Statistics and Probability Letters 143 (2018) 37–42

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Best look-alike prediction: Another look at the Bayesian classifier and beyond

Hanmei Sun [a], Jiming Jiang [b,*], Thuan Nguyen [c], Yihui Luan [a]

[a] Shandong University, China
[b] University of California, Davis, USA
[c] Oregon Health & Science University, USA

ARTICLE INFO

Article history:
Received 23 September 2017
Received in revised form 13 July 2018
Accepted 17 July 2018
Available online 1 August 2018

Keywords:
BLAP
Categorical outcome
Mixed logistic model
Zero-inflated random variable
Small area estimation

ABSTRACT

A criterion of optimality in prediction is proposed that requires the predictor to assume the same type of values as the random variable it is predicting. In the case of categorical responses, the method leads to the Bayesian classifier with a uniform prior. However, the method extends to other cases, such as zero-inflated observations, as well. The method, called best look-alike prediction (BLAP), justifies an "usual practice" from a theoretical standpoint. Application of BLAP to small area estimation is considered. A real-data example is discussed.

1. Introduction

Many practical problems are related to prediction, where the main interest is at subject (e.g., precision medicine) or sub-population (e.g., small area estimation) level. The traditional concept of best prediction (BP) is in terms of mean squared prediction error (MSPE). Under such a framework, the best predictor (BP) is known as the conditional expectation of the random variable to be predicted, say, α, given the observed data, say, Y, that is, $\mathrm{E}(\alpha|Y)$. Based on the BP, a number of more specialized prediction methods have been developed, such as best linear prediction (BLP) and best linear unbiased prediction (BLUP). See, for example, sec. 2.3 of Jiang (2007). In particular, mixed model prediction, that is, prediction based on mixed effects models, has a fairly long history starting with Henderson's early work in animal breeding (Henderson, 1948). See, for example, Robinson (1991), Jiang and Lahiri (2006), Jiang (2007, sec. 2.3), and Rao and Molina (2015).

In spite of its dominance in prediction theory, and overwhelming impact in practice, the BP can have a very different look than the random variable it is trying to predict. This is particularly the case when the random variable is discrete, categorical, or has some features related to a discrete or categorical random variable. For example, if α is a binary random variable taking the values 1 or 0, its BP, $\mathrm{E}(\alpha|Y)$, is typically not equal to 1 or 0; instead, the value of $\mathrm{E}(\alpha|Y)$ usually lies strictly between 0 and 1. Such a feature of the BP is sometimes unpleasant, or inconvenient, for a practitioner because the values 1 and 0 correspond directly to outcomes of scientific, social, or economic interest; there is no such an outcome that corresponds to, say, 0.35, or at least not directly.

Let α be an unobserved, possibly vector-valued random variable for which we wish to predict. The prediction will be based on the observed data, denoted by Y. A predictor, say $\tilde{\alpha}$, is said to be *look-alike* with respect to α if it has the same set

* Corresponding author.
E-mail address: jimjiang@ucdavis.edu (J. Jiang).

https://doi.org/10.1016/j.spl.2018.07.014

of possible values as α. We derive the best predictor under this framework and a suitable criterion of optimality, which is different than the BP. We refer the method as best look-alike prediction, or BLAP.

It should be noted that, in the case of predicting categorical outcomes, BLAP leads to the same solution as what is known as the Bayesian classifier under a uniform prior (e.g., Nurty and Devi, 2011). However, under our framework there is no prior; instead, the unknown parameters involved in BLAP are estimated from the data, leading to the empirical BLAP, or EBLAP. Furthermore, the BLAP method applies to other situations, such as zero-inflated observations, for which there is no Bayesian classifier.

The derivation of BLAP is given in Section 2 for the case of prediction of a discrete or categorical random variables. In Section 3, we derive BLAP for zero-inflated random variables. In Section 4 we consider an application of BLAP to small area estimation (e.g., Rao and Molina, 2015) with zero-inflated random effects. Some real-data applications are discussed in Section 5. Further details and results can be found in an online supplement.

2. BLAP for discrete/categorical random variable

Let α denote a discrete or categorical random variable (r.v.) that we wish to predict. Without loss of generality, we can assume that α is a discrete r.v. whose values are nonnegative integers. Let S denote the set of possible values of α. Let $\tilde{\alpha}$ be a predictor of α based on the observed data, Y. $\tilde{\alpha}$, is look-alike (with respect to α) if it also has S as its set of possible values. The performance of $\tilde{\alpha}$ is measured by the probability of mismatch:

$$\mathrm{p}(\tilde{\alpha} \neq \alpha) = \sum_{k\in S} \mathrm{P}(\tilde{\alpha} = k, \alpha \neq k). \tag{1}$$

$\tilde{\alpha}$ is said to be the best look-alike predictor, or BLAP, if it minimizes the probability of mismatch, (1). The following theorem defines a BLAP.

Theorem 1. *A BLAP of α is given by*

$$\tilde{\alpha}^* = \min\left\{ i \in S : \mathrm{P}(\alpha = i|Y) = \max_{k\in S} \mathrm{P}(\alpha = k|Y) \right\}, \tag{2}$$

provided that the right side of (2) is computable.

Note. In the context of BP, it is well known that the expression of BP is very similar to the Bayesian posterior mean. Similarly, the expression of BLAP in this case, that is, (2), resembles that of the Bayesian classifier under a uniform prior (e.g., Nurty and Devi, 2011). Of course, there is no prior distribution in our consideration.

Proof of Theorem 1. Note that the right side of (1) can be expressed as

$$\begin{aligned}
\sum_{k\in S} \mathrm{E}\left\{1_{(\tilde{\alpha}=k)}\mathrm{P}(\alpha \neq k|Y)\right\} &= \sum_{k\in S} \mathrm{E}\left[1_{(\tilde{\alpha}=k)}\{1 - \mathrm{P}(\alpha = k|Y)\}\right] \\
&\geq \sum_{k\in S} \mathrm{E}\left[1_{(\tilde{\alpha}=k)}\left\{1 - \max_{k\in S} \mathrm{P}(\alpha = k|Y)\right\}\right] \\
&= \sum_{k\in S} \mathrm{E}\left[1_{(\tilde{\alpha}^*=k)}\{1 - \mathrm{P}(\alpha = k|Y)\}\right] \\
&= \mathrm{P}(\tilde{\alpha}^* \neq \alpha).
\end{aligned} \tag{3}$$

The second-to-last equation in (3) is because, when $\tilde{\alpha}^* = k$, one has, by the definition, $\mathrm{P}(\alpha = k|Y) = \max_{k\in S}\mathrm{P}(\alpha = k|Y)$; the last equation in (3) is, again, due to (1) and the first equation in (3). This completes the proof.

A special case of Theorem 1 is the binary case, which we state as a corollary.

Corollary 1. *Suppose that δ is a binary r.v. taking the values 1 and 0. Then, the BLAP of δ is given by $\tilde{\delta}^* = 1_{\{\mathrm{P}(\delta=1|Y)\geq 1/2\}}$, provided that $\mathrm{P}(\delta = 1|Y)$ is computable.*

Typically, the conditional probability, $\mathrm{P}(\alpha = k|Y)$, depends on some unknown parameters, say, θ. It is customary to replace the θ by $\hat{\theta}$, a consistent estimator of θ, on the right side of (2). The result is called an empirical BLAP, or EBLAP, denoted by $\hat{\alpha}^*$.

3. BLAP of a zero-inflated random variable

A zero-inflated random variable, α, has a mixture distribution with one mixture component being 0 and the other mixture component being an a.s. nonzero random variable. Suppose that $\alpha = \delta\xi$, where δ is a binary random variable such that

$\mathrm{P}(\delta = 1) = p = 1 - \mathrm{P}(\delta = 0)$; ξ is a random variable such that $\mathrm{P}(\xi \neq 0) = 1$, and δ, ξ are independent. Then, α is a zero-inflated random variable with the nonzero component being ξ. A predictor, $\tilde{\alpha}$, is look-alike (with respect to α) if it is also zero-inflated.

To find the BLAP of α, note that the latter has two components: a binary component and a continuous one. Let us first focus on the binary component. Ideally, $\tilde{\alpha}$ should be zero whenever α is zero, and nonzero whenever α is nonzero. Denote $A = \{\tilde{\alpha} = 0\}$, $B = \{\alpha = 0\}$. Then, similar to the proof of Theorem 1, we have

$$\begin{aligned} \mathrm{P}(A \triangle B) &= \mathrm{P}(A \cap B^{c}) + \mathrm{P}(A^{c} \cap B) \\ &= \mathrm{P}(\tilde{\alpha} = 0, \alpha \neq 0) + \mathrm{P}(\tilde{\alpha} \neq 0, \alpha = 0) \\ &= \mathrm{E}\{1_{(\tilde{\alpha}=0)}\mathrm{P}(\delta = 1|Y)\} + \mathrm{E}\{1_{(\tilde{\alpha}\neq 0)}\mathrm{P}(\delta = 0|Y)\} \\ &= \mathrm{E}[\mathrm{P}(\delta = 1|Y) + \{\mathrm{P}(\delta = 0|Y) - \mathrm{P}(\delta = 1|Y)\}1_{(\tilde{\alpha}\neq 0)}] \\ &= \mathrm{P}(\delta = 1) + \mathrm{E}[\{\mathrm{P}(\delta = 0|Y) - \mathrm{P}(\delta = 1|Y)\}1_{(\tilde{\alpha}\neq 0)}]. \end{aligned} \tag{4}$$

where Y denotes the observed data. The last expression in (4) shows that to minimize $\mathrm{P}(A \triangle B)$ it suffices to allow $\tilde{\alpha} \neq 0$ whenever $\mathrm{P}(\delta = 0|Y) - \mathrm{P}(\delta = 1|Y) \leq 0$, that is,

$$\tilde{\alpha} \neq 0 \ \text{ iff } \ \mathrm{P}(\delta = 1|Y) \geq \frac{1}{2}. \tag{5}$$

It follows that any optimal $\tilde{\alpha}$ must have the expression

$$\tilde{\alpha} = 1_{\{\mathrm{P}(\delta=1|Y)\geq 1/2\}}\tilde{\xi}. \tag{6}$$

Recall that $\alpha = \delta\xi$ and, according to Corollary 1, the indicator function in (6) is the BLAP of δ. Therefore, $\tilde{\xi}$ corresponds to a predictor of ξ.

Now let us consider the continuous component, ξ. The following theorem states that the optimal $\tilde{\xi}$ in (6) is the BP of (not ξ but) α.

Theorem 2. *The optimal $\tilde{\xi}$, in the sense of minimizing the MSPE among all predictors satisfying (6), is $\tilde{\xi}^* = \mathrm{E}(\alpha|Y)$; therefore, the BLAP of α is given by*

$$\tilde{\alpha}^* = 1_{\{\mathrm{P}(\delta=1|Y)\geq 1/2\}}\mathrm{E}(\alpha|Y), \tag{7}$$

provided that the right side of (7) is computable.

Note. There is an interesting interpretation of (7): The BLAP of α is a product of the BLAP of δ, $1_{\{\mathrm{P}(\delta=1|Y)\geq 1/2\}}$, and the BP of α, $\mathrm{E}(\alpha|Y)$.

Proof of Theorem 2. We have, by (6),

$$\begin{aligned} \mathrm{E}(\tilde{\alpha} - \alpha)^2 &= \mathrm{E}\left[(\tilde{\xi} - \alpha)^2 1_{\{\mathrm{P}(\delta=1|Y)\geq 1/2\}} + \alpha^2 1_{\{\mathrm{P}(\delta=1|Y)<1/2\}}\right] \\ &= \mathrm{E}\left[1_{\{\mathrm{P}(\delta=1|Y)\geq 1/2\}}\mathrm{E}\{(\tilde{\xi} - \alpha)^2|Y\}\right] + \mathrm{E}[\alpha^2 1_{\{\mathrm{P}(\delta=1|Y)<1/2\}}]. \end{aligned}$$

Note that $\mathrm{E}\{(\tilde{\xi} - \alpha)^2|Y\} = \tilde{\xi}^2 - 2\tilde{\xi}\mathrm{E}(\alpha|Y) + \mathrm{E}(\alpha^2|Y) = \{\tilde{\xi} - \mathrm{E}(\alpha|Y)\}^2 + \mathrm{var}(\alpha|Y) \geq \mathrm{var}(\alpha|Y)$ with the equality holding if and only if $\tilde{\xi} = \mathrm{E}(\alpha|Y)$. Thus, we have

$$\mathrm{E}(\tilde{\alpha} - \alpha)^2 \geq \mathrm{E}[1_{\{\mathrm{P}(\delta=1|Y)\geq 1/2\}}\mathrm{var}(\alpha|Y)] + \mathrm{E}[\alpha^2 1_{\{\mathrm{P}(\delta=1|Y)<1/2\}}], \tag{8}$$

and the equality in (8) holds when $\tilde{\xi} = \mathrm{E}(\alpha|Y)$. Therefore, in conclusion, the BLAP of α is given by (7). This completes the proof.

We can extend the result of Theorem 2 to zero-inflated vector-valued random variable. This is defined as $\delta\xi$, where δ is the same as before but ξ is a random vector such that $\mathrm{P}(\xi = 0) = 0$, where the 0 inside the probability means the zero vector. A BLAP of α is defined as a predictor, $\tilde{\alpha}$, such that (i) it is zero-inflated vector-valued; (ii) it minimizes $\mathrm{P}(A \triangle B)$ of (4), where the 0 means zero vector; and (iii) it minimizes the MSPE, $\mathrm{E}(|\tilde{\alpha} - \alpha|^2)$, among all predictors satisfying (i) and (ii). By a similar argument, the following can be proved.

Theorem 3. *The BLAP of a zero-inflated vector-valued α is given by (7), where $\mathrm{E}(\alpha|Y)$ is the vector-valued conditional expectation, assumed computable.*

As in Section 2, if the right side of (7) involves a vector of unknown parameters, θ, an EBLAP is obtained by replacing θ by $\hat{\theta}$, a consistent estimator of θ.

The BLAP developed in this section has potential applications in many fields. For example, Jiang et al. (2016) considered misspecified mixed model analysis for genome-wide association study, in which the majority of the random effects are identical to zero. Another application of BLAP is considered in the next section.

4. Small area estimation with zero-inflated random effects

Datta et al. (2011) considered a Fay–Herriot model (Fay and Herriot, 1979) for small area estimation (e.g., Datta, 2009, Rao and Molina, 2015), in which the variance of the area-specific random effects is potentially zero. This is equivalent to that the random effects themselves are zero. Datta and Mandal (2015) further considered a Fay–Herriot model with zero-inflated random effects. This is different from Datta et al. (2011) in that some of the random effects may be zero, while the others are not. A slightly modified version of the Datta–Mandal model can be expressed as

$$Y_i = x_i'\beta + \alpha_i + \epsilon_i, \quad i = 1, \dots, m, \tag{9}$$

where x_i is a vector of known covariates, β is a vector of unknown parameters, α_i is an area-specific random effect, and ϵ_i is a sampling error. It is assumed that $\alpha_i = \delta_i v_i$ such that the following hold: (i) $\delta_i, v_i, \epsilon_i, i = 1, \dots, m$ are independent; (ii) δ_i is binary such that $\mathrm{P}(\delta_i = 1) = \rho = 1 - \mathrm{P}(\delta_i = 0)$, $\rho \in [0, 1]$ being an unknown probability; and (iii) $v_i \sim N(0, A)$, A being an unknown variance; and (iv) $\epsilon_i \sim N(0, D_i)$, D_i being a known sampling variance. Let $\theta = (\beta', A, \rho)$ denote the vector of unknown parameters. The BLAP developed in the previous section can be applied to prediction of the random effects, α_i. Detailed derivations can be found in the online supplement.

To use the BLAP in practice, we need to replace θ by a consistent estimator. Here we consider using the maximum likelihood estimator (MLE). Write $f_i(Y_i|\theta) = (1-\rho)\phi(Y_i - x_i'\beta, D_i) + \rho\phi(Y_i - x_i'\beta, A + D_i)$, where $\phi(\cdot, \sigma^2)$ denotes the pdf of $N(0, \sigma^2)$. The following theorem states that, under some regularity conditions, the MLE is consistent. Note that this result does not come as a consequence of the standard asymptotic theory for LMM (e.g., Jiang, 2007, sec. 1.8.3).

Theorem 4. *Suppose that (i)* $A > 0$, $D_i > 0$, $1 \le i \le m$; *(ii)*

$$\liminf_{m\to\infty} \lambda_{\min}\left[\frac{1}{m}\sum_{i=1}^{m} \mathrm{E}\left\{\frac{\partial^2}{\partial\theta\partial\theta'}\log f(Y_i|\theta)\bigg|_{\theta_0}\right\}\right] > 0,$$

where $\lambda_{\min}$ *denotes the smallest eigenvalue, and* θ_0 *the true* θ; *(iii) define* $\sigma_{i,st}(\theta) = (\partial^2/\partial\theta_s\partial\theta_t)\log f(Y_i|\theta)$, $1 \le s, t \le r = \dim(\theta)$, *then* $\lim_{m\to\infty} m^{-2}\sum_{i=1}^{m}\mathrm{E}\{\sigma_{i,st}^2(\theta_0)\} = 0$ *for all* s, t; *and (iv)* $m^{-1}\sum_{i=1}^{m}[\sigma_{i,st}(\theta) - \mathrm{E}\{\sigma_{i,st}(\theta)\}]$ *converges uniformly to zero in a neighborhood of* θ_0 *for all* s, t. *Then, with probability tending to one, there is* $\hat\theta$ *which is a solution to the likelihood equation,* $(\partial/\partial\theta)\log f(Y|\theta) = 0$, *such that* $\hat\theta \xrightarrow{\mathrm{P}} \theta_0$ *as* $m \to \infty$.

The proof of Theorem 4 follows by verifying the conditions of Theorem 2 of Foutz (1977). The detail is omitted. In the next section we consider an application of BLAP to a small area estimation problem.

5. Real-data examples

Our first real-data example is regarding thromboembolic or hemorrhagic complications (e.g., Glass et al., 1997), which occur in as many as 60% of patients who underwent extracorporeal membrane oxygenation (ECMO), an invasive technology used to support children during periods of reversible heart or lung failure (e.g., Muntean, 2002). Over half of pediatric patients on ECMO are currently receiving antithrombin (AT) to maximize heparin sensitivity. In a retrospective, multi-center, cohort study of children ($\le$ 18 years of age) who underwent ECMO between 2003 and 2012, 8601 subjects participated in 42 free-standing children's hospitals across 27 U.S. states and the District of Columbia known as Pediatric Health Information System (PHIS). Many of the outcome variables were binary, such as the bleed_binary variable (BB), which is a main outcome variable indicating hemorrhage complication of the treatment; and the DischargeMortalit1Flag variable (DM1F), which is associated with mortality. Here the treatment refers to AT. The data are also potentially clustered, with the clusters corresponding to the children's hospitals.

In addition to the treatment indicator, there were 20 other covariate variables, for which information were available. A step-wise variable selection procedure for logistic regression was applied that reduced the total number of covariate variables to 12, including 2 continuous variables and 10 dummy variables. The details about the variables are omitted due to space limit. The 12 variables were used as predictors for future outcomes associated with the two outcome variables, BB and DM1F, mentioned above, under a mixed logistic model for each outcome variable (see Example 1 in the Supplementary Material.). We compare BLAP with the standard logistic regression prediction (SLRP), which predicts the binary outcome using a similar decision rule as the BLAP, except without incorporating the random effects associated with the clusters, that is, hospitals. More specifically, the SLRP predicts the outcome as 1 is the estimated probability for the outcome equal to 1 is $\ge 1/2$; otherwise, the outcome is predicted as 0. The performance of BLAP and SLRP are compared using a delete-1 cross validation. Namely, we take out one subject each time, and use the rest of the data to make predictions of the deleted outcomes, using both BLAP and SLRP.

We obtained empirical probability of correct prediction (EPCP) by SLRP and BLAP for the 42 hospitals for both the BB outcome and the DM1F outcome. It is seen that BLAP performs better, overall, for both outcome variables. More specifically, not counting the ties, BLAP has higher EPCP for 25 out of 35 hospitals, or 71.4%, while SLRP has higher EPCP for 10 out of the 35 hospitals, or 28.6%, for the BB outcome. As for the DM1F outcome, not counting the ties, BLAP has higher EPCP for 24 out of 36 hospitals, or 66.7%, while SLRP has higher EPCP for 12 out of the 36 hospitals, or 33.3%. The empirical results are consistent with what the theory (Corollary 1) predicts regarding the superiority of the BLAP. The detailed results can be found in the online supplement.

Our second real-data example is regarding small area estimation. Datta et al. (2002) considered a data set regarding median income of four-person families for the fifty states of U.S. and the District of Columbia using cross-sectional and time series modeling. The primary source of data is the annual supplement to the March Sample of the Current Population Survey (CPS), which provides individual annual income data categorized into intervals of $2500. The direct survey estimates were obtained from the CPS using linear interpolation. Two secondary sources of data were also available. The first is the U.S. decennial censuses (Census) which produce median incomes for the 50 states and D.C. based on the "long form" filled out by approximately one-sixth of the U.S. population. These census median income estimates are believed to be free of sampling errors. The second is per-capita income estimates produced by the Bureau of Economic Analysis (BEA) division of the U.S. Department of Commerce. Since the per-capita income estimates are not based on any sampling techniques, they do not have any sampling errors associated with them. From the Census and BEA data, an adjusted census median income (adjusted Census) is obtained by multiplying the preceding census median income by the ratio of BEA per-capita income for the current year to that of the preceding census year.

As an illustration, we consider data collected in the year of 1982. Here, $y_i, i = 1, \dots, 51$ are the CPS data. Two covariates are considered. The first is adjusted Census, denoted by x_{1i}; the second is Census, denoted by x_{2i}. The CPS sampling variances, D_i, are available from the data. We adjust the scales by dividing y_i, x_{1i}, x_{2i} by 10^4, and D_i by 10^8. A Fay–Herriot model with zero-inflated random effects can be expressed as $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \alpha_i + e_i, \quad i = 1, \dots, 51$, where α_i, e_i satisfy the assumptions in Section 4 [below (9)]. Using the EBLAP method in Section 4, we obtain the MLE of the model parameters, $\theta = (\beta_0, \beta_1, \beta_2, A, \rho)$, and the EBLAPs of the random effects. From the latter, we obtain the EBLAPs of the small area means, $\mu_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \alpha_i$, as $\hat{\mu}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \hat{\alpha}_i$, where $\hat{\beta}_j$ is the MLE of $\beta_j, j = 0, 1, 2$, and $\hat{\alpha}_i$ is the EBLAP of α_i. As a comparison, we also computed the corresponding values of the standard empirical best linear unbiased predictors (EBLUPs; e.g., (Rao and Molina, 2015)) of the random effects and small area means, denoted by $\hat{\alpha}_i^*$ and $\hat{\mu}_i^*$, respectively. The EBLUPs are based on the Prasad–Rao estimate of A (Prasad and Rao, 1990).

It is seen that the EBLAPs of 25 out of the 51 state-specific random effects are exactly zero; in contrast, none of the EBLUPs of the random effects is exactly zero. It is also observed that, when both EBLAP and EBLUP are nonzero, the former tends to be larger in absolute value than the latter; and, when EBLAP is zero, the corresponding value of EBLUP tends to be small in absolute value. An interpretation is that, because EBLUP assumes that all of the random effects are nonzero, it has to "split" the random effect more "evenly" in order to match the overall for EBLAP. This can also be seen from the estimates of A under the two different models. In fact, the estimates of $\sqrt{A}$ under the Fay–Herriot model with zero-inflated random effects (corresponding to EBLAP) and under the Fay–Herriot model (corresponding to the EBLUP) are 0.925 and 0.155, respectively. The detailed results can be found in the online supplement.

Concluding remark. Random effects models are often used in practice to overcome limitations of regression or generalized linear models, or other fully-specified parametric models. The idea is to use the unspecified random effects to "capture" variations in the response that are not captured by the covariates through the parametric model. In fact, random effects models, or mixed effects models, have offered an alternative to nonparametric modeling. In this regard, the EBLAP method is useful in addressing problems of complex data structure by providing an intuitive and conceptually simple approach.

Acknowledgments

The research of Hanmei Sun and Yihui Luan was supported by the Natural Science Foundation of China Grants 11371227, 61432010 and 11626247. The research of Thuan Nguyen and Jiming Jiang are partially supported by the NSF grants DMS-1509557 and DMS-1512084, respectively. The authors are grateful to the Associate Editor and referees' constructive comments.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.spl.2018.07.014.

References

Datta, G.S., 2009. Model-based approach to small area estimation. In: Pfeffermann, D., Rao, C.R. (Eds.), HandBook of Statistics: Sample Surveys: Inference and Analysis, vol. 29B. North-Holland, pp. 251–288.

Datta, G.S., Hall, P., Mandal, A., 2011. Model selection by testing for the presence of small-area effects, and applications to area-level data. J. Amer. Statist. Assoc. 106, 361–374.

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. J. Statist. Plann. Inference 102, 83–97.

Datta, G.S., Mandal, A., 2015. Small area estimation with uncertain random effects. J. Amer. Statist. Assoc. 110, 1735–1744.

Fay, R.E., Herriot, R.A., 1979. Estimates of income for small places: an application of James–Stein procedures to census data. J. Amer. Statist. Assoc. 74, 269–277.

Foutz, R.V., 1977. On the unique consistent solution to the likelihood equations. J. Amer. Statist. Assoc. 72, 147–148.

Glass, P., Bulas, D.I., Wagner, A.E., et al., 1997. Severity of brain injury following neonatal extracorporeal membrane oxygenation and outcome at age 5 years. Dev. Med. Child Neurol. 39, 441–448.
Henderson, C.R., 1948. Estimation of General, Specific and Maternal Combining Abilities in Crosses Among Inbred Lines of Swine (Ph. D. Thesis), Iowa State Univ. Ames, Iowa.
Jiang, J., 2007. Linear and Generalized Linear Mixed Models and their Applications. Springer, New York.
Jiang, J., Lahiri, P., 2006. Mixed model prediction and small area estimation. TEST 15, 1–96.
Jiang, J., Li, C., Paul, D., Yang, C., Zhao, H., 2016. On high-dimensional misspecified mixed model analysis in genome-wide association study. Ann. Statist. 44, 2127–2160.
Muntean, W., 2002. Fresh frozen plasma in the pediatric age group and in congenital coagulation factor deficiency. Thromb. Res. 107, S29–S32 [0049-3848 (Print); 0049-3848 (Linking)].
Nurty, M.N., Devi, V.S., 2011. Pattern Recognition: An Algorithmic Approach. Springer-Verlag, London.
Prasad, N.G.N., Rao, J.N.K., 1990. The estimation of mean squared errors of small area estimators. J. Amer. Statist. Assoc. 85, 163–171.
Rao, J.N.K., Molina, I., 2015. Small Area Estimation, seocnd ed.. Wiley, New York.
Robinson, G.K., 1991. That BLUP is a good thing: The estimation of random effects (with discussion). Statist. Sci. 6, 15–51.