CIs for the small area mean are presented. Basic CIs, based on the normal approximation of the parameter have empirical coverages lower than the desired nominal level. The pivot-type bootstrap CIs have smaller coverage error than the percentile bootstrap CIs. Double bootstrap CIs perform well, but do not improve the coverage accuracy compared to the level one bootstrap CIs; results consistent with the results in Chang and Hall (2014).

A general purpose procedure is proposed for CIs. The procedure is based on a multiple of a Student-t distribution. The degrees of freedom for the Student-t distribution and the standard error of the small area mean prediction are produced. Given the degrees of freedom for the Student-t distribution and the standard error of the small area mean prediction, a CI for the small area mean can be constructed in the common form  $(\hat{\theta}_i \pm \zeta_{1-\alpha/2,i} se(\hat{\theta}_i))$ , where  $\zeta_{1-\alpha/2,i,df_i}$

is the  $100(1 - \alpha/2)^{th}$  quantile of the Student-t distribution with given degrees of freedom  $df_i$ ,  $i$  denotes the area,  $1 - \alpha$  is the desired level, and  $se(\hat{\theta}_i)$  is the given standard error of  $\hat{\theta}_i$ . The coverage of the general bootstrap CI is comparable to the coverage of the level specific bootstrap CI.

## 4.8 Appendix A. Estimated Variance of the Variance of Random Area Effects

### 4.8.1 Model for $\tilde{x}$

Let  $r_i := \bar{u}_{ix} - (\mu_x - \hat{\mu}_x)$ , for  $\bar{u}_{ix}$  defined for (4.15). Then  $E(r_i^2) = \sigma_\delta^2 + a_{i.x}^{-1}\sigma_\epsilon^2$  and  $Var(r_i^2) = 2(\sigma_\delta^2 + a_{i.x}^{-1}\sigma_\epsilon^2)^2$ , where  $a_{i.x} = \sum_{j=1}^{n_i} a_{ijx}$ , as defined for (4.14). Following Wang and Fuller (2003),

$$\hat{\sigma}_\delta^2 = \sum_{i=1}^m w_{2i} [m(m-1)^{-1}(\hat{r}_i^2 - a_{i.x}^{-1}\hat{\sigma}_\epsilon^2),]$$

where  $w_{2i}$  are area specific weights and  $\hat{r}_i := \bar{u}_{ix}$ . Let

$$M_{bb} := \sum_{i=1}^m w_{2i} [m(m-1)^{-1}\hat{r}_i^2,]$$

be the component of  $\hat{\sigma}_\delta^2$  that does not depend on the estimated variance component  $\hat{\sigma}_\epsilon^2$ . If we ignore the variance of  $\hat{\sigma}_\epsilon^2$ , then  $\hat{V}(\hat{\sigma}_\delta^2) \approx \hat{V}(M_{bb})$  and

$$\hat{V}(M_{bb}) = \sum_{i=1}^m w_{2i}^2 [m(m-1)^{-1}2(\hat{\sigma}_\delta^2 + a_{i.x}^{-1}\hat{\sigma}_\epsilon^2)^2].$$

The optimal choice of weights  $w_{2i}$  is

$$w_{2i,opt} = \left( \sum_{i=1}^m [(\hat{\sigma}_\delta^2 + a_{i.x}^{-1}\hat{\sigma}_\epsilon^2)^{-2}] \right)^{-1} [(\hat{\sigma}_\delta^2 + a_{i.x}^{-1}\hat{\sigma}_\epsilon^2)^{-2}].$$

The maximum likelihood estimator is approximately equal to the estimator with optimal weights. Using  $w_{2i,opt}$ , the estimated variance  $\hat{V}(\hat{\sigma}_\delta^2)$  becomes

$$\hat{V}(\hat{\sigma}_\delta^2) = 2(m-1)^{-1}m \left( \sum_{i=1}^m [(\hat{\sigma}_\delta^2 + a_{i.x}^{-1}\hat{\sigma}_\epsilon^2)^{-2}] \right)^{-1}.$$

### 4.8.2 Model for $y$

Let  $a_i := \bar{u}_{yi} - \bar{h}\beta_i(\beta - \hat{\beta})$ . Then  $E(a_i^2) \approx \bar{h}_{b_i,i}^2\sigma_b^2 + n_i^{-1}\sigma_\epsilon^2$  and  $Var(a_i^2) \approx 2(\bar{h}_{b_i,i}^2\sigma_b^2 + n_i^{-1}\sigma_\epsilon^2)^2$ . Following Wang and Fuller (2003),

$$\hat{\sigma}_b^2 = \sum_{i=1}^m w_{2i} [m(m-1)^{-1}(\hat{a}_i^2 - \tilde{\sigma}_{ei}^2)\bar{h}_{b_i,i}^{-2},]$$

where  $w_{2i}$  are area specific weights,  $\hat{\alpha}_i := \bar{u}_{yi} - \bar{h}_{\beta,i}(\hat{\beta} - \hat{\beta}) = \bar{u}_{yi}$  and  $\tilde{\sigma}_{ei}^2 = n_i^{-1}\tilde{\sigma}_e^2$ . Let

$$M_{bb} := \sum_{i=1}^m w_{2i} \left[ m(m-1)^{-1} \hat{\alpha}_i^2 \bar{h}_{b,i}^{-2} \right],$$

be the component of  $\hat{\sigma}_b^2$  that does not depend on the estimated variance component  $\tilde{\sigma}_{ei}^2$ . If we ignore the variance of  $\tilde{\sigma}_{ei}^2$ , then  $\hat{V}(\hat{\sigma}_b^2) \approx \hat{V}(M_{bb})$  and

$$\hat{V}(M_{bb}) = \sum_{i=1}^m w_{2i}^2 \left[ m(m-1)^{-1} 2(\bar{h}_{b,i}^2 \hat{\sigma}_b^2 + n_i^{-1} \tilde{\sigma}_e^2)^2 \bar{h}_{b,i}^{-4} \right]$$

The optimal choice of weights  $w_{2i}$  is

$$w_{2i,opt} = \left( \sum_{i=1}^m [(\bar{h}_{b,i}^2 \hat{\sigma}_b^2 + n_i^{-1} \tilde{\sigma}_e^2)^{-2} \bar{h}_{b,i}^4] \right)^{-1} [(\bar{h}_{b,i}^2 \hat{\sigma}_b^2 + n_i^{-1} \tilde{\sigma}_e^2)^{-2} \bar{h}_{b,i}^4].$$

The maximum likelihood estimator is approximately equal to the estimator with optimal weights. Using  $w_{2i,opt}$ , the estimated variance  $\hat{V}(\hat{\sigma}_b^2)$  becomes

$$\hat{V}(\hat{\sigma}_b^2) = 2(m-1)^{-1} m \left( \sum_{i=1}^m [(\bar{h}_{b,i}^2 \hat{\sigma}_b^2 + n_i^{-1} \tilde{\sigma}_e^2)^{-2} \bar{h}_{b,i}^4] \right)^{-1}.$$

## 4.9 Appendix B. Alternative Double Bootstrap Methods

We outline two double bootstrap calibration methods for constructing symmetric  $(1 - \alpha)$ -level CIs. One method estimates the bias in the confidence interval level by the average biases in level one and level two levels, and the other is a calibration method based on local fitting for the estimation of cutoff points.

For this methods, the  $T$  statistics are scaled by the level one bootstrap standard deviation of  $|T_i^*|$

$$\hat{\sigma}_{i,B}^* = 2^{-1} |T^*|_{i,([.95B_1]+1)},$$

where  $|T^*|_{i,([.95B_1]+1)}$  denotes the 95% quantile point of  $|T_i^*|$ . Let the pivot-type statistics be

$$\hat{T}_{i,k,1}^* := \hat{T}_{i,k}^* / \hat{\sigma}_{i,B}^*, \quad \hat{T}_{i,k,1}^{**} := \hat{T}_{i,k}^{**} / \hat{\sigma}_{i,B}^*.$$

*Calibration - averaging biases.*

Let  $F^*(\zeta)_i = B_1^{-1} \sum_{k=1}^{B_1} I(|T_{i,k,1}^*| < \zeta)$  be the estimated distribution of the level one statistic  $T_{i,1}^*$  and let  $F^{**}(\zeta)_i = B_1^{-1} \sum_{k=1}^{B_1} I(|T_{i,k,1}^{**}| < \zeta)$  be the estimated distribution of the level two statistic  $T_{i,1}^{**}$ . Let a CI be

$$I_{22} := \left( \hat{\theta}_i \pm \hat{\zeta}_{i,C} \hat{\sigma}_{i,T} \hat{\sigma}_{i,B}^* \right), \quad (4.73)$$

where

$$\begin{aligned} x_{1,i,\alpha} &= F_i^{*-1}(1 - \alpha), \\ x_{2,i,\alpha} &= F_i^{*-1} [F_i^{**}(F_i^{*-1}(1 - \alpha))] = F_i^{*-1}(F_i^*(x_1) - d_1), \\ d_{1,i,\alpha} &= F_i^*(x_1) - F_i^{**}(x_1), \\ d_{2,i,\alpha} &= F_i^*(x_2) - F_i^{**}(x_2), \\ d_{3,i,\alpha} &= 0.5(d_1 + d_2), \\ \hat{\zeta}_{i,C} &= F_i^{*-1}(F_i^*(x_1) - d_3). \end{aligned}$$

Notice that  $x_1$  is the estimated cutoff point  $\hat{\zeta}_{i,\alpha}$  in (4.59),  $d_{1,i,\alpha}$  is the estimated bias in the level one confidence interval level in (4.64), and  $x_2$  is the estimated cutoff point  $\hat{\zeta}_{i,C}$  in (4.63).

*Calibration - local fitting.*

We estimate the distribution functions of the scaled bootstrap statistics  $T_{i,1}^*$  and  $T_{i,1}^{**}$  at five points,  $\zeta = 1.8, 1.9, 2.0, 2.1, 2.2$ . Let  $x_{k_i}^*$  be the normal quantiles for  $F_i^*(\zeta)$  and let  $x_{k_i}^{**}$  be the normal quantiles for  $F_i^{**}(\zeta)$ .

We denote the double bootstrap interval by

$$I_3 := \left( \hat{\theta}_i \pm \hat{\zeta}_{i,C} \hat{\sigma}_{i,T} \hat{\sigma}_{i,B}^* \right), \quad (4.74)$$

where  $\hat{\zeta}_{i,C}$  is the area specific double bootstrap bias corrected cutoff point constructed in the local fitting method.

*Algorithm* (Extended to 4 or 5 points).

If  $x_{k,i,\zeta=2}^* \leq x_{k,i,\zeta=2}^{**}$ , fit the regression  $E((\mathbf{x}_i^*, \mathbf{x}_i^{**})' | \zeta) = (1_3 \otimes I_{(2 \times 2)}, \zeta_{1:3} \otimes 1_2) \boldsymbol{\beta}$  and estimate the parameters  $\boldsymbol{\beta}$ .

If  $x_{k,i,\zeta=2}^* > x_{k,i,\zeta=2}^{**}$ , fit the regression  $E((\mathbf{x}_i^*, \mathbf{x}_i^{**})' | \zeta) = (1_3 \otimes I_{(2 \times 2)}, \zeta_{3:5} \otimes 1_2) \boldsymbol{\beta}$  and estimate the parameters  $\boldsymbol{\beta}$ .

Let  $\mathbf{x}_{i,C}^{**} = 2\hat{\beta}_{0,i} - \hat{\beta}_{1,i} + \hat{\beta}_{2,i}\zeta$  be the double bootstrap calibrated cutoff points that define the double bootstrap estimated distribution function. The final  $\hat{\zeta}_{i,C}$  cutoff point is the solution to  $1.96 = 2\hat{\beta}_{0,i} - \hat{\beta}_{1,i} + \hat{\beta}_{2,i}\zeta_C$ .

The estimated standard error of the prediction error is  $\hat{c}_i = 1.96^{-1} \hat{\zeta}_{i,C} \hat{\sigma}_{i,T} \hat{\sigma}_{i,B}^*$  and the length of the interval  $I_3$  is  $2\hat{c}_i$ .



## 4.10 Appendix C. Two-sided $(1 - \alpha)$ -level CIs for other Model Parameters

We compute two-sided  $(1 - \alpha)$ -level CIs for the variance component  $\sigma_b^2$  and for the vector of fixed effects coefficients,  $\beta$ . Let  $(\hat{\beta}, \hat{\sigma}_b^2)$  be the estimated vector of parameters  $(\beta, \sigma_b^2)$ . Level one bootstrap samples are generated using the vector  $(\hat{\beta}, \hat{\sigma}_b^2)$  and bootstrap estimates are denoted by  $(\hat{\beta}_k^*, \hat{\sigma}_{b_k}^{2*})$ , for  $k = 1, \dots, B_1$ . Similarly, double bootstrap samples are generated using the vector  $(\hat{\beta}_k^*, \hat{\sigma}_{b_k}^{2*})$  and double bootstrap estimates are denoted by  $(\hat{\beta}_k^{**}, \hat{\sigma}_{b_k}^{2**})$ , for  $k = 1, \dots, B_1$ .

### 4.10.1 Random Effects Variance Component

The  $100(1 - \alpha)\%$  Wald CI for  $\sigma_b^2$  would be  $\hat{\sigma}_b^2 \pm z_{\alpha/2} \sqrt{\hat{V}(\sigma_b^2)}$ , symmetric about  $\hat{\sigma}_b^2$ , where  $\hat{V}(\sigma_b^2)$  is defined in (4.52). However, under the assumption of normality of the data, the distribution of a variance component is a chi-squared distribution, which is not symmetric. Moreover, the number of degrees of freedom for the estimation of  $\sigma_b^2$  is small. Hence, we construct Satterthwaite-type CIs. The pivot-type CIs are constructed using the statistic  $\hat{R}_\sigma := \frac{\hat{\sigma}_b^2}{\sigma_b^2}$ .

Basic CIs for  $\sigma_b^2$  are constructed by approximating the distribution of  $\hat{R}_\sigma$  by a Chi-squared distribution, with  $\nu$  degrees of freedom. The degrees of freedom parameter is estimated as  $\hat{\nu} = 2 \left( \hat{\sigma}_b^2 / \sqrt{\hat{V}(\sigma_b^2)} \right)^2$  and the basic  $(1 - \alpha)$ -level CI for  $\sigma_b^2$  is

$$\left( \frac{\hat{\nu} \hat{\sigma}_b^2}{\chi_{\hat{\nu}, 1-\alpha/2}^2}, \frac{\hat{\nu} \hat{\sigma}_b^2}{\chi_{\hat{\nu}, \alpha/2}^2} \right). \quad (4.75)$$

Bootstrap pivot-type CIs are based on the bootstrap distributions of the single and double bootstrap estimates of the pivot-type statistic  $\hat{R}_{\sigma,k}^* := \frac{\hat{\sigma}_{b_k}^{2*}}{\hat{\sigma}_b^2}$  and  $\hat{R}_{\sigma,k}^{**} := \frac{\hat{\sigma}_{b_k}^{2**}}{\hat{\sigma}_{b_k}^{2*}}$ , respectively. Also, percentile CIs for  $\sigma_b^2$  are constructed, based on the distribution of the bootstrap prediction errors  $\hat{\sigma}_{b_k}^{2*} - \hat{\sigma}_b^2$ .

### 4.10.2 Fixed Effects Coefficients

Percentile CIs for  $\beta$  are constructed, based on the bootstrap distributions of the bootstrap prediction errors  $\hat{\beta}_k^* - \hat{\beta}$ ,  $\hat{\beta}_k^{**} - \hat{\beta}_k^*$ . Also, pivot-type CIs for  $\beta$  are constructed, using the statistic  $\hat{T}_i := (\hat{\beta} - \beta) \hat{V}^{-1/2}(\hat{\beta})$ , where  $\hat{V}(\hat{\beta})$  is the estimated variance of the estimated vector

of coefficients  $\beta$  from the model fit for  $\mathbf{y}$ , using  $R$ . Bootstrap pivot-type CIs are based on the bootstrap distributions of the single and double bootstrap estimates of the pivot-type statistic  $\hat{T}_k^* := (\hat{\beta}_k^* - \hat{\beta})\hat{V}^{-1/2}(\hat{\beta}_k^*)$  and  $\hat{T}_{i,k}^{**} := (\hat{\beta}_k^{**} - \hat{\beta}_k^*)\hat{V}^{-1/2}(\hat{\beta}_k^{**})$ , respectively.

### 4.10.3 Simulation Results: Confidence Intervals for the Model Parameters

We construct 90%, 95% and 99% confidence intervals for  $(\sigma_b^2, \beta)$  using different bootstrap methods described in Section 4.5. Also, Wald-type CI based on the pivot-like method, denoted  $I_W$  are constructed. Recall that  $I_1$  is the level one pivot-type bootstrap CI. The empirical coverages are given in Table 4.13. The CIs for the parameters  $\sigma_b^2, \beta_0$  lead to undercoverage. Basic CIs for  $\sigma_b^2$  given in (4.75) perform best, bootstrap CIs lead to undercoverage. Basic CIs for  $\beta$  given in perform well, but level one bootstrap CIs perform best.

Table 4.13: Empirical Coverages for Bootstrap CIs  
( $\mu_{xi}$  known,  $B_1 = 400$  and 800 MC Samples)

$100(1 - \alpha)\%$		$I_W$	$I_1$
90%	$\beta_0$	86.25	88.00
	$\beta_1$	89.75	90.25
	$\sigma_b^2$	93.00	86.50
95%	$\beta_0$	92.00	94.00
	$\beta_1$	93.75	95.00
	$\sigma_b^2$	94.50	88.25
99%	$\beta_0$	97.50	97.25
	$\beta_1$	99.00	99.00
	$\sigma_b^2$	97.75	91.50

### 4.11 Appendix D. Numerical Integration

There are many ways to approximate the integrals in (4.2, 4.17, 4.19, 4.21). The algorithms available in  $R$  are very slow. We create a finite discrete approximation using a set of numbers  $z_k, k = 1, 2, \dots, K$  and a set of weights  $w_k, k = 1, 2, \dots, K$  such that

$$\sum_{k=1}^K w_k (z_k, z_k^2, z_k^3, z_k^4) \approx (0, 1, 0, 3). \quad (4.76)$$

The  $\{z_k, w_k\}$  is an approximation for the normal distribution. See Erciulescu and Fuller (2013) for  $K = 50$  and  $w_k = 1/K$ , for all  $k$ . We reduce the error in the numerical approximation, by considering unequal weights  $w_k$  and a set of  $K = 26$  points. Let the set of points be

$$z = \pm(0.1, 0.306, 0.525, 0.758, 0.998, 1.259, 1.519, 1.69, 1.80, 1.95, 2.26, 2.5, 2.9)$$

and let the set of weights  $w$  be the set

$$(0.08, 0.08, 0.08, 0.07, 0.06, 0.05, 0.03, 0.01, 0.01, 0.01, 0.008, 0.007, 0.005),$$

with each of the elements repeated twice.

The approximations for the integral expressions in (4.2, 4.17, 4.19, 4.21), when the parameters are estimated, are:

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

$$\theta_i = \sum_{k=1}^K w_k g(\mathbf{x}_{ik}^*, \boldsymbol{\beta}, b_i), \quad (4.77)$$

where  $x_{ik}^* = \mu_{xi} + z_k \sigma_\epsilon$ .

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

$$\hat{\theta}_i = \frac{\sum_{k=1}^K w_k \sum_{k=1}^K w_k g(\mathbf{x}_{ik}^*, \boldsymbol{\beta}, b_k^*) \prod_{t=1}^{n_i} f(y_{it}|x_{it}, b_k^*)}{\sum_{k=1}^K w_k \prod_{t=1}^{n_i} f(y_{it}|x_{it}, b_k^*)}, \quad (4.78)$$

where  $x_{ik}^* = \mu_{xi} + z_k \hat{\sigma}_\epsilon$  and  $b_k^* = z_k \hat{\sigma}_b$ .

(iii) predicted small area mean of  $\mathbf{y}$  with unknown, fixed  $\mu_{xi}$  and auxiliary information  $\tilde{\boldsymbol{\mu}}_x$

$$\hat{\theta}_i = \frac{\sum_{k=1}^K w_k \sum_{k=1}^K w_k g(\mathbf{x}_{ik}^*, \boldsymbol{\beta}, b_k^*) \prod_{t=1}^{n_i} f(y_{it}|x_{it}, b_k^*)}{\sum_{k=1}^K w_k \prod_{t=1}^{n_i} f(y_{it}|x_{it}, b_k^*)}, \quad (4.79)$$

where  $x_{ik}^* = \hat{\mu}_{xi} + z_k \hat{\sigma}_\epsilon$  and  $b_k^* = z_k \hat{\sigma}_b$ .

(iv) predicted small area mean of  $\mathbf{y}$  with unknown random covariate mean and auxiliary information

$\tilde{\boldsymbol{\mu}}_x$

$$\hat{\theta}_i = \frac{\sum_{k=1}^K w_d \sum_{k=1}^K w_k \sum_{k=1}^K w_k g(\mu_x + \delta_k^* + \epsilon_k^*, \boldsymbol{\beta}, b_k^*) \prod_{t=1}^{n_i} f(y_{it}|x_{it}, b_k^*) f(x_{it}|\delta_k^*) f(\tilde{\mu}_{xi}|\delta_k^*)}{\sum_{k=1}^K w_k \sum_{k=1}^K w_k \prod_{t=1}^{n_i} f(y_{it}|x_{it}, b_k^*) f(x_{it}|\delta_k^*) f(\tilde{\mu}_{xi}|\delta_k^*)}, \quad (4.80)$$

where  $\epsilon_k^* = z_k \hat{\sigma}_\epsilon$ ,  $b_k^* = z_k \hat{\sigma}_b$  and  $\delta_k^* = z_k \hat{\sigma}_\delta$ .

## 4.12 Appendix E. Additional Simulation Results

Table 4.14: Monte Carlo Properties of Prediction MSE Estimators,  
 $\mu_{xi}$  known ( $B_1 = 400, B_2 = 1, N = 800$  , Variances Multiplied by  $10^3$ )

$n_i$		Taylor	Level 1		Level 2	
		$\hat{\alpha}^T$	$\hat{\alpha}^*$	$\hat{\alpha}_{\Delta}^*$	$\hat{\alpha}_T^{**}$	$\hat{\alpha}_{\Delta,T}^{**}$
2	$V(\hat{\theta} - \theta)$	9.7234	9.7234	9.7234	9.7234	9.7234
	RelBias	-0.0515	-0.1260	-0.0295	-0.0339	-0.0067
	$CV(\hat{\alpha})$	0.5614	0.4503	0.5573	0.4998	0.6888
	Bias/sd	-0.0923	-0.2802	-0.0534	-0.0680	-0.0103
	Bias/se	-1.8462	-5.6047	-1.0686	-1.3599	-0.2064
10	$V(\hat{\theta} - \theta)$	7.2419	7.2419	7.2419	7.2419	7.2419
	RelBias	-0.0203	-0.1355	-0.0168	-0.0352	0.0164
	$CV(\hat{\alpha})$	0.4443	0.3425	0.4485	0.3815	0.5930
	Bias/sd	-0.0460	-0.3961	-0.0376	-0.0927	0.0276
	Bias/se	-0.9208	-7.9221	-0.7518	-1.8546	0.5530
40	$V(\hat{\theta} - \theta)$	3.5974	3.5974	3.5974	3.5974	3.5974
	RelBias	0.0552	-0.1155	-0.0353	-0.0248	-0.0139
	$CV(\hat{\alpha})$	0.3238	0.2176	0.3599	0.2419	0.6024
	Bias/sd	0.1711	-0.5308	-0.0975	-0.1029	-0.0227
	Bias/se	3.4220	-10.6155	-1.9502	-2.0576	-0.4540

Table 4.15: Monte Carlo Properties of Prediction MSE Estimators,  
 $\mu_{xi}$  known ( $B_1 = 400, B_2 = 1, N = 400$ , Variances Multiplied by  $10^3$ )

$n_i$		Taylor	Level 1		Level 2	
		$\hat{\alpha}^T$	$\hat{\alpha}^*$	$\hat{\alpha}_{\Delta}^*$	$\hat{\alpha}_T^{**}$	$\hat{\alpha}_{\Delta,T}^{**}$
10	$V(\hat{\theta} - \theta)$	7.2575	7.2575	7.2575	7.2575	7.2575
	RelBias	0.0108	-0.1146	-0.0131	-0.0297	0.0114
	$CV(\hat{\alpha})$	0.3910	0.2841	0.3773	0.3057	0.5304
	Bias/sd	0.0277	-0.4034	-0.0348	-0.0971	0.0215
	Bias/se	0.5540	-8.0681	-0.6965	-1.9425	0.4302
20	$V(\hat{\theta} - \theta)$	5.3147	5.3147	5.3147	5.3147	5.3147
	RelBias	-0.0060	-0.0911	-0.0116	-0.0056	-0.0025
	$CV(\hat{\alpha})$	0.3170	0.2364	0.3216	0.2538	0.4648
	Bias/sd	-0.0190	-0.3852	-0.0360	-0.0221	-0.0054
	Bias/se	-0.3795	-7.7048	-0.7198	-0.4415	-0.1089
40	$V(\hat{\theta} - \theta)$	3.3877	3.3877	3.3877	3.3877	3.3877
	RelBias	0.1110	-0.0462	0.0229	0.0317	0.0340
	$CV(\hat{\alpha})$	0.2825	0.1904	0.3011	0.2060	0.5041
	Bias/sd	0.3930	-0.2427	0.0759	0.1537	0.0675
	Bias/se	7.8593	-4.8544	1.5179	3.0741	1.3493

Table 4.16: Monte Carlo Properties of Prediction MSE Estimators,  
 $\mu_{xi}$  unknown, random,  $\tilde{\mu}_{xi}$  observed  
( $B_1 = 400, B_2 = 1, N = 400$ , Variances Multiplied by  $10^3$ )

$n_i$		Taylor $\hat{\alpha}^T$	Level 1 $\hat{\alpha}^*$		Level 2 $\hat{\alpha}_T^{**}$	
				$\hat{\alpha}_\Delta^*$		$\hat{\alpha}_{\Delta,T}^{**}$
2	$V(\hat{\theta} - \theta)$	10.7758	10.7758	10.7758	10.7758	10.7758
	RelBias	-0.0182	-0.1057	-0.0145	-0.0197	0.0064
	$CV(\hat{\alpha})$	0.5070	0.4142	0.4957	0.4594	0.5771
	Bias/sd	-0.0360	-0.2551	-0.0292	-0.0428	0.0111
	Bias/se	-0.7196	-5.1017	-0.5841	-0.8559	0.2226
10	$V(\hat{\theta} - \theta)$	7.8867	7.8867	7.8867	7.8867	7.8867
	RelBias	-0.0023	-0.1181	-0.0098	-0.0213	0.0170
	$CV(\hat{\alpha})$	0.4198	0.3199	0.4202	0.3574	0.5550
	Bias/sd	-0.0055	-0.3691	-0.0233	-0.0595	0.0307
	Bias/se	-0.1097	-7.3830	-0.4662	-1.1907	0.6133
40	$V(\hat{\theta} - \theta)$	3.9183	3.9183	3.9183	3.9183	3.9183
	RelBias	0.0474	-0.1168	-0.0418	-0.0312	-0.0224
	$CV(\hat{\alpha})$	0.3135	0.2042	0.3465	0.2286	0.5848
	Bias/sd	0.1513	-0.5719	-0.1207	-0.1365	-0.0384
	Bias/se	3.0260	-11.4380	-2.4148	-2.7303	-0.7673

Table 4.17: Properties of Estimated Length of 95% Wald-type CIs  
 Bootstrap Estimated Prediction MSE  
 ( $B_1 = 400, B_2 = 1$ , Entries Multiplied by  $10^3$ )

	$n_i$	$I_{B,\Delta}$	$I_{DBT,\Delta}$	$I_B$	$I_{DBT}$
$\mu_{xi}$ known	2	352.4	352.0	347.3	328.2
	sd	29.3	29.1	28.8	26.8
800 MC Samples	10	307.0	306.1	302.4	282.5
	sd	20.4	20.1	19.9	18.5
	40	221.1	220.5	218.9	205.7
	sd	9.2	9.1	8.9	8.7
$\mu_{xi}$ unknown fixed	2	382.2	381.9	377.1	358.2
	$\tilde{\mu}_{xi}$ observed	sd	27.3	27.1	26.8
400 MC Samples	10	326.4	325.5	321.9	303.3
	sd	19.1	18.9	18.7	17.4
	40	231.3	230.7	229.0	216.7
	sd	8.9	8.7	8.6	8.3
$\mu_{xi}$ unknown random	2	379.0	378.8	373.3	354.2
	$\tilde{\mu}_{xi}$ observed	sd	27.6	27.4	27.0
400 MC Samples	10	325.0	324.1	320.4	301.3
	sd	19.3	19.1	18.8	17.4
	40	230.9	230.2	228.7	216.1
	sd	9.0	8.8	8.6	8.4

Table 4.18: Properties of Estimated Length of 95% Wald-type CIs,  
 $\mu_{xi}$  known, Bootstrap Estimated Prediction MSE  
 ( $B_1 = 400, B_2 = 1, N = 400$  MC Samples, Entries Multiplied by  $10^3$ )

$n_i$	$I_{B,\Delta}$	$I_{DBT,\Delta}$	$I_B$	$I_{DBT}$
10	313.1	312.1	309.1	293.1
sd	16.7	16.5	16.4	15.9
20	272.3	271.1	269.4	255.8
sd	12.1	11.9	11.7	11.6
40	223.3	222.5	221.3	211.2
sd	7.6	7.4	7.4	7.5



Table 4.19: Properties of Estimated Length of CIs,  
 $\mu_{xi}$  known ( $B_1 = 400, B_2 = 1, N = 800$ , Entries Multiplied by  $10^3$ )

$100(1 - \alpha)\%$	$n_i$	Wald-type	Level 1	Level 1	Level 2	Level 2
		Pivot	Percentile	Pivot	Percentile	Pivot
		$I_W$	$I_{p1}$	$I_1$	$I_{p2T}$	$I_{2T}$
90%	2	300.8	289.3	374.6	301.3	398.9
	sd	27.9	24.0	29.2	26.2	33.1
	10	268.6	252.3	309.2	265.7	332.8
	sd	19.6	16.7	20.9	18.7	24.4
	40	199.7	182.8	202.9	192.3	215.8
	sd	9.9	7.7	10.5	9.1	12.6
95%	2	358.4	345.5	487.5	350.2	500.5
	sd	33.2	28.6	38.1	31.3	43.7
	10	320.1	301.7	394.3	311.0	415.5
	sd	23.4	20.0	27.0	22.8	32.6
	40	238.0	218.7	249.8	226.9	263.1
	sd	11.8	9.2	13.1	11.2	16.2
99%	2	471.0	454.3	795.2	440.1	750.5
	sd	43.7	38.2	73.3	41.9	82.0
	10	420.7	398.2	624.0	396.4	625.1
	sd	30.7	26.7	50.4	31.4	63.1
	40	312.8	288.6	362.5	293.1	375.3
	sd	15.5	12.3	22.1	16.2	30.2

Table 4.20: Properties of Estimated Length of CIs,  
 $\mu_{xi}$  known ( $B_1 = 400, B_2 = 1, N = 400$ , Entries Multiplied by  $10^3$ )

$100(1 - \alpha)\%$	$n_i$	Wald-type	Level 1	Level 1	Level 2	Level 2
		Pivot	Percentile	Pivot	Percentile	Pivot
		$I_W$	$I_{p1}$	$I_1$	$I_{p2T}$	$I_{2T}$
90%	10	275.6	257.9	298.7	268.9	315.2
	sd	17.0	13.8	16.1	15.1	18.1
	20	235.4	224.8	250.4	234.5	263.9
	sd	12.1	10.0	12.1	11.2	13.9
	40	199.8	184.8	199.1	191.8	207.9
	sd	8.3	6.4	8.2	7.4	9.7
95%	10	328.4	308.3	373.1	316.4	387.5
	sd	20.2	16.6	19.7	18.5	23.1
	20	280.5	268.9	309.2	277.5	323.0
	sd	14.4	11.9	15.0	13.7	17.9
	40	238.0	221.2	242.7	227.7	251.9
	sd	9.9	7.7	10.1	9.2	12.2
99%	10	431.5	406.2	569.2	406.1	571.4
	sd	26.6	22.4	35.1	26.2	45.3
	20	368.7	355.1	458.5	359.0	472.0
	sd	18.9	16.3	27.1	20.1	38.0
	40	312.8	291.2	343.4	294.8	353.0
	sd	13.0	10.5	17.2	13.7	24.2

Table 4.21: Properties of Estimated Length of CIs,  
 $\mu_{xi}$  unknown, random,  $\tilde{\mu}_{xi}$  observed ( $B_1 = 400, B_2 = 1, N = 400$ , Entries Multiplied by  $10^3$ )

$100(1 - \alpha)\%$	$n_i$		Wald-type Pivot $I_W$	Level 1 Percentile $I_{p1}$	Level 1 Pivot $I_1$	Level 2 Percentile $I_{p2T}$	Level 2 Pivot $I_{2T}$
90%	2	mean	326.2	310.6	373.1	322.8	392.9
		sd	25.9	22.4	27.5	24.5	30.6
	10	mean	284.4	267.2	311.9	280.8	332.4
		sd	18.9	15.7	20.2	17.7	23.1
	40	mean	207.9	191.0	208.6	200.1	220.6
		sd	9.9	7.4	10.4	8.9	12.4
95%	2	mean	388.8	370.4	467.9	376.0	479.4
		sd	30.9	26.8	34.7	29.1	38.6
	10	mean	338.8	319.4	388.2	329.8	407.0
		sd	22.6	18.9	25.4	21.6	29.8
	40	mean	247.8	228.6	255.0	236.9	267.2
		sd	11.8	8.9	12.9	10.9	15.7
99%	2	mean	510.9	487.2	700.8	475.7	675.0
		sd	40.6	35.9	58.7	39.3	65.4
	10	mean	445.3	420.4	576.0	420.7	583.0
		sd	29.6	25.5	42.7	30.0	53.4
	40	mean	325.6	300.8	359.5	305.4	370.1
		sd	15.5	12.1	20.5	16.0	27.5

Table 4.22: Properties of Estimated Correction in Double Bootstrap CIs,  
 $\mu_{xi}$  known ( $B_1 = 400, B_2 = 1, N = 800$ , Entries Multiplied by  $10^3$ )

$n_i$		90%	95%	99%
2	mean	-4.1406	-0.7906	1.7964
	sd	24.2739	17.2583	8.1619
	se	0.8582	0.6102	0.2886
10	mean	-5.1445	-1.0937	1.8839
	sd	23.3405	16.7615	8.1570
	se	0.8252	0.5926	0.2884
40	mean	-5.4354	-1.4211	1.9721
	sd	20.4117	15.0533	7.4614
	se	0.7217	0.5322	0.2638

Table 4.23: Properties of Estimated Correction in Double Bootstrap CIs,  
 $\mu_{xi}$  known ( $B_1 = 400, B_2 = 1, N = 400$ , Entries Multiplied by  $10^3$ )

$n_i$		90%	95%	99%
10	mean	2.3292	3.9630	3.6922
	sd	18.5940	14.0947	7.5062
	se	0.9297	0.7047	0.3753
20	mean	0.6792	2.6703	3.5312
	sd	18.7010	13.9082	7.3816
	se	0.9351	0.6954	0.3691
40	mean	-0.4760	1.8370	3.1318
	sd	17.7089	13.1627	6.9726
	se	0.8854	0.6581	0.3486

Table 4.24: Properties of Estimated Correction in Double Bootstrap CIs,  
 $\mu_{xi}$  unknown, random,  $\tilde{\mu}_{xi}$  observed ( $B_1 = 400, B_2 = 1, N = 400$ , Entries Multiplied by  $10^3$ )

$n_i$		90%	95%	99%
2	mean	-4.7036	-1.9922	1.0443
	sd	18.4035	13.2826	6.3164
	se	0.9202	0.6641	0.3158
10	mean	-5.1042	-1.8594	1.2703
	sd	19.6573	14.3928	6.8010
	se	0.9829	0.7196	0.3400
40	mean	-4.7833	-1.5099	1.7562
	sd	19.1674	13.9510	6.8994
	se	0.9584	0.6976	0.3450

Table 4.25: Properties of the Length of Bootstrap CIs, Estimated Distribution of  $|T^*|$ ,  
 $\mu_{xi}$  known ( $B_1 = 400, B_2 = 1, N = 800$ )

$n_i$	Level 1			Level 2		
	90%	95%	99%	90%	95%	99%
2	375.6	485.6	796.9	398.0	499.1	759.2
sd	29.0	38.1	72.6	33.2	42.6	80.0
10	309.0	394.4	624.5	331.4	415.5	631.6
sd	20.8	26.9	49.8	24.5	31.5	61.5
40	202.4	250.3	363.6	214.0	263.7	381.8
sd	10.4	13.0	21.7	12.6	15.7	28.9

Table 4.26: Properties of the Length of Bootstrap CIs, Estimated Distribution of  $|T^*|$ ,  $\mu_{xi}$  known ( $B_1 = 400, B_2 = 1, N = 400$ )

$n_i$	Level 1			Level 2		
	90%	95%	99%	90%	95%	99%
10	297.7	374.3	569.5	312.7	389.0	578.0
sd	16.0	19.6	34.5	17.9	22.3	43.4
20	249.4	310.4	459.4	261.6	324.4	478.4
sd	12.0	14.9	26.5	13.8	17.4	36.2
40	198.3	243.2	345.5	205.9	252.5	360.4
sd	8.2	10.0	16.7	9.5	11.8	22.7

Table 4.27: Properties of the Length of Bootstrap CIs, Estimated Distribution of  $|T^*|$ ,  $\mu_{xi}$  unknown,  $\tilde{\mu}_{xi}$  observed ( $B_1 = 400, B_2 = 1, N = 400$ )

	$n_i$	Level 1			Level 2		
		90%	95%	99%	90%	95%	99%
$\mu_{xi}$ fixed	2	374.7	470.1	708.9	392.6	483.0	694.4
	sd	28.9	36.9	61.7	32.1	40.3	67.4
	10	312.9	390.8	582.6	331.7	409.6	595.5
	sd	20.4	25.8	43.0	23.4	29.5	51.7
	40	208.8	256.4	364.8	219.4	268.7	381.9
	sd	10.5	12.9	20.4	12.4	15.3	26.1
$\mu_{xi}$ random	2	373.1	467.4	702.1	390.2	479.5	687.0
	sd	27.4	34.8	58.0	30.4	38.2	63.4
	10	311.4	388.5	577.2	330.1	407.5	592.1
	sd	20.1	25.3	42.2	23.1	29.1	51.7
	40	208.1	255.1	361.6	218.6	267.4	378.6
	sd	10.3	12.8	20.1	12.3	15.2	26.2

Table 4.28: Monte Carlo Properties of Estimated Parameters for the Student-t Distribution,  $\mu_{xi}$  known ( $B_1 = 400, B_2 = 1, N = 800$ )

	$n_i$	Level1	Level 2	Level1	Level 2
		$\hat{d}f_{i,B}$	$\hat{d}f_{i,DB,T}$	$\hat{\tau}_{i,B}$	$\hat{\tau}_{i,DB,T}$
5 <sup>th</sup> quantile	2	2.5917	2.7925	0.5822	0.6241
	10	2.8811	2.7729	0.6209	0.6442
	40	3.9344	3.2953	0.7215	0.7044
50 <sup>th</sup> quantile	2	4.5886	8.9194	0.7500	0.8786
	10	5.2679	8.9194	0.7882	0.9048
	40	9.4680	15.9400	0.8910	0.9724
% greater than 500	2	1.91	21.15		
	10	2.75	20.73		
	40	10.13	29.12		

Table 4.29: Monte Carlo Properties of Estimated Parameters for the Student-t Distribution,  
 $\mu_{xi}$  known ( $B_1 = 400, B_2 = 1, N = 400$ )

	$n_i$	Level1	Level 2	Level1	Level 2
		$\hat{d}f_{i,B}$	$\hat{d}f_{i,DB,T}$	$\hat{\tau}_{i,B}$	$\hat{\tau}_{i,DB,T}$
5 <sup>th</sup> quantile	10	3.0797	2.9619	0.6498	0.6612
	20	3.3748	2.9558	0.6785	0.6718
	40	4.2962	3.5061	0.7456	0.7194
50 <sup>th</sup> quantile	10	6.5805	10.3262	0.8354	0.9239
	20	8.2943	12.5356	0.8763	0.9490
	40	13.5966	21.5902	0.9296	0.9824
% greater than 500	10	5.7917	22.6458		
	20	9.6042	25.7083		
	40	16.7917	33.3750		

Table 4.30: Monte Carlo Properties of Estimated Parameters for the Student-t Distribution,  
 $\mu_{xi}$  unknown, random,  $\tilde{\mu}_{xi}$  observed ( $B_1 = 400, B_2 = 1, N = 400$ )

	$n_i$	Level1	Level 2	Level1	Level 2
		$\hat{d}f_{i,B}$	$\hat{d}f_{i,DB,T}$	$\hat{\tau}_{i,B}$	$\hat{\tau}_{i,DB,T}$
5 <sup>th</sup> quantile	2	3.5632	3.7073	0.6911	0.7209
	10	3.6564	3.3276	0.7035	0.7127
	40	4.6181	3.8659	0.7684	0.7446
50 <sup>th</sup> quantile	2	7.2393	19.3179	0.8462	0.9536
	10	7.6596	15.6088	0.8611	0.9585
	40	12.7143	28.7083	0.9200	0.9913
% greater than 500	2	4.4583	31.5417		
	10	6.2292	28.6667		
	40	14.5833	36.0208		

#### 4.12.1 Results for Model 1, Different Variance of Random Effects

Two simulation studies using the sample set of 400 Monte Carlo samples and the same two sets of 200 bootstrap samples as described in the simulation set-up in the introduction to Section 4.5 are considered, for two different random effects variance parameters,  $\sigma_b^2 = 0.0064$  and  $\sigma_b^2 = 0.4900$ . For the first choice, 63% of the sample estimates  $\hat{\sigma}_b^2$  are equal to the lower bound value. For the second choice, no sample estimates  $\hat{\sigma}_b^2$  are equal to the lower bound value. Empirical coverages for the general bootstrap pivot-type CIs for the area means are presented in Table 4.31.

The 90% and 95% bootstrap CIs have larger coverage error when  $\sigma_b^2 = 0.0064$  than the CIs constructed when  $\sigma_b^2 = 0.25$ , but 99% bootstrap CIs have coverage closer to the nominal level when  $\sigma_b^2 = 0.0064$  than the coverage of the CIs constructed when  $\sigma_b^2 = 0.25$ . The bootstrap CIs constructed when  $\sigma_b^2 = 0.49$  have overcoverage of about 0.2% to 1.9%.

Table 4.31: Empirical Coverages for Level one Bootstrap CIs, Estimated Distribution of  $|T^*|$  ( $\mu_{xi}$  known, 400 MC Samples,  $B_1 = 400$ ,  $B_2 = 1$ )

	$n_i$	90%	95%	99%
$\sigma_b^2 = 0.0064$	2	89.0	93.8	98.5
	10	88.7	93.4	98.3
	40	91.0	95.2	98.9
$\sigma_b^2 = 0.49$	2	91.8	96.9	99.6
	10	91.1	96.6	99.6
	40	90.2	95.4	99.2

## 4.13 Appendix F. Plots for the Simulation Results for Model 2

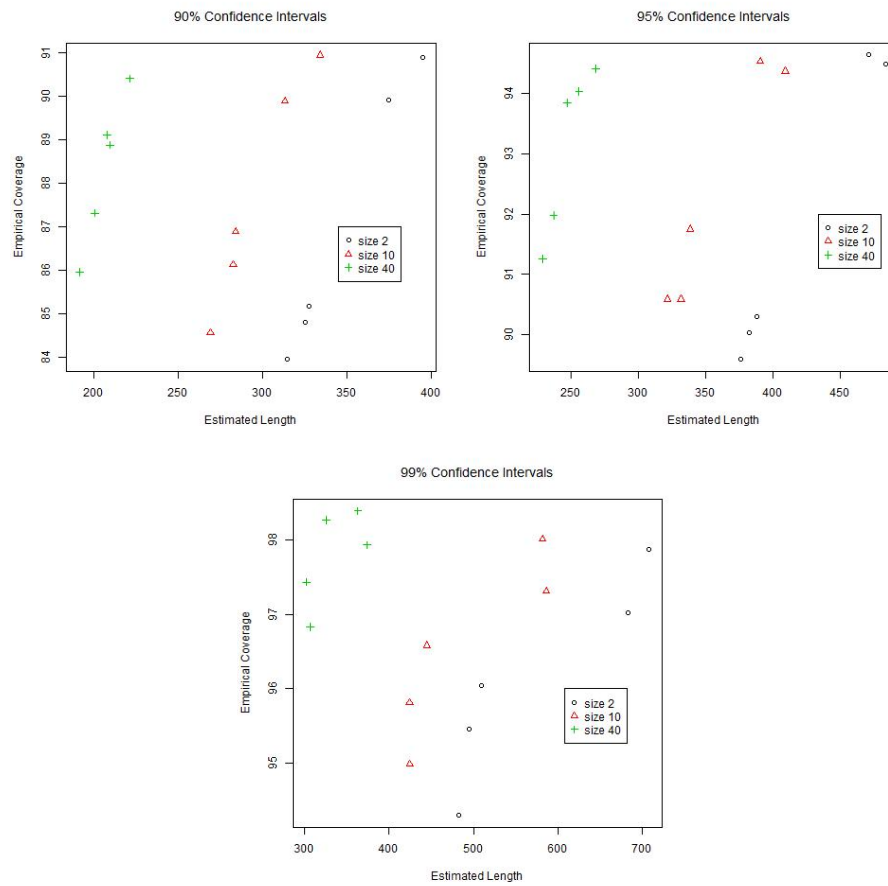


Figure 4.1: CIs for the Small Area Means: Empirical Coverage versus Estimated Length



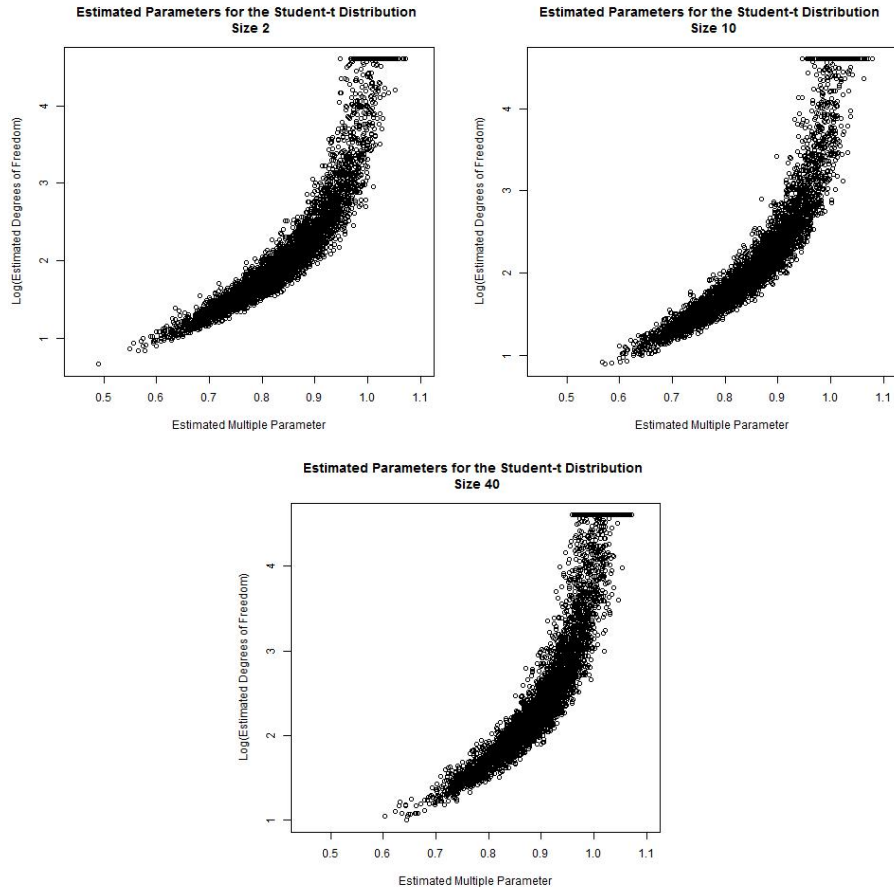


Figure 4.2: General CIs for the Small Area Means: Estimated Parameters for the Student-t Distribution

## Bibliography

- [1] Bates, D. (2008), "Fitting Mixed-Effects Models Using the lme4 Package in R," *International Meeting of the Psychometric Society*.
- [2] Chang, J. and Hall, P. (2014), "Double-bootstrap methods that use a single double-bootstrap simulation," *Biometrika*, submitted.
- [3] Chatterjee, S., Lahiri, P., Li, H. (2008), "Parametric Bootstrap Approximation to the Distribution of EBLUP and Related Prediction Intervals in Linear Mixed Models," *The Annals of Statistics*, 36, 3, 1221-1245.
- [4] Dass, S., Maiti, T., Ren, H., Sinha, S. (201), "Confidence interval estimation of small area parameters shrinking both means and variances," *Survey Methodology*, 38, 2, 173-187.
- [5] Datta, G.S. and Lahiri, P. (2000), "A unified measure of uncertainty of best linear unbiased predictors in small area estimation problems," *Statistica Sinica*, 10, 613-627.
- [6] Datta, G.S., Rao, J.N.K. and Smith, D. (2005) "On measuring the variability of small area estimators under a basic area level model. " *Biometrika*, 92, 183-196.
- [7] Datta, G.S., Rao, J.N.K. and Smith, D. (2012) "Amendments and Corrections: On measuring the variability of small area estimators under a basic area level model, " *Biometrika*, 99, 2, 509.
- [8] 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.

- [9] Davidson, R. and MacKinnon, J.G. (2007), "Improving the reliability of bootstrap tests with the fast double bootstrap," *Computational Statistics and Data Analysis*, 51, 3259-3281.
- [10] Diao, L., Smith, D., Datta, G.S., Maiti, T., Opsomer, J. (2014), "Accurate Confidence Interval Estimation of Small Area Parameters Under the Fay-Herriot Model," *Scandinavian Journal of Statistics*, 41, 497-515.
- [11] Erciulescu, A.L. and Fuller, W.A. (2013), "Small Area Prediction of the Mean of a Binomial Random Variable," *Survey Research Methods Section, JSM Proceedings*, 855-863.
- [12] Erciulescu, A.L. and Fuller, W.A. (2014), "Small Area Prediction under Alternative Model Specifications," *Statistics in Transition and Survey Methodology*. Manuscript submitted for publication.
- [13] Fay, R.E. and Herriot, R.A. (1979), "Estimates of income for small places: an application of James-Stein procedures to census data," *Journal of the American Statistical Association*, 74, 269-277.
- [14] Fuller, W.A. (1990), "Prediction of True Values for the Measurement Error Model," *Statistical Analysis of Measurement Error Models and Applications, Contemporary Mathematics*, 112, 41-57.
- [15] Fuller, W.A. and Harter, R.M. (1987), "The multivariate components of variance model for small area estimation," *Small Area Statistics: An International Symposium*, Platek, R., Rao, J.N.K., Sarndal, C.E. and Singh, M.P. (Eds.), John Wiley, New York, 103-123.
- [16] Ghosh, M. and Sinha, K. (2007), "Empirical Bayes estimation in finite population sampling under functional measurement error models," *Journal of Statistical Planning and Inference*, 137, 2759-2773.
- [17] Ghosh, M., Sinha, K. and Kim, D. (2006), "Empirical and Hierarchical Bayesian Estimation in Finite Population Sampling under Structural Measurement Error Models," *Scandinavian Journal of Statistics*, 33, 3, 591-608.

- [18] Giacomini, R., Politis, D. and White H. (2013), "A Warp-Speed Method for Conducting Monte Carlo Experiments Involving Bootstrap Estimators," *Econometric Theory*, 29, 567-589.
- [19] Hall P.(1986), "On the bootstrap and confidence intervals," *Annals of Statistics*, 14, 4, 1431-1452.
- [20] Hall P. and Maiti T. (2006), "On parametric bootstrap methods for small area prediction," *J.R. Statist. Soc. B*, 68, 2, 221-238.
- [21] Jiang, J. and Lahiri, P. (2001), "Empirical best prediction for small area inference with binary data," *Annals of the Institute of Statistical Mathematics*, 53, 2, 217-243..
- [22] Lahiri, P. and Rao, J. (1995), " Robust estimation of mean squared error of small area estimators," *Journal of the American Statistical Association*, 82, 758-766.
- [23] Liu, B. and Diallo, M. (2013), "Parametric Bootstrap Confidence Intervals for Survey-Weighted Small Area Proportions," *JSM 2013- SRMS Proceedings*, 109-121.
- [24] Pfeffermann, D. and Correa, S. (2012), "Empirical bootstrap bias correction and estimation of prediction mean square error in small area estimation." *Biometrika*, 99, 457-472.
- [25] Prasad, N.G.N. and Rao, J.N.K. (1990), "The estimation of mean squared error of small area estimators.," *Journal of the American Statistical Association*, 85, 163-171.
- [26] Rao J.N.K. (2003), "Small Area Estimation," *Wiley Series in Survey Methodology*.
- [27] Wang, J. and Fuller, W.A. (2003), "The mean squared error of small area predictors constructed with estimated area variances," *Journal of the American Statistical Association*, 98, 716-723.
- [28] Ybarra, L.M.R. and Lohr, S.L. (2008), " Small area estimation when auxiliary information is measured with error," *Biometrika*, 95, 919-931.

## CHAPTER 5. SUMMARY

This dissertation investigates unit level models when the covariate mean is measured with error. Different cases of auxiliary information are considered. Prediction methods for the small area mean, estimation of the prediction mean squared error (MSE) and confidence intervals (CIs) for the small area means are presented for the case when the response variable is binary. These results are important to agencies and policy makers interested in constructing reliable estimates for areas with small sample sizes.

The small area estimation literature has grown considerably in the past few decades. However, most investigations are for linear models with known covariate mean. Small area studies for cases when the covariate mean is measured with error are summarized in Chapter 1. In Chapter 2, two methods for constructing small area mean predictions are compared. The first method is based on the conditional distribution of the random area effects given the response variables. The second method, called the 'plug-in method' is based on the direct substitution of the predicted random area effects into the small area mean expression. The 'plug-in method' is easier to implement and has been used in many studies. The MMSE prediction involves numerical integration, but given the recent computational power this should not be difficult. In a simulation study, we show that the 'plug-in' predictor for the small area mean can have sizeable bias, while the bias in the MMSE prediction error was small, relative to the standard errors of the prediction for the small area mean.

In Chapter 3, we demonstrate the efficiency gains associated with the random specification for the auxiliary variable measured with error. The prediction mean squared error (MSE) is smaller when the area mean for the covariates is predicted based on a random mean model

for the covariates than when area mean for the covariates is estimated based on a fixed mean model for the covariates. The prediction MSE is smaller when additional auxiliary information is available and included in the estimation. It is shown that the effect of including auxiliary information, if available, in the estimation is smaller for the random mean model than for the fixed mean model for the covariates.

There is challenge in estimating the prediction MSE for the small area mean of a nonnormal response variable because there is no closed form expression for the prediction MSE. In Chapter 3 we propose a parametric fast double bootstrap procedure to estimate the prediction MSE for the small area mean of a binary response variable. Analysis of variance results are presented for the variance of the estimated prediction MSE, using a simulation study. The proposed procedure has smaller bootstrap error than a classical fast double bootstrap procedure with the same number of samples. In Chapter 4 we investigate a different bootstrap bias correction method that reduces the bias in the level one bootstrap MSE estimators, at the cost of increasing the variance.

The studies in the literature described in Chapter 1 on small area models for nonnormal response propose prediction MSE estimators for the area means, but do not investigate the constructions of CIs for the small area means. In Chapter 4 we present two sided confidence intervals (CIs) for the small area means of a binary response variable. The basic CIs are constructed using a normal approximation for the distribution of the parameter. It is shown that the basic CIs for the small area mean, constructed using different estimators for the prediction MSE, have empirical coverages lower than the desired nominal level. Pivot-type bootstrap CIs perform better than the percentile bootstrap CIs, with respect to the coverage errors. Double bootstrap CIs perform well, but do not improve the coverage accuracy compared to the level one bootstrap CIs. A method for constructing bootstrap CIs for a general level is proposed. The user is given a degrees of freedom for the Student-t distribution and a standard error of the small area mean prediction. This enables the user to construct CIs for the small area mean for any desired nominal level without having to repeat the computationally intensive bootstrap

procedure. The coverage of the general bootstrap CI is comparable to the coverage of the level specific bootstrap CI.