# Validity Testing for Coverage Properties of Small Area Models for Cell Proportions*

Aaron Gilary[1], Jerry Maples[1], Eric V. Slud[1,2]
[1]U.S. Census Bureau, CSRM, 4600 Silver Hill Road, Washington, DC 20233
[2]Mathematics Department, University of Maryland, College Park, MD 20742

**Abstract**

In Gilary, Maples, and Slud (2012), we compared three possible models for small proportions in small survey domains: Fay-Herriot (Fay and Herriot 1979), GLMM (Jiang and Lahiri 2006), and Beta-Binomial (Prentice 1986). The comparisons used bootstrap-based confidence intervals which were justified by asymptotic theory or established only for large samples. Here we build upon that work by conducting simulations of small area data in moderate-sample settings, for two purposes: to evaluate the performance of different analysis methods when using each of the three simulation models; and to assess the validity of predictors and Wald-type confidence interval coverage properties for each method.

**Key Words:** arcsine square root transformation, beta-binomial regression, confidence bound, parametric bootstrap, prediction interval, transformed Fay-Herriot model

*This report is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

# 1. Introduction

Gilary, Maples, and Slud (2012) addressed two similar problems relating to the estimation of proportions in small survey domains. The first one was to provide bounding intervals of small estimated proportions in American Community Survey (ACS) tabulations. The second was to construct a measure of uncertainty for county and place level estimates of Erroneous Enumeration (EE) rates among Housing Units in the 2010 Census Coverage Measurement (CCM) program.

These problems have a common, salient feature: they require interval estimates corresponding to survey point estimates with values near zero. This feature commands special treatment because the straightforward design-based variance estimators of small proportions yield interval estimators unrealistically close to zero as a function of the point estimate.

Gilary, Maples, and Slud (2012) discussed three general model-based small-area techniques for one-sided interval estimation of small survey proportions, and compared the resulting estimates to 'cell-based' or direct interval estimates that do not borrow strength across small domains or cells. The methods were applied to ACS 2009 data in Slud (2012) and to the EE rates for housing units in the 2010 CCM in Gilary, Maples, and Slud (2012).

This paper presents a medium-scale simulation study (which we disconnect from the applications) to measure the accuracy of predictions for each of the three models studied, and the benefits of each type in bracketing the unknown proportions for these areas. The performance is assessed by reducing the mean error and matching the empirical variance to the theoretical variance. The paper also examines the empirical coverage properties of Wald-type 90 percent confidence intervals using these simulations, and studies the sensitivity of the results to different choices of parameters.

## 1.1 Models and Background

The small area models used in this investigation were defined by Gilary, Maples, and Slud (2012). The details are restated below with some points condensed.

The motivating setting is a data structure in which $m$ small domains (called 'areas' here) $i = 1, \ldots, m$ are fitted with survey-weighted (ratio) estimators $\hat{\pi}_i \equiv \hat{Y}_i/\hat{N}_i$, where $\hat{Y}_i$ and $\hat{N}_i$ are the area $y$-indicator total and population-total estimates, respectively. The survey-estimated proportion $\hat{\pi}_i$ is defined equal to the further generalized expressions:

$$\hat{\pi}_i = \frac{\hat{Y}_i}{\hat{N}_i} = \frac{y_i}{n_i} \tag{1}$$

The denominator $n_i$ represents the number of people or units sampled in area $i$, which is defined by (1) and interpreted as the counts of $y$-indicators of 1 in the sample. In survey estimation contexts $n_i$ may be defined as an "effective" sample size, used as a measure of how many 'independent' units are represented in the estimate (e.g., Kish, 1965).

All of the models used involve a nonlinear transformation from input parameters to the probability scale.

#### *1.1.1 Small Area Models*

The three models explored here are the transformed Fay-Herriot, the GLMM, and the Beta-Binomial. All of these models contain a fixed-effect component $\eta_i$ (a linear combination $\mathbf{x}_i^{'}\beta$ ), an area-level random effect which is explicit $(u_i)$ in the Fay-Herriot and GLMM models and implicit in the Beta-Binomial, and a design-based error term which is explicit $(\epsilon_i)$ in the Hay-Herriot model and implicit in GLMM and Beta-Binomial through the binomial distribution. We jointly specify a form for both $\pi_i$ and the observation $y_i/n_i$. First, we consider a *transformed Fay-Herriot* model based on the classic model of Fay and Herriot (1979):

$$\mathbf{FHtr}: \begin{cases} \theta_i^{raw} = & \mathbf{x}_i^{'}\beta + u_i \\ \theta_i & = \max(0, \min(\theta_i^{raw}, \frac{\pi}{2})) \\ y_i/n_i & = \sin^2(\max(0, \min(\theta_i^{raw} + \epsilon_i), 1)) \\ \pi_i & = \sin^2 \theta_i \end{cases} \tag{2}$$

The target parameter $\theta_i$ is $\mathbf{x}_i^{'}\beta + u_i$ before transformation and possible truncation. Then the response fraction maps and truncates the target parameter $\pi_i$ on the probability scale. The point of the transformation within this model is to stabilize the variance, so that for binomial or simple random sampled data, the variance of this transformation of the sample mean does not depend on the underlying rate $\pi_i$.

Note that because values outside the range $(0,\ \pi/2)$ are possible with the raw target $\theta_i^{raw}$, those values are truncated at 0 and $\pi/2$. We refer to this as truncation although in other contexts it might be called censoring (see Slud and Maiti, 2011, for the distinction). This process yields a mass of values at the zero bound, which can reflect the motivating data examples if there is a sizeable number of small areas with observed zeroes. The resulting model is not precisely a Fay-Herriot model.

The other small-area models both treat $y_i$ as binomial with $n_i$ trials and success-probability $\pi_i$, with the mean of the latter expressed in terms of the logistic distribution function $L(x) = e^x/(1+e^x)$. One of these models is the *random-intercept logistic*, discussed by Jiang and Lahiri (2006):

$$\mathbf{GLMM}: \begin{cases} y_i \sim \text{Binom}(n_i, \pi_i) \\ \pi_i = L(\mathbf{x}_i^{'}\beta + v_i)\ , \quad v_i \sim \mathbf{N}(0, \sigma_v^2) \end{cases} \tag{3}$$

We generate raw targets on the logit scale that are neither transformed nor truncated, and response fractions on the probability scale. Unknown parameters $\beta$ and $\sigma_v^2$ are computed through maximum likelihood estimation.

Our final model is the *Beta-Binomial* employing a logit link specifically studied in Prentice (1986). The Beta-Binomial can be used to explain extra variation beyond what is modeled directly by the linear fixed effects $\mathbf{x}_i^{'}\beta$:

$$\mathbf{BBIN}: \begin{cases} y_i \sim \text{Binom}(n_i, \pi_i) \\ \pi_i \sim \text{Beta}\Big(L(\mathbf{x}_i^{'}\beta)\cdot\tau,\ (1 - L(\mathbf{x}_i^{'}\beta))\cdot\tau\Big) \end{cases} \tag{4}$$

where $\tau$ is used as a precision parameter.

The Beta-Binomial distribution does not have an explicit random effect, but there is an implicit effect upon drawing $\pi_i$ from the Beta distribution. The target for the Beta-Binomial is $\pi_i$ but there is no explicit expression of this target as in the other models.

Within all three small area models, the regression coefficients $\beta$ are unknown and are estimated by maximum likelihood simultaneously with respective random-area-effect dispersion parameters $\sigma_u^2$, $\sigma_v^2$, or $\tau$.

The first goal for this investigation is to study the estimates under each type of generation and analysis model, and in particular how they would be affected by zeroes arising from truncation in the Fay-Herriot model. Another is to validate confidence interval properties under large-sample normal distribution theory for parameter estimators. Medium-scale simulations will study the application of the normal-theory based confidence intervals that underlie our bounds from previous work. Do these estimated intervals have suitable coverage and widths?

### 1.2 Small Area Point Estimators

The point estimates of small-area proportions can be expressed through their empirical best predictor (EB) for each of the three models. For the Fay-Herriot model, the EB is transformed. These EBs $\widehat{\pi}_i^{\rm BP}$ are as follows (Rao 2003; Jiang and Lahiri 2006; Prentice 1986):

$$\textbf{FHtr:} \quad \arcsin(\sqrt{\widehat{\pi}_i^{\rm BP}}) \;=\; \mathbf{x}_i'\hat{\beta} + \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + (4n_i)^{-1}} \Big( \arcsin(\sqrt{\frac{y_i}{n_i}}) - \mathbf{x}_i'\hat{\beta} \Big) \tag{5}$$

$$\textbf{GLMM:} \qquad \widehat{\pi}_i^{\rm BP} \;=\; g(y_i+1,\, n_i+1,\, \mathbf{x}_i'\hat{\beta},\, \hat{\sigma}_v^2)/g(y_i,\, n_i,\, \mathbf{x}_i'\hat{\beta},\, \hat{\sigma}_v^2) \tag{6}$$

where

$$g(k,n,\eta,\sigma^2) \;\equiv\; \int \frac{e^{(\eta+\sigma z)k}}{(1+e^{\eta+\sigma z})^n}\,\phi(z)dz \quad , \quad \phi(\cdot) \sim \mathcal{N}(0,1)$$

$$\phi(z) = \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}$$

$$\textbf{BBIN:} \qquad \widehat{\pi}_i^{\rm BP} \;=\; \{y_i + \hat{\tau}\, L(\mathbf{x}_i'\hat{\beta})\} \,/\, \{n_i + \hat{\tau}\} \tag{7}$$

where recall that $L(\cdot)$ denotes the logistic distribution function.

Now we examine the EB variance estimator $V$ for the difference between the predictor and the target. For the Fay-Herriot method, $V$ is given through the 'higher-order correct' sum of terms $g_1, g_2, g_3$ (see Rao, 2003, 6.2.2). For GLMM, the estimator is:

$$V \;=\; g(2,2,\eta_i,\sigma_v^2) - \sum_{k=0}^{n_i} \binom{n_i}{k} \frac{\{g(k+1,n_i+1,\eta_i,\sigma_v^2)\}^2}{g(k,n_i,\eta_i,\sigma_v^2)} \tag{8}$$

and for Beta-Binomial:

$$V \;=\; \frac{\tau\, e^{x_i'\beta}}{(\tau+1)(1+e^{x_i'\beta})^2(n_i+\tau)}. \tag{9}$$

The EB is used to construct naïve, first principles confidence intervals bounded by $\hat{\pi}_i \pm 1.645 * Var(\hat{\pi}_i)$. Ultimately, we will use bootstrap techniques as a means of bracketing our confidence intervals, as asserted in Gilary, Maples, and Slud (2012), but we evaluate the naïve intervals here to examine their effectiveness in practice.

## 2. Simulation Studies

The simulations involved generating estimates using each of the models, and applying each model for analysis to the same data set. The two parameters of interest here are the small-area targets $\pi_i$ and responses $y_i$.

In order to generate the models we need input parameters. The sampling variances for these models borrow the pattern used by Datta, Rao, and Smith (2005). Their pattern used groups of five area error variances $s_i$ which are set equal to $\frac{1}{4n_i}$. Each simulation of $m$ small areas will repeat each variance in the group $m/5$ times. Our simulations primarily used $s_i$ values defined as multiples (not depending on $i$) of the variance sets (.3, .4, .5, .6, .7).

The parameters tested for the Fay-Herriot model were designed to produce probability scale proportions $\pi_i$ ranging between 0 and .3. Then we selected GLMM and Beta-Binomial parameters that would produce similar proportions once *they* are transformed to the probability scale. A series of different arrangements were tested to find a suitable set of parameters. The chosen parameters were as follows:

| | | | | |
|---|---|---|---|---|
| **FHtr:** | $\mu$=.25, | $\sigma^2$=.1, | $n_i = 40$, | (on arcsine scale) |
| **GLMM:** | $\mu$=-2, | $\sigma^2$=.5, | $n_i = 40$, | (on logit scale) |
| **BBIN:** | $\mu$=-1.8, | $\tau$=5, | $n_i = 40$, | (on logit scale). |

Other input parameter choices produced similar simulation results. We intend to expand the simulation to allow nonconstant regression covariates as the $\overline{x}_i$s, but have not done so yet. We input a constant mean term as the m-vector $\mathbf{1}$ of 1s.

The parameters $\mu$ and $\sigma_u^2$ were chosen under the general guideline that they produce 10 to 40 percent zeroes. The corresponding means $\pi_i$ needed to be sufficiently different from 0 that $y_i$ would only seldom fall outside the interval (0,1) on the probability scale. The simulations began with 15 small areas and 1,000 replications.

### 2.1 Addressing Values $y_i = 0$ in the Fay-Herriot Model

The first issue to examine is truncation in the Fay-Herriot model. This investigation began with studying the normality of the difference between target and predictor, which exposed a concentration of mass near zero for the Fay-Herriot point predictor. Out of the small-proportion transformed Fay-Herriot cases where $\theta_i^{raw}$ is less than 0 and must be truncated at 0, most are associated with $\theta_i = 0$. See Figures 1 and 2 which show the distribution of differences between predictor $\pi_i$ and target $\theta_i$ both before and after removing the zero targets, for a series of 1,000 simulations in one small area. The density curve in Figure 2 is the same as that in Figure 1, while the total area in the histogram bars in Figure 2 is reduced due to the removal of cases with $\theta_i = 0$. Overall, 85 percent of the $\theta_i^{raw} \leq 0$ had $\theta_i = 0$.

Figures 3, 4, and 5 show the target $\theta_i$ (on the x-axis, on arcsine-square root scale) plotted against the predictor $\hat{\pi}_i^{BP}$ (on the y-axis) when generation and analysis models are of the same type, for Fay-Herriot, GLMM, and Beta-Binomial respectively. The tail of zeroes along the x-axis on Figure 3 shows the truncated cases where the target equals zero. These plots indicate that the predictors generally estimate the target well, with little bias.

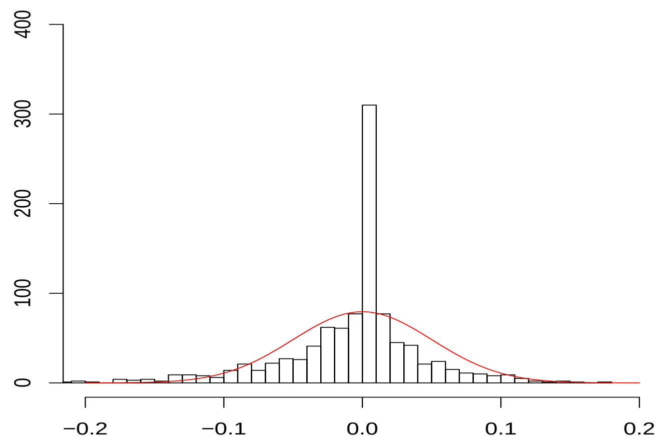


**Figure 1**: Histogram of $\theta_i$ Minus $\pi_i$ for Fay-Herriot Model, as Expressed in (2), Before Removal of Cases with $\theta_i = 0$. The Superimposed Line Represents the Normal Curve with Zero Mean and Empirical Variance.

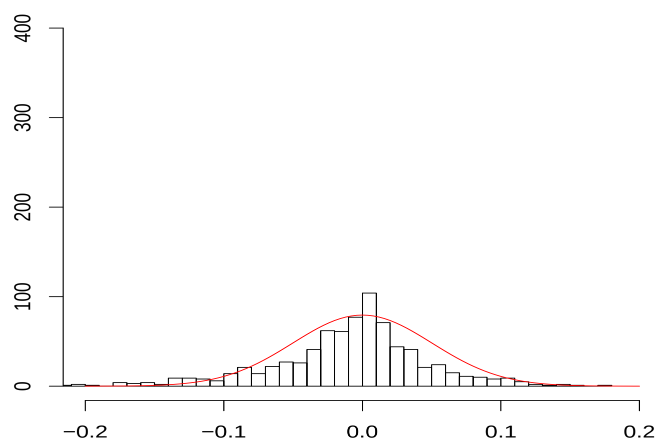


**Figure 2**: Histogram of $\theta_i$ Minus $\pi_i$ for Fay-Herriot Model, as Expressed in (2), After Removal of Cases with $\theta_i = 0$. The Superimposed Line Represents the Normal Curve with Zero Mean and Empirical Variance.

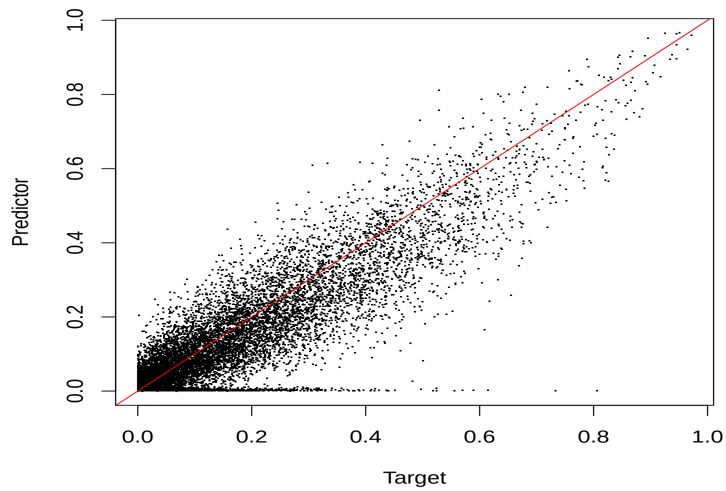


**Figure 3**: Predictor $\hat{\pi}_i^{BP}$ Versus Target $\pi_i$, for the Truncated Fay-Herriot Model on Probability Scale. For 1,000 Simulations Across 15 Small Areas.

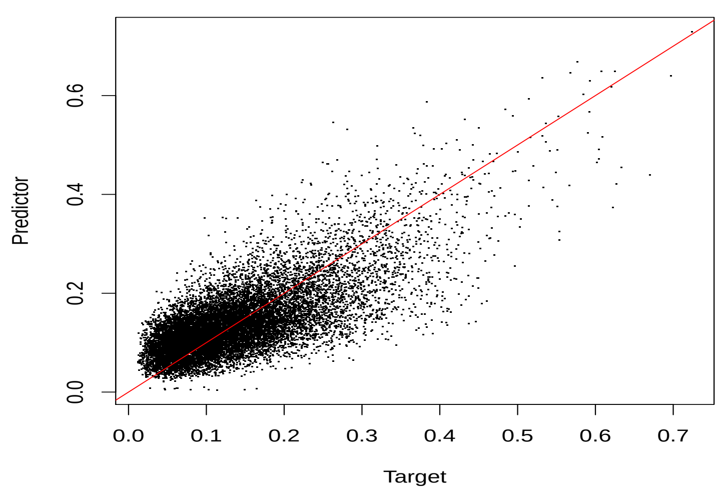


**Figure 4**: Predictor $\hat{\pi}_i^{BP}$ Versus Target $\pi_i$, for the GLMM Model on Probability Scale. For 1,000 Simulations Across 15 Small Areas.

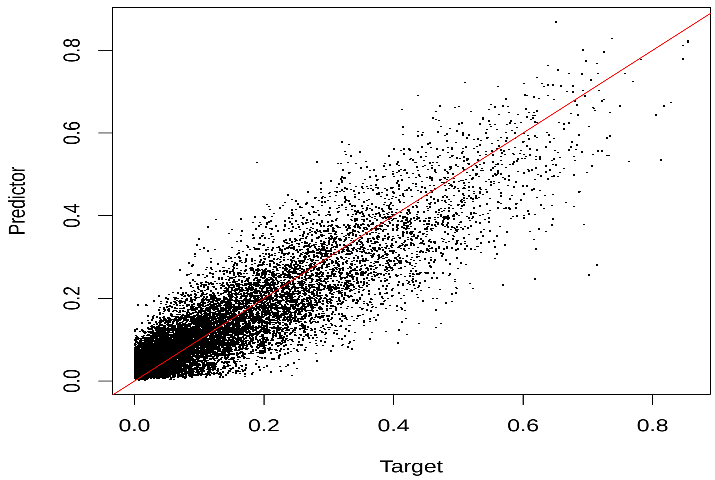


**Figure 5**: Predictor $\hat{\pi}_i^{BP}$ Versus Target $\pi_i$, for the Beta-Binomial Model on Probability Scale. For 1,000 Simulations Across 15 Small Areas.

## 2.2 Estimation Performance

To study estimation performance we set up a 3x3 cross-design for first generating the data under each of the three models, and then analyzing the generated sets of targets using each of the models for analysis. All of these tables show results from a single simulation of 1,000 values in each of 15 small areas. The simulation results are consistent to within three decimal places over independent, repeated runs.

Table 1 shows estimated probability-scale parameters for empirical average bias and squared error under truncated Fay-Herriot generation, under each of three types of analysis models. This table and the following ones can help to compare bias and variance across models and also look at overall performance of the simulation. The empirical average bias shows the difference between target and predictor, while the empirical and theoretical $\sigma^2$ columns show how the observed mean squared error from the simulations compares to the theoretical mean squared error $(V)$, as defined in (8) and (9) using the true values. In this case, the average bias is small for each method, but is slightly smaller for the GLMM and Beta-Binomial analysis methods than for the Fay-Herriot. This speaks to the effect of truncation for the Fay-Herriot model and to the strong estimation properties of the other two.

The empirical variance is only slightly above the theoretical variance for all methods, which is a promising result. Because the theoretical variance of the difference was calculated on the arcsine scale, there is no comparable theoretical variance here. On the arcsine scale, the Fay-Herriot empirical variance is slightly above the theoretical variance.

Table 2 shows the estimated parameters for GLMM generation under the same analysis models. This table shows that GLMM and Beta-Binomial have much smaller average biases than the Fay-Herriot, and once again all methods have comparable (and generally strong) empirical variances.

Table 3 shows the estimated parameters for Beta-Binomial generation under the same analysis models. These results are similar to the GLMM results, and show

**Table 1**: Probability-Scale Estimation Parameters for Bias and Squared Error, for each Analysis Model, Under Fay-Herriot Generation

| | | F-H Generation Model | | |
|---|---|---|---|---|
| | | Emp. Avg. Bias | Emp. MSE | Theor. MSE |
| Analysis Model | F-H | -.0087 | .0633 | – |
| | GLMM | .0062 | .0634 | .0590 |
| | BBIN | .0062 | .0628 | .0599 |

**Table 2**: Probability-Scale Estimation Parameters for Bias and Squared Error, for each Analysis Model, Under GLMM Generation

| | | GLMM Generation Model | | |
|---|---|---|---|---|
| | | Emp. Avg. Bias | Emp. MSE | Theor. MSE |
| Analysis Model | F-H | -.0138 | .0627 | – |
| | GLMM | .0005 | .0603 | .0499 |
| | BBIN | .0005 | .0606 | .0492 |

the benefits of coverage under GLMM and Beta-Binomial analysis.

When the number of small areas was expanded to 75, the results were generally the same. The empirical variance shrunk slightly, approaching the theoretical variance.

### 2.3 Coverage Performance of Bounds

Finally, we wanted to explore the general coverage properties of naïve, normal-theory confidence intervals. Because these sample sizes are moderate, the central limit theorem will not mandate normality.

Table 4 shows the percentage of targets covered by naïve 90 percent confidence intervals when they are implemented under the cross design. The results range from overcoverage (with the Fay-Herriot) to undercoverage (with GLMM and Beta-Binomial). The results are affected both by truncation in the Fay-Herriot and non-normality in the other two distributions. Overall, these naïve intervals do not perform as well for the GLMM or Beta-Binomial as for the more conservative Fay-Herriot model.

There is no disadvantage found in the overall coverage proportion from the analysis model being of a different type. The table indicates that the nominal coverage rates are even better in such an instance. (Because of zero-truncation, the generation and analysis models are never technically the same, even when they are of the same type.)

**Table 3**: Probability-Scale Estimation Parameters for Bias and Squared Error, for each Analysis Model, Under Beta-Binomial Generation

| | | BBIN Generation Model | | |
|---|---|---|---|---|
| | | Emp. Avg. Bias | Emp. MSE | Theor. MSE |
| Analysis Model | F-H | -.0150 | .0672 | – |
| | GLMM | .0005 | .0660 | .0602 |
| | BBIN | -.0004 | .0656 | .0606 |

**Table 4**: Confidence Interval Coverage Under Different Generation and Analysis Model Combinations

| Analysis Method | FH-gen | GLMM-gen | BBIN-gen |
|---|---|---|---|
| Fay-Herriot | .919 | .867 | .880 |
| GLMM | .884 | .778 | .873 |
| Beta-Binomial | .885 | .773 | .874 |

## 3. Conclusion

### 3.1 Summary

GLMM and Beta-Binomial analysis results track very closely. They are similar models handled in a similar style for this project, but the average biases under those analysis models are less than .01, and the difference between their empirical and theoretical variances is similarly small. These methods are robust to the choice of generation model.

The inadequacy of normal-theory confidence intervals can be seen here. In some results they cover close to 90 percent of estimates, but on the whole coverage is inadequate especially for matching model types. Nor did testing different input parameters yield the specified coverage level. There is enough deviation from normal theory in the implementation of these complex models that one should be skeptical of normality-based confidence intervals in similar contexts.

Matching model type for generation and analysis makes less difference than anticipated. In some cases, unmatched model types provided smaller average biases and variances, or better overall coverage rates, as seen in the coverage shown in Table 4.

### 3.2 Priorities for Future Research

We plan to increase the number of small areas, and complexity of small-area models. We have begun this effort by expanding to 75 small areas, but there is more to do. We would also like to base these models on estimated regression coefficients, and incorporate effective sample sizes as well.

To supplement our work here, we are working on bootstrap results which are still in progress. We plan to examine bootstrap-based confidence intervals from the GLMM and Beta-Binomial models. We will extend the work on Fay-Herriot bootstrap methods covered by Chatterjee, Lahiri, and Li (2008). We will test to see if parametric bootstraps can yield coverage properties accurate to within a small interval given as a function of the number of small areas $m$. Once we generate bootstrap-interval simulations, we will be able to assess the merits of different modes of analysis through their confidence intervals.

## References

Chatterjee, S., Lahiri, P. and Li, H. (2008), *Parametric Bootstrap Approximation to the Distribution of EBLUP and Related Prediction Intervals in Small-area Estimation.* Annals of Statistics, 36, 1221-1245.

Datta, G., Rao, J., and Smith, D. (2005), *On Measuring the Variability of Small Area Estimators Under a Basic Area Level Model.* Biometrika, 92, 183-196.

Fay, R. and Herriot, R. (1979), *Estimation of Income from Small Places: an Application of James-Stein Procedures to Census Data.* Journal of the American Statistical Association, 74, 269-277.

Gilary, A., Maples, J. and Slud, E. (2012), *Small Area Confidence Bounds on Small Cell Proportions in Survey Populations.* Proceedings of the American Statistical Association, Survey Research Methods Section, 3541-3555.

Jiang, J. and Lahiri, P. (2006), *Mixed Model Prediction and Small-area Estimation (with Discussion).* Test 15, 1-96.

Kish, L. (1965), **Survey Sampling.** Wiley and Sons.

Prentice, R. (1986), *Binary Regression using an Extended Beta-binomial Distribution, with Discussion of Correlation Induced by Covariate Measurement Errors.* Journal of the American Statistical Association, 81, 321-327.

R Development Core Team (2011), *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Rao, J.N.K. (2003), **Small Area Estimation.** Wiley.

Slud, E. (2012), *Assessment of Zeroes in Survey-Estimated Tables via Small-Area Confidence Bounds.* Journal of the Indian Society of Agricultural Statistics, 66, 157-169.

Slud, E., and Maiti, T. (2011), *Small Area Estimation Based on Survey Data from a Left-Censored Fay-Herriot Model.* Journal of Statistical Planning and Inference, 141(11), 3520-3535.