Estimation of Poverty at the School District Level Using Hierarchical Bayes Modeling*

Sam Hawala† Partha Lahiri‡

Abstract
Direct estimates of the poverty level for more than 25% of the school-districts in the U.S., are generally not available from the American Community Survey (ACS) data that are compiled annually. The Small Area Income and Poverty Estimates (SAIPE) program has depended on coalescing information from the decennial census, the Internal Revenue Service (IRS), from linking IRS records to Census geography, and from poverty estimates at the county level, to produce poverty estimates for every school district. Income data from the decennial census in particular played an important role, but are no longer available starting from the 2010 census. This research is part of the Census Bureau's effort to consider alternative estimation procedures. We consider Hierarchical Bayes models that combine information from the ACS, and from the IRS, to obtain annual estimates of most current and most relevant poverty rates.

Key Words: Small Area Estimation, SAIPE, MCMC, Hierarchical Bayes

1. Introduction

In this paper, we explore the estimation of school district poverty rates for related school-age children in families. Poverty estimates at the school district level are part of the Census Bureau data products that are ultimately used to allocate federal funds under Title I of the 'No-Child-Left-Behind Act of 2001' for education programs to aid disadvantaged children.

The Small Area Income and Poverty Estimates (SAIPE) program at the U.S. Census Bureau strives to deliver estimates that are more precise and more reliable than estimates from single-year American Community Survey (ACS) direct estimates. The aim for single year ACS estimates is to provide reliable data for areas with population size of 65,000 or more. A substantial number of small school districts have very few samples or even no sample since the total sample size is fixed by the survey budget. Figure 1 shows the number of the 2009 ACS sample cases in school districts among those with sample sizes less than 30. For example, 1,861 school districts (or 13% of the total 13,204 school districts) each have less than 5 children in the ACS sample.

Although the five-year ACS data are designed to produce estimates at all levels of geography including school districts under 65,000 total population, these 5-year estimates target an average over a 5-year period in the population, not an annual most current, most relevant measure of poverty.

The SAIPE program has used a synthetic technique called the "minimum change" method to create estimates of the number of related school-age (5-17) children in poverty for school districts. This minimum change method combines IRS tax data,

*This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed on statistical, methodological, technical, or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.

†U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233

‡JPSM, University of Maryland, and U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233

Census 2000 data, and the official county poverty estimates to obtain estimates of school district poverty shares. Poverty shares are the proportions of the numbers of poor school-age children in each school district that was wholly or partially contained in that county.

There are two concerns with the current minimum change method. First, it relies on the decennial census that, starting in 2010, no longer includes long form data collection on income and detailed population and housing characteristics. The ACS for a far smaller sample than the decennial census long form is now the main source for this kind of information on an annual basis. Secondly, an appreciable number of tax returns cannot be assigned to the school districts in which the addresses, appearing on the tax returns, are located. To this degree we are unable to geocode many IRS tax records and therefore unable to use them for prediction. The county geocoding rate however, could be useful.

The SAIPE program is researching new ways to estimate poverty at the school district level. We began our research evaluating a reweighting method of Schirm and Zaslavsky (1997) whereby we distribute the weight for a person among other school districts based on the similarity of the sample person with those in the other school districts. Schirm and Zaslavsky have shown that a redistribution of weights could substantially improve the precision by allowing the use of many more observations in the estimation of a school district. However, our reweighting algorithm converged only for a subset of school districts. For many of the others, we were unable to find intial values that would lead to the successful conclusion of our two-step optimization procedure. Therefore we put the reweighting approach aside, for now, and began working on this Hierarchical Bayesian estimation effort, which became the focus of our paper.

The rest of the paper is organized as follows. Sections 1.1 and 1.2 further discuss the current methodology and issues therein. Section 2 presents the models tested. Section 3 details the Hierarchical Bayesian aspect of the estimation. In section 4 we discuss the results, and in section 5 we outline some future research.

1.1 The Geocoding Problem

There are difficulties in determining the school district to which some individual tax returns should be assigned. Most addresses are "city-style", which usually means there is a house number and street name; these addresses can nearly always be designated to be in a particular census block, and therefore a particular school district. The inability to designate which school district an address is in, i.e. low geocoding rate, is usually due either to "rural route"-style addresses and P.O. boxes (where the location of the addressee's residence is not known precisely enough to assign it to a block), or to new housing/developments which have not made it into the Census Bureau's map databases yet. Usually information that is coded to the block level is combined for all the blocks within a county or school district.

This geocoding problem takes some predictive power away from the IRS data. However, the geocoding rates, which are at the county level, may, in and of themselves contribute some information on the poverty level of the county as a whole and of the school districts within. In fact the degree to which the county is rural or urban may portend it's poverty level.

1.2 The Minimum Change Method

The current SAIPE estimation method for school districts relies on the "Minimum Change" algorithm, see Maples and Bell (2007). A rough description of this algorithm is that it allocates the exemptions from each county's non-geocoded tax returns to the school districts in that county, in such a way as to minimize the change in the shares from the previous census. It sets the proportion of poor school-age related children in each school district as (i) the tax share if those tax shares are greater than Census 2000 shares, or (ii) the tax share plus a portion of the non-geocoded tax records: the non-geocoded tax records are distributed according to the distribution of the Census 2000 shares that remain that are greater than the corresponding tax shares. The algorithm also insures that the assignments are done in such a way that the shares add up to 100%.

2. Model

Our proposed models relate the single-year ACS school district direct estimates, to the school district IRS tax data and the county geocoding rate. The total IRS tax exemptions provide a variable related to population, while the poor exemptions, which are the exemptions on returns with adjusted gross income below the poverty threshold, provide a variable related to poverty. The county non-geocoding rate is the proportion of the county tax records, which cannot be assigned to their respective school districts.

Let θ_i denote the true poverty rate in school district i and let y_i denote the observed estimated poverty rate in school district i. $y_i =$ (ACS estimated number in poverty)/(ACS estimated number in poverty universe). For 2009, we have ACS data for 13,204 school districts.

We attempted two different sampling models that relate the true poverty rate θ_i to covariates just mentioned: the school district tax poverty rate and the county non-geocoding rate.

Let $\boldsymbol{\beta}$ denote a vector of three regression coefficients, including an intercept, so that $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$, $\mathbf{x}_i = (1, x_{i1}, x_{i2})$ are the covariates.

The linking models, that relate the true poverty rate to the covariates, have a parametric structure specifying a linear relationship between θ, x, and β. In one of the models we consider a logit link, in the other we consider the identity link after an arcsin transformation on the observed rate. This transformation has the effect of stabilizing the variance.

2.1 Sampling Model I

We model the poverty rate y_i via the normal distribution, with mean θ_i and variance $\frac{\theta_i(1-\theta_i)}{\widetilde{n}_i}$

$$p(y_i|\theta_i) \quad \sim \quad \mathcal{N}\left(\theta_i, \quad \frac{\theta_i(1-\theta_i)}{\widetilde{n}_i}\right) \tag{1}$$

$\widetilde{n}_i$ is the effective sample size: $\widetilde{n}_i = n_i/\mathrm{deff_i}$, where n_i is the ACS sample size for school district i. The model assumes $\widetilde{n}_i$ known.

The design effect $\mathrm{deff_i}$ reflects the effect of the complex sample design (Kish, 1965). For now, we approximate deff_i by a design effect for a stratified simple random sample with a negligible sampling fraction in all strata. The deff for a school district is the deff for the county that contains it. Let c_i be the county that contains school district i.

$$\text{deff}_i = n_{c_i} \sum_h \tilde{W}^2_{c_i h} / n_{c_i h}$$

$\tilde{W}_{c_i h} = \tilde{N}_{c_i h} / \tilde{N}_{c_i}, \quad \tilde{N}_{c_i h} = \sum_{j \in h} w_{ij}, \quad \tilde{N}_{c_i} = \sum_{h \in c_i} \tilde{N}_{c_i h}, \quad n_{c_i} = \sum_h n_{c_i h}.$
w_{ij} is the weight for child j in household h, which is in county c_i. $n_{c_i h}$ is the number of children in household h in county c_i.

The $\tilde{W}_{c_i h}, \quad \tilde{N}_{c_i h}, \quad$ and $\quad \tilde{N}_{c_i}$ are sample estimates of the corresponding population quantities. The approach using the population quantities was discussed in Liu & al. (2007). Ongoing work on design effects, (Lahiri (2010) unpublished report,) which uses the results discussed in Hawala and Lahiri (2010), will provide better estimates for the design effect.

Let

- model I(a) - linking model

$$\text{logit}\,(\theta_i) \quad \sim \quad \mathcal{N}\left(\mathbf{x}_i'\boldsymbol{\beta}, \quad \tau^2\right) \tag{2}$$

- model I(b) - linking model

$$\begin{aligned} \text{logit}\,(\theta_i) \quad &= \quad \mathbf{x}_i'\boldsymbol{\beta} \quad + u \\ u \quad &\sim \quad t(0, \tau, \nu) \end{aligned} \tag{3}$$

  where $t(0, \tau, \nu)$ is the t-distribution with ν degrees of freedom, and precision parameter τ

2.2 Sampling Model II

We obtain alternative models by making a variance stabilizing transformation, as suggested by Carter and Rolph (1974), and initially studied by Box and Cox (1964).

We set $z_i = arcsine\left(\sqrt{y_i}\right)$, and

- model II(a) - Normal sampling model

$$p(z_i|\theta_i) \quad \sim \quad \mathcal{N}\left(\pi_i, \quad \frac{1}{4\tilde{n}_i}\right) \tag{4}$$

- model II(b) - t-distribution sampling model

$$p(z_i|\theta_i) \quad \sim \quad t\left(\pi_i, \xi_i, \nu_z\right) \tag{5}$$

We set the precision parameter $\xi_i = \frac{4\tilde{n}_i \nu_z}{\nu_z - 2}$, to keep the same variance $\frac{1}{4\tilde{n}_i}$ as in model II(a). In both cases II(a) and II(b) we use the following linking model

$$\pi_i \quad \sim \quad \mathcal{N}\left(\mathbf{x}_i'\boldsymbol{\beta}, \quad \tau^2\right) \tag{6}$$

In both models II(a) and II(b), the true rates we're estimating are $\theta_i = \sin^2(\pi_i)$. For all four models we call the parameters $\boldsymbol{\beta}, \quad \tau$, and ν the structural parameters.

3. Hierarchical Bayes Analysis

In this section, we apply Hierarchical Bayes (HB) analysis to the models introduced in Section 2. Estimates of the posterior mean and variance of parameters are obtained from Markov Chain Monte Carlo (MCMC) simulation.

3.1 Prior Distributions

In a hierarchical Bayesian framework, we assume independent diffuse prior distributions for the hyperparameters. Let $\boldsymbol{\beta}$ have a (locally) uniform distribution with $p(\boldsymbol{\beta}) \propto 1$. Independently, for models I(a), II(a), and II(b) we set $\tau^2 \sim Unif(a_\tau, b_\tau)$, whereas for model I(b) wet set $\frac{1}{\tau^2} \sim Unif(a_\tau, b_\tau)$. Here $Unif$ denotes the uniform distribution and a_τ, b_τ are known positive constants. The constants are set to be large to reflect vague knowledge about the parameters.

The degrees of freedom ν for the t-distribution was treated as an unknown random variable in the model. We assume a uniform prior over the following degrees of freedom 2, 4, 6, 8, 10, 12, 15, 20, 30, and 50. The value 2 was not used for the prior on ν_z in model II(b).

3.2 Posterior Estimates

The posterior distribution of the poverty rate and the variance parameter τ can be approximated by MCMC simulations, which we ran using a combination of BUGS and R in our application. We produced 3 parallel Markov chains.

Estimates of the posterior means, and standard deviations of the structural parameters: the regression coefficients $\boldsymbol{\beta}$, the variance parameter τ, and the degrees of freedom ν are given in the Appendix. See Tables 2 and 3.

4. Discussion

In the context of estimating poverty rates in U.S. school districts, one of the major problems with direct survey estimates is the prevalence of districts with very small sample sizes (less than 30), see figure 1. We are proposing hierarchical Bayesian models to produce more reliable estimators than the direct survey (ratio) estimator.

We explored a few models in succession, adjusting the assumptions along the way as some computation results suggested. All the figures and tables are placed in the appendix.

Results for the deviance information criterion (DIC) (Spiegelhalter et al., 2002), which is often used for comparing non-nested models, are given in Table 1. Model II(a), with the smallest DIC, seems to perfom best among the 4 models considered. Model II(a) may not be the true model but it is closer to it than the other three models. These results show that the arcsine transformatioin with a linear link function provides a better fit than no transformation with a logit link.

In table 2 and table 3, we provide the posterior means of the structural parameters along with their posterior standard deviations.

The 95% credible intervals for the coefficient of the non-geocoding rate b_3 are: for model I(a) $(-0.5426, -0.0338)$, for model I(b) $(0.0466, 0.6664)$, for model II(a) $(-0.2379, 0.0341)$, and for model II(b) $(-0.2628, 0.0084)$. We notice that, in all four models, b_3 is small. It is not significantly different from zero for models II(a), and II(b) indicating that the non-geocoding rate could be taken out without much loss to the accuracy of the estimated poverty rates.

Figure 5 shows how Model II(a) resulted in estimates that are close to the direct estimates, specifically for large areas, where we tend to believe the direct survey estimates. The use of the t distribution in level 2 of model I(b) (figure 4), replacing the normal distribution of model I(a), corrected the bias in the average posterior mean. If we average out the posterior means across all the small areas and subtract the average of the direct estimates, we get the bias for model I(a): $0.4247 - .174 = .2507$, vs for model I(b) $.2077 - .174 = .0337$. However, both models I(a) and I(b) show a very poor fit to the direct estimates and the posterior variance is very small, in the order of 10^{-5}. In both figure 3 and figure 4 the shaded histogram is tightly clustered around the mean.

We wish to further confirm that the combination of the variance stabilizing arcsine transformation performed on the direct survey estimates of the poverty rates, and the use of Hierarchical Bayes estimator outperforms other model choices, even in the absence of covariates with strong predictive power. Figure 2 shows weak correlations between the direct estimate and the covariates. Moreover, we wish to motivate the use of the direct estimates from the ACS-5-year data, into our model checking strategy, as well as a predictor in our models.

5. Future Research

We are pursuing several research tracks. Direct estimates of sampling variances are unreliable for many of the small areas. A Generalized Variance Function (GVF) method was proposed in Hawala-Lahiri (2010) for states. We will be extending this work to the counties. Another track we are pursuing is estimating a Bayesian beta regression model that specifically models the probability of observing a rate of zero or a one.

As more data become available, namely, five consecutive years of ACS data and SAIPE program input data (including IRS data), we will consider a hierarchical time series model that combines all the available information. This method will borrow strength from prior years of ACS, other small areas, and covariates.

5.1 GVF model for the sampling variance

One problem with the sampling model we used is that $\widehat{\theta_i}$ is very imprecise when the sample size n_i is small, or when θ_i is either close to 0 or close to 1. We have some ongoing work to estimate a generalized variance function that will be used to estimate the sampling variance of θ_i,

5.2 Zero-one-inflated Bayesian beta regression

We are also conducting a zero-one-inflated Bayesian beta regression modelling effort. This is being done at the county level for the time being, but we hope to extend it to the school district level. This latter model takes into account the non-negligible number of survey estimates of rates that turn out as zero or one.

5.3 Time Series Model

A more realistic model, is a hierarchical time series model, as in Datta et. al (2002). This model would borrow strength from past years of ACS data, other small areas and covariates. Let

y_{it} : ACS survey weighted proportion for school district i and year t.

$\mathbf{x}_{it}$: a vector of covariates available for school district i and year t.
n_{it} : the effective sample size for school district i and year t.
Areas are grouped in K groups G_k, by population size, $k \in 1, \dots, K$

1. $y_i|\theta_{i,t} \quad \sim \quad \mathcal{N}\left(\theta_{i,t}, \quad \frac{\theta_{i,t}(1-\theta_{i,t})}{\widetilde{n}_{i,t}}\right)$
2. $g(\theta_{i,t})|\theta_{i,t-1} \quad \sim \quad \mathcal{N}\left(\mathbf{x}'_{it}\boldsymbol{\beta}_{kt} + \alpha_{kt}, \quad \tau^2_{kt}\right)$
3. $\alpha_{i,t}|\alpha_{i,t-1} \quad \sim \quad \mathcal{N}\left(\alpha_{i,t-1}, \quad \sigma^2_{kt}\right)$

We will likely transform the observed rates y_i. As was shown, the *arcsine* transformation seems to be beneficial. We will also likely use the t-distribution in level 1, and we will estimate the parameters in a similar manner, as we did for models II(b).

Acknowledgments

The authors wish to thank the Census Bureau's SAIPE team for helping with data preparation. We wish also to thank Donald Luery, Andrew Zbikoswki, and Jasen Taciak, for their helpful comments.

REFERENCES

Box, G.E.P., and Cox, D.R. (1964), "An Analysis of Transformations", *Journal of the Royal Statistical Society. Series B,* Vol. 26, No. 2, 211-252.

Carter, G.M., Rolph, J.E. (1974). "Empirical Bayes Methods Applied to Estimating Fire Alarm Probabilities." *Journal of the American Statistical Association,* Vol. 69, No. 348, pp. 880-885.

Datta G.S., Lahiri P., Maiti T., (2002). "Empirical Bayes estimation of median income of four-person families by state using time series and cross-sectional data." *Journal of Statistical Planning and Inference*, 102, 83-97.

Hawala, S., and Lahiri, P. (2010), "Variance Modeling in the U.S. Small Area Income and Poverty Estimates Program for the American Community Survey," *Proceedings, Section on Bayesian Statistical Science, Section on Survey Research Methods, AS[CD-ROM]*, Alexandria, VA: American Statistical Association.

Kish, L. (1965). *Survey Sampling.* New York: John Wiley.

Lahiri, P., and Pramanik, S. (2010), "Estimation of Average Design-based Mean Squared Error of Synthetic Small Area Estimators," *Statistics Canada International Symposium Series: Proceedings*, Ottawa, Ontario: Statistics Canada.

Liu, B., Lahiri, P., and Kalton, G. (2007), "Hierarchical Bayes Modeling of Survey-Weighted Small Area Proportions", *Proceedings, Section on Bayesian Statistical Science*, Section on Survey Research Methods, American Statistical Association.

Maples, J. and Bell, W. (2007), "Small Area Estimation of School District Child Population and Poverty: Studying Use of IRS Income Tax Data", *Research Report Series, Statistics*, #2007-11, U.S. Census Bureau. `http://www.census.gov/srd/papers/pdf/rrs2007-11.pdf`

Schirm, A., and Zaslavsky, A. (1997), Reweighting Households to Develop Microsimulation Estimates for States", *Proceedings, Section on Bayesian Statistical Science*, Section on Survey Research Methods, American Statistical Association.

Spiegelhalter, D., N. Best, B. Carlin, and A. van der Linde (2002), Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society Series B*, 64, 583-639.

A. Appendix: Figures and Tables

Figure 1: Number of school districts by ACS-2009 number of related children

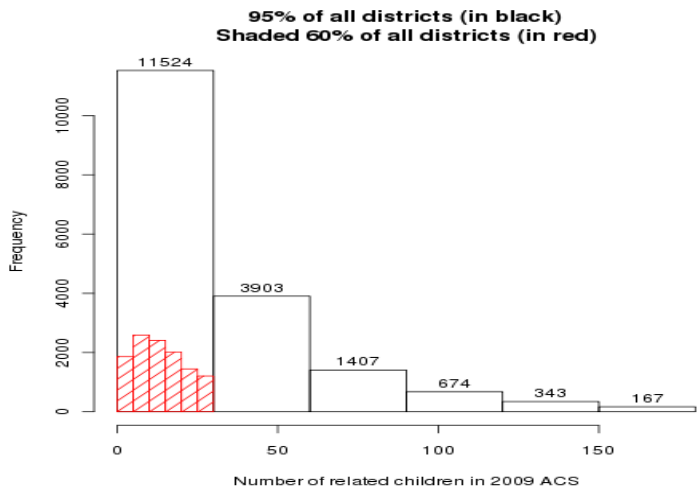

Figure 2: Direct estimate vs. predictors

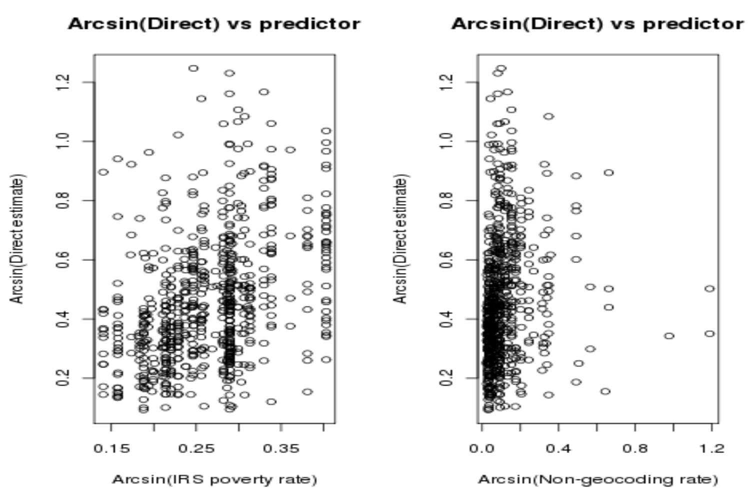

Figure 3: Observed and Estimated Rates - Model I(a)

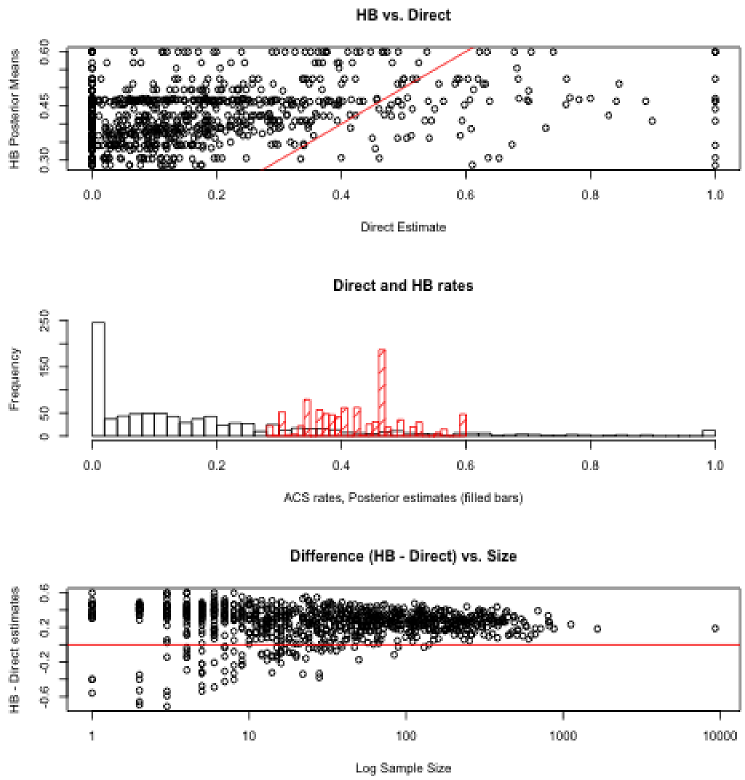

Figure 4: Observed and Estimated Rates - Model I(b)

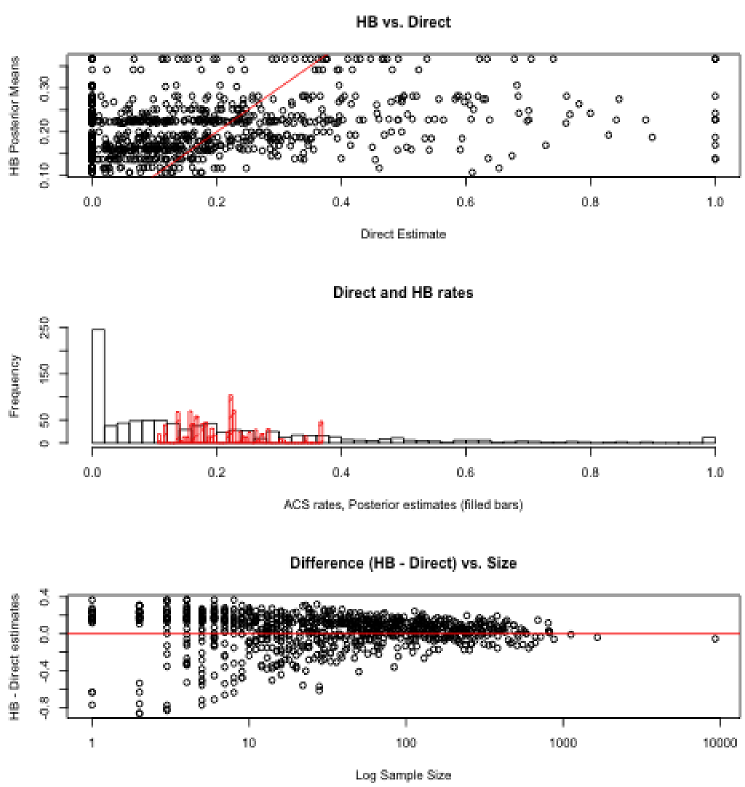

Figure 5: Observed and Estimated Rates - Model II(a)

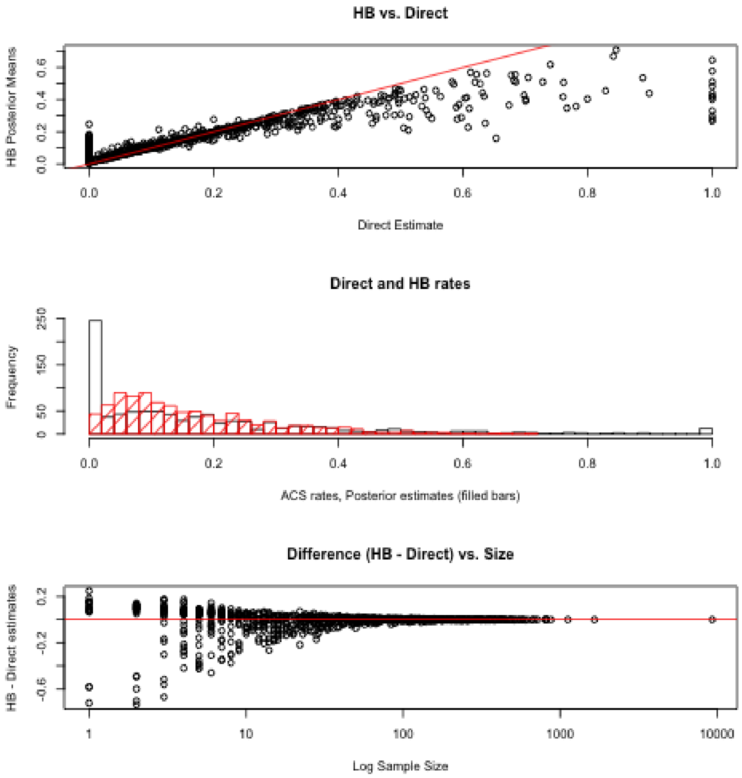

Figure 6: Observed and Estimated Rates - Model II(b)

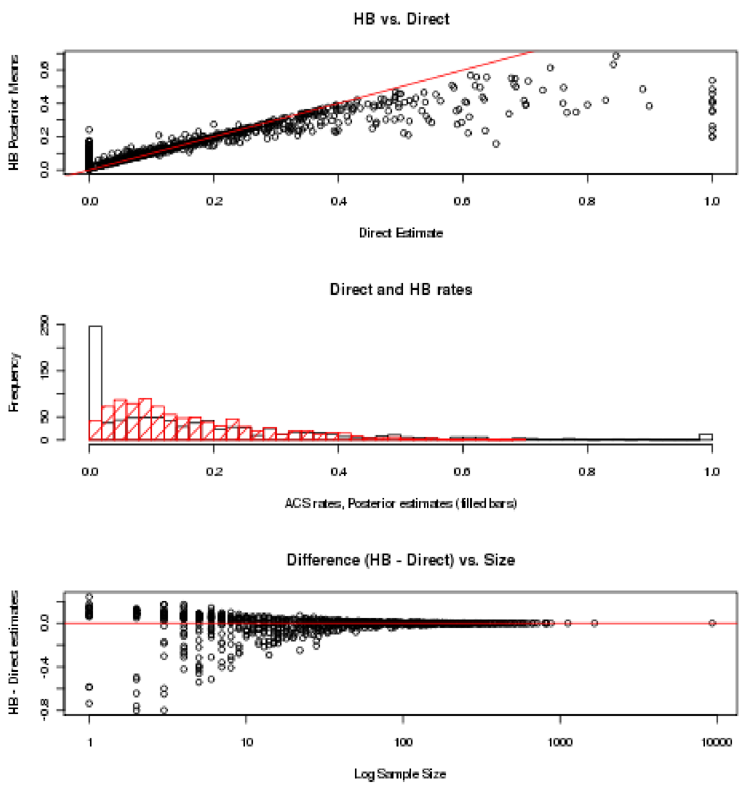

Table 1: Deviance Information Criterion (DIC)

summary	
Model I(a)	3817
Model I(b)	1677
Model II(a)	-1096
Model II(b)	-817.7

Table 2: Models I(a), I(b) - Posterior Means and Standard Deviations of structural parameters

Model	b_1	b_2	b_3	τ	ν
I(a) - Means	-0.5682	5.3313	-0.2882	5.5294	
I(a) - St. Devs	1.2236	0.1568	0.1298	2.6446	
I(b) - Means	-3.7442	6.0828	0.3565	1.8322	15.73
I(b) - St. Devs	2.2451	0.2042	0.1581	4.389	13.9

Table 3: Models II(a), II(b), - Posterior Means and Standard Deviations of structural parameters and ν_z

Model	b_1	b_2	b_3	τ	ν_z
II(a) - Means	-0.0014	1.5081	-0.1019	5.2866	
II(a) - St.Devs	0.0338	0.1302	0.0694	0.1947	
II(b) - Means	-0.0047	1.5208	-0.1272	28.64	19.11
II(b) - St. Devs	0.0339	0.1313	0.0692	2.106	11.85