CIRJE-F-911

Parametric Transformed Fay-Herriot Model for Small Area Estimation

Shonosuke Sugasawa
Graduate School of Economics, The University of Tokyo

Tatsuya Kubokawa
The University of Tokyo

December 2013

Parametric Transformed Fay-Herriot Model for Small Area Estimation

Shonosuke Sugasawa* and Tatsuya Kubokawa†

University of Tokyo

December 23, 2013

Abstract

Consider the small area estimation when positive area-level data like income, revenue, harvests or production are available. Although a conventional method is the log-transformed Fay-Herriot model, the log-transformation is not necessarily appropriate. Another popular method is the Box-Cox transformation, but it has drawbacks that the maximum likelihood estimator (ML) of the transformation parameter is not consistent and that the transformed data are truncated.

In this paper, we consider parametric transformed Fay-Herriot models, and clarify conditions on transformations under which the ML estimator of the transformation is consistent. It is shown that the dual power transformation satisfies the conditions. Based on asymptotic properties for estimators of parameters, we derive a second-order approximation of the prediction error of the empirical best linear unbiased predictors (EBLUP) and obtain a second-order unbiased estimator of the prediction error. Finally, performances of the proposed procedures are investigated through simulation and empirical studies.

Key words and phrases: Asymptotically unbiased estimator, Box-Cox transformation, dual power transformation, Fay-Herriot model, linear mixed model, mean squared error, parametric bootstrap, small area estimation, variable selection.

1 Introduction

The linear mixed models (LMM) with both random and fixed effects have been extensively and actively studied from both theoretical and applied aspects in the literature. As specific normal linear mixed models, the Fay-Herriot model and the nested error regression models have been used in small-area estimation, where direct estimates like sample means for small areas have unacceptable estimation errors because sample sizes of small areas are small. Then the

*Graduate School of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN, E-Mail: shonosuke622@gmail.com

†Faculty of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN, E-Mail: tatsuya@e.u-tokyo.ac.jp

model-based shrinkage methods such as the empirical best linear unbiased predictor (EBLUP) have been utilized for providing reliable estimates for small-areas with higher precisions by borrowing data in the surrounding areas. For a good survey on this topic, see Ghosh and Rao (1994) and Rao (2003).

In this paper, we address the problem of transformation when positive area-level data like income, revenue, harvests or production are available. To handle such positive data, one usually makes the logarithmic transformation $\log(y)$ for positive y and applies log-transformed data to the Fay-Herriot model (Slud and Maiti, 2006). Although this approach based on the log-transformed Fay-Herriot model has been used as a standard method, the log-transformation is not necessarily appropriate. An alternative method is the Box-Cox power transformation (Box and Cox, 1964) given by

$$h^{BC}(y,\lambda) = \begin{cases} (y^{\lambda}-1)/\lambda, & \lambda \neq 0, \\ \log y, & \lambda = 0, \end{cases}$$

and one can make a flexible transformation for the positive data through estimation of the transformation parameter. However, we are faced with shortcomings that the maximum likelihood (ML) estimator of the transformation parameter λ is not consistent and that the transformed data are truncated. It is not necessarily suitable that one applies the truncated data to normally distributed models.

A feasible recipe for treating these issues is the dual power transformation suggested by Yang (2006), which is given by

$$h^{DP}(y,\lambda) = \begin{cases} (y^{\lambda}-y^{-\lambda})/2\lambda, & \lambda > 0, \\ \log y, & \lambda = 0. \end{cases}$$

This transformation possesses properties similar to the Box-Cox transformation, but does not suffer from the truncation problem, since it is a monotone function from $\mathbb{R}_+$ onto $\mathbb{R}$, where $\mathbb{R}$ and $\mathbb{R}_+$ denote the sets of real numbers and positive real numbers, respectively.

In this paper, we consider a class of parametric transformations including the dual power transformation and suggest a parametric transformed Fay-Herriot model. The model has three unknown parameters $\boldsymbol{\beta}$, A and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_q)'$ which are regression coefficients, variance of a random effect and the transformation parameter, respectively. Given $\boldsymbol{\lambda}$, the inference on the parameters $\boldsymbol{\beta}$ and A and the prediction can be reduced to the results given in the literature. Given $\boldsymbol{\lambda}$, $\boldsymbol{\beta}$ is estimated by the generalized least squares estimator $\widehat{\boldsymbol{\beta}}(\widehat{A}(\boldsymbol{\lambda}), \boldsymbol{\lambda})$ for estimator $\widehat{A}(\boldsymbol{\lambda})$ of A. For estimation of A, we have several methods, and well-known estimators $\widehat{A}(\boldsymbol{\lambda})$ are the maximum likelihood (ML) estimator, the restricted maximum likelihood (REML) estimator, the Fay-Herriot estimator and the Prasad-Rao estimator. The transformation parameter $\boldsymbol{\lambda}$ can be estimated through the likelihood functions based on $\widehat{\boldsymbol{\beta}}(\widehat{A}(\boldsymbol{\lambda}), \boldsymbol{\lambda})$ and $\widehat{A}(\boldsymbol{\lambda})$. Thus, the consistency of estimators $\widehat{\boldsymbol{\lambda}}$ of $\boldsymbol{\lambda}$ depends on the transformation $h(y,\boldsymbol{\lambda})$ and the estimator $\widehat{A}(\boldsymbol{\lambda})$. In this paper, we shall clarify conditions on $h(y,\boldsymbol{\lambda})$ and $\widehat{A}(\boldsymbol{\lambda})$ under which the estimator $\widehat{\boldsymbol{\lambda}}$ is consistent. Based on asymptotic properties of these estimators, we derive a second-order approximation of the mean squared error of the empirical best linear unbiased predictor (EBLUP) and its second-order unbiased estimator using the parametric bootstrap method.

The paper is organized as follows: In Section 2, the parametric transformed Fay-Herriot model is described and the assumptions on the transformation are provided. As a useful transformation, the dual power transformation is discussed. In Section 3, the methods for estimating the transformation parameter are given, and the consistency and some asymptotic properties which will be used for evaluation of the prediction error are shown. In Section 4, we provide EBLUP for the small-area estimation and derive a second-order approximation of the mean squared error (MSE) of EBLUP. It is interesting to note that the approximation of MSE consists of five terms: the first three terms are the known terms of the MSE approximation given $\boldsymbol{\lambda}$ and the last two terms come from randomness of estimation of $\boldsymbol{\lambda}$. We also derive a second-order unbiased estimator of the MSE using the parametric bootstrap method.

In Section 5, we investigate finite-sample performances of estimators of the parameters, MSE of EBLUP and estimators of MSE through simulation. The suggested procedures are applied to the data in the Survey of Family Income and Expenditure (SFIE) in Japan. The parametric transformation suggested in the paper has parameters for adjustment, which enables us to flexibly analyze the small-area positive data. In fact, the survey data treated there show that the estimate of the transformation parameter is far from the log-transformed model. The concluding remarks are given in Section 6, and the technical proofs are given in Appendix.

2 Parametric Transformed Fay-Herriot Models

2.1 Transformed Fay-Herriot models

Let $h(y,\boldsymbol{\lambda})$ be a monotone transformation from $\mathbb{R}_+$ to $\mathbb{R}$ for positive y. It is noted that the transformation involves unknown parameter $\boldsymbol{\lambda}=(\lambda_1,\ldots,\lambda_q)'$, where $\boldsymbol{\lambda}'$ denotes the transpose of $\boldsymbol{\lambda}$. It is assumed that positive data $y_1,\ldots,y_m$ are available, where y_i is an area-level data like a sample mean for the i-th small area. For $i=1,\ldots,m$, assume that the transformed observation $h(y_i,\boldsymbol{\lambda})$ has a linear mixed model suggested by Fay and Herriot (1979), given by

$$h(y_i,\boldsymbol{\lambda})=\boldsymbol{x}_i'\boldsymbol{\beta}+v_i+\varepsilon_i, \tag{1}$$

where $\boldsymbol{x}_i$ is a p-dimensional known vector, $\boldsymbol{\beta}$ is a p-dimensional unknown vector of regression coefficients, v_i is a random effect associated with the area i and ε_i is an error term. It is assumed that v_i, ε_i, $i=1,\ldots,m$, are mutually independently distributed as $v_i\sim\mathcal{N}(0,A)$ and $\varepsilon_i\sim\mathcal{N}(0,D_i)$, where A is an unknown common variance and $D_1,\ldots,D_m$ are known variances of the error terms. It is noted that this model is interpreted as a Bayesian model when the distribution of v_i is regarded as a prior distribution.

When $h(y,\boldsymbol{\lambda})=y$, the model (1) is called the Fay-Herriot model suggested by Fay and Herriot (1979). When y takes a value in real numbers $\mathbb{R}$, the Fay-Herriot model may be appropriate. Since y is positive, however, we need to transform y. A standard method is the logarithmic transformation $h(y,\boldsymbol{\lambda})=\log(y)$. Slud and Maiti (2006) studied the small-area estimation problem in the log-transformed Fay-Heriot model. However, the logarithmic transformation is not necessarily suitable.

Alternative method is parametric transformations, and the well known method is the Box-

Cox power transformation described by

$$h^{BC}(y,\lambda) = \begin{cases} (y^\lambda - 1)/\lambda, & \lambda \neq 0, \\ \log y, & \lambda = 0. \end{cases}$$

Although the Box-Cox transformation is very popular, it has drawbacks that the maximum likelihood estimator of λ is not consistent and that $h^{BC}(y,\lambda)$ is truncated as $h^{BC}(y,\lambda) \geq -1/\lambda$ for $\lambda > 0$ and $h^{BC}(y,\lambda) \leq -1/\lambda$ for $\lambda < 0$. This implies that the Box-Cox transformation is incompatible with the normality assumption due to the truncation problem, unless the transformation parameter λ equals to zero.

Taking it into account that transformed observations in (1) satisfy the normality assumption, in this paper, we consider a class of transformations satisfying Assumption 1 given below. For notational convenience, we use the notations described as

$$\begin{aligned} h_y(y,\boldsymbol{\lambda}) =& \frac{\partial h(y,\boldsymbol{\lambda})}{\partial y}, \quad h_\lambda(y,\boldsymbol{\lambda}) = \frac{\partial h(y,\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}}, \quad h_{\lambda\lambda}(y,\boldsymbol{\lambda}) = \frac{\partial^2 h(y,\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}\partial \boldsymbol{\lambda}'}, \\ h_{y\lambda}(y,\boldsymbol{\lambda}) =& \frac{\partial^2 h(y,\boldsymbol{\lambda})}{\partial y \partial \boldsymbol{\lambda}}, \quad h_{y\lambda\lambda}(y,\boldsymbol{\lambda}) = \frac{\partial^3 h(y,\boldsymbol{\lambda})}{\partial y \partial \boldsymbol{\lambda}\partial \boldsymbol{\lambda}'}. \end{aligned} \tag{2}$$

Assumption 1. The following are assumed for the transformation $h(y,\boldsymbol{\lambda})$:

(A.1) $h(y,\boldsymbol{\lambda})$ is an monotone function of y ($y > 0$) and its range is $\mathbb{R}$.

(A.2) The partial derivatives given in (2) exist and they are continuous.

(A.3) Transformation function $h(y,\boldsymbol{\lambda})$ satisfies following integrability conditions.

$$\begin{aligned} E\big[\{h(y,\boldsymbol{\lambda}) - \mu\} h_\lambda(y,\boldsymbol{\lambda})\big] = \boldsymbol{O}(1), \quad E\big[h_\lambda(y,\boldsymbol{\lambda})\big] = \boldsymbol{O}(1) \\ E\big[h_{\lambda\lambda}(y,\boldsymbol{\lambda})\big] = O(1), \quad E\Big[\frac{d}{d\boldsymbol{\lambda}}\Big(\frac{h_{y\lambda}(y,\boldsymbol{\lambda})}{h_y(y,\boldsymbol{\lambda})}\Big)\Big] = O(1), \end{aligned}$$

where $h(y,\boldsymbol{\lambda})$ is normally distributed with mean μ and variance σ^2, and the notation $\boldsymbol{O}(1)$ means that each component in $\boldsymbol{O}(1)$ is of order $O(1)$.

Assumption (A.1) means that the transformation is a one-to-one and onto function from $\mathbb{R}_+$ to $\mathbb{R}$. Clearly, (A.1) is not satisfied by the Box-Cox transformation, but by $\log(y)$. Assumptions (A.2) and (A.3) will be used to show consistency of estimators of $\boldsymbol{\lambda}$ and to evaluate asymptotically prediction errors of the empirical best linear unbiased predictors. A example of parametric transformations with two parameters λ_1 and λ_2 is the sifted dual power transformation given by

$$h^{SDP}(y,\lambda_1,\lambda_2) = \begin{cases} \left\{(y+\lambda_1)^{\lambda_2} - (y+\lambda_1)^{-\lambda_2}\right\}/2\lambda_2, & \lambda_2 > 0, \\ \log(y+\lambda_1), & \lambda_2 = 0. \end{cases}$$

Clearly, this transformation satisfies (A.1). In the case of $\lambda_1 = 0$, it is the dual power transformation which will be treated in the next subsection.

2.2 Dual power transformation

A parametric transformation treated as an example in this paper is the dual power transformation suggested by Yang (2006), which is described as

$$h^{DP}(y,\lambda) = \begin{cases} (y^{\lambda} - y^{-\lambda})/2\lambda, & \lambda > 0, \\ \log y, & \lambda = 0. \end{cases} \tag{3}$$

Although the Box-Cox transformation does not satisfy assumption (A.1), the dual power transformation satisfies (A.1). As shown later, the maximum likelihood estimator of λ for $h^{DP}(y,\lambda)$ is consistent, while the MLE of the transformation parameter in the Box-Cox transformation is not consistent. Thus, the dual power transformation is useful, and we provide some explanations on it.

It is noted that for $z = h^{DP}(y,\lambda)$, the inverse transformation is expressed as

$$y = \left(\lambda z + \sqrt{\lambda^2 z^2 + 1}\right)^{1/\lambda}$$

for $\lambda \neq 0$, and $y = e^z$ for $\lambda = 0$. Also note that some derivatives of $h^{DP}(y,\lambda)$ related to (A.2) are written as

$$\begin{aligned} h_{\lambda}^{DP}(y,\lambda) &= \frac{y^{\lambda} + y^{-\lambda}}{2\lambda} \log y + \frac{h^{DP}(y,\lambda)}{\lambda}, \quad h_{y}^{DP}(y,\lambda) = \frac{1}{2}(y^{\lambda-1} + y^{-\lambda-1}), \\ h_{y\lambda}^{DP}(y,\lambda) &= \frac{1}{2}\log y(y^{\lambda-1} - y^{-\lambda-1}), \quad h_{\lambda\lambda}^{DP}(y,\lambda) = h^{DP}(y,\lambda)(\log y)^2, \\ \frac{d}{d\lambda}\left(\frac{h_{y\lambda}^{DP}(y,\lambda)}{h_{y}^{DP}(y,\lambda)}\right) &= \frac{4(\log y)^2}{(y^{\lambda} + y^{-\lambda})^2}. \end{aligned}$$

It is clear that $h^{DP}(y,\lambda)$ $(\lambda > 0)$ satisfies (A.1) and (A.2), since it is continuous and differentiable. We here check whether the dual power transformation satisfies the integrability conditions in (A.3). Let $z (= h^{DP}(y,\lambda))$ be a random variable normally distributed with mean μ and variance σ^2. Then,

$$\begin{aligned} \left|E[h_{\lambda}^{DP}(y,\lambda)]\right| &= \frac{1}{\lambda^2}\left|E\left[\sqrt{1+\lambda^2 z^2}\log\left(\lambda z + \sqrt{1+\lambda^2 z^2}\right) + \lambda z\right]\right| \\ &< \frac{1}{\lambda^2}\left|E[(1+\lambda^2 z^2)(\lambda z + \lambda^2 z^2)]\right| + \frac{1}{\lambda}|E[z]| = O(1), \end{aligned}$$

$$\begin{aligned} &\left|E\left[\{h^{DP}(y,\lambda) - \mu\}h_{\lambda}^{DP}(y,\lambda)\right]\right| \\ &= \frac{1}{\lambda^2}\left|E\left[(z-\mu)\sqrt{1+\lambda^2 z^2}\log\left(\lambda z + \sqrt{1+\lambda^2 z^2}\right) + \lambda z(z-\mu)\right]\right| \\ &< \frac{1}{\lambda^2}\left|E\left[(z-\mu)(1+\lambda^2 z^2)(\lambda z + \lambda^2 z^2)\right]\right| + \frac{1}{\lambda}\left|E\left[z(z-\mu)\right]\right| = O(1), \end{aligned}$$

$$\begin{aligned} \left|E\left\{h_{\lambda\lambda}^{DP}(y,\lambda)\right\}\right| &= \left|E\left[h^{DP}(y,\lambda)(\log y)^2\right]\right| \\ &= \frac{1}{\lambda^2}\left|E\left[z\left\{\log\left(\lambda z + \sqrt{1+\lambda^2 z^2}\right)\right\}^2\right]\right| \\ &< E\left[|z|^3(1+\lambda z)^2\right] = O(1), \end{aligned}$$

and

$$0 < E\Big[\frac{d}{d\lambda}\Big(\frac{h_{y\lambda}^{DP}(y,\lambda)}{h_y^{DP}(y,\lambda)}\Big)\Big] = E\Big[\frac{4(\log y)^2}{(y^\lambda + y^{-\lambda})^2}\Big]$$
$$= E\Big[\frac{2}{\lambda^2\sqrt{1+\lambda^2 z^2}}\Big\{\log\big(\lambda z + \sqrt{1+\lambda^2 z^2}\big)\Big\}^2\Big]$$
$$< E\big[2z^2(1+\lambda z)^2\big] = O(1).$$

These evaluations show that the dual power transformation satisfies (A.3).

3 Consistent Estimators of Parameters

In this section, we derive consistent estimators of the parameters $\boldsymbol{\beta}$, A and $\boldsymbol{\lambda}$ in model (1). Especially, it is important how to obtain consistent estimators for the transformation parameter $\boldsymbol{\lambda}$ under Assumption 1. It is noted that the maximum likelihood estimator of the transformation parameter in the Box-Cox transformation is not consistent, since it has the truncation problem as mentioned before.

3.1 Estimation of $\boldsymbol{\beta}$ and A given $\boldsymbol{\lambda}$

We begin by estimating $\boldsymbol{\beta}$ and A when $\boldsymbol{\lambda}$ is given. In this case, the conventional procedures given in the literature for the Fay-Herriot model can be inherited to the transformed model. Thus, for given A and $\boldsymbol{\lambda}$, the maximum likelihood (ML) or generalized least square (GLS) estimator of β is

$$\widehat{\boldsymbol{\beta}}(A,\boldsymbol{\lambda}) = \Big\{\sum_{j=1}^m (A+D_j)^{-1}\boldsymbol{x}_j\boldsymbol{x}_j'\Big\}^{-1}\sum_{j=1}^m (A+D_j)^{-1}\boldsymbol{x}_j h(y_j,\boldsymbol{\lambda}). \tag{4}$$

Concerning estimation of A given $\boldsymbol{\lambda}$, we consider a class of estimators $\widehat{A}(\boldsymbol{\lambda})$ satisfying the following assumption:

Assumption 2. The following are assumed for the estimator $\widehat{A}(\boldsymbol{\lambda})$ of A:

(A.4) $\widehat{A}(\boldsymbol{\lambda}) = A + O_p(m^{-1/2})$, where the notation $\boldsymbol{O}_p(m^{-1/2})$ means that each component in $\boldsymbol{O}_p(m^{-1/2})$ is of order $O(m^{-1/2})$.

(A.5) $\partial\widehat{A}(\boldsymbol{\lambda})/\partial\boldsymbol{\lambda} = \boldsymbol{O}_p(1)$,

(A.6) $\partial\widehat{A}(\boldsymbol{\lambda})/\partial\boldsymbol{\lambda} - E\big[\partial\widehat{A}(\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}\big] = \boldsymbol{O}_p(m^{-1/2})$.

Assumption (A.1) implies that the estimator $\widehat{A}(\boldsymbol{\lambda})$ is consistent. Assumptions (A.2) and (A.3) will be used for approximating prediction errors of EBLUP. Let us define $\widehat{\boldsymbol{\beta}}(\boldsymbol{\lambda})$ by

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\lambda}) = \widehat{\boldsymbol{\beta}}(\widehat{A}(\boldsymbol{\lambda}),\boldsymbol{\lambda}),$$

which is provided by substituting $\boldsymbol{A}(\boldsymbol{\lambda})$ into $\widehat{\boldsymbol{\beta}}(A,\boldsymbol{\lambda})$ in (4). Asymptotic properties of $\widehat{\boldsymbol{\beta}}(\boldsymbol{\lambda})$ can be investigated under conditions on D_i and $\boldsymbol{x}_i$.

Assumption 3. The following are assumed for D_i and $\boldsymbol{x}_i$:

(A.7) $m^{-1}\sum_{j=1}^m \boldsymbol{x}_j\boldsymbol{x}_j'$ converges to a positive definite matrix.

(A.8) There exist constants $\underline{D}$ and $\overline{D}$ such that $\underline{D} \le D_i \le \overline{D}$ for $i = 1, \dots, m$, and $\underline{D}$ and $\overline{D}$ are positive constants independent of m.

Since $\widehat{\boldsymbol{\beta}}(A,\boldsymbol{\lambda}) \sim \mathcal{N}_p(\boldsymbol{\beta}, \{\sum_{j=1}^m (A+D_j)^{-1}\boldsymbol{x}_j\boldsymbol{x}_j'\}^{-1})$, it is clear that $\widehat{\boldsymbol{\beta}}(A,\boldsymbol{\lambda})$ is consistent and $\widehat{\boldsymbol{\beta}}(A,\boldsymbol{\lambda}) - \boldsymbol{\beta} = \boldsymbol{O}_p(m^{-1/2})$ under Assumption 3. Some asymptotic properties on $\widehat{\boldsymbol{\beta}}(\boldsymbol{\lambda}) = \widehat{\boldsymbol{\beta}}(\widehat{A}(\boldsymbol{\lambda}),\boldsymbol{\lambda})$ are given in the following lemma which will be proved in Appendix. This lemma will be used in Lemma 2 for showing the condition (A.6) for each estimator of A.

Lemma 1. *Assume the conditions* (A.4) *and* (A.5) *in Assumption 2 and Assumption 3. Then it holds that* $\widehat{\boldsymbol{\beta}}(\boldsymbol{\lambda}) - \boldsymbol{\beta} = O_p(m^{-1/2})$ *and*

$$\partial\widehat{\boldsymbol{\beta}}(\boldsymbol{\lambda})/\partial\boldsymbol{\lambda} - E\Big[\partial\widehat{\boldsymbol{\beta}}(\widehat{A}(\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}\Big] = \boldsymbol{O}_p(m^{-1/2}).$$

We here demonstrate that several estimators $\widehat{A}(\boldsymbol{\lambda})$ suggested in the literature satisfy Assumption 2. A simple moment estimator of A due to Prasad and Rao (1990) is given by

$$\widehat{A}_{PR}(\boldsymbol{\lambda}) = (m-p)^{-1}\Big\{\sum_{j=1}^m (h(y_j,\boldsymbol{\lambda}) - \boldsymbol{x}_j'\widehat{\boldsymbol{\beta}}^{OLS})^2 - \sum_{j=1}^m D_j\left\{1 - \boldsymbol{x}_j'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_j\right\}\Big\}, \tag{5}$$

where $\boldsymbol{X} = (\boldsymbol{x}_1,\dots,\boldsymbol{x}_m)'$, and $\boldsymbol{\beta}^{OLS}$ is the ordinary least squares (OLS) estimator $\widehat{\boldsymbol{\beta}}_{LS} = (\sum_{j=1}^m \boldsymbol{x}_j\boldsymbol{x}_j')^{-1}\sum_{j=1}^m \boldsymbol{x}_j h(y_j,\boldsymbol{\lambda})$. Another moment estimator due to Fay and Herriot (1979), denoted by $\widehat{A}_{FH}(\boldsymbol{\lambda})$, is given as the solution of the equation

$$\sum_{j=1}^m (A+D_j)^{-1}\left\{h(y_j,\boldsymbol{\lambda}) - \boldsymbol{x}_j'\widehat{\boldsymbol{\beta}}(A,\boldsymbol{\lambda})\right\}^2 = m-p. \tag{6}$$

The maximum likelihood estimator (ML) of A, denoted by $\widehat{A}_{ML}(\boldsymbol{\lambda})$, is obtained as the solution of the equation

$$\sum_{j=1}^m (A+D_j)^{-2}\left\{h(y_j,\boldsymbol{\lambda}) - \boldsymbol{x}_j'\widehat{\boldsymbol{\beta}}(A,\boldsymbol{\lambda})\right\}^2 = \sum_{j=1}^m (A+D_j)^{-1}. \tag{7}$$

Also the restricted maximum likelihood estimator (REML) of A, denoted by $\widehat{A}_{REML}(\boldsymbol{\lambda})$, is given as the solution of the equation

$$\sum_{j=1}^m \frac{\left\{h(y_j,\boldsymbol{\lambda}) - \boldsymbol{x}_j'\widehat{\boldsymbol{\beta}}(A,\boldsymbol{\lambda})\right\}^2}{(A+D_j)^2} = \sum_{j=1}^m \frac{1}{A+D_j} - \sum_{j=1}^m \frac{\boldsymbol{x}_j'\left\{\sum_{k=1}^m (A+D_k)^{-1}\boldsymbol{x}_k\boldsymbol{x}_k'\right\}^{-1}\boldsymbol{x}_j}{(A+D_j)^2}. \tag{8}$$

Then, it can be verified that the above four estimators satisfy Assumption 2. The proof will be given in Appendix.

Lemma 2. *Under Assumption 3, the estimators* $\widehat{A}_{PR}(\boldsymbol{\lambda})$, $\widehat{A}_{FH}(\boldsymbol{\lambda})$, $\widehat{A}_{ML}(\boldsymbol{\lambda})$ *and* $\widehat{A}_{REML}(\boldsymbol{\lambda})$ *satisfy Assumption 2.*

3.2 Estimation of transformation parameter $\boldsymbol{\lambda}$

We provide a consistent estimator of the transformation parameter $\boldsymbol{\lambda}$. For estimating $\boldsymbol{\lambda}$, we use the log-likelihood function, which is expressed as

$$L(\boldsymbol{\lambda}, A, \boldsymbol{\beta}) \propto -\frac{1}{2}\sum_{j=1}^{m}\log(A+D_j) - \frac{1}{2}\sum_{j=1}^{m}\frac{\{h(y,\boldsymbol{\lambda}) - \boldsymbol{x}_i'\boldsymbol{\beta}\}^2}{A+D_j} + \sum_{j=1}^{m}\log h_y(y_j,\boldsymbol{\lambda}). \tag{9}$$

The derivative with respect to $\boldsymbol{\lambda}$ is written as

$$\begin{aligned} F(\boldsymbol{\lambda}, A, \boldsymbol{\beta}) &\equiv \frac{\partial L(\boldsymbol{\lambda}, A, \boldsymbol{\beta})}{\partial \boldsymbol{\lambda}} \\ &= \sum_{j=1}^{m}\frac{h_{y\lambda}(y_j,\boldsymbol{\lambda})}{h_y(y_j,\boldsymbol{\lambda})} - \sum_{j=1}^{m}(A+D_j)^{-1}\{h(y_j,\boldsymbol{\lambda}) - \boldsymbol{x}_j\boldsymbol{\beta}\}\, h_\lambda(y_j,\boldsymbol{\lambda}). \end{aligned}$$

Then, we get an estimator $\widehat{\boldsymbol{\lambda}}$ as the solution of the equation:

$$F(\widehat{\boldsymbol{\lambda}}, \widehat{A}(\widehat{\boldsymbol{\lambda}}), \widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\lambda}})) = 0, \tag{10}$$

where $\widehat{A}(\boldsymbol{\lambda})$ is an estimator of A satisfying Assumption 2.

When $\widehat{A}(\boldsymbol{\lambda})$ is the ML estimator of A, the resulting estimator from (10) is the ML estimator of $\boldsymbol{\lambda}$, which implies that it has consistency and asymptotic normality under suitable conditions. For the variance component A, however, the Prasad-Rao, Fay-Herriot and REML estimators have been used in the literature instead of the ML of A. In this case, it is not necessarily guaranteed that the estimator derived from (10) is consistent. The following lemma shows that the estimator derived from (10) is consistent provided $\widehat{A}(\boldsymbol{\lambda})$ satisfies Assumption 2. The proof will be given in Appendix.

Lemma 3. *Let $\widehat{\boldsymbol{\lambda}}$ be the solution of* (10). *Then, $\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda} = O_p(m^{-1/2})$ and $E[\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}] = \boldsymbol{O}(m^{-1})$ under Assumptions 1 and 2.*

4 EBLUP and Evaluation of the Prediction Error

We now provide the empirical best linear unbiased predictor (EBLUP) for small-area estimation and evaluate asymptotically the prediction error of EBLUP. Since EBLUP includes the estimator of the transformation parameter in the transformed Fay-Herriot model, it is harder to evaluate the prediction error than in the non-transformed Fay-Herriot model. To this end, the asymptotic results derived in the previous sections are heavily used.

4.1 EBLUP

We here consider the problem of predicting $\eta_i = \boldsymbol{x}_i'\boldsymbol{\beta} + v_i$, which is the conditional mean of the transformed data given v_i, namely, $E[h(y_i,\boldsymbol{\lambda})|v_i]$. The best predictor of η_i corresponds to the Bayes estimator given by

$$\hat{\eta}_i^B(\boldsymbol{\beta}, A, \boldsymbol{\lambda}) = \boldsymbol{x}_i'\boldsymbol{\beta} + \frac{A}{A+D_i}\{h(y_i,\boldsymbol{\lambda}) - \boldsymbol{x}_i'\boldsymbol{\beta}\}. \tag{11}$$

Since $\boldsymbol{\beta}$, A and $\boldsymbol{\lambda}$ are unknown, we need to use the estimators suggested in Section 3. Substituting $\widehat{\boldsymbol{\beta}}(A,\boldsymbol{\lambda})$, given in (4), into $\hat{\eta}_i^B(\boldsymbol{\beta},A,\boldsymbol{\lambda})$ yields the estimator

$$\hat{\eta}_i^{EB0}(A,\boldsymbol{\lambda}) = \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}(A,\boldsymbol{\lambda}) + A(A+D_i)^{-1}\big\{h(y_i,\boldsymbol{\lambda}) - \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}(A,\boldsymbol{\lambda})\big\},$$

which is the best linear unbiased predictor (BLUP) as a function of $h(y_i,\boldsymbol{\lambda})$, $i=1,\ldots,m$.

For the parameters A and $\boldsymbol{\lambda}$, we can use the estimators $\widehat{A}(\widehat{\boldsymbol{\lambda}})$ and $\widehat{\boldsymbol{\lambda}}$ suggested in Section 3. Substituting their estimators into the BLUP, we get the empirical best linear unbiased predictor (EBLUP)

$$\hat{\eta}_i^{EB} = \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\lambda}}) + \frac{\widehat{A}(\widehat{\boldsymbol{\lambda}})}{\widehat{A}(\widehat{\boldsymbol{\lambda}})+D_i}\big\{h(y_i,\widehat{\boldsymbol{\lambda}}) - \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\lambda}})\big\}. \tag{12}$$

In the Bayesian context, it corresponds to the empirical Bayes estimator of η_i.

4.2 Second-order approximation of the prediction error

As prediction error of EBLUP, we employ the mean squared error (MSE) of $\hat{\eta}_i^{EB}$ defined as (12), which is defined as

$$\mathrm{MSE}_i(A,\boldsymbol{\lambda}) = E\big[(\hat{\eta}_i^{EB}-\eta_i)^2\big],$$

for $i=1,\ldots,m$. It is seen that the MSE can be decomposed as

$$\begin{aligned} E\big[(\hat{\eta}_i^{EB}-\eta_i)^2\big] &= E\big[(\hat{\eta}_i^{EB}-\hat{\eta}_i^{B})^2\big] + E\big[(\hat{\eta}_i^{B}-\eta_i)^2\big] \\ &= E\big[(\hat{\eta}_i^{EB}-\hat{\eta}_i^{EB1})^2\big] + 2E\big[(\hat{\eta}_i^{EB}-\hat{\eta}_i^{EB1})(\hat{\eta}_i^{EB1}-\hat{\eta}_i^{B})\big] \\ &\quad + E\big[(\hat{\eta}_i^{EB1}-\hat{\eta}_i^{B})^2\big] + E\big[(\hat{\eta}_i^{B}-\eta_i)^2\big], \end{aligned} \tag{13}$$

where

$$\hat{\eta}_i^{EB1} = \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}(\boldsymbol{\lambda}) + \frac{\widehat{A}(\boldsymbol{\lambda})}{\widehat{A}(\boldsymbol{\lambda})+D_i}\big\{h(y_i,\boldsymbol{\lambda}) - \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}(\boldsymbol{\lambda})\big\}.$$

It is noted that the first two terms in the r.h.s. of (13) are affected by estimation error of $\widehat{\boldsymbol{\lambda}}$, but the last two terms are not affected, namely, $E[(\hat{\eta}_i^{EB1}-\hat{\eta}_i^{B})^2]$ and $E[(\hat{\eta}_i^{B}-\eta_i)^2]$ do not depend on randomness of $\widehat{\boldsymbol{\lambda}}$. Thus, it follows from the well-known result in small area estimation (Datta, Rao and Smith (2005)) that under Assumption 3,

$$E[(\hat{\eta}_i^{EB1}-\hat{\eta}_i^{B})^2] + E[(\hat{\eta}_i^{B}-\eta_i)^2] = g_{1i}(A) + g_{2i}(A) + g_{3i}(A) + O(m^{-3/2}), \tag{14}$$

where $g_{1i}(A) = AD_i/(A+D_i)$, $g_{2i}(A) = D_i(A+D_i)^{-2}\boldsymbol{x}_i'\big(\sum_{j=1}^m \boldsymbol{x}_j\boldsymbol{x}_j'(A+D_j)^{-1}\big)^{-1}\boldsymbol{x}_i$ and $g_{3i}(A) = 2^{-1}D_i(A+D_i)^{-2}\mathrm{Var}(\widehat{A})$. Thus, we need to evaluate the first two terms.

Note that $\widehat{\boldsymbol{\lambda}}-\boldsymbol{\lambda} = \boldsymbol{O}_p(m^{-1/2})$ given in Lemma 3. Then, the first term can be approximated as

$$E[(\hat{\eta}_i^{EB}-\hat{\eta}_i^{EB1})^2] = E\Big[(\widehat{\boldsymbol{\lambda}}-\boldsymbol{\lambda})'\Big(\frac{\partial}{\partial\boldsymbol{\lambda}}\hat{\eta}_i^{EB1}\Big)\Big(\frac{\partial}{\partial\boldsymbol{\lambda}}\hat{\eta}_i^{EB1}\Big)'(\widehat{\boldsymbol{\lambda}}-\boldsymbol{\lambda})\Big] + O(m^{-3/2}).$$

To estimate this term, the following lemma is helpful.

Lemma 4. *Under Assumptions 1, 2 and 3, the derivative of* $\hat{\eta}_i^{EB1}$ *is approximated as*

$$\frac{\partial}{\partial \boldsymbol{\lambda}}\hat{\eta}_i^{EB1} = \boldsymbol{R}_{1i} + \boldsymbol{O}_p(m^{-1/2}),$$

where

$$\begin{aligned}\boldsymbol{R}_{1i} =& \frac{A}{A+D_i}h_\lambda(y_i,\boldsymbol{\lambda}) + \frac{D_i}{A+D_i}\boldsymbol{x}_i'\Big(\sum_{j=1}^m \frac{\boldsymbol{x}_j\boldsymbol{x}_j'}{A+D_j}\Big)^{-1}\sum_{j=1}^m \frac{\boldsymbol{x}_j}{A+D_j}E[h_\lambda(y_j,\boldsymbol{\lambda})]\\ &+ \frac{D_i}{(A+D_i)^2}\left\{h(y_i,\boldsymbol{\lambda}) - \boldsymbol{x}_i'\boldsymbol{\beta}\right\}\boldsymbol{r}(A),\end{aligned}$$

and $\boldsymbol{r}(A)$ *is a leading term of* $E\big[\partial \widehat{A}(\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}\big]$.

It follows from Lemma 4 that $E[(\hat{\eta}_i^{EB} - \hat{\eta}_i^{EB1})^2] = g_{4i}(A,\boldsymbol{\lambda}) + O(m^{-3/2})$, where

$$g_{4i}(A,\boldsymbol{\lambda}) = E\big[(\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})'\boldsymbol{R}_{1i}\boldsymbol{R}_{1i}'(\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})\big]. \tag{15}$$

For some specific cases, we can calculate values of $\boldsymbol{r}(A)$. For $\widehat{A}_{FH}(\boldsymbol{\lambda})$, $\widehat{A}_{ML}(\boldsymbol{\lambda})$ and $\widehat{A}_{REML}(\boldsymbol{\lambda})$, the values of $\boldsymbol{r}(A)$ are given by

$$\boldsymbol{r}(A) = \Big(\sum_{j=1}^m (A+D_j)^{-k}\Big)^{-1}\Big(\sum_{j=1}^m (A+D_j)^{-k}E\left[\left\{h(y_j,\boldsymbol{\lambda}) - \boldsymbol{x}_j'\boldsymbol{\beta}\right\}h_\lambda(y_j,\boldsymbol{\lambda})\right]\Big),$$

where $k=1$ corresponds to $\widehat{A}_{FH}(\boldsymbol{\lambda})$, and $k=2$ corresponds to $\widehat{A}_{ML}(\boldsymbol{\lambda})$ and $\widehat{A}_{REML}(\boldsymbol{\lambda})$. For $\widehat{A}_{PR}(\boldsymbol{\lambda})$, the value of $\boldsymbol{r}(A)$ is given by

$$\boldsymbol{r}(A) = \frac{2}{m-p}\sum_{j=1}^m E\left[\left\{h(y_j,\boldsymbol{\lambda}) - \boldsymbol{x}_j'\boldsymbol{\beta}\right\}h_\lambda(y_j,\boldsymbol{\lambda})\right].$$

For the second term, note that $\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda} = \boldsymbol{O}_p(m^{-1/2})$, $\widehat{A}(\boldsymbol{\lambda}) - A = O_p(m^{-1/2})$ and $\widehat{\boldsymbol{\beta}}(\boldsymbol{\lambda}) - \boldsymbol{\beta} = \boldsymbol{O}_p(m^{-1/2})$. Then it follows from Lemma 4 that

$$\begin{aligned}&2E[(\hat{\eta}_i^{EB} - \hat{\eta}_i^{EB1})(\hat{\eta}_i^{EB1} - \hat{\eta}_i^{B})] \qquad (16)\\ =&2E\Big[\Big(\frac{\partial}{\partial\boldsymbol{\lambda}}\hat{\eta}_i^{EB1}\Big)'(\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})\Big\{\Big(\frac{\partial\hat{\eta}_i^B}{\partial\boldsymbol{\beta}}\Big)'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + \Big(\frac{\partial\hat{\eta}_i^B}{\partial A}\Big)(\widehat{A} - A)\Big\}\Big] + O(m^{-3/2})\\ =&2E\Big[(\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})'\boldsymbol{R}_{1i}\Big(\frac{\partial\hat{\eta}_i^B}{\partial\boldsymbol{\beta}}\Big)'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\Big] + 2E\Big[\boldsymbol{R}_{1i}'\Big(\frac{\partial\hat{\eta}_i^B}{\partial A}\Big)(\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})(\widehat{A} - A)\Big] + O(m^{-3/2})\\ =&g_{5i}(A,\boldsymbol{\lambda}) + O(m^{-3/2}), \qquad (17)\end{aligned}$$

where

$$g_{5i}(A,\boldsymbol{\lambda}) = 2E[(\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})'\boldsymbol{R}_{i1}\boldsymbol{R}_{i2}'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})] + 2E[\boldsymbol{R}_{i1}R_{i3}(\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})(\widehat{A} - A)]$$

for

$$\boldsymbol{R}_{i2} = \frac{\partial\hat{\eta}_i^B}{\partial\boldsymbol{\beta}} = \frac{D_i}{A+D_i}\boldsymbol{x}_i', \quad R_{i3} = \frac{\partial\hat{\eta}_i^B}{\partial A} = \frac{D_i}{(A+D_i)^2}\left\{h(y_i,\boldsymbol{\lambda}) - \boldsymbol{x}_i'\boldsymbol{\beta}\right\}.$$

It is noted that $g_{4i}(A,\boldsymbol{\lambda})$ and $g_{5i}(A,\boldsymbol{\lambda})$ are of order $O(m^{-1})$ and that $g_{4i}(A,\boldsymbol{\lambda})$ and $g_{5i}(A,\boldsymbol{\lambda})$ generally cannot be expressed explicitly. Combining the above calculations gives the following theorem.

Theorem 1. *Under Assumptions 1, 2 and 3, the prediction error of EBLUP given in* (12) *is approximated as*

$$\mathrm{MSE}_i = g_{1i}(A) + g_{2i}(A) + g_{3i}(A) + g_{4i}(A, \boldsymbol{\lambda}) + g_{5i}(A, \boldsymbol{\lambda}) + O(m^{-3/2}),$$

where g_{ki}, $k = 1, \dots 5$ are defined in (14), (15) *and* (17).

4.3 Second-order unbiased estimator of the prediction error

For practical applications, we need to estimate the mean squared error of EBLUP. Although $g_{4i}(A, \boldsymbol{\lambda})$ and $g_{5i}(A, \boldsymbol{\lambda})$ are not expressed explicitly, we can provide their estimators using the parametric bootstrap method..

Corresponding to model (1), random variable z_i^* can be generated as $z_i^* = \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}} + v_i^* + \epsilon_i^*$ for $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}(\widehat{A}(\widehat{\boldsymbol{\lambda}}), \widehat{\boldsymbol{\lambda}})$, where v_i^*'s and ε_i^*'s are mutually independently distributed random errors such that $v_i^* | \boldsymbol{y} \sim \mathcal{N}(0, \widehat{A})$ and $\varepsilon_i^* \sim \mathcal{N}(0, D_i)$. Let us define y_i^* as the solution of the equation

$$z_i^* = h(y_i^*, \widehat{\boldsymbol{\lambda}}), \quad i = 1, \dots, m.$$

The estimators $\widehat{\boldsymbol{\lambda}}^*$, $\widehat{\boldsymbol{\beta}}^*$ and $\widehat{A}^*$ can be obtained from y_i^*, $i = 1, \dots, m$, by using the same manners as used in $\widehat{\boldsymbol{\lambda}}$, $\widehat{\boldsymbol{\beta}}$ and $\widehat{A}$.

Since $g_{2i}(A) + g_{3i}(A) = O(m^{-1})$, it is seen that $g_{2i}(\widehat{A}) + g_{3i}(\widehat{A})$ is a second order unbiased estimator of $g_{2i}(A) + g_{3i}(A)$, namely $E[g_{2i}(\widehat{A}) + g_{3i}(\widehat{A})] = g_{2i}(A) + g_{3i}(A) + O(m^{-3/2})$.

For estimation of $g_{1i}(A)$, $g_{1i}(\widehat{A})$ has a second-order bias, since $g_{1i}(A) = O(1)$. Thus, we need to correct the bias up to second order. By the Taylor series expansion of $g_{1i}(\widehat{A}(\widehat{\boldsymbol{\lambda}}))$, it is observed that

$$\begin{aligned} E\big[g_{1i}(\widehat{A}(\widehat{\boldsymbol{\lambda}}))\big] &= E\Big[g_{1i}(A) + \{\widehat{A}(\widehat{\boldsymbol{\lambda}}) - A\}\frac{d}{dA}g_{1i}(A)\Big] + O(m^{-1}) \\ &= g_{1i}(A) + E\big[\widehat{A}(\widehat{\boldsymbol{\lambda}}) - A\big]\frac{D_i^2}{(A + D_i)^2} + O(m^{-1}), \end{aligned}$$

and that

$$\begin{aligned} \widehat{A}(\widehat{\boldsymbol{\lambda}}) - A &= \widehat{A}(\boldsymbol{\lambda}) - A + (\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})\frac{\partial}{\partial \boldsymbol{\lambda}}\widehat{A}(\boldsymbol{\lambda}) + O_p(m^{-1}) \\ &= (\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})\Big\{\frac{\partial}{\partial \boldsymbol{\lambda}}\widehat{A}(\boldsymbol{\lambda}) - E\Big[\frac{\partial}{\partial \boldsymbol{\lambda}}\widehat{A}(\boldsymbol{\lambda})\Big]\Big\} + (\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})E\Big[\frac{\partial}{\partial \boldsymbol{\lambda}}\widehat{A}(\boldsymbol{\lambda})\Big] + O_p(m^{-1}). \end{aligned}$$

Then it follows from Assumption 2 and Lemma 3 that $E\big[\widehat{A}(\widehat{\boldsymbol{\lambda}}) - A\big] = O(m^{-1})$, which implies that

$$E\big[g_{1i}(\widehat{A}(\widehat{\boldsymbol{\lambda}}))\big] = g_{1i}(A) + b_i(A, \boldsymbol{\lambda}) + O(m^{-3/2}),$$

where $b_i(A, \boldsymbol{\lambda})$ is a bias with order $O(m^{-1})$. Hence, based on the parametric bootstrap, we get a second-order unbiased estimator of $g_{1i}(\widehat{A}(\widehat{\boldsymbol{\lambda}}))$ given by

$$\overline{g_{1i}}(\widehat{A}, \widehat{\boldsymbol{\lambda}}) = 2g_{1i}(\widehat{A}(\widehat{\boldsymbol{\lambda}})) - E^*\big[g_{1i}(\widehat{A}^*)|\boldsymbol{y}\big]. \tag{18}$$

In fact, it can be verified that $E[\overline{g_{1i}}(\widehat{A}, \widehat{\boldsymbol{\lambda}})] = g_{1i}(A) + O(m^{-3/2})$, since $E^*[g_{1i}(\widehat{A}^*)|\boldsymbol{y}] = g_{1i}(\widehat{A}(\widehat{\boldsymbol{\lambda}})) + b_i(\widehat{A}(\widehat{\boldsymbol{\lambda}}), \widehat{\boldsymbol{\lambda}}) + O_p(m^{-3/2})$.

For $g_{4i}(A, \boldsymbol{\lambda})$ and $g_{5i}(A, \boldsymbol{\lambda})$, their estimators based on the parametric bootstrap are given by

$$
\begin{aligned}
\overline{g_{4i}}(\widehat{A}, \widehat{\boldsymbol{\lambda}}) =& E_*\big[(\hat{\eta}_i^{EB*} - \hat{\eta}_i^{EB1*})^2 \big| \boldsymbol{y}\big], \\
\overline{g_{5i}}(\widehat{A}, \widehat{\boldsymbol{\lambda}}) =& 2E_*\big[(\hat{\eta}_i^{EB*} - \hat{\eta}_i^{EB1*})(\hat{\eta}_i^{EB1*} - \hat{\eta}_i^{B*}) \big| \boldsymbol{y}\big],
\end{aligned}
$$

where

$$
\begin{aligned}
\hat{\eta}_i^{B*} =& \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\lambda}}) + \frac{\widehat{A}(\widehat{\boldsymbol{\lambda}})}{\widehat{A}(\widehat{\boldsymbol{\lambda}}) + D_i}\big\{h(y_i^*, \widehat{\boldsymbol{\lambda}}) - \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\lambda}})\big\}, \\
\hat{\eta}_i^{EB1*} =& \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}^*(\widehat{\boldsymbol{\lambda}}) + \frac{\widehat{A}^*(\widehat{\boldsymbol{\lambda}})}{\widehat{A}^*(\widehat{\boldsymbol{\lambda}}) + D_i}\big\{h(y_i^*, \widehat{\boldsymbol{\lambda}}) - \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}^*(\widehat{\boldsymbol{\lambda}})\big\}, \\
\hat{\eta}_i^{EB*} =& \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}^*(\widehat{\boldsymbol{\lambda}}^*) + \frac{\widehat{A}^*(\widehat{\boldsymbol{\lambda}}^*)}{\widehat{A}^*(\widehat{\boldsymbol{\lambda}}^*) + D_i}\big\{h(y_i^*, \widehat{\boldsymbol{\lambda}}^*) - \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}^*(\widehat{\boldsymbol{\lambda}}^*)\big\}.
\end{aligned}
$$

Combining the above estimators yields the estimator of MSE_i given by

$$
\widehat{\mathrm{MSE}_i}^* = \overline{g_{1i}}(\widehat{A}, \widehat{\boldsymbol{\lambda}}) + g_{2i}(\widehat{A}) + g_{3i}(\widehat{A}) + \overline{g_{4i}}(\widehat{A}, \widehat{\boldsymbol{\lambda}}) + \overline{g_{5i}}(\widehat{A}, \widehat{\boldsymbol{\lambda}}). \tag{19}
$$

Theorem 2. *Under Assumptions 1, 2 and 3, $\widehat{\mathrm{MSE}_i}^*$ is a second order unbiased estimator of MSE_i, that is*

$$
E[\widehat{\mathrm{MSE}_i}^*] = \mathrm{MSE}_i + O\left(m^{-3/2}\right).
$$

5 Simulation and Empirical Studies

In this section, we investigate finite-sample performances of estimators of the parameters, MSE of EBLUP and estimators of MSE through simulation. We also applied the suggested procedures to the data in the Survey of Family Income and Expenditure (SFIE) in Japan.

5.1 Finite sample behaviors of estimators

We first investigate finite sample performances of the estimators of λ, A through simulation in the model (1) without covariates, namely $\boldsymbol{x}'\boldsymbol{\beta} = \mu$, where the transformation function is a dual power transformation given in (3). In the simulation implemented here, we set $\mu = 0$.

In the simulation experiments, we generate 100,000 data sets of $y_i = h^{-1}(v_i + \varepsilon_i, \lambda)$, $i = 1, \ldots, m$, where $m = 30$ and $\lambda = 0.2$, 0.6 and 1.0 to investigate performances of estimators. The random effect v_i is generated from $\mathcal{N}(0, 1)$ with $A = 1$, and the sampling error ε_i is generated from $\mathcal{N}(0, D_i)$. For D_i's, we treat the three patterns:

(a) $0.1, 0.2, 0.3, 0.4, 0.5$; (b) $0.1, 0.3, 0.5, 0.8, 1.0$; (c) $0.1, 0.4, 0.7, 1.1, 1.5$.

There are five groups $G_1, \ldots, G_5$ and six small areas in each group. The error variance D_i is common in the same group.

For estimation of A and λ, we use four methods of the maximum likelihood estimator (ML), restricted maximum likelihood estimator (REML), Prasad–Rao estimator (PR) and Fay–Herriot estimator (FH). We also apply the log-transformed model for the simulated data, which corresponds to the case of $\lambda = 0$ in the dual power transformation. For estimation A and μ of this model, we use the maximum likelihood method.

The average values of estimates of λ, A and μ are reported in Table 1. For small λ such as $\lambda = 0.2$, the logarithmic transformation gives slightly better estimates for A. For relatively large λ such as $\lambda = 0.6$ and $\lambda = 1.0$, the estimates of A in the logarithmic transformed case tends to underestimate and their performances are not as good as the performances in the parametric transformed case. Comparing the four methods for estimating A, we can see that the REML method gives the closest estimates for the true value of A. Concerning estimation of λ, the suggested estimator gives good estimates for $\lambda = 1$. Although it overestimates λ slightly for $\lambda = 0.2$, it seems to work as a whole.

5.2 Numerical properties of MSE and its estimators

We next investigate MSE of EBLUP $\hat{\eta}_i^{EB}$ and performances of estimators of MSE, where REML estimator is used for A. The simulation experiments are implemented in the same framework as used in the previous subsection. Let $\{Y_i^{(s)}, i = 1, \ldots, m\}$ be simulated data in the s-th replication for $s = 1, \ldots, S = 100,000$. Let $\hat{\eta}_i^{EB(s)}$ be EBLUP and let $\hat{\eta}_i^{B(s)}$ be the best predictor for the s-th replication. Also let $h(y_i^{(s)}, \widehat{\boldsymbol{\lambda}}^{(s)})$ be the direct predictor for the s-th replication. Then the true values of MSE of EBLUP and the direct predictor $h(y_i, \widehat{\boldsymbol{\lambda}})$ can be numerically obtained by

$$MSE(\hat{\eta}_i^{EB}) \approx S^{-1} \sum_{s=1}^{S} \left(\hat{\eta}_i^{EB(s)} - \hat{\eta}_i^{B(s)} \right)^2 + AD_i/(A + D_i),$$

$$MSE(h(y_i, \widehat{\boldsymbol{\lambda}})) \approx S^{-1} \sum_{s=1}^{S} \left(h(y_i^{(s)}, \widehat{\boldsymbol{\lambda}}^{(s)}) - \hat{\eta}_i^{B(s)} \right)^2 + AD_i/(A + D_i),$$

and their averages over six small areas within group G_i are denoted by $MSE_{EBLUP}(G_i)$ and $MSE_{DP}(G_i)$ for $i = 1, \ldots, 5$. The true values of $MSE_{EBLUP}(G_i)$ and the percentage relative gain in MSE defined by $100 \times \{MSE_{DP}(G_i) - MSE_{EBLUP}(G_i)\}/MSE_{DP}(G_i)$ are reported in Table 2, where values of percentage relative gain in MSE are given in parentheses. It is noted that EBLUP is a shrinkage predictor and $h(y_i, \widehat{\boldsymbol{\lambda}})$ is the non-shrinkage direct predictor. Thus, large values of the relative gain in MSE mean that the improvements of EBLUP over the direct predictor are large. Table 2 reveals that for all groups, the prediction error of EBLUP is smaller than that of the direct predictor. Especially, the improvement of EBLUP seems significant in G_3, G_4 and G_5. This implies that EBLUP works well still in the transformed Fay-Herriot model.

The averages of estimates of MSE are obtained based on 5,000 simulated datasets with 1,000 replication for bootstrap, where the estimator of MSE is given in (19). Then the bias and the

Table 1: Average values of the estimators of μ, A and λ for $m = 30$, $\mu = 0$, $A = 1$, D_i-patterns (a), (b) and (c). We used four types of estimators of A and log-transformed mode (the value of $\hat{\mu}$ is multiplied by 100)

	Pattern (a)			Pattern (b)			Pattern (c)		
	$\hat{\lambda}$	$\widehat{A}$	$\hat{\mu}$	$\hat{\lambda}$	$\widehat{A}$	$\hat{\mu}$	$\hat{\lambda}$	$\widehat{A}$	$\hat{\mu}$
$\lambda = 0.2$									
ML	0.40	1.25	-0.12	0.37	1.26	-0.22	0.35	1.24	0.23
REML	0.36	1.14	-0.12	0.34	1.14	-0.21	0.32	1.14	0.23
PR	0.41	1.32	-0.39	0.38	1.33	0.17	0.36	1.38	-0.18
FH	0.41	1.25	0.21	0.37	1.26	0.05	0.35	1.24	-0.26
log	—	0.90	-0.04	—	0.88	-0.10	—	0.86	0.04
$\lambda = 0.6$									
ML	0.72	1.26	-0.13	0.71	1.26	-0.21	0.70	1.26	0.24
REML	0.67	1.11	-0.13	0.67	1.12	-0.19	0.67	1.11	0.24
PR	0.72	1.31	-0.39	0.71	1.32	0.15	0.70	1.37	-0.20
FH	0.72	1.26	0.20	0.71	1.26	0.05	0.70	1.25	-0.20
log	—	0.66	-0.04	—	0.61	-0.06	—	0.58	0.08
$\lambda = 1.0$									
ML	1.12	1.30	-0.14	1.12	1.31	-0.20	1.11	1.31	0.24
REML	1.06	1.12	-0.15	1.07	1.13	-0.19	1.07	1.13	0.24
PR	1.10	1.24	-0.41	1.15	1.36	0.14	1.11	1.41	-0.20
FH	1.12	1.30	0.20	1.12	1.30	0.05	1.11	1.30	-0.30
log	—	0.46	-0.03	—	0.41	-0.03	—	0.38	0.11

relative bias of the MSE estimator are reported in Table 3. From this table, it seems that the MSE estimator gives good estimates for MSE of EBLUP although it tends to overestimate.

5.3 Application to the survey data

We now apply the suggested procedures to the data in the Survey of Family Income and Expenditure (SFIE) in Japan. In this study, we use the data of the disbursement item 'Education' in the survey in November 2011. The average disbursement (scaled by 10,000 Yen) at each capital city of 47 prefectures in Japan is obtained by y_i for $i = 1, \ldots, 47$, and each variance of D_i is appropriately calculated based on the data of the disbursement 'Education' at the same city every November in the past ten years. Although the average disbursements in SFIE are reported every month, the sample size are around 100 for most prefectures, and data of the item 'Education' have high variability. On the other hand, we have data in the National Survey of Family Income and Expenditure (NSFIE) for 47 prefectures. Since NSFIE is based on much larger sample than SFIE, the average disbursements in NSFIE are more reliable, but

Table 2: True values of MSE of EBLUP multiplied by 100 and percentage relative gain in MSE for $m = 30$, $\mu = 0$, $A = 1$ and D_i-patterns (a), (b) and (c) (values of percentage relative gain in MSE are given in parentheses).

	Pattern (a)			Pattern (b)			Pattern (c)		
λ	0.2	0.6	1.0	0.2	0.6	1.0	0.2	0.6	1.0
G_1	12.9	14.3	16.1	12.8	14.3	15.6	12.7	14.1	15.5
	(13.8)	(9.5)	(6.0)	(12.9)	(7.6)	(10.1)	(12.7)	(8.3)	(6.9)
G_2	21.2	23.0	25.0	28.6	31.2	32.9	35.2	37.7	40.0
	(20.8)	(16.3)	(13.5)	(25.5)	(19.5)	(19.2)	(28.7)	(24.5)	(22.9)
G_3	28.4	30.6	32.7	40.3	43.4	45.7	50.0	53.0	55.8
	(26.4)	(20.5)	(17.8)	(34.5)	(29.1)	(27.4)	(40.2)	(36.7)	(33.3)
G_4	34.6	36.9	39.3	53.2	56.5	59.0	62.9	66.3	68.9
	(31.1)	(26.0)	(22.9)	(44.8)	(39.6)	(37.8)	(51.4)	(48.2)	(45.8)
G_5	39.9	42.4	45.0	59.5	63.3	65.4	71.4	74.7	77.2
	(35.4)	(29.4)	(27.5)	(50.1)	(44.8)	(43.5)	(59.0)	(55.7)	(54.0)

Table 3: Average of estimates of MSE multiplied by 100 and their relative biases for $m = 30$, $\mu = 0$, $A = 1$ and D_i-patterns (a), (b) and (c) (percentage relative biases of MSE estimators are given in parentheses).

	Pattern (a)			Pattern (b)			Pattern (c)		
λ	0.2	0.6	1.0	0.2	0.6	1.0	0.2	0.6	1.0
G_1	15.1	16.5	18.7	15.1	16.6	18.8	14.5	16.3	18.9
	(17.2)	(16.0)	(16.4)	(18.2)	(16.1)	(20.4)	(15.8)	(15.5)	(22.0)
G_2	23.8	25.4	27.8	31.4	33.3	35.9	37.5	39.8	43.1
	(11.9)	(10.1)	(11.8)	(9.9)	(7.2)	(8.9)	(6.7)	(5.6)	(7.8)
G_3	31.2	32.9	35.6	43.8	45.7	48.3	52.6	55.1	58.7
	(9.9)	(7.4)	(8.8)	(8.5)	(5.4)	(5.7)	(5.4)	(3.9)	(5.3)
G_4	37.7	39.3	42.1	57.2	59.1	61.5	66.1	68.5	72.3
	(8.8)	(6.7)	(6.9)	(7.4)	(4.5)	(4.3)	(5.1)	(3.4)	(4.9)
G_5	43.2	44.9	47.7	63.9	65.7	67.8	75.1	77.4	81.2
	(8.2)	(5.8)	(5.8)	(7.4)	(3.8)	(3.6)	(5.2)	(3.5)	(5.2)

this survey has been implemented every five years. In this study, we use the data of the item 'Education' of NSFIE in 2009, which is denoted by X_i for $i = 1, \ldots, 47$. Thus, we apply the dual power transformed Fay-Herriot model (1) described in Section 2, that is

$$\frac{y_i^{\lambda} - y_i^{-\lambda}}{2\lambda} = \boldsymbol{x}_i'\boldsymbol{\beta} + v_i + \varepsilon_i, \qquad i = 1, \ldots, 47,$$

where $\boldsymbol{x}_i' = (1, X_i), \boldsymbol{\beta} = (\beta_1, \beta_2)'$.

We used the REML estimators for estimation of A and λ, and their estimates are $\hat{\lambda} = 1.44$ and $\widehat{A} = 0.11$. The GLS estimates of β_1 and β_2 are $\widehat{\boldsymbol{\beta}}_1 = -1.09$ and $\widehat{\boldsymbol{\beta}}_2 = 0.75$, so that the regression coefficient on X_i is positive, namely there is a positive correlation between y_i and X_i. The values of EBLUP in seven prefectures around Tokyo are reported in Table 4 with the estimates of their MSEs based on (19). Note that the estimate of λ is 1.44, which is far away from 0. This means that the logarithmic transformation does not seem appropriate for analyzing the data treated here.

It is interesting to investigate what happen when one uses the log-transformed model for the same data. The log-transformed model is

$$\log(y_i) = \boldsymbol{x}_i'\boldsymbol{\beta} + v_i + \varepsilon_i, \quad i = 1, \ldots, 47.$$

When the REML estimator is used for estimation of A and $\boldsymbol{\beta}$, their estimates are given by $\widehat{A} = 0.06$, $\widehat{\boldsymbol{\beta}}_1 = -0.90$ and $\widehat{\boldsymbol{\beta}}_2 = 0.61$. Note that the estimate of A in the log-transformed model is smaller than that in the dual power transformed model. Remember that $\widehat{A}$ determines the rate of shrinkage of y_i toward $\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}$, namely, the rate is larger as the value of $\widehat{A}$ is larger. Thus, y_i in the log-transformed model are not shrunken as much as in the dual power transformed model.

Table 4: Values of second EBLUP and their estimated MSE.

prefecture	D_i	$h(y_i, \hat{\lambda})$	$\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}$	$\hat{\eta}_i^{2EB}$	$\widehat{\text{MSE}}_i$
Ibaraki	0.112	-0.215	-0.161	-0.188	0.075
Tochigi	0.444	0.002	-0.158	-0.125	0.111
Gunma	0.110	-0.752	-0.092	-0.429	0.073
Saitama	0.056	0.213	0.461	0.294	0.058
Chiba	0.536	1.681	0.187	0.451	0.120
Tokyo	0.026	0.464	0.315	0.437	0.030
Kanagawa	0.188	1.068	0.235	0.551	0.097

6 Concluding Remarks

In this paper, we have suggested the parametric transformed Fay-Herriot model motivated from analysis of data with positive values like income, revenue, harvests or production. We have provided the estimation procedures for unknown parameters including the transformation parameter as well as the regression coefficients and the variance component. Their consistency and some other asymptotic properties have been shown and used to evaluate the prediction error of EBLUP asymptotically. It has been illustrated through simulation that the proposed procedures work in the estimation of the parameters, the performance of MSE of EBLUP and the estimation of MSE.

Although the well-known transformation is the Box-Cox transformation, it has drawbacks that the ML estimator of the transformation parameter is not consistent and that the transformed data are truncated. As an alternative method, the dual power transformation has been suggested, and it has been shown that all the results including consistency derived in this paper can be applied.

A conventional model for analyzing positive data is the log-transformed Fay-Herriot model. However, the logarithmic transformation is not necessarily appropriate. The parametric transformation suggested here has parameters for adjustment, which enables us to flexibly analyze the small-area positive data. In fact, the survey data treated in this paper show that the estimate of the transformation parameter is far from the log-transformed model.

Although the uncertainty of the predictor is measured in this paper based on the prediction error $E\{(\hat{\eta}_i^{EB} - \eta_i)^2\}$, it may be natural to measure the uncertainty with $E[\{h^{-1}(\hat{\eta}_i^{EB}, \hat{\lambda}) - h^{-1}(\eta_i, \hat{\lambda})\}^2]$, which is not tractable, however. Concerning this respect, it seems to us that MSE behaviors due to $E\{(\hat{\eta}_i^{EB} - \eta_i)^2\}$ are associated with performances due to $E[\{h^{-1}(\hat{\eta}_i^{EB}, \hat{\lambda}) - h^{-1}(\eta_i, \hat{\lambda})\}^2]$. Thus, we can guess a feature of basic performances of EBLUP in terms of $E[\{h^{-1}(\hat{\eta}_i^{EB}, \hat{\lambda}) - h^{-1}(\eta_i, \hat{\lambda})\}^2]$ from the behavior based on $E\{(\hat{\eta}_i^{EB} - \eta_i)^2\}$.

Acknowledgments.
Research of the second author was supported in part by Grant-in-Aid for Scientific Research (21540114 and 23243039) from Japan Society for the Promotion of Science.

Appendix

A.1 Proof of Lemma 1. Since it can be easily seen that $\widehat{\boldsymbol{\beta}}(\widehat{A}(\boldsymbol{\lambda}), \boldsymbol{\lambda}) - \boldsymbol{\beta} = O_p(m^{-1/2})$, we here give the proof of the second part. Straightforward calculation shows that

$$\frac{\partial \widehat{\boldsymbol{\beta}}(\widehat{A}(\boldsymbol{\lambda}), \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} = \Big(\sum_{j=1}^m \frac{\boldsymbol{x}_j \boldsymbol{x}_j'}{\widehat{A}(\boldsymbol{\lambda}) + D_j}\Big)^{-1} \sum_{j=1}^m \frac{\boldsymbol{x}_j \boldsymbol{x}_j'}{(\widehat{A}(\boldsymbol{\lambda}) + D_j)^2} (\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^*) \Big(\frac{\partial \widehat{A}(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}}\Big) + \Big(\sum_{j=1}^m \frac{\boldsymbol{x}_j \boldsymbol{x}_j'}{\widehat{A} + D_j}\Big)^{-1} \sum_{j=1}^m \frac{\boldsymbol{x}_j}{\widehat{A} + D_j} h_\lambda(y_j, \boldsymbol{\lambda}), \tag{20}$$

where

$$\widehat{\boldsymbol{\beta}}^* = \Big\{\sum_{j=1}^m \frac{\boldsymbol{x}_j \boldsymbol{x}_j'}{(\widehat{A} + D_j)^2}\Big\}^{-1} \sum_{j=1}^m \frac{\boldsymbol{x}_j}{(\widehat{A} + D_j)^2} h(y_j, \boldsymbol{\lambda}). \tag{21}$$

Since $\widehat{\boldsymbol{\beta}}^* - \boldsymbol{\beta} = \boldsymbol{O}_p(m^{-1/2})$, it is seen that

$$\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^* = \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} - (\widehat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}) = \boldsymbol{O}_p(m^{-1/2}).$$

Thus from Assumption 2, the expectation of the first term in (20) is $\boldsymbol{O}(m^{-1/2})$. For the second term in (20), we have

$$E\Big[\Big(\sum_{j=1}^m \frac{\boldsymbol{x}_j \boldsymbol{x}_j'}{\widehat{A} + D_j}\Big)^{-1} \sum_{j=1}^m \frac{\boldsymbol{x}_j}{\widehat{A} + D_j} h_\lambda(y_j, \boldsymbol{\lambda})\Big] = \Big(\sum_{j=1}^m \frac{\boldsymbol{x}_j \boldsymbol{x}_j'}{A + D_j}\Big)^{-1} \sum_{j=1}^m \frac{\boldsymbol{x}_j}{A + D_j} E[h_\lambda(y_j, \boldsymbol{\lambda})] + \boldsymbol{O}(m^{-1/2}),$$

where the order of the leading term of the last formula is $\boldsymbol{O}_p(1)$. Then,

$$E[\partial\widehat{\boldsymbol{\beta}}(\widehat{A}(\boldsymbol{\lambda}),\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}] = \Big(\sum_{j=1}^m \frac{\boldsymbol{x}_j\boldsymbol{x}_j'}{A+D_j}\Big)^{-1}\sum_{j=1}^m \frac{\boldsymbol{x}_j}{A+D_j}E[h_\lambda(y_j,\boldsymbol{\lambda})] + \boldsymbol{O}(m^{-1/2}). \tag{22}$$

Therefore we obtain

$$\begin{aligned}
&\sqrt{m}\Big\{\partial\widehat{\boldsymbol{\beta}}(\widehat{A}(\boldsymbol{\lambda}),\boldsymbol{\lambda})/\partial\boldsymbol{\lambda} - E[\partial\widehat{\boldsymbol{\beta}}(\widehat{A}(\boldsymbol{\lambda}),\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}]\Big\}\\
&= \Big(\frac{1}{m}\sum_{j=1}^m \frac{\boldsymbol{x}_j\boldsymbol{x}_j'}{A+D_j}\Big)^{-1}\Big(\frac{1}{m}\sum_{j=1}^m \frac{\boldsymbol{x}_j\boldsymbol{x}_j'}{(A+D_j)^2}\Big)\sqrt{m}(\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\beta}}^*)\Big(\frac{\partial\widehat{A}(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}}\Big)\\
&+ \Big(\frac{1}{m}\sum_{j=1}^m \frac{\boldsymbol{x}_j\boldsymbol{x}_j'}{A+D_j}\Big)^{-1}\frac{1}{\sqrt{m}}\sum_{j=1}^m \frac{\boldsymbol{x}_j}{A+D_j}\Big\{h_\lambda(y_j,\boldsymbol{\lambda}) - E[h_\lambda(y_j,\boldsymbol{\lambda})]\Big\} + \boldsymbol{O}_p(1).
\end{aligned} \tag{23}$$

Since $\partial\widehat{A}(\boldsymbol{\lambda})/\partial\boldsymbol{\lambda} = \boldsymbol{O}_p(1)$ from (A.5) in Assumption 2, the first term in (23) has $\boldsymbol{O}_p(1)$. For the second term in (23), from the central limit theorem, we have

$$\frac{1}{\sqrt{m}}\sum_{j=1}^m \frac{\boldsymbol{x}_j}{A+D_j}\Big\{h_\lambda(y_j,\boldsymbol{\lambda}) - E\big[h_\lambda(y_j,\boldsymbol{\lambda})\big]\Big\} = \boldsymbol{O}_p(1),$$

which, together with Assumption 3, implies that the second term in (23) is of order $\boldsymbol{O}_p(1)$. Therefore we can conclude that $\partial\widehat{\boldsymbol{\beta}}(\widehat{A}(\boldsymbol{\lambda}),\boldsymbol{\lambda})/\partial\boldsymbol{\lambda} - E[\partial\widehat{\boldsymbol{\beta}}(\widehat{A}(\boldsymbol{\lambda}),\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}] = \boldsymbol{O}_p(m^{-1/2})$. □

A.2 Proof of Lemma 2. It is clear that the condition (A.4) is satisfied for the estimators of A from the results given in the literature, so that we shall verify the conditions (A.5) and (A.6) of Assumption 2.

For $\widehat{A}_{PR}$ defined in (5), it is seen that

$$\begin{aligned}
\frac{\partial\widehat{A}_{PR}(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}} &= \frac{2}{m-p}\sum_{j=1}^m \{h(y_j,\boldsymbol{\lambda}) - \boldsymbol{x}_j'\boldsymbol{\beta}\}h_\lambda(y_j,\boldsymbol{\lambda}) - \frac{2}{m-p}\sum_{j=1}^m \boldsymbol{x}_j'(\widehat{\boldsymbol{\beta}}^{OLS}-\boldsymbol{\beta})h_\lambda(y_j,\boldsymbol{\lambda})\\
&- \frac{2}{m-p}\sum_{j=1}^m \{h(y_j,\boldsymbol{\lambda}) - \boldsymbol{x}_j'\boldsymbol{\beta}\}\boldsymbol{x}_j'\Big(\frac{\partial\widehat{\boldsymbol{\beta}}^{OLS}}{\partial\boldsymbol{\lambda}}\Big) + \frac{2}{m-p}\sum_{j=1}^m \boldsymbol{x}_j'(\widehat{\boldsymbol{\beta}}^{OLS}-\boldsymbol{\beta})\boldsymbol{x}_j'\Big(\frac{\partial\widehat{\boldsymbol{\beta}}^{OLS}}{\partial\boldsymbol{\lambda}}\Big),
\end{aligned}$$

and that

$$\frac{\partial\widehat{\boldsymbol{\beta}}^{OLS}}{\partial\boldsymbol{\lambda}} = \Big(\frac{1}{m}\sum_{j=1}^m \boldsymbol{x}_j\boldsymbol{x}_j'\Big)^{-1}\frac{1}{m}\sum_{j=1}^m \boldsymbol{x}_j h_\lambda(y_j,\boldsymbol{\lambda})' = O_p(1)$$

by the law of large numbers. Since $\widehat{\boldsymbol{\beta}}^{OLS} - \boldsymbol{\beta} = \boldsymbol{O}_p(m^{-1/2})$, we have $\partial\widehat{A}_{PR}(\boldsymbol{\lambda})/\partial\boldsymbol{\lambda} = \boldsymbol{O}_p(1)$, which shows (A.5). For (A.6), note that

$$E\Big[\frac{\partial\widehat{A}_{PR}(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}}\Big] = \frac{2}{m-p}\sum_{j=1}^m E\big[\{h(y_j,\boldsymbol{\lambda}) - \boldsymbol{x}_j'\boldsymbol{\beta}\}h_\lambda(y_j,\boldsymbol{\lambda})\big] + O(m^{-1/2}). \tag{24}$$

Then, it is observed that

$$\sqrt{m}\Big\{\frac{\partial\widehat{A}_{PR}(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}} - E\Big[\frac{\partial\widehat{A}_{PR}(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}}\Big]\Big\} = \frac{2}{m-p}\sum_{j=1}^m Z_j + O_p(m^{-1/2})$$

where

$$Z_j = \{h(y_j, \boldsymbol{\lambda}) - \boldsymbol{x}_j'\boldsymbol{\beta}\}h_\lambda(y_j, \boldsymbol{\lambda}) - E\big[\{h(y_j, \boldsymbol{\lambda}) - \boldsymbol{x}_j'\boldsymbol{\beta}\}h_\lambda(y_j, \boldsymbol{\lambda})\big]. \tag{25}$$

Since it is clear that $E[Z_j] = 0$, $j = 1, \dots, m$, and $Z_1, \dots, Z_j$ are independent, by central limit theorem, we have

$$\sqrt{m}\Big\{\frac{\partial \widehat{A}_{PR}(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} - E\Big[\frac{\partial \widehat{A}_{PR}(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}}\Big]\Big\} = O_p(1),$$

which shows (A.6), and Assumption 2 is satisfied for $\widehat{A}_{PR}$.

We next show Lemma 2 for $\widehat{A}_{FH}$, $\widehat{A}_{ML}$ and $\widehat{A}_{REML}$. For the proofs, we first demonstrate that the condition (A.5) is satisfied. Then we can use Lemma 1, which is guaranteed under (A.4), (A.5) and Assumption 3. Using Lemma 1, we next show the condition (A.6) for the estimators.

Since $\widehat{A}_{FH}$, $\widehat{A}_{ML}$ and $\widehat{A}_{REML}$ are defined as the solutions of the equations (6), (7) and (8), it follows from the implicit function theorem that

$$\frac{\partial}{\partial \boldsymbol{\lambda}}\widehat{A}(\boldsymbol{\lambda}) = -\frac{G_\lambda(\boldsymbol{\lambda}, \widehat{A})}{G_A(\boldsymbol{\lambda}, \widehat{A})}, \tag{26}$$

where $G(\boldsymbol{\lambda}, A) = 0$ is the equation defining the estimator of A, and

$$G_\lambda(\boldsymbol{\lambda}, \widehat{A}) = \frac{\partial}{\partial \boldsymbol{\lambda}} G(\boldsymbol{\lambda}, A)\Big|_{A=\widehat{A}}, \qquad G_A(\boldsymbol{\lambda}, \widehat{A}) = \frac{\partial}{\partial A} G(\boldsymbol{\lambda}, A)\Big|_{A=\widehat{A}}$$

For $\widehat{A}_{FH}$, $\widehat{A}_{ML}$ and $\widehat{A}_{REML}$, the function $G_\lambda(\boldsymbol{\lambda}, \widehat{A})$ has the following form:

$$\begin{aligned}
&G_\lambda(\boldsymbol{\lambda}, \widehat{A})\\
&= \frac{\partial}{\partial \boldsymbol{\lambda}}\Big\{\sum_{j=1}^m (\widehat{A} + D_j)^{-k}\{h(y_j, \boldsymbol{\lambda}) - \boldsymbol{x}_j'\widehat{\boldsymbol{\beta}}\}^2\Big\}\\
&= 2\sum_{j=1}^m (\widehat{A} + D_j)^{-k}\{h(y_j, \boldsymbol{\lambda}) - \boldsymbol{x}_j'\boldsymbol{\beta}\}h_\lambda(y_j, \boldsymbol{\lambda}) - 2\sum_{j=1}^m (\widehat{A} + D_j)^{-k}\boldsymbol{x}_j'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})h_\lambda(y_j, \boldsymbol{\lambda})\\
&\quad - 2\sum_{j=1}^m (\widehat{A} + D_j)^{-k}\{h(y_j, \boldsymbol{\lambda}) - \boldsymbol{x}_j'\boldsymbol{\beta}\}\boldsymbol{x}_j'\Big(\frac{\partial}{\partial \boldsymbol{\lambda}}\widehat{\boldsymbol{\beta}}\Big) + 2\sum_{j=1}^m (\widehat{A} + D_j)^{-k}\boldsymbol{x}_j'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\boldsymbol{x}_j'\Big(\frac{\partial}{\partial \boldsymbol{\lambda}}\widehat{\boldsymbol{\beta}}\Big),
\end{aligned} \tag{27}$$

where the case of $k = 1$ corresponds to $\widehat{A}_{FH}$, and the case of $k = 2$ corresponds to $\widehat{A}_{ML}$ and $\widehat{A}_{REML}$. Note that $\partial\widehat{\boldsymbol{\beta}}/\partial\boldsymbol{\lambda}$ is expressed as (20), which contains $\partial\widehat{A}/\partial\boldsymbol{\lambda}$. Then solving (26) on $\partial\widehat{A}/\partial\boldsymbol{\lambda}$, we have

$$\frac{\partial \widehat{A}(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} = \frac{I_1(\boldsymbol{y})}{I_2(\boldsymbol{y})}, \tag{28}$$

where

$$\begin{aligned}
I_1(\boldsymbol{y}) &= 2\sum_{j=1}^m (\widehat{A} + D_j)^{-k}\{h(y_j, \boldsymbol{\lambda}) - \boldsymbol{x}_j'\boldsymbol{\beta}\}h_\lambda(y_j, \boldsymbol{\lambda}) - 2\sum_{j=1}^m (\widehat{A} + D_j)^{-k}\boldsymbol{x}_j'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})h_\lambda(y_j, \boldsymbol{\lambda})\\
&\quad - 2\sum_{j=1}^m (\widehat{A} + D_j)^{-k}\{h(y_j, \boldsymbol{\lambda}) - \boldsymbol{x}_j'\boldsymbol{\beta}\}\boldsymbol{x}_j' \cdot J_1(\boldsymbol{y}) + 2\sum_{j=1}^m (\widehat{A} + D_j)^{-k}\boldsymbol{x}_j'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\boldsymbol{x}_j' \cdot J_1(\boldsymbol{y}),
\end{aligned}$$

and

$$I_2(\boldsymbol{y}) = -G_A(\boldsymbol{\lambda}, \widehat{A}) + 2\sum_{j=1}^{m}(\widehat{A} + D_j)^{-k}\{h(y_j, \boldsymbol{\lambda}) - \boldsymbol{x}_j'\boldsymbol{\beta}\}\boldsymbol{x}_j \cdot J_2(\boldsymbol{y})$$
$$- 2\sum_{j=1}^{m}(\widehat{A} + D_j)^{-k}\boldsymbol{x}_j'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\boldsymbol{x}_j' \cdot J_2(\boldsymbol{y}),$$

for $J_1(\boldsymbol{y})$ and $J_2(\boldsymbol{y})$ defined as

$$\partial\widehat{\boldsymbol{\beta}}(\widehat{A}(\boldsymbol{\lambda}), \boldsymbol{\lambda})/\partial\boldsymbol{\lambda} = J_2(\boldsymbol{y})\Big(\frac{\partial\widehat{A}(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}}\Big) + J_1(\boldsymbol{y}),$$

which is obtained from (20). Using the expression of (28), we show that $\partial\widehat{A}(\boldsymbol{\lambda})/\partial\boldsymbol{\lambda} = \boldsymbol{O}_p(1)$, which is sufficient to verify that $I_1(\boldsymbol{y})/m = \boldsymbol{O}_p(1)$ and $I_2(\boldsymbol{y})/m = O_p(1)$. For this purpose, the following facts are useful:

$$\frac{1}{m}\sum_{j=1}^{m}(\widehat{A} + D_j)^{-k}\{h(y_j, \boldsymbol{\lambda}) - \boldsymbol{x}_j'\boldsymbol{\beta}\}h_\lambda(y_j, \boldsymbol{\lambda}) = \boldsymbol{O}_p(1), \tag{29}$$

$$\frac{1}{m}\sum_{j=1}^{m}(\widehat{A} + D_j)^{-k}\boldsymbol{x}_j'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})h_\lambda(y_j, \boldsymbol{\lambda}) = \boldsymbol{O}_p(m^{-1/2}), \tag{30}$$

$$\frac{1}{m}\sum_{j=1}^{m}(\widehat{A} + D_j)^{-k}\{h(y_j, \boldsymbol{\lambda}) - \boldsymbol{x}_j'\boldsymbol{\beta}\}\boldsymbol{x}_j' = \boldsymbol{O}_p(m^{-1/2}), \tag{31}$$

$$\frac{1}{m}\sum_{j=1}^{m}(\widehat{A} + D_j)^{-k}\boldsymbol{x}_j'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\boldsymbol{x}_j' = \boldsymbol{O}_p(m^{-1/2}), \tag{32}$$

where $k = 0, 1, 2$. These facts can be verified by noting that $\widehat{A} - A = O_p(m^{-1/2})$, $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \boldsymbol{O}_p(m^{-1/2})$ and using the law of large numbers and the central limit theorem. Now we consider the order of $I_1(\boldsymbol{y})/m$ and $I_2(\boldsymbol{y})/m$. From the proof of Lemma 1, we have $J_1(\boldsymbol{y}) = O_p(1)$ and $J_2(\boldsymbol{y}) = O_p(m^{-1/2})$. If we assume that $m^{-1}G_A(\boldsymbol{\lambda}, \widehat{A}) = O_p(1)$ (this is actually proved for each estimators in the end of the proof), it is immediate from (31) and (32) that

$$I_2(\boldsymbol{y})/m = O_p(1).$$

Similarly from (29)∼(32), we have

$$I_1(\boldsymbol{y})/m = O_p(1),$$

and we obtain $\partial\widehat{A}(\boldsymbol{\lambda})/\partial\boldsymbol{\lambda} = \boldsymbol{O}_p(1)$. Hence, it has been shown that the condition (A.5) is satisfied by $\widehat{A}_{FH}$, $\widehat{A}_{ML}$ and $\widehat{A}_{REML}$.

We next show that the condition (A.6) is satisfied by $\widehat{A}_{FH}$, $\widehat{A}_{ML}$ and $\widehat{A}_{REML}$. Since (A.4) and (A.5) are satisfied, we can use Lemma 1. Then,

$$\partial\widehat{\boldsymbol{\beta}}(\widehat{A}(\boldsymbol{\lambda}), \boldsymbol{\lambda})/\partial\boldsymbol{\lambda} - E\big[\partial\widehat{\boldsymbol{\beta}}(\widehat{A}(\boldsymbol{\lambda}), \boldsymbol{\lambda})/\partial\boldsymbol{\lambda}\big] = \boldsymbol{O}_p(m^{-1/2}).$$

To show $\partial\widehat{A}(\boldsymbol{\lambda})/\partial\boldsymbol{\lambda} - E\{\partial\widehat{A}(\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}\} = \boldsymbol{O}_p(m^{-1/2})$, we use the expression (27) (we can also use the expression (28), but using the expression (27) is more simple since we can use Lemma 1 such that we know the order of $\partial\widehat{\boldsymbol{\beta}}/\partial\boldsymbol{\lambda}$). From (29)$\sim$(32) and Lemma 1, we can evaluate (27) as

$$\begin{aligned}\frac{1}{m}G_\lambda(\boldsymbol{\lambda},\widehat{A}) &= \frac{2}{m}\sum_{j=1}^m(\widehat{A}+D_j)^{-k}\big\{h(y_j,\boldsymbol{\lambda})-\boldsymbol{x}_j'\boldsymbol{\beta}\big\}h_\lambda(y_j,\boldsymbol{\lambda})+\boldsymbol{O}_p(m^{-1/2})\\ &= \frac{2}{m}\sum_{j=1}^m(A+D_j)^{-k}\big\{h(y_j,\boldsymbol{\lambda})-\boldsymbol{x}_j'\boldsymbol{\beta}\big\}h_\lambda(y_j,\boldsymbol{\lambda})+\boldsymbol{O}_p(m^{-1/2})\end{aligned}$$

since $\widehat{A}-A=O_p(m^{-1/2})$. Here we assume that

$$-m^{-1}G_A(\boldsymbol{\lambda},\widehat{A})=c(A)+O_p(m^{-1/2}), \tag{33}$$

where $c(A)$ is a constant depending on A. This will be proved for each estimator in the end of this proof. Then we have

$$\begin{aligned}E\Big[\frac{\partial\widehat{A}(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}}\Big] &= E\Big[-\frac{m^{-1}G_\lambda(\boldsymbol{\lambda},\widehat{A})}{m^{-1}G_A(\boldsymbol{\lambda},\widehat{A})}\Big] \\ &= c(A)^{-1}\cdot\frac{2}{m}\sum_{j=1}^m(A+D_j)^{-k}E\big[\big\{h(y_j,\boldsymbol{\lambda})-\boldsymbol{x}_j'\boldsymbol{\beta}\big\}h_\lambda(y_j,\boldsymbol{\lambda})\big]+\boldsymbol{O}(m^{-1/2}).\end{aligned} \tag{34}$$

Therefore we have

$$\begin{aligned}\sqrt{m}\Big\{\frac{\partial\widehat{A}(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}}-E\Big[\frac{\partial\widehat{A}(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}}\Big]\Big\} &= \frac{G_\lambda(\boldsymbol{\lambda},\widehat{A})/\sqrt{m}}{G_A(\boldsymbol{\lambda},\widehat{A})/m}-E\Big[\frac{G_\lambda(\boldsymbol{\lambda},\widehat{A})/\sqrt{m}}{G_A(\boldsymbol{\lambda},\widehat{A})/m}\Big]\\ &= c(A)^{-1}\frac{2}{\sqrt{m}}\sum_{j=1}^m(A+D_j)^{-k}Z_j+\boldsymbol{O}_p(1),\end{aligned}$$

where Z_j is defined in (25), and by the central limit theorem, we have

$$\sqrt{m}\big[\partial\widehat{A}(\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}-E\big\{\partial\widehat{A}(\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}\big\}\big]=\boldsymbol{O}_p(1).$$

Consequently, we have proved for $\widehat{A}_{FH},\widehat{A}_{ML}$ and $\widehat{A}_{REML}$.

It remains to show that $-m^{-1}G_A(\boldsymbol{\lambda},\widehat{A})=c(A)+O_p(m^{-1/2})$ for $\widehat{A}_{FH},\widehat{A}_{ML}$ and $\widehat{A}_{REML}$.

For $\widehat{A}_{FH}$, from (6), we have

$$\begin{aligned}G_A(\boldsymbol{\lambda},\widehat{A}) = &-\sum_{j=1}^m(\widehat{A}+D_j)^{-k-1}\big\{h(y_j,\boldsymbol{\lambda})-\boldsymbol{x}_j'\widehat{\boldsymbol{\beta}}\big\}^2\\ &-2\sum_{j=1}^m(\widehat{A}+D_j)^{-k}\big\{h(y_j,\boldsymbol{\lambda})-\boldsymbol{x}_j'\widehat{\boldsymbol{\beta}}\big\}\boldsymbol{x}_j\Big(\frac{\partial}{\partial A}\widehat{\boldsymbol{\beta}}(A)\Big),\end{aligned}$$

where

$$\frac{\partial}{\partial A}\widehat{\boldsymbol{\beta}}(A)=\Big(\sum_{j=1}^m\frac{\boldsymbol{x}_j\boldsymbol{x}_j'}{\widehat{A}(\boldsymbol{\lambda})+D_j}\Big)^{-1}\sum_{j=1}^m\frac{\boldsymbol{x}_j\boldsymbol{x}_j'}{(\widehat{A}(\boldsymbol{\lambda})+D_j)^2}(\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\beta}}^*),$$

where $\widehat{\boldsymbol{\beta}}^*$ is defined in (21). Note that $\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\beta}}^*=\boldsymbol{O}_p(m^{-1/2})$ and from the low of large numbers, we have

$$\frac{\partial}{\partial A}\widehat{\boldsymbol{\beta}}(A)=\Big(\frac{1}{m}\sum_{j=1}^m\frac{\boldsymbol{x}_j\boldsymbol{x}_j'}{\widehat{A}(\boldsymbol{\lambda})+D_j}\Big)^{-1}\Big[\frac{1}{m}\sum_{j=1}^m\frac{\boldsymbol{x}_j\boldsymbol{x}_j'}{(\widehat{A}(\boldsymbol{\lambda})+D_j)^2}\Big](\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\beta}}^*)=\boldsymbol{O}_p(m^{-1/2}).$$

Thus we have

$$\begin{aligned}\frac{1}{m}G_A(\boldsymbol{\lambda},\widehat{A}) &= -\frac{1}{m}\sum_{j=1}^m (A+D_j)^{-2}\big\{h(y_j,\boldsymbol{\lambda})-\boldsymbol{x}_j'\boldsymbol{\beta}\big\}^2 \\ &\quad -2\Big[\frac{1}{m}\sum_{j=1}^m (A+D_j)^{-1}\big\{h(y_j,\boldsymbol{\lambda})-\boldsymbol{x}_j'\boldsymbol{\beta}\big\}\boldsymbol{x}_j\Big]\Big(\frac{\partial}{\partial A}\widehat{\boldsymbol{\beta}}(A)\Big)+\boldsymbol{O}_p(m^{-1/2}) \\ &= -\frac{1}{m}\sum_{j=1}^m (A+D_j)^{-2}\big\{h(y_j,\boldsymbol{\lambda})-\boldsymbol{x}_j'\boldsymbol{\beta}\big\}^2+\boldsymbol{O}_p(m^{-1/2}).\end{aligned}$$

Since $E\big[\big\{h(y_j,\boldsymbol{\lambda})-\boldsymbol{x}_j'\boldsymbol{\beta}\big\}^2\big]=A+D_j$, by the low of large numbers, we have

$$\frac{1}{m}G_A(\boldsymbol{\lambda},\widehat{A}) = -\frac{1}{m}\sum_{j=1}^m (A+D_j)^{-1}+\boldsymbol{O}_p(m^{-1/2}), \tag{35}$$

where the order of the leading term is $O(1)$, corresponding with $c(A)$.

Similarly, for $\widehat{A}_{ML}$ and $\widehat{A}_{REML}$ defined in (7) and (8), straight calculation (almost the same as the case of $\widehat{A}_{FH}$) shows that

$$\frac{1}{m}G_A(\boldsymbol{\lambda},\widehat{A}) = -\frac{1}{m}\sum_{j=1}^m (A+D_j)^{-2}+\boldsymbol{O}_p(m^{-1/2}), \tag{36}$$

where the order of the leading term is $O(1)$, corresponding with $c(A)$. □

A.3 Proof of Lemma 3. We begin by showing that $\hat{\boldsymbol{\lambda}}-\boldsymbol{\lambda}=O_p(m^{-1/2})$. By Taylor series expansion of equation (10), we have

$$\hat{\boldsymbol{\lambda}}-\boldsymbol{\lambda} = -F(\boldsymbol{\lambda},\widehat{A},\widehat{\boldsymbol{\beta}})\Big(\partial F(\boldsymbol{\lambda},\widehat{A},\widehat{\boldsymbol{\beta}})/\partial\boldsymbol{\lambda}'|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^*}\Big)^{-1},$$

where

$$\begin{aligned}&\partial F(\boldsymbol{\lambda},\widehat{A}(\boldsymbol{\lambda}),\widehat{\boldsymbol{\beta}}(\boldsymbol{\lambda}))/\partial\boldsymbol{\lambda}' \\ &= \sum_{j=1}^m\frac{h_{y\lambda\lambda}(y_j,\boldsymbol{\lambda})}{h_y(y_j,\boldsymbol{\lambda})}-\sum_{j=1}^m\frac{h_{y\lambda}(y_j,\boldsymbol{\lambda})'h_{y\lambda}(y_j,\boldsymbol{\lambda})}{(h_y(y_j,\boldsymbol{\lambda}))^2}-\sum_{j=1}^m\frac{h(y_j,\boldsymbol{\lambda})-\boldsymbol{x}_j'\widehat{\boldsymbol{\beta}}(\widehat{A}(\boldsymbol{\lambda}),\boldsymbol{\lambda})}{\widehat{A}(\boldsymbol{\lambda})+D_j}h_{\lambda\lambda}(y_j,\boldsymbol{\lambda}) \\ &\quad -\sum_{j=1}^m\frac{h_\lambda(y_j,\boldsymbol{\lambda})-\boldsymbol{x}_j'(\partial\widehat{\boldsymbol{\beta}}(\widehat{A}(\boldsymbol{\lambda}),\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}')}{\widehat{A}(\boldsymbol{\lambda})+D_j}h_\lambda(y_j,\boldsymbol{\lambda})'+\sum_{j=1}^m\frac{h(y_j,\boldsymbol{\lambda})-\boldsymbol{x}_j'\widehat{\boldsymbol{\beta}}(\widehat{A}(\boldsymbol{\lambda}),\boldsymbol{\lambda})}{(\widehat{A}(\boldsymbol{\lambda})+D_j)^2}\frac{\partial\widehat{A}(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}}h_\lambda(y_j,\boldsymbol{\lambda})' \\ &= K_1+K_2+K_3+K_4, \quad \text{(say)}\end{aligned}$$

where $\boldsymbol{\lambda}^*$ is a vector satisfying $\boldsymbol{\lambda}<\boldsymbol{\lambda}^*<\hat{\boldsymbol{\lambda}}$. It is here noted that for $\boldsymbol{a}=(a_1,\ldots,a_p)$ and $\boldsymbol{b}=(b_1,\ldots,b_p)$, $\boldsymbol{a}<\boldsymbol{b}$ denotes $a_j<b_j$ for $j=1,\ldots,p$. For K_1, from Assumption 1, we have

$$E\Big[\frac{h_{y\lambda\lambda}(y_j,\boldsymbol{\lambda})}{h_y(y_j,\boldsymbol{\lambda})}-\frac{h_{y\lambda}(y_j,\boldsymbol{\lambda})'h_{y\lambda}(y_j,\boldsymbol{\lambda})}{(h_y(y_j,\boldsymbol{\lambda}))^2}\Big]=E\Big[\frac{\partial}{\partial\boldsymbol{\lambda}}\Big(\frac{h_{y\lambda}(y_j,\lambda)}{h_y(y_j,\lambda)}\Big)\Big]=O(1)$$

for $j=1,\ldots,m$. Since $y_1,\ldots,y_m$ are independent, by the law of large numbers, we have

$$\frac{1}{m}K_1=\frac{1}{m}\Big\{\sum_{j=1}^m\frac{h_{y\lambda\lambda}(y_j,\boldsymbol{\lambda})}{h_y(y_j,\boldsymbol{\lambda})}-\sum_{j=1}^m\frac{h_{y\lambda}(y_j,\boldsymbol{\lambda})'h_{y\lambda}(y_j,\boldsymbol{\lambda})}{(h_y(y_j,\boldsymbol{\lambda}))^2}\Big\}=O_p(1).$$

Under Assumptions 1 and 2, we have

$$
\begin{aligned}
\frac{1}{m}K_2 &= \frac{1}{m}\sum_{j=1}^m \frac{h(y_j,\boldsymbol{\lambda}) - \boldsymbol{x}_j'\widehat{\boldsymbol{\beta}}(\widehat{A}(\boldsymbol{\lambda}),\boldsymbol{\lambda})}{\widehat{A}(\boldsymbol{\lambda})+D_j} h_{\lambda\lambda}(y_j,\boldsymbol{\lambda}) \\
&= \frac{1}{m}\sum_{j=1}^m \frac{h(y_j,\boldsymbol{\lambda}) - \boldsymbol{x}_j'\boldsymbol{\beta}}{A+D_j} h_{\lambda\lambda}(y_j,\boldsymbol{\lambda}) - \frac{1}{m}(\widehat{A}-A)\sum_{j=1}^m \frac{h(y_j,\boldsymbol{\lambda}) - \boldsymbol{x}_j'\boldsymbol{\beta}}{(A^*+D_j)^2} h_{\lambda\lambda}(y_j,\boldsymbol{\lambda}) \\
&\quad - \frac{1}{m}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\sum_{j=1}^m \frac{\boldsymbol{x}_j}{A+D_j} h_{\lambda\lambda}(y_j,\boldsymbol{\lambda}) + \frac{1}{m}(\widehat{A}-A)(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\sum_{j=1}^m \frac{\boldsymbol{x}_j}{(A^*+D_j)^2} h_{\lambda\lambda}(y_j,\boldsymbol{\lambda}) \\
&= \frac{1}{m}\sum_{j=1}^m \frac{h(y_j,\boldsymbol{\lambda}) - \boldsymbol{x}_j'\boldsymbol{\beta}}{A+D_j} h_{\lambda\lambda}(y_j,\boldsymbol{\lambda}) + O_p(m^{-1/2}) = O_p(1).
\end{aligned}
$$

Similarly, we can evaluate K_3 as

$$
\begin{aligned}
\frac{1}{m}K_3 &= \frac{1}{m}\sum_{j=1}^m \frac{h_\lambda(y_j,\boldsymbol{\lambda}) - \boldsymbol{x}_j'(\partial\widehat{\boldsymbol{\beta}}(\widehat{A}(\boldsymbol{\lambda}),\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}')}{\widehat{A}(\boldsymbol{\lambda})+D_j} h_\lambda(y_j,\boldsymbol{\lambda})' \\
&= \frac{1}{m}\sum_{j=1}^m \frac{h_\lambda(y_j,\boldsymbol{\lambda}) - \boldsymbol{x}_j'(\partial\widehat{\boldsymbol{\beta}}(\widehat{A}(\boldsymbol{\lambda}),\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}')}{A+D_j} h_\lambda(y_j,\boldsymbol{\lambda})' \\
&\quad - (\widehat{A}-A)\cdot\frac{1}{m}\sum_{j=1}^m \frac{h_\lambda(y_j,\boldsymbol{\lambda}) - \boldsymbol{x}_j'(\partial\widehat{\boldsymbol{\beta}}(\widehat{A}(\boldsymbol{\lambda}),\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}')}{(A^*+D_j)^2} h_\lambda(y_j,\boldsymbol{\lambda})' \\
&= O_p(1)
\end{aligned}
$$

under Assumptions 1 and 2. Moreover,

$$
\begin{aligned}
\frac{1}{m}K_4 =& \frac{1}{m}\sum_{j=1}^m \frac{h(y_j,\boldsymbol{\lambda}) - \boldsymbol{x}_j'\widehat{\boldsymbol{\beta}}(\widehat{A}(\boldsymbol{\lambda}),\boldsymbol{\lambda})}{(\widehat{A}(\boldsymbol{\lambda})+D_j)^2} \frac{\partial\widehat{A}(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}} h_\lambda(y_j,\boldsymbol{\lambda})' \\
=& \frac{1}{m}\Big(\frac{\partial\widehat{A}(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}}\Big)\sum_{j=1}^m \frac{h(y_j,\boldsymbol{\lambda}) - \boldsymbol{x}_j'\boldsymbol{\beta}}{(A+D_j)^2} h_\lambda(y_j,\boldsymbol{\lambda})' \\
&- (\widehat{A}-A)\Big(\frac{\partial\widehat{A}(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}}\Big)\frac{1}{m}\sum_{j=1}^m \frac{h(y_j,\boldsymbol{\lambda}) - \boldsymbol{x}_j'\boldsymbol{\beta}}{2(A^*+D_j)^3} h_\lambda(y_j,\boldsymbol{\lambda})' \\
&- \Big(\frac{\partial\widehat{A}(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}}\Big)(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\frac{1}{m}\sum_{j=1}^m \frac{\boldsymbol{x}_j}{(A+D_j)^2} h_\lambda(y_j,\boldsymbol{\lambda})' \\
&+ (\widehat{A}-A)\Big(\frac{\partial\widehat{A}(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}}\Big)(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\frac{1}{m}\sum_{j=1}^m \frac{\boldsymbol{x}_j}{2(A^*+D_j)^3} h_\lambda(y_j,\boldsymbol{\lambda})',
\end{aligned}
$$

which is of order $O_p(1)$. As a result, we have

$$
\frac{1}{m}\Big\{\frac{\partial F(\boldsymbol{\lambda},\widehat{A}(\boldsymbol{\lambda}),\widehat{\boldsymbol{\beta}}(\boldsymbol{\lambda}))}{\partial\boldsymbol{\lambda}'}\Big|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^*}\Big\} = O_p(1).
$$

Furthermore, by Assumption 1, we have

$$
\begin{aligned}
&F(\boldsymbol{\lambda}, \widehat{A}(\boldsymbol{\lambda}), \widehat{\boldsymbol{\beta}}(\boldsymbol{\lambda})) \\
&= \sum_{j=1}^{m} \frac{h_{y\lambda}(y_j, \boldsymbol{\lambda})}{h_y(y_j, \boldsymbol{\lambda})} - \sum_{j=1}^{m} \frac{h(y_j, \boldsymbol{\lambda}) - \boldsymbol{x}_j' \widehat{\boldsymbol{\beta}}(\widehat{A}(\boldsymbol{\lambda}), \boldsymbol{\lambda})}{\widehat{A}(\boldsymbol{\lambda}) + D_j} h_\lambda(y_j, \boldsymbol{\lambda}) \\
&= \sum_{j=1}^{m} \frac{h_{y\lambda}(y_j, \boldsymbol{\lambda})}{h_y(y_j, \boldsymbol{\lambda})} - \sum_{j=1}^{m} \frac{h(y_j, \boldsymbol{\lambda}) - \boldsymbol{x}_j' \boldsymbol{\beta}}{A + D_j} h_\lambda(y_j, \boldsymbol{\lambda}) - (\widehat{A} - A) \sum_{j=1}^{m} \frac{h(y_j, \boldsymbol{\lambda}) - \boldsymbol{x}_j' \boldsymbol{\beta}}{(A^* + D_j)^2} h_\lambda(y_j, \boldsymbol{\lambda}) \\
&\quad - (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \sum_{j=1}^{m} \frac{\boldsymbol{x}_j}{A + D_j} h_\lambda(y_j, \boldsymbol{\lambda}) + (\widehat{A} - A)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \sum_{j=1}^{m} \frac{\boldsymbol{x}_j}{(A^* + D_j)^2} h_\lambda(y_j, \boldsymbol{\lambda}),
\end{aligned}
$$

which is evaluated as

$$
\sum_{j=1}^{m} \Big\{ \frac{h_{y\lambda}(y_j, \boldsymbol{\lambda})}{h_y(y_j, \boldsymbol{\lambda})} - \frac{h(y_j, \boldsymbol{\lambda}) - \boldsymbol{x}_j' \boldsymbol{\beta}}{A + D_j} h_\lambda(y_j, \boldsymbol{\lambda}) \Big\} + O_p(m^{1/2}).
$$

For all $j = 1, \ldots, m$, we have

$$
E\Big[\frac{h_{y\lambda}(y_j, \boldsymbol{\lambda})}{h_y(y_j, \boldsymbol{\lambda})} - (A + D_j)^{-1} \big\{ h(y_j, \boldsymbol{\lambda}) - \boldsymbol{x}_j \boldsymbol{\beta} \big\} h_\lambda(y_j, \boldsymbol{\lambda}) \Big] = E\Big[\frac{\partial \boldsymbol{\lambda} \log f(Y_j; \boldsymbol{\lambda}, \boldsymbol{\beta}, A)}{\partial \boldsymbol{\lambda}} \Big] = \boldsymbol{0},
$$

where $f(y_j; \boldsymbol{\lambda}, \boldsymbol{\beta}, A)$ is the density function of observation y_j defined in (1). To show $\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda} = O_p(m^{-1/2})$, by the central limit theorem, we have

$$
\frac{1}{\sqrt{m}} F(\boldsymbol{\lambda}, \widehat{A}(\boldsymbol{\lambda}), \widehat{\boldsymbol{\beta}}(\boldsymbol{\lambda})) = \boldsymbol{O}_p(1).
$$

Therefore we have

$$
\sqrt{m}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}) = -\frac{1}{\sqrt{m}} F(\boldsymbol{\lambda}, \widehat{A}(\boldsymbol{\lambda}), \widehat{\boldsymbol{\beta}}(\boldsymbol{\lambda})) \Big\{ \frac{1}{m} \Big(\partial F(\boldsymbol{\lambda}, \widehat{A}, \widehat{\boldsymbol{\beta}}) / \partial \boldsymbol{\lambda}' |_{\boldsymbol{\lambda} = \boldsymbol{\lambda}^*} \Big) \Big\}^{-1} = \boldsymbol{O}_p(1),
$$

and we conclude that $\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda} = \boldsymbol{O}_p(m^{-1/2})$.

We next show that $E[\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}] = \boldsymbol{O}(m^{-1})$. From the first part of Lemma 3, we have $\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda} = \boldsymbol{O}_p(m^{-1/2})$. Then expanding (10) shows that

$$
\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda} = -F(\boldsymbol{\lambda}, \widehat{A}, \widehat{\boldsymbol{\beta}}) \big(\partial F(\boldsymbol{\lambda}, \widehat{A}, \widehat{\boldsymbol{\beta}}) / \partial \boldsymbol{\lambda}' \big)^{-1} + \boldsymbol{O}_p(m^{-1}).
$$

We have to show that the expectation of the first term is $\boldsymbol{O}(1/m)$. We obtain

$$
\begin{aligned}
&E\Big[\frac{1}{m} F(\boldsymbol{\lambda}, \widehat{A}, \widehat{\boldsymbol{\beta}}) \Big\{ \frac{1}{m} \Big(\partial F(\boldsymbol{\lambda}, \widehat{A}, \widehat{\boldsymbol{\beta}}) / \partial \boldsymbol{\lambda}' \Big) \Big\}^{-1} \Big] \\
&\quad = E\Big[\Big\{ \frac{1}{m} F(\boldsymbol{\lambda}, \widehat{A}, \widehat{\boldsymbol{\beta}}) \Big\} \Big\{ \frac{1}{m} E\big[\partial F(\boldsymbol{\lambda}, \widehat{A}, \widehat{\boldsymbol{\beta}}) / \partial \boldsymbol{\lambda}' \big] + O_p(m^{-1/2}) \Big\}^{-1} \Big] \\
&\quad = E\Big\{ \frac{1}{m} F(\boldsymbol{\lambda}, \widehat{A}, \widehat{\boldsymbol{\beta}}) \Big\} \Big\{ \frac{1}{m} E\big[\partial F(\boldsymbol{\lambda}, \widehat{A}, \widehat{\boldsymbol{\beta}}) / \partial \boldsymbol{\lambda}' \big] \Big\}^{-1} + O(m^{-1}),
\end{aligned}
$$

which is of order $O(m^{-1})$, since

$$
\begin{aligned}
E\Big[\frac{1}{m} F(\boldsymbol{\lambda}, \widehat{A}, \widehat{\boldsymbol{\beta}}) \Big] &= -E(\widehat{A} - A) \cdot \frac{1}{m} \sum_{j=1}^{m} \frac{E\big[\big\{ h(y_j, \boldsymbol{\lambda}) - \boldsymbol{x}_j' \boldsymbol{\beta} \big\} h_\lambda(y_j, \boldsymbol{\lambda}) \big]}{(A^* + D_j)^2} \\
&\quad - E(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \cdot \frac{1}{m} \sum_{j=1}^{m} \frac{\boldsymbol{x}_j}{A + D_j} E\big[h_\lambda(y_j, \boldsymbol{\lambda}) \big] + \boldsymbol{O}(m^{-1}),
\end{aligned}
$$

which is of order $\boldsymbol{O}(m^{-1})$, where we used the fact that $E[\widehat{A}-A]=O(m^{-1})$ and $E[\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}]=\boldsymbol{O}(m^{-1})$, and since

$$\frac{1}{m}E\big[\partial F(\boldsymbol{\lambda},\widehat{A},\widehat{\boldsymbol{\beta}})/\partial\boldsymbol{\lambda}'\big]=O(1).$$

Thus $E[\hat{\boldsymbol{\lambda}}-\boldsymbol{\lambda}]=\boldsymbol{O}(m^{-1})$ follows. □

A.4 Proof of Lemma 4. By the Taylor series expansion of $\hat{\eta}_i^{EB1}$, we have

$$\begin{aligned}\hat{\eta}_i^{EB1}-\hat{\eta}_i^{B}=&\frac{D_i}{A+D_i}\boldsymbol{x}_i'(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})+\frac{D_i}{(A+D_i)^2}(\widehat{A}-A)\big\{h(y_i,\boldsymbol{\lambda})-\boldsymbol{x}_i'\boldsymbol{\beta}\big\}\\&-\frac{D_i}{(A^*+D_i)^2}\boldsymbol{x}_i'(\widehat{A}-A)(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})-\frac{D_i}{(A^*+D_i)^3}\big\{h(y_i,\boldsymbol{\lambda})-\boldsymbol{x}_i'\boldsymbol{\beta}^*\big\}(\widehat{A}-A)^2,\end{aligned}$$

where A^* is an intermediate value of A and $\widehat{A}$ and $\boldsymbol{\beta}^*$ is an intermediate vector of $\boldsymbol{\beta}$ and $\widehat{\boldsymbol{\beta}}$. Differentiating the both sides by $\boldsymbol{\lambda}$, we have

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\lambda}}\hat{\eta}_i^{EB1}=&\frac{\partial}{\partial\boldsymbol{\lambda}}\hat{\eta}_i^{B}+\frac{D_i}{A+D_i}\boldsymbol{x}_i'\Big(\frac{\partial}{\partial\boldsymbol{\lambda}}\widehat{\boldsymbol{\beta}}(\boldsymbol{\lambda})\Big)+\frac{D_i}{(A+D_i)^2}\Big(\frac{\partial}{\partial\boldsymbol{\lambda}}\widehat{A}(\boldsymbol{\lambda})\Big)\big\{h(y_i,\boldsymbol{\lambda})-\boldsymbol{x}_i'\boldsymbol{\beta}\big\}\\&-\frac{D_i}{2(A^*+D_i)^2}\boldsymbol{x}_i'\Big(\frac{\partial}{\partial\boldsymbol{\lambda}}\widehat{A}(\boldsymbol{\lambda})\Big)(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})-\frac{D_i}{2(A^*+D_i)^2}\boldsymbol{x}_i'(\widehat{A}-A)\Big(\frac{\partial}{\partial\boldsymbol{\lambda}}\widehat{\boldsymbol{\beta}}(\boldsymbol{\lambda})\Big)\\&-\frac{2D_i}{(A^*+D_i)^3}h_\lambda(y_i,\boldsymbol{\lambda})(\widehat{A}-A)\Big(\frac{\partial}{\partial\boldsymbol{\lambda}}\widehat{A}(\boldsymbol{\lambda})\Big)\\=&\frac{\partial}{\partial\boldsymbol{\lambda}}\hat{\eta}_i^{B}+\frac{D_i}{(A+D_i)^2}E\Big[\frac{\partial}{\partial\boldsymbol{\lambda}}\widehat{A}(\boldsymbol{\lambda})\Big]\big\{h(y_i,\boldsymbol{\lambda})-\boldsymbol{x}_i'\boldsymbol{\beta}\big\}\\&+\frac{D_i}{A+D_i}\boldsymbol{x}_i'E\Big[\frac{\partial}{\partial\boldsymbol{\lambda}}\widehat{\boldsymbol{\beta}}(\boldsymbol{\lambda})\Big]+\boldsymbol{O}_p(m^{-1/2}),\end{aligned}$$

by using Lemmas 1 and 2. Also from Lemmas 1 and 2, we already know that

$$E\Big[\frac{\partial}{\partial\boldsymbol{\lambda}}\widehat{\boldsymbol{\beta}}(\boldsymbol{\lambda})\Big]=\Big(\sum_{j=1}^m\frac{\boldsymbol{x}_j\boldsymbol{x}_j'}{A+D_j}\Big)^{-1}\sum_{j=1}^m\frac{\boldsymbol{x}_j}{A+D_j}E\big[h_\lambda(y_j,\boldsymbol{\lambda})\big]+\boldsymbol{O}(m^{-1/2}),\tag{37}$$

and

$$E\Big[\frac{\partial}{\partial\boldsymbol{\lambda}}\widehat{A}(\boldsymbol{\lambda})\Big]=\Big(\sum_{j=1}^m(A+D_j)^{-k}\Big)^{-1}\Big(\sum_{j=1}^m\frac{E[\{h(y_j,\boldsymbol{\lambda})-\boldsymbol{x}_j'\boldsymbol{\beta}\}h_\lambda(y_j,\boldsymbol{\lambda})]}{(A+D_j)^k}\Big)+\boldsymbol{O}(m^{-1/2}),\tag{38}$$

where $k=1$ corresponds to $\widehat{A}_{FH}$ and $k=2$ corresponds to $\widehat{A}_{ML}$ and $\widehat{A}_{REML}$. The formula (37) comes from (22), and the formula (38) is obtained by combining (33), (35) and (36). For $\widehat{A}_{PR}$, from (24),

$$E\Big[\frac{\partial\widehat{A}_{PR}(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}}\Big]=\frac{2}{m-p}\sum_{j=1}^m E\big[\big\{h(y_j,\boldsymbol{\lambda})-\boldsymbol{x}_j'\boldsymbol{\beta}\big\}h_\lambda(y_j,\boldsymbol{\lambda})\big]+\boldsymbol{O}(m^{-1/2}),$$

which completes the proof. □

References

[1] Box, G.E.P, and Cox, D.R. (1964). An analysis of transformation (with discussion). *Journal of Royal Statistical Society B*, **26**, 211-252.

[2] 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.

[3] Fay, R. and Herriot, R. (1979). Estimators of income for small area places: an application of James–Stein procedures to census. *Journal of American Statistical Association*, **74**, 341-353.

[4] Ghosh, M. and Rao, J.N.K. (1994). Small area estimation: An appraisal. *Statist. Science*, **9**, 55-93.

[5] Prasad, N. and Rao, J. N. K. (1990). The estimation of mean-squared errors of small-area estimators. *Journal of American Statistical Association*, **90**, 758-766.

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

[7] Slud, E.V. and Maiti, T. (2006). Mean-squared error estimation in transformed Fay-Herriot models. *Journal of Royal Statistical Society B*, **68**, 239-257.

[8] Yang, Z. L. (2006). A modified family of power transformations. *Economics Letters*, **92**, 14-19.