Journal of Multivariate Analysis 101 (2010) 882–892

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

An adjusted maximum likelihood method for solving small area estimation problems

Huilin Li [a,*], P. Lahiri [b]

[a] Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD 20892, United States

[b] Joint Program of Survey Methodology, University of Maryland, College Park, MD 20742, United States

ARTICLE INFO

Article history:
Received 25 September 2008
Available online 31 October 2009

AMS subject classifications:
62F12
62F40
62J99

Keywords:
Adjusted density maximization estimator
The Fay–Herriot model
Parametric bootstrap
Prediction intervals

ABSTRACT

For the well-known Fay–Herriot small area model, standard variance component estimation methods frequently produce zero estimates of the strictly positive model variance. As a consequence, an empirical best linear unbiased predictor of a small area mean, commonly used in small area estimation, could reduce to a simple regression estimator, which typically has an overshrinking problem. We propose an adjusted maximum likelihood estimator of the model variance that maximizes an adjusted likelihood defined as a product of the model variance and a standard likelihood (e.g., a profile or residual likelihood) function. The adjustment factor was suggested earlier by Carl Morris in the context of approximating a hierarchical Bayes solution where the hyperparameters, including the model variance, are assumed to follow a prior distribution. Interestingly, the proposed adjustment does not affect the mean squared error property of the model variance estimator or the corresponding empirical best linear unbiased predictors of the small area means in a higher order asymptotic sense. However, as demonstrated in our simulation study, the proposed adjustment has a considerable advantage in small sample inference, especially in estimating the shrinkage parameters and in constructing the parametric bootstrap prediction intervals of the small area means, which require the use of a strictly positive consistent model variance estimate.

1. Introduction

The Fay–Herriot model [1], widely used in small area estimation, consists of two levels. In Level 1, we have the sampling model,

$$Y_i|\theta_i \sim N(\theta_i, D_i), \quad i = 1, \ldots, m,$$

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

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

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 $x_i = (x_{i1}, \ldots, x_{ip})'$, 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

* Corresponding author.
E-mail addresses: lih5@mail.nih.gov (H. Li), plahiri@survey.umd.edu (P. Lahiri).

0047-259X/$ – see front matter. Published by Elsevier Inc.
doi:10.1016/j.jmva.2009.10.009

estimators. Some earlier applications of the Fay–Herriot model include the estimation of: (i) false alarm probabilities in New York city [2]; (ii) the batting averages of major league baseball players [3]; and (iii) prevalence of toxoplasmosis in El Salvador [3]. More recently, the Fay–Herriot model was used: to estimate poverty rates for the US states, counties, and school districts [4] and to estimate proportions at the lowest level of literacy for states and counties [5]. For a comprehensive review of the theory and applications of the above model, see [6, Chapter 7].

The best predictor (BP) of θ_i and the associated mean squared prediction error (MSPE) are given by

$$\hat{\theta}_i^{\text{BP}} = Y_i - B_i(Y_i - x_i'\beta),$$

and

$$\text{MSPE}[\hat{\theta}_i^{\text{BP}}] = \text{E}[\hat{\theta}_i^{\text{BP}} - \theta_i]^2 = g_{1i}(B_i),$$

where $0 < B_i = D_i/(A+D_i) < 1$, $i = 1, \ldots, m$; E is the expectation with respect to the joint distribution of Y and θ induced by the Fay–Herriot model and $g_{1i}(B_i) = D_i(1 - B_i)$.

The best predictor shrinks the direct estimator Y_i towards the regression surface $x_i'\beta$, the amount of shrinkage being determined by the shrinkage factor B_i. The closer the value of B_i to 1, the greater the strength of the Level 2 model and hence the greater the efficiency of the best predictor, as reflected by a smaller value of the mean squared prediction error of the best predictor. When $A = 0$, that is when the Level 2 model is perfect, $B_i = 1$ for all $i = 1, \ldots, m$. In this case, the best predictor is identical to the regression estimator. This situation, however, is unrealistic, since Level 2 modeling, just like any modeling, cannot be perfect, that is A should be always greater than 0. Thus, throughout the paper we assume $A > 0$.

In practice, both β and the B_i's are unknown and need to be estimated from the data. The regression parameter β is estimated by the weighted least square estimator $\tilde{\beta}^{\text{w}} = (\sum_{j=1}^m x_j x_j' B_j/D_j)^{-1} \sum_{j=1}^m x_j Y_j B_j/D_j$. When this estimator of β is plugged into the best predictor, the best linear unbiased predictor (BLUP) of θ_i is obtained and is denoted by $\hat{\theta}_i^{\text{BLUP}}$. It is now clear that the shrinkage factors B_i are important parameters to estimate. They are needed for a good evaluation of the Level 2 model and to carry out the necessary prediction analyses. When estimates of B_i are plugged into the best linear unbiased predictor formula, one obtains an empirical best linear unbiased predictor (EBLUP) of θ_i, denoted by $\hat{\theta}_i^{\text{EBLUP}}$.

From Jensen's inequality and the convexity of B_i as a function of A, it follows that $\hat{B}_i$ overestimates B_i even when an exactly unbiased estimator of A is used, and the extent of the overestimation may be severe for small m. In addition, standard methods of estimation of A considered in the literature, including using the Prasad–Rao simple method-of-moments estimator, $\hat{A}^{\text{PR}}$ [7], the Fay–Herriot method-of-moments estimator, $\hat{A}^{\text{FH}}$ [1,8], the maximum likelihood estimator, $\hat{A}^{\text{ML}}$, and the residual maximum likelihood estimator, $\hat{A}^{\text{RE}}$, are all subject to zero estimates resulting in undesirable estimates $\hat{B}_i = 1$ for all $i = 1, \ldots, m$. In real life data analyses, this problem is quite frequent (see, e.g., [9] and [10]).

In Section 2, we introduce an adjustment to the maximum (profile or residual) likelihood estimator of A in order to produce a strictly positive estimate of A, for small m. The proposed adjustment increases the order of bias of the residual maximum likelihood estimator, but not the mean squared error, up to the order $O(m^{-1})$. However, the mean squared error or the bias property of the maximum profile likelihood estimator of A remains unaffected, up to order $O(m^{-1})$. In terms of the estimation of the shrinkage factors B_i, the adjustment does not increase the order of the bias or the mean squared error, irrespective of whether a profile or residual likelihood function is used for the adjustment. While there is no clear advantage of using the proposed adjusted maximum likelihood methods for large m, they have a clear edge over the standard methods for small m in terms of preventing the full shrinkage.

Morris [11] proposed a method, known as the adjusted density maximization (ADM) method, as an intermediary step in approximating a hierarchical Bayes solution. Recently, Morris and Tang [12] (also see [13]) pursued the ADM method for the Fay–Herriot model. The ADM approximations to the posterior means of A and B_i, under an (improper) uniform prior on β and superharmonic prior [14] on A, are identical to the corresponding adjusted maximum residual likelihood estimators given in this paper. However, unlike Morris and Tang [12], we consider a classical prediction approach, which does not assume a prior distribution for β and A, in measuring the uncertainty of the proposed EBLUP and the associated prediction interval. Moreover, for the Fay–Herriot model, Morris and Tang [12] did not suggest the adjusted maximum profile likelihood estimator, which appears to perform better than the adjusted maximum residual likelihood estimator in our simulation study.

The mean squared prediction errors of empirical best linear unbiased predictors of θ_i that use the proposed adjusted maximum likelihood estimators are presented in Section 3. In this section, we also provide the second-order (or nearly) unbiased estimators of the mean squared prediction errors of empirical best linear unbiased predictors when the proposed adjusted maximum likelihood estimators are used. The use of the proposed adjusted maximum likelihood estimators of A does not affect the mean squared prediction errors of empirical best linear unbiased predictors, up to the order $O(m^{-1})$. However, the expressions for the proposed nearly unbiased mean squared prediction error estimators are different for different methods of estimating A.

Cox [15] and Morris [16] proposed normality based empirical Bayes confidence intervals of θ_i. The coverage errors of such intervals are typically of order $O(m^{-1})$. Chatterjee, Lahiri and Li [17] proposed an improved interval estimation using the parametric bootstrap method. The method requires repeated generation of a pivotal quantity from several bootstrap samples. A strictly positive estimate of A is absolutely needed for this method since the pivotal quantity is undefined when the A estimate is zero. A crude fix is to take a small positive number whenever the A estimate turns out to be zero. But, in