Generalized Maximum Likelihood Method in Linear Mixed Models with an Application in Small-Area Estimation

P.Lahiri and Huilin Li

University of Maryland, College Park, and National Cancer Institute

Abstract

Standard methods frequently produce zero estimates of dispersion parameters in the underlying linear mixed model. As a consequence, the EBLUP estimate of a small area mean reduces to a simple regression estimate. In this paper, we consider a class of generalized maximum residual likelihood estimators that covers the well-known profile maximum likelihood and the residual maximum likelihood estimators. The general class of estimators has a number of different estimators for the dispersion parameters that are strictly positive and enjoy good asymptotic properties. In addition, the mean squared error of the corresponding EBLUP estimator is asymptotically equivalent to those of the profile maximum likelihood and residual maximum likelihood estimators in the higher order asymptotic sense. However, the strictly positive generalized maximum likelihood estimators have an advantage over the standard methods in estimating the shrinkage parameters and in constructing the parametric bootstrap prediction intervals of the small area means. We shall illustrate our methodology using an example from the SAIPE program of the US Census Bureau.

KEYWORDS: EBLUP; linear mixed model; mean squared error; variance components.

1   Introduction

Linear mixed models have been frequently used in various small area estimation application. Most of the small area models can be considered as a special case of the following general linear mixed model of the form

$$y = X\beta + Zv + e, \tag{1.1}$$

where y is an $n \times 1$ vector of sample observations; X and Z are known matrices; β is a $p \times 1$ vector of unknown parameters (fixed effects); v and e are independent following normal distributions with means 0 and covariance matrices $G = G(\sigma)$ and $R = R(\sigma)$, respectively, depending on some unknown vector of dispersion parameters, $\sigma = (\sigma_1, \sigma_1, \ldots, \sigma_q)' \in \Theta = \{\sigma : \sigma_i \geq 0, i = 1, \cdots, q\}$. We assume that p is fixed and X is of full rank p $(< n)$. Note that $\text{cov}(y) = \Sigma = R + ZGZ'$.

The maximum likelihood estimator of σ is obtained by maximizing a given likelihood $L(\sigma)$, or equivalently the corresponding log-likelihood $l(\sigma)$, with respect to σ in the parametric space Θ. Different maximum likelihood estimators of σ are obtained by varying the choice of $L(\sigma)$. The profile maximum likelihood (PML) and residual maximum likelihood (REML) estimators are among the well-known maximum likelihood estimators considered in the literature. The maximum likelihood estimators enjoy good asymptotic properties

(see, e.g., Jiang 1996; Das et al. 2004, Jiang 2007). However, even when the dispersion parameters are strictly positive, the standard maximum likelihood estimators, including the PML and REML, do not ensure strictly positive estimates of the dispersion parameters for a given dataset.

In section 2, we propose a rich class of generalized maximum likelihood ratio estimators of σ, which includes the PML and REML as special cases. The class also includes consistent estimators that are strictly positive. In this section, we consider different generalized maximum likelihood estimators for the well-known Fay-Herriot model and present higher order asymptotic expressions for the bias and mean square error of the generalized maximum likelihood estimators of a non-linear function of the model variance. In section 3, we consider the related prediction problem. We present the expression for the empirical best linear unbiased prediction (EBLUP) estimator of a general mixed effect. The mean square error of EBLUP and the related prediction interval problem are discussed for the Fay-Herriot model. In section 4, we analyze data from the Small Area Income and Poverty Estimation (SAIPE) project of the US Census Bureau.

2 The Generalized Maximum Likelihood Estimation

We define the generalized likelihood as

$$L(\sigma) = h(\sigma) \times L_R(\sigma),$$

where $L_R(\sigma)$ denote the residual likelihood of σ and $h(\sigma)$ is a suitably chosen smooth function of A that is free of y. The generalized maximum likelihood (GML) estimator of σ, if exists, is obtained by maximizing the generalized likelihood $L(\sigma)$ with respect to σ in the parameter space Θ. Since there is a lot of flexibility in the choice of $h(\sigma)$, we can cover a wide range of likelihoods (e.g., profile maximum likelihood, residual likelihood, etc.) considered in the literature. The generalized maximum likelihood estimator of σ can be made strictly positive by imposing the condition that $h(\sigma) = 0$ when any of the dispersion parameters is zero. The asymptotic properties of the generalized maximum likelihood estimators can be obtained using the standard mathematical tools used to prove similar properties for the profile maximum likelihood and residual maximum likelihood estimators.

We now illustrate the generalized maximum likelihood estimation with the help of the well-known Fay-Herriot model. The Fay-Herriot model (Fay & Herriot, 1979) consists of two levels. In Level 1, the sampling model,

$$y_i|\theta_i \sim N(\theta_i, D_i), \ i = 1, \cdots, m,$$

independently for each i. In Level 2, the linking model,

$$\theta_i \sim N(x_i'\beta, A), \ i = 1, \cdots, m,$$

also independently for each i.

Level 1 accounts for the sampling variability of the regular survey estimates y_i of true small area means θ_i. Level 2 links θ_i to a vector of p known auxiliary variables $x_i = (x_{i1}, \cdots, x_{ip})'$, often obtained from administrative and census records. The sampling variances D_i are assumed to be known.

The Fay-Herriot model has been widely used in small area estimation and related problems for a variety of reasons, including its simplicity, its ability to protect confidentiality of microdata and its ability to produce design-consistent estimators. Some earlier applications of the Fay-Herriot model include the estimation of: (i) false alarm probabilities in New York city (Carter and Rolph, 1974); (ii) the batting averages of major league baseball players (Efron and Morris, 1975); and (iii) prevalence of toxoplasmosis in El Salvador (Efron and Morris, 1975). More recently, the Fay-Herriot model was used: to estimate poverty rates for the U.S. states, counties, and school districts (Citro et al., 1997) and to estimate proportions at the lowest level of literacy for states and counties (Mohadjer et al., 2007). For a comprehensive review of the theory and applications of the above model, see Rao (2003, Chapter 7).

Define $X = (x_1, \cdots, x_m)'$, $v = (v_1, \cdots, v_m)'$, $e = (e_1, \cdots, e_m)'$, and $y = (y_1, \cdots, y_m)'$. The Fay-Herriot model can be rewritten as

$$y = X\beta + v + e, \tag{2.1}$$

a special case of the general linear mixed model with block diagonal variance-covariance structure: $\Sigma = D + A$, where $D = \text{diag}(D_1, \cdots, D_m)$. Thus, for this application $\sigma = A$.

The generalized likelihood of A is given by

$$L(A) = h(A) \times L_R(A),$$

where

$$L_R(A) = c|X'\Sigma^{-1}X|^{-1/2}|\Sigma|^{-1/2}\exp\{-\frac{1}{2}y'Py\},$$

where c is a generic constant free from A and $P = \Sigma^{-1} - \Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}$. The generalized maximum likelihood estimator of A, denoted by $\hat{A}$, is obtained by maximizing the proposed generalized likelihood $L(A)$, or the corresponding generalized log-likelihood function of A given by

$$l(A) = c - \frac{1}{2}\left[\log(|X'\Sigma^{-1}X|) + \log(|\Sigma|) + y'Py\right] + \log h(A).$$

Let $g(A)$ be a general non-linear smooth function of A, which could depend on area i. We estimate $g(A)$ by $g(\hat{A})$. Under mild regularity conditions, it can be shown that the mean squared error (MSE) of $g(\hat{A})$ is the same as that of the residual maximum likelihood estimator $g(\hat{A}_R)$, up to the order $O(m^{-1})$, and is given by

$$E[g(\hat{A}) - g(A)]^2 = \left[\frac{\partial g}{\partial A}\right]^2 \frac{2}{\text{tr}(\Sigma^{-2})} + o(m^{-1}). \tag{2.2}$$

In other words, the mean squared error of $g(\hat{A})$ does not depend on the adjustment factor $h(A)$, up to the order $O(m^{-1})$. However, the biases of different estimators of $g(\hat{A})$ depend on the adjustment factor $h(A)$. Since the MSE of $g(\hat{A})$ does not depend on the adjustment factor $h(A)$, up to the order $O(m^{-1})$, we could compare the bias ratio

$$\frac{\text{Bias}[g(\hat{A})]}{\text{Var}[g(\hat{A})]} = \left(\frac{\partial g}{\partial A}\right)^{-1} \frac{\partial \log h(A)}{\partial A} + \frac{1}{2} \frac{\partial^2 g}{\partial A^2} \left(\frac{\partial g}{\partial A}\right)^{-2} + o(1).$$

For the case of the shrinkage factors $g(A) = B_i = D_i/(A + D_i)$, we have

$$\frac{\text{Bias}[B_i(\hat{A})]}{\text{Var}[B_i(\hat{A})]} = \frac{1}{1 - \gamma_i} \left[1 - \frac{\phi}{\gamma_i}\right],$$

where

$$\gamma_i = \frac{A}{A + D_i} \quad \text{and} \quad \phi = A \frac{\text{Bias}(\hat{A})}{\text{Var}(\hat{A})}.$$

For the following choice of the adjustment term:

$$h_1(A) = A^{\frac{1}{2}c_1} \left(|\Sigma|^{\frac{1}{m}}\right)^{\frac{1}{2}c_2} |X'\Sigma^{-1}X|^{\frac{1}{2}c_3},$$

we get

$$\phi = \frac{1}{2}[c_1 + c_2\bar{\gamma} - c_3\bar{\gamma}_w] + o(m^{-1}),$$

where $\bar{\gamma} = m^{-1}\sum_{j=1}^m \gamma_j$, $\bar{\gamma}_w = \sum_{j=1}^m \gamma_j h_{jj}$, and $h_{jj} = \gamma_j x_j' \left(\sum_{j=1}^m \gamma_j x_j x_j'\right)^{-1} x_j$, under mild regularity conditions. Note that the adjustment factor $h(A)$ is quite general and generates a variety of likelihood functions considered in the literature. For example, the choices (i) $c_1 = c_2 = 0$, $c_3 = 1$, (ii) $c_1 = c_2 = c_3 = 0$, and (iii) $c_1 = 2, c_2 = c_3 = 0$ result in the profile likelihood, residual likelihood and an adjusted residual likelihood considered by Morris (2006).

We are specially interested in $h_1(A)$ with $c_1 > 0$ since this restriction will ensure a strictly positive estimate of A. To fix ideas, consider $c_1 = q$, $c_2 = c_3 = 0$ for which $\phi = q/2 + o(1)$. Thus, in this case the generalized likelihood is obtained when the residual likelihood is multiplied by $A^{\frac{1}{2}q}$. Note that the generalized likelihood reduces to the Morris' ADM for the choice $q = 2$. Unlike the REML, Morris ADM is strictly positive. But,

compared to the REML, the Morris ADM overestimates A as the leading term of ϕ for Morris's ADM is 1 while it is just zero for the REML. When the parameters of interest are the shrinkage factors, the bias ratio expression suggests some good choices for the generalized maximum likelihood estimator.

3   The Prediction of Mixed Effects

In small area estimation, we are interested in estimating $\mu = l'\beta + m'v$, for specified vectors of constants l and m. The best linear unbiased prediction (BLUP) estimator of μ, under model (1.1), is given by

$$t(\sigma) = l'\tilde{\beta} + s(\sigma)'(y - X\tilde{\beta}) , \tag{3.1}$$

where

$$\tilde{\beta} = \tilde{\beta}(\sigma) = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y$$

is the generalized least squares estimator of β, and $s(\sigma) = \Sigma^{-1}ZGm$. An empirical best linear unbiased prediction (EBLUP) estimator $t(\hat{\sigma})$ of μ is obtained by plugging in an estimator $\hat{\sigma}$ for σ in the BLUP expression. We refer interested readers to Rao (2003) and Jiang (2007) for a detailed discussions on BLUP and EBLUP.

The mean squared error (MSE) of EBLUP is defined as

$$\text{MSE}[t(\hat{\sigma})] = \text{E}[t(\hat{\sigma}) - \mu]^2,$$

where E is over the model (1.1). An estimator, denoted as $\text{mse}[t(\hat{\sigma})]$, is called a second-order unbiased (or nearly unbiased) estimator of $\text{MSE}[t(\sigma)]$ if

$$\text{E}\left\{\text{mse}[t(\hat{\sigma})]\right\} = \text{MSE}[t(\hat{\sigma})] + o(m^{-1}).$$

Some important papers that use the Taylor series linearization method in obtaining second-order unbiased MSPE estimators include Datta and Lahiri (2000), Das et al. (2004), and Datta et al. (2005). See Rao (2003) and Jiang (2007) for a detailed account of the Taylor linearization method in small area estimation. Under certain regularity conditions, Das et al. (2004) obtained a second-order unbiased (or nearly unbiased) Taylor linearization estimator of $\text{MSE}[t(\hat{\sigma})]$ when the dispersion parameters are estimated by standard "score function" methods. It is straightforward to extend their results when the dispersion parameters are estimated by the proposed generalized maximum likelihood method, under suitable regularity conditions on the adjustment factor $h(\sigma)$.

For the Fay-Herriot model, an EBLUP estimator of θ_i is given by

$$\hat{\theta}_i^{\text{EBLUP}} = y_i - \hat{B}_i(y_i - x_i'\hat{\beta}),$$

where $\hat{B}_i = D_i/(\hat{A} + D_i), i = 1, \cdots, m$, and $\hat{\beta}^{\text{w}} = \left(\sum_{j=1}^m x_j x_j' \hat{B}_j/D_j\right)^{-1} \sum_{j=1}^m x_j y_j \hat{B}_j/D_j$, the weighted least square estimator of β with B_j replaced by the generalized maximum likelihood estimators $\hat{B}_j$.

Using algebra similar to that of Li and Lahiri (2009), we obtain the following second-order (or, nearly) unbiased estimator of $\text{MSE}(\hat{\theta}_i^{\text{EBLUP}})$:

$$\text{mse}(\hat{\theta}_i^{\text{EBLUP}}) = g_{1i}(\hat{B}_i) + g_{2i}(\hat{B}) + 2g_{3i}(\hat{B}) - (\hat{B}_i)^2 \widehat{\text{Bias}}(\hat{A}),$$

where

$$
\begin{aligned}
g_{1i}(B_i) &= D_i(1 - B_i), \\
g_{2i}(B) &= B_i^2 x_i' \left(\sum_{j=1}^m \frac{B_j}{D_j} x_j x_j'\right)^{-1} x_i, \\
g_{3i}(B) &= 2\frac{B_i^3}{D_i} \left\{\sum_{j=1}^m \left(\frac{B_j}{D_j}\right)^2\right\}^{-1},
\end{aligned}
$$

with $B = (B_1, \cdots, B_m)'$ and $\widehat{\text{Bias}}(\hat{A})$ is an estimator of bias of the generalized maximum likelihood estimator (see section 2).

Let us now discuss the prediction interval problem. Following Cox (1975), one can propose the following empirical best prediction interval for θ_i: $\hat{\theta}_i^{\text{EBLUP}} \pm z_{\alpha/2} \hat{\sigma}_i$, where $\hat{\sigma}_i^2 = g_{1i}(\hat{B}_i)$ and $z_{\alpha/2}$ is the upper $100(1 - \alpha/2)\%$ point of the standard normal distribution. This prediction interval is asymptotically correct in the sense that the coverage probability converges to $1 - \alpha$, for large m and certain regularity conditions. However, for small m, this is not efficient since the coverage error of this interval is of order $O(m^{-1})$, which is not accurate enough for most small area applications. See Chatterjee et al. (2008) for a review of different attempts to improve on the Cox's prediction interval.

For a general linear mixed model, Chatterjee et al. (2008) proposed a parametric bootstrap method to obtain a prediction interval directly from the bootstrap histogram of the certain pivot quantity. The method is a significant improvement over the Cox empirical Bayes prediction interval. One of the critical requirements of the parametric bootstrap method is that the dispersion parameter estimates must be strictly positive. This paper thus strengthens the parametric bootstrap methodology of Chatterjee et al. (2008) by suggesting methods to generate strictly positive dispersion parameters. For the Fay-Herriot model, the

$100(1 - \alpha)\%$ parametric bootstrap prediction interval for θ_i is given by:

$$[\hat{\theta}_i^{\text{EBLUP}} - q_1 \hat{\sigma}_i, \hat{\theta}_i^{\text{EBLUP}} + q_2 \hat{\sigma}_i],$$

where

$$\mathcal{L}_i^*(q_2) - \mathcal{L}_i^*(q_1) = 1 - \alpha,$$

$\mathcal{L}_i^*$ being the parametric bootstrap approximation to $\mathcal{L}_i$, the distribution of $\hat{\sigma}_i^{-1}(\theta_i - \hat{\theta}_i^{\text{EBLUP}})$. See Li and Lahiri (2009) for details.

4   SAIPE Data Analysis

Governmental policies increasingly demand income and poverty estimates for small areas. In the U.S. more than \$130 billion of federal funds per year are allocated based on these estimates. In addition, states utilize these small area estimates to divide federal funds and their own funds to areas within the state. These funds cover a wide range of community necessities and services including education, public health, and numerous others. Therefore, there is a growing need to refine manners in which these estimates are taken to provide an increased level of precision. In response to the growing need for these estimates, the Census Bureau formed a committee on Small Area Income and Poverty Estimates (SAIPE). This committee was created in the early 1990's with the goal of providing more timely and precise estimates than those from the data in the decennial census in accordance with the need for these estimates in order to allocate governmental funds. For example, the "Improving America's Schools Act" requires SAIPE estimates of poor school-age children (under 18) for counties as well as school districts in order to allocate more than \$7 billion annually for the educationally disadvantaged students. The SAIPE provides estimates of income and poverty at state, county, and district levels. For example, it provides estimates of all poor persons, poor children under the age of 5, poor children under the age of 18, poor children aged 5-17, and median income of households for states. See Citro et al. (1997) and Bell (1999), for details.

In this section, we study the problem of estimating annual poverty rates of school-aged (5-17) children for the 50 states and District of Columbia of the United States, an important component of the Census Bureau's Small Area Income and Poverty Estimates (SAIPE) program. For the state level SAIPE, the Census Bureau has switched from the Current Population Survey (CPS) to the American Community Survey (ACS). But, for the sake of illustration, we will use the old CPS based estimation. Fay and Train (1997) developed a Fay-Herriot model for this estimation problem. In our notations, y_i denotes the direct estimate of the true

poverty rate θ_i, and x_i is a 5×1 vector consisting of unity and four known auxiliary variables derived from administrative records. The sampling variances D_i were obtained from a sampling error model of Otto and Bell (1995) that involved fitting a generalized variance function (GVF) to five years of direct variance and covariance estimates for each state produced by Fay and Train (1995). U.S. Internal Revenue Service income tax return files were used to obtain two auxiliary variables: an analogue to state child poverty rates and also state rates of non-filing for income taxes. Data from the U.S. Department of Agriculture were used to develop a variable reflecting state participation rates in the food stamp poverty assistance program. In addition, x_i includes the residual from regressing 5-17 state poverty rates from the previous 1990 decennial census on the other regression variables for 1989 (the census income reference year). For further details on SAIPE, the readers are referred to Bell(1999) and the website: http://www.census.gov/hhes/www/saipe.html.

Table 1 displays different estimates of A for five years, from 1989 to 1993. We consider four standard methods — the Prasad-Rao (PR), Fay-Herriot (FH), REML and PML. In addition, we consider two new generalized maximum likelihood methods obtained by choosing $h_1(A) = A$ (AR) and $h_1(A) = A|X'\Sigma^{-1}X|^{\frac{1}{2}}$ (AM). It is interesting to note that all the standard methods considered in Table 1 are subject to zero estimate. For the year 1992, the standard methods fail to produce a positive estimate. The REML and PML estimates are all zero except for the year 1993. In contrast, all ADM estimates are positive supporting our theory. The results for the PML and REML was first noted by Bell (1999, 2001). When A is estimated at zero, the EBLUP estimator puts all the weight to the synthetic estimator and none to the direct estimator. So for this important application, we need a variance component estimation method that can always provide a positive estimate with desirable statistical properties.

In Figure 1, we draw the profile log-likelihood function of A for the year 1992. We can readily see that the function attains its maximum on the boundary, which gives the zero PML estimate. On the other hand, AM maximizes an adjusted profile likelihood obtained by multiplying the profile likelihood by A. The multiplier A pushes the mode of the profile likelihood to the right resulting in strictly positive estimate of A. For the year 1993, the PML estimate of A is 0.43. In this case also, the multiplier A pushes the mode of the profile likelihood to the right yielding a larger adjusted PML estimate (2.42). Similar conclusions can be drawn for REML and adjusted REML.

Overestimation of the adjusted methods makes up for underestimation of the shrinkage estimator B_i due to its convexity. We again consider the years 1992 and 1993 to compare the impact of the REML, PML and two adjusted likelihood estimators of A on B_i. Figure 2 displays B_i versus states, where states are sorted

in increasing order of the D_i's. In the graphs, ADM RE and ADM ML are simply the adjusted REML and adjusted PML estimators that we discussed earlier. Since for 1992 both RE and PML estimates of A are zero, we have $\hat{B}_i = 1$ for all the states. On the other hand, the adjusted likelihood estimates of A yield large $\hat{B}_i$ for the states with large D_i and small $\hat{B}_i$ for the states with small D_i. When D_i is small, the direct estimate y_i is reasonable. Thus, in this situation a sensible EBLUP method should put more weight to y_i than the synthetic regression part. In this sense, the REML and PML do not provide a sensible result; but the adjusted likelihood methods do. For the year 1993, all the methods produce non-zero estimates of A, and yield large $\hat{B}_i$ for the states with large D_i and small $\hat{B}_i$ for the states with small D_i. However, estimates of B_i based on the adjusted likelihood estimates of A are much smaller than the corresponding REML and PML estimates and thus the adjusted likelihood methods put more weight to the direct part and are more conservative than the corresponding EBLUP based on REML or ML.

For the year 1992, all the standard methods provide very unreliable interval estimates. For example, the the length of the Cox empirical Bayes (EB) interval is typically much narrower than the ideal length, which causes severe undercoverage problem. So it will be interesting to examine the lengths of the Cox empirical Bayes intervals for our application. For 1992, the Cox EB interval produces a prediction interval of length 0. To understand the difference between prediction intervals produced by a traditional method and a parametric bootstrap method, we consider the year 1993 where all methods produce non-zero estimates of A. As an illustration, we compare the parametric bootstrap prediction intervals with the Cox prediction interval that uses the Fay-Herriot method of estimating A. For this data set, the length of the parametric bootstrap prediction intervals based on the adjusted maximum likelihood methods are always more than those of the Cox prediction intervals for all the states, inflation in lengths ranging from 18% to 83%. For details, see Li (2007).

References

[1] Bell, W. (1999). Accounting for uncertainty about variances in small area estimation. *Bulletin of the International Statistical Institute*, 52nd Session, Helsinki, 1999.

[2] Bell, W. (2001), Discussion with "Jackknife in the Fay-Herriott model with an example," *Proceeding of the Seminar on Funding Opportunity in Survey Research* , 98-104.

[3] Carter, G. and Rolph, J. (1974), Empirical Bayes Methods Applied to estimating fire alarm probabilities.

Journal of the American Statistical Association. **69**, 880-885.

[4] Chatterjee, A., Lahiri, P., and Li, H. (2008), Parametric bootstrap approximation to the distribution of EBLUP, and related prediction intervals in linear mixed models. *Annals of Statisitcs.* **36**, 1221-1245

[5] Citro, C.F., Cohen, M.L., Kalton, G.,and West, K.K. eds. (1997), *Small-area estimates of school-age children in poverty: Interim report I-evaluation of 1993 county estimates for title I allocations.* National Academy Press, Washington, D.C.

[6] Cox, D.R. (1975), Prediction intervals and empirical Bayes confidence intervals. *Perspectives in Probability and Statistics,* papers in honor of M.S. Bartlett. Ed: J. Gani, Academic Press, 47-55.

[7] Das, K., Jiang, J. and Rao, J.N.K. (2004), Mean squared error of empirical predictor. *Annals of Statisitcs.* **32**, 818-840.

[8] Datta, G. S. and Lahiri, P. (2000), A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statist. Sinica.* **10**, 613-627.

[9] Datta, G.S., Rao, J.N.K. and Smith, D.D. (2005), On measuring the variability of small area estimators under a basic area level model. *Biometrika.* **92**, 183-196.

[10] Efron, B., and Morris, C. (1975), Data analysis using Stein's estimator and its generalizations. *J. Amer. Statist. Asso..* **70**, 311-9.

[11] Fay, R.E., and Herriot, R.A. (1979), Estimates of income for small places: an application of James-Stein procedure to Census data. *Journal of the American Statistical Association.* **74**, 269-277.

[12] Fay, R.E. and Train, G.F. (1997), "Small Domain Methodology for Estimating Income and Poverty Characteristics for States in 1993," *American Statistical Association, Proceedings of the Social Statistics Section*, 183-188.

[13] Jiang, J. (1996). REML estimation: asymptotic behavior and related topics. *Annals Statistics* **24**, 255-286.

[14] Jiang, J. (2007), *Linear and generalized linear mixed models and their applications.* Springer.

Table 1: Different Variance Estimation Methods

Year	PR	FH	REML	PML	AR	AM
1989	0.00	0.06	0.00	0.00	1.83	1.15
1990	1.13	0.85	0.00	0.00	2.31	1.42
1991	0.00	0.00	0.00	0.00	1.67	1.14
1992	0.00	0.00	0.00	0.00	1.74	1.17
1993	5.85	2.67	1.70	0.43	3.61	2.42

[15] Li, H.(2007), Small area estimation: an empirical best linear unbiased prediction approach. Ph.D. Thesis, Dept. of Mathematics, University of Maryland, College Park.

[16] Li, H. Lahiri, P.(2009), Adjusted Maximum Likelihood Method in Small Area Estimation Problems, under revision in *Journal of Multivariate Analysis.*

[17] Mohadjer, L., Rao, J.N.K., Liu, B., Krenzke, T., and VanDekerckhove, W. (2007), Hierarchical Bayes small area estimates of adult literacy using unmatched sampling and linking models, to appear in *Proceedings of the Survey Research Methods Section, American Statistical Association.*

[18] Morris, C.N. (2006), Mixed model prediction and small area estimation (with discussions). *Test.* **15**, 1, 72-76.

[19] Otto, M. and Bell, W. (1995), "Sampling Error Modelling of Poverty and Income Statistics for States," *Proceedings of the Section on Survey Research Methods*, Washington, D.C. American Statistical Association.

[20] Rao, J.N.K. (2003), *Small Area Estimation*, Wiley, New York.
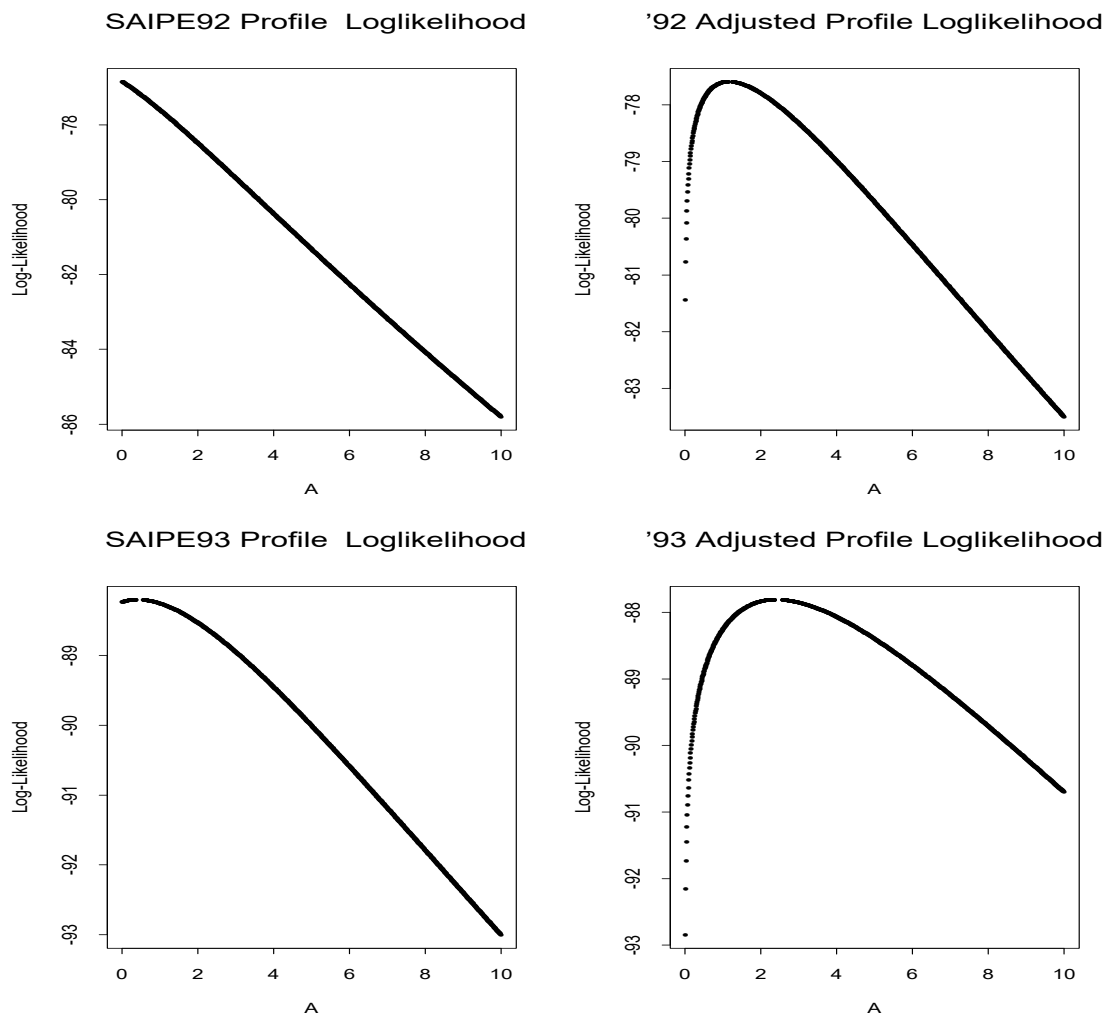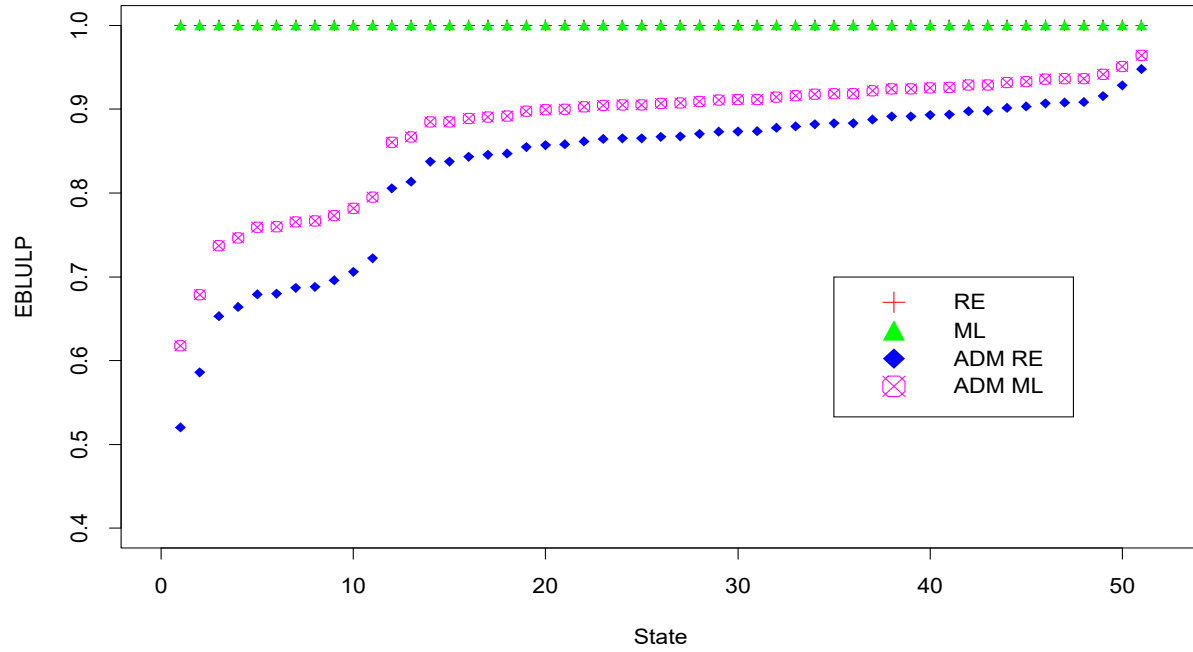
Figure 1: SAIPE 92 and 93 Log Profile Likelihood

Figure 2: SAIPE 92 and 93 B_i versus State