Small-area estimation under a nonlinear transformed area-level model

R. Priam *

November 15, 2023

Abstract

It is proposed a new generalized estimator for small area with a Fay-Herriot model and a nonlinear transformation of the outcome. Two different bias corrections and the estimation of the variance or mean squared error are presented. The method is related to an integral approach via an expansion from the sum of exponential functions which allows to approximate the back-transformation. Current methods for mean estimation in small area from survey data are intensively studied since a few decades but are often relevant only for linear models. The Box-Cox transformation associated with and without a back-transformation illustrates the proposed approach with perspectives.

1 Introduction

Small area methods [1] are often seen as a way to reduce the variance for the estimator of a mean from a dataset in comparison to the more usual synthetic estimators. They uses independent variables which bring information to the estimator and allows a reduction of the variance at the cost to need some bias correction approach in order to avoid a bad estimator of the mean. Thus in small area, looking for a bias correction may be as important as computing the mean squared error. If both may be available via computational algorithms, analytical solutions bring additional information in the domain.

Herein, the studied case involves a nonlinear transformation of the outcome with eventually bounded values, such that $y_i \in [a, b]$ where y_i is an outcome while a and b are the bounds. This seems not studied in the related literature such that [2, 3, 4], except in [5] and some alternative distributions such as the Gamma one which may be also relevant but out of the scope herein. Such missing constraints are very sad because lower and upper bounds may be required in some domains of applied research such that in psychology which often introduces data surveys where some answers have limited values, typically zero to ten or minus hundred to hundred for instance. This justifies the need for a bounded transformation, the one considered herein is based on the logistic transformation bounded [6] associated to the Box-Cox one [7]. Next after this is often this function called generalized logistic-bounded box-cox transformation which is involved in the estimations but any other more relevant transformation may be preferred.

*rpriam@gmail.com.

2 FH model under transformation

The area-level model is reviewed, from its definition to the estimation of the parameters, before the nonlinear transformation is introduced. The resulting back-transformation is explained and the correction of the bias discussed.

2.1 Small area estimation (SAE)

When there are predictive variables, it has been introduced in the literature the mixed models in survey theory via the small area methods. This allows to reduce the variance of the mean for an outcome, by fitting a model with these additional variables. There are mainly two models, one directly at the level of the area, and one at the level of the units. This is the first one which is involved next but the proposed generalized approach may be extended to other models.

Model

The general model comes from the bayesian statistics. Matricially,

$$\mathbf{z} \;=\; \mathbf{X}\boldsymbol{\beta} + \mathbf{v} + \mathbf{e}\,.$$

It is written for the Fay-Herriot model [8], the following particular case:

$$z_i = \mathbf{x}_i^T\boldsymbol{\beta} + u_i + \epsilon_i$$

For the matricial notation, one denotes the vector of regression coefficients $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)^T$, the vector of random effect $\mathbf{v} = (u_1, \cdots, u_m)^T$, and the vector of sampling noises $\mathbf{e} = (\epsilon_i, \cdots, \epsilon_m)$. Thus, the random effects are $u_i \sim N(0, \sigma_u^2)$ and the sampling errors are $\epsilon_i \sim N(0, \sigma_{e_i}^2)$, both are independent. Matricially, this leads to the covariance matrix for $\mathbf{y}$ equal to $\boldsymbol{\Sigma}(\sigma_u^2) = diag(\sigma_u^2 + \sigma_{e_1}^2, \cdots, \sigma_u^2 + \sigma_{e_m}^2)$, where σ_u^2 is unknown. For the estimation of $\boldsymbol{\beta}$ and σ_u^2, several approaches have been proposed in the literature, called REML and ML for the two main methods currently used most of the time. Another way to see the model is via the distribution of the data: $\mu_i \sim N(\mathbf{x}_i^T\boldsymbol{\beta}, \sigma_u^2)$, $z_i|\mu_i \sim N(\mu_i, \sigma_{e_i}^2)$, and thus, $z_i \sim N(\mathbf{x}_i^T\boldsymbol{\beta}, \sigma_{e_i}^2 + \sigma_u^2)$. The problem with this notation is to not show the random effect explicitely.

Estimators

Let denote the fraction:

$$\gamma_i = \gamma_i(\sigma_u^2) = \sigma_{e_i}^2(\sigma_u^2 + \sigma_{e_i}^2)^{-1}\,.$$

The best (or Bayes) estimator of μ_i is under squared error loss as follows, for σ_u^2 given, say the mean estimator from small area:

$$\vartheta_{iz} = E(\mu_i|z_i) = z_i - \frac{\sigma_{e_i}^2}{\sigma_u^2 + \sigma_{e_i}^2}(z_i - \mathbf{x}_i^T\boldsymbol{\beta}) = (1-\gamma_i)z_i + \gamma_i\mathbf{x}_i^T\boldsymbol{\beta}\,.$$

Here, the bayesian theory gave that,

$$\mu_i|z_i \sim N(\vartheta_{iz}, g_{1i}(\sigma_u^2)) \text{ where } g_1(\sigma_u^2) = (1-\gamma_i)\sigma_{e_i}^2\,.$$

Following [9], it is deduced the theoritical mean squared error for the linear model:

$$MSE(\vartheta_i) \quad = \quad g_{1i}(\sigma_u^2)\,.$$

This induces that the variance of the mean in the area is reduced by a factor $(1-\gamma_i)$, with the value $g_{1i}(\sigma_u^2)$ instead of $\sigma_{e_i}^2$, justifying the approach of the small areas.

In practice with a data sample, the parameters $\boldsymbol{\beta}$ and σ_u^2 are unknown, such that they are estimated. It is usual to denote then $\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\sigma_u^2)$ as a function of the parameter σ_u^2, and thus the parameter vector of regression coefficients is denoted $\hat{\boldsymbol{\beta}}$ when σ_u^2 is replaced by an estimation of σ_u^2, say $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\hat{\sigma}_u^2)$. When $\tilde{\boldsymbol{\beta}}$ replaces $\boldsymbol{\beta}$ the best (or Bayes) estimator ϑ_{iz} becomes the first-step empirical best (or empirical Bayes) estimator of μ_i such that it is denoted $\tilde{\vartheta}_i$. When this is the estimation of the $\tilde{\sigma}_u^2$ replaces σ_u^2 in this first step estimator, with the notation $\hat{\gamma}_i = \gamma_i(\hat{\sigma}_u^2)$, one gets the final EB estimator,

$$\hat{\vartheta}_{iz} \quad = (1-\hat{\gamma}_i)z_i + \hat{\gamma}_i \mathbf{x}_i^T \hat{\boldsymbol{\beta}}\,.$$

This is called more generally the empirical best linear unbiased predictor (EBLUP). This is the area-level estimator involved herein while unit level models are for a future work.

Algorithm for parameters inference

For the estimation, several methods exist. For a given σ_u^2, one computes the vector of regression coefficients as follows, as in an usual weighted regression, thus with heterodescacity. From the estimation $\hat{\sigma}_u^2$, it is deduced $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\hat{\sigma}_u^2)$, such that:

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= (\mathbf{X}^T\hat{\boldsymbol{\Sigma}}\mathbf{X})^{-1}\mathbf{X}^T\hat{\boldsymbol{\Sigma}}\mathbf{y} \\
&= \left\{\sum_{i=1}^m \frac{1}{\hat{\sigma}_u^2 + \sigma_{e_i}^2}\mathbf{x}_i\mathbf{x}_i^T\right\}^{-1} \sum_{i=1}^m \frac{1}{\hat{\sigma}_u^2 + \sigma_{e_i}^2}\mathbf{x}_i y_i\,.
\end{aligned}$$

Let denote $\mathbf{xx}_i = \mathbf{x}_i^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i$, and $\hat{\boldsymbol{\beta}}_{LS}$ the solution from the usual unweighted least squared regression with $\mathbf{X}$ and $\mathbf{y}$ as respectively the design matrix and the vector of outcomes. For the estimation of σ_u^2, the usual different approaches are as follows.

- Method of Prasad-Rao with an exact solution $\hat{\sigma}^{*2}_{u,\mathrm{PR}}$,

$$\hat{\sigma}^{*2}_{u,\mathrm{PR}} \quad = \quad \frac{1}{m-p}\sum_{i=1}^m \left[(y_i - \mathbf{x}_i^T\hat{\boldsymbol{\beta}}_{LS})^2 - \sigma_{e_i}^2(1-\mathbf{xx}_i)\right]\,.$$

- Method of Fay-Herriot with a numerical solution $\hat{\sigma}^{*2}_{u,\mathrm{FH}}$,

$$\sum_{i=1}^m \frac{1}{\sigma_u^2 + \sigma_{e_i}^2}\left[y_i - \mathbf{x}_i^T\tilde{\boldsymbol{\beta}}(\sigma_u^2)\right]^2 = m-p\,.$$

- Method of ML with a numerical solution $\hat{\sigma}^{*2}_{u,\mathrm{ML}}$,

$$\sum_{i=1}^m \frac{\left[y_i - \mathbf{x}_i^T\tilde{\boldsymbol{\beta}}(\sigma_u^2)\right]^2}{(\sigma_u^2 + \sigma_{e_i}^2)^2} - \frac{1}{(\sigma_u^2 + \sigma_{e_i}^2)} = 0\,.$$

- Method of REML with a numerical solution $\hat{\sigma}^{*2}_{u,\mathrm{REML}}$,

$$\sum_{i=1}^{m} \frac{\left[y_i - \mathbf{x}_i^T \tilde{\boldsymbol{\beta}}(\sigma_u^2)\right]^2 - (\sigma_u^2 + \sigma_{e_i}^2) + \mathbf{x}_i^T \left(\sum_l \frac{1}{\sigma_u^2 + \sigma_{e_l}^2} \mathbf{x}_l \mathbf{x}_l^T \right)^{-1} \mathbf{x}_i}{(\sigma_u^2 + \sigma_{e_i}^2)^2} = 0\,.$$

These different methods for the estimation of σ_u^2 were proposed in the literature by different authors, see [9] for instance. In the four cases, one ends with a positive solution by choosing respectively, $\hat{\sigma}^2_{u,\mathrm{PR}} = max(0, \hat{\sigma}^{*2}_{u,\mathrm{PR}})$, $\hat{\sigma}^2_{u,\mathrm{FH}} = max(0, \hat{\sigma}^{*2}_{u,\mathrm{FH}})$, $\hat{\sigma}^2_{u,\mathrm{ML}} = max(0, \hat{\sigma}^{*2}_{u,\mathrm{ML}})$ and $\hat{\sigma}^2_{u,\mathrm{REML}} = max(0, \hat{\sigma}^{*2}_{u,\mathrm{REML}})$. Once the estimation $\hat{\sigma}_u^2$ is available, this leads to $\hat{\boldsymbol{\beta}}$ which enters the empirical estimator of the mean, as explained above. As these estimators are all random variables depending on the available sample, the variability is measured by computing a variance or a mean squared error. When the linear model is not enough, a nonlinear transformation is introduced as explained next after.

2.2 Nonlinear transformations with eventual bounds

The transformation, back-transformation and bias correction are presented in this part for several functions when eventually some constrained are required for the minimum and the maximum of the outcome variable. A typical example is for a probability which belongs to the interval $[0, 1]$, such that one does not want to predict values outside.

Transformation of the target variable

Below κ_ξ^{-1} denotes a transformation with parameters ξ, typically herein the box-cox. Transformations without parameters such as squared-root for instance may be preferred but are included. Let write the transformation as follows for the usual linear regression with fixed effects.

$$z_i = \kappa_\xi^{-1} \left(\frac{y_i - a}{b - c\, y_i} \right) .$$

Hence, this is the transformed outcome variable which is a linear combination of the independent variables instead of the untransformed one,

$$z_i = \beta_0 + \sum_j \beta_j x_{ij} + \epsilon_i\,.$$

With bounds, for instance, in the case logarithmique, it is retrieved the bounds: for infinite positive values of $\hat{z}_i$, the predicted outcome $\hat{z}_i$ becomes equal to b, while for infinite negative values, it becomes equal to a, as expected. This remains true for the bounds in $[0; 1]$ such that the case of a probability or rescaled value, $y_i/\max_y - \min_y \in [0; 1]$, when $\max_y$ and $\min_y$ are respectively the minimum and maximum values of the outcome rescaled. With $[a, b]$ the interval for the bounds, this induces for the inverse of the Box-Cox transformation for $\lambda \neq 0$ and of the log when $\lambda = 0$,

$$\hat{y}_i = \begin{cases} \dfrac{a + b e^{\hat{z}_i}}{1 + c e^{\hat{z}_i}} & \text{for } \lambda = 0 \\ \dfrac{a + b(1 + \lambda \hat{z}_i)^{1/\lambda}}{1 + c(1 + \lambda \hat{z}_i)^{1/\lambda}} & \text{for } \lambda \neq 0 \end{cases}$$

Table 1: Generalized transformation for skewed and bounded outcomes

Name	a	b	c	$\kappa_\xi^{-1}(y)$	$\kappa_\xi(z)$ for a=b=0, c=1
Logarithmic	0	1	0	$\log(y)$	$\exp(z)$
Box-cox	0	1	0	$\frac{y^\lambda-1}{\lambda}$	$(1+\lambda z)^{1/\lambda}$
Box-cox (bis)	0	1	0	y^λ	$z^{1/\lambda}$
Logistic	0	1	1	$\log(y)$	$\exp(z)$
Logistic-bounded	a	b	1	$\log(y)$	$\exp(z)$
Log-shift	0	1	0	$\log(y+s)$	$\exp(z)-s$
Dual-shift	0	1	0	$\frac{(y+s)^\lambda-(y+s)^{-\lambda}}{2\lambda}$	$(\lambda z+\sqrt{1+\lambda^2 z^2})^{1/\lambda}-s$
Box-Cox-shift	0	1	0	$\frac{(y+s)^\lambda-1}{\lambda}$	$(1+\lambda z)^{1/\lambda}-s$
Inv-sinus-sqrt	0	1	0	$\sin^{-1}(\sqrt{y})$	$\sin^2(z)$
Inv-sinus-pow	0	1	0	$\sin^{-1}(y^{\frac{1}{\lambda}})$	$\sin^\lambda(z)$
Sinh-sinh	0	1	0	$\sinh(s_2\sinh^{-1}(y)-s_1)$	$\sinh(\frac{\sinh^{-1}(z)+s_1}{s_2})$

For our proposed generalized review, examples of transformations from or adapted from the literature [10, 11] are listed in Table 1 with the bounds added for limited outcomes. Next after, the transformation is denoted κ^{-1} with $z_i = \kappa^{-1}(y_i)$, such that the bounds are implicit and included in the nonlinear function for a lighter notation.

Back-transformation and bias correction

When a transformation is involved, the inverse of the transformation leads to another estimator suitable for the former outcome:

$$\vartheta_{iy} = \kappa(\vartheta_{iz}) \text{ and } \hat{\vartheta}_{iy} = \kappa(\hat{\vartheta}_{iz})\,.$$

Because, $E[\hat{\vartheta}_{iy}] \neq \vartheta_{iy}$, the back-transformation needs to add a multiplicative bias correction denoted $\hat{\rho}$, such that now the estimator becomes:

$$\hat{\vartheta}^*_{iy} \;=\; \hat{\rho}_i\kappa(\hat{\vartheta}_{iz})\,.$$

This is with a similar notation as for the logarithmic transformation in [3], such that $\hat{\rho}_i$ is an estimator (by replacing the true unknown quantities β and σ_u^2 by their estimations) of ρ_i. The later one is defined with the quotient and estimated quotient:

$$\rho_i = \frac{E(\kappa(\vartheta_{iz}))}{E(\kappa(\hat{\vartheta}_{iz}))} \text{ and } \hat{\rho}_i = \frac{\hat{E}(\kappa(\vartheta_{iz}))}{\hat{E}(\kappa(\hat{\vartheta}_{iz}))}\,.$$

Such correction may do the job because one can write then the wanted result as follows:

$$E(\hat{\vartheta}^*_{iy}) = E(\hat{\rho}_i\kappa(\hat{\vartheta}_{iz})) \doteq E(\hat{\rho}_i)E(\kappa(\hat{\vartheta}_{iz})) \doteq E(\kappa(\vartheta_{iz}))\,.$$

In the case of the logarithmic transformation, the expression for the correction in the back-transformation is computed in [3] as follows,

$$\hat{\rho}_{i,\mathrm{SM}} = \exp(0.5(1-\hat{\gamma}_i)\hat{\sigma}_u^2) \text{ with } \kappa() = \exp()\,.$$

But also more advanced expression in [4], while the case box-cox transformation or the logistic-bounded transformation for instance were not adressed in the literature to our knowledge.

Alternative nonlinearities without back-transformation

As seen before, the transformation of the outcome asks for a back-transformation and a bias correction. This seems more justified for an unit-level model where each observed outcome value of each unit is transformed: in the area-level model, the nonlinearities may be directly introduced in the model as made possible by nonlinear mixed models [12, 13, 14, 15, 16]. This is written for instance as follows for a mixed model.

$$\begin{aligned} \mathbf{y} &= \kappa(\mathbf{z}) \\ &= \kappa\left(\mathbf{X}\boldsymbol{\beta} + \mathbf{v}\right) + \mathbf{e}\,. \end{aligned}$$

For the Fay-Herriot model, one may get:

$$\begin{aligned} y_i &= \kappa(z_i) \\ &= \kappa\left(\mathbf{x}_i^T\boldsymbol{\beta} + u_i\right) + \epsilon_i\,. \end{aligned}$$

Because the latent error is differently hidden here, one may prefer to add another error term within the nonlinear function κ, or at least consider a quantile regression (or median one) instead of a normal noise, for more flexibility. This kind of model has been studied in mixed models but seems not considered corrently for small area to our knowledge. Such approach needs to be tested further at least for the Fay-Herriot model in order to compare with the available ones.

Next after, the classical way is discussed in order to correct for the nonlinear transformation of the outcome but in two stages: from the back-transformation itself and from the sampling error of the parameter estimation.

2.3 Bias correction via expectation for the back-transformation

Bias corrections were mostly proposed for unit-level small area models but some attempts exist for area-level models too. Thus, diverse methods for bias correction have been introduced in the literature such as a recent one via integration, but they are often not general on the contrary to our approach.

Expectation from a gaussian distribution

Recently, it has been suggested a bias-corrected estimator after back-transformation in [11] inspired from a recent method in [17]. When denoting $\hat{\sigma}_{iz}^2 = (1-\hat{\gamma}_i)\hat{\sigma}_u^2$, an analytical solution may be available for this inverse sine transformation as follows:

$$\begin{aligned} \hat{\vartheta}_{i,\mathrm{SIN}\frac{1}{2}}^{**} &= E_{\hat{\vartheta}_{iz}}(\sin^2(\hat{\vartheta}_{i,z})) \\ &= \int_{-\infty}^{+\infty} \sin^2(t)\frac{1}{\sqrt{2\pi\hat{\sigma}_{iz}^2}}\exp\left(-\frac{(t-\hat{\vartheta}_{i,z})^2}{2\hat{\sigma}_{iz}^2}\right)dt \\ &= \exp(-2\hat{\sigma}_{iz}^2)\,\sin^2(\hat{\vartheta}_{i,z}) + 0.5\left(1-\exp(-2\hat{\sigma}_{iz}^2)\right)\,. \end{aligned}$$

Here, the closed-form solution is found after classical trigonometry followed by a classical calculus for integrals. One may prefer to avoid a range with the whole real line by restricting the integral to $[0, 0.5\pi]$ as in the available r implementation for this estimator, but the gaussian function may be already near zero for the values outside of the bounds for some parameters: the serie resulting of the periodicity of sinus may be cut by keeping only this interval. The approach of bias correction via an integral was introduced at first for bias correction for the dual power (and eventually Box-Cox) transformation in [17]. A nearly similar term is retrieved for this alternative transformation in the conditional setting as one may write that for the log-transformation:

$$\begin{aligned}
\hat{\vartheta}^{**}_{iy,\mathrm{LOG}} &= E_{\hat{\vartheta}_{iz}}(\exp(\hat{\vartheta}_{i,z})) \\
&= \int_{-\infty}^{+\infty} \exp(t) \frac{1}{\sqrt{2\pi\hat{\sigma}^2_{iz}}} \exp\left(-\frac{(t-\hat{\vartheta}_{i,z})^2}{2\hat{\sigma}^2_{iz}}\right) dt \\
&= \exp(0.5\hat{\sigma}^2_{iz}) \exp(\hat{\vartheta}_{i,z}) \,.
\end{aligned}$$

It is interesting to note the exponential term with a variance which appears in both cases. As explained in [17], this leads to the usual bias correction for this different transformation. More generally for other transformations, an approximated closed-form expression of the integral may be computed analytically from a sum of exponential decaying (or increasing) functions, when denoting $K = 2K'$ without loss of generality,

$$\begin{aligned}
\hat{\vartheta}^{**,int}_{iy,\mathrm{GEN}} &= E_{\hat{\vartheta}_{iz}}(\kappa(\hat{\vartheta}_{i,z})) \\
&= \int_{-\infty}^{+\infty} t \frac{1}{\sqrt{2\pi\hat{\sigma}^2_{iz}}} \exp\left(-\frac{(\kappa^{-1}(t)-\hat{\vartheta}_{i,z})^2}{2\hat{\sigma}^2_{iz}}\right) \left.\frac{d\kappa^{-1}(u)}{du}\right|_{u=t} dt \\
&\doteq \sum_{\ell=1}^{\ell=L} \frac{\alpha_{\ell,\kappa}}{\sqrt{2\pi\hat{\sigma}^2_{iz}}} \int_{-\infty}^{+\infty} \exp\left(-\frac{(t-\hat{\vartheta}_{i,z})^2+2\hat{\sigma}^2_{iz}\omega_{\ell,\kappa}\, t}{2\hat{\sigma}^2_{iz}}\right) dt \\
&= \sum_{\ell=1}^{\ell=L} \alpha_{\ell,\kappa} \exp\left(-\omega_{\ell,\kappa}\hat{\vartheta}_{iz}\right) \exp\left(0.5\omega^2_{\ell,\kappa}\hat{\sigma}^2_{iz}\right) \\
&= \sum_{\ell=1}^{\ell=L} \alpha_{\ell,\kappa} \exp\left(-\omega_{\ell,\kappa}\hat{\vartheta}_{iz}\right) \left\{1 + \sum_{k=1}^{k=K'} \tfrac{1}{k!}(0.5\omega^2_{\ell,\kappa}\hat{\sigma}^2_{iz})^k\right\} + o(\hat{\sigma}^K_{iz}) \\
&= \kappa(\hat{\vartheta}_{iz}) + \sum_{k=1}^{k=K'} \tfrac{1}{(2k)!} \left.\tfrac{d^{2k}\kappa(u)}{du^{2k}}\right|_{u=\hat{\vartheta}_{iz}} (2k-1)!!\,\hat{\sigma}^{2k}_{iz} + o(\hat{\sigma}^K_{iz})
\end{aligned}$$

At the last row, it is recognized the central moments from the normal (or gaussian) distribution for powers even as they are zero otherwise. As explained next paragraph, a small value of $\hat{\sigma}^2_{iz}$ is required only if the first terms are kept. It it not required the hypothesis that the estimate before back-transformation was small because the two exponentials terms come as a product before the linear approximation, but the parameters $\omega_{\ell,\kappa}$ are supposed to be no too big actually. The back-transformation introduced by these authors may be computed with the expression just before for a strictly positive derivative of the back-transformation. For transformations such as the dual power and sinh sinh ones, the derivative may be positive, such that this is relevant because the gaussian function may be again zero for the values outside of definition range of $\kappa()$. This is with the parameters $(\alpha_{\ell,\kappa})$ and $(\omega_{\ell,\kappa})$ herein, with say $\kappa(t) = \sum_\ell \alpha_{\ell,\kappa} e^{-\omega_{\ell,\kappa} t}$. It may be also retrieved the case of the log-transformation when only the first decaying exponential function is kept with $\alpha_{1,\kappa} = 1$ and $\omega_{1,\kappa} = -1$, thus the other terms are removed from the sum. This leads to a general bias-corrected estimator for a back-transformation $\kappa()$ when an enough sharp approximation [18] with a sum of exponentials is available for the function.

Note that because this is the exponential function, and considering the gaussian, one may supposed that there exists indeed an integer K where the expansion is almost equal to the original integral such that anyway the expansion above and the integral are perfectly equivalent. The exponential form allows a higher order expansion in order to retrieve the integration almost exactly, because in the case of the exponential function it is well known that the Taylor serie when taken infinite is true for any value on the real line, with for a truncated serie the remaining term is easy to deal with. Thus one may write that:

$$
\begin{aligned}
& \left| \hat{\vartheta}^{**,int}_{iy,\mathrm{GEN}} - \left[\kappa(\hat{\vartheta}_{iz}) + \sum_{k=1}^{k=K'} \frac{1}{k!} \left. \frac{d^k \kappa(u)}{du^k} \right|_{u=\hat{\vartheta}_{iz}} \hat{\sigma}^{2k}_{iz} \right] \right| \\
= & \left| \left[\sum_{\ell=1}^{\ell=L} \alpha_{\ell,\kappa} \exp\left(-\omega_{\ell,\kappa}(u - 0.5\hat{\sigma}^2_{iz}\omega_{\ell,\kappa})\right) \right]_{u=\hat{\vartheta}_{i,z}} - \left[\kappa(\hat{\vartheta}_{iz}) + \sum_{k=1}^{k=K'} \frac{1}{k!} \left. \frac{d^k \kappa(u)}{du^k} \right|_{u=\hat{\vartheta}_{iz}} \hat{\sigma}^{2k}_{iz} \right] \right| \\
\leq & \sum_{\ell=1}^{\ell=L} \alpha_{\ell,\kappa} \exp\left(-\omega_{\ell,\kappa}\hat{\vartheta}_{iz}\right) \left| \exp\left(0.5\omega^2_{\ell,\kappa}\hat{\sigma}^2_{iz}\right) - \left[1 + \sum_{k=1}^{k=K'} \frac{1}{k!} (0.5\omega^2_{\ell,\kappa}\hat{\sigma}^2_{iz})^k \right] \right| \\
\leq & \kappa(\hat{\vartheta}_{iz}) \frac{(0.5\omega^2_{*,\kappa}\hat{\sigma}^2_{iz})^{K'+1} \exp\left(0.5\omega^2_{*,\kappa}\hat{\sigma}^2_{iz}\right)}{(K'+1)!} \, .
\end{aligned}
$$

With $\omega_{*,\kappa} = \max_{1 \leq \ell \leq L} \omega_{\ell,\kappa}$, the last term converges to zero when K' grows as expected, such that the serie may be truncated for a finite number of term, achieving an accurate approximation of the integral.

Expectation from Taylor serie and first moments

An alternative approximation for a generalized bias correction is with respect to the second order or higher order approximation of the transformation function (see also [19]). This comes from a polynomial approximation of the estimator around the true value, which is obtained with the terms of the Taylor-Young serie. Hence, this suggests an alternative estimator with the central moments involved,

$$
E_{\hat{\vartheta}_{iz}}[\kappa(\hat{\vartheta}_{iz})] \;=\; \kappa(\vartheta_{iz}) + \sum_{k=1}^{k=K} \frac{1}{k!} \left. \frac{d^k \kappa(u)}{du^k} \right|_{u=\vartheta_{iz}} E_{\hat{\vartheta}_{iz}}[(\hat{\vartheta}_{iz} - \vartheta_{iz})^k] + o(E_{\hat{\vartheta}_{iz}}[(\vartheta_{iz} - \hat{\vartheta}_{iz})^K]) \, .
$$

This may be estimated for an alternative estimator of bias correction for the back-transformation as follows:

$$
\begin{aligned}
\hat{E}_{\hat{\vartheta}_{iz}}[\kappa(\hat{\vartheta}_{iz})] \;&=\; \kappa(\hat{\vartheta}_{iz}) + \sum_{k=1}^{k=K} \frac{1}{k!} \left. \frac{d^k \kappa(u)}{du^k} \right|_{u=\hat{\vartheta}_{iz}} \hat{E}_{\hat{\vartheta}_{iz}}[(\hat{\vartheta}_{iz} - \vartheta_{iz})^k] \\
&=\; \left(1 + \frac{1}{\kappa(\hat{\vartheta}_{iz})} \sum_{k=1}^{k=K} \frac{1}{k!} \left. \frac{d^k \kappa(u)}{du^k} \right|_{u=\hat{\vartheta}_{iz}} \hat{E}_{\hat{\vartheta}_{iz}}[(\hat{\vartheta}_{iz} - \vartheta_{iz})^k] \right) \kappa(\hat{\vartheta}_{iz}) \, .
\end{aligned}
$$

Typically the estimations for the moments may be found after the normal asymptotic theory from central limit theorem but there is also available some parametric boostrap [20] instead. This allows several possible estimates such that:

$$
\hat{E}_{\hat{\vartheta}_{iz}}[(\hat{\vartheta}_{iz} - \vartheta_{iz})^k] = \begin{cases} E_*[(\hat{\vartheta}_{iz} - E_*[\hat{\vartheta}_{iz}])^k] & \text{for a (parametric) boostrap estimate for any } k \leq K \, , \\ (2k'-1)!\,!\,\hat{\sigma}^{2k'}_{iz} & \text{for an asymptotic (gaussian) estimate if } k = 2k' \, , \\ 0 & \text{for an asymptotic (gaussian) estimate if } k = 2k'+1 \, . \end{cases}
$$

Here $E_*[.]$ is for the averaging from the b generated versions of $\hat{\vartheta}_{iz}$ according to the methodology from this boostrap. For the exponential form of the integral above, this is different because the asymptotic normality is supposed since the beginning of the definition of the integral such that only one expansion is available. This suggests that the proposed approach with a direct expansion for a generalized back-transformation is more flexible than an integral approach, this is the estimator which is kept in the next sections.

When noting that the estimator in [3] is actually obtained directly from the statistics of interest for the log-transformation as the expectation of $\exp(\vartheta_{iz})$, hence as the expectation of $\kappa(\vartheta_{iz})$ in the general case and considering the other existing bias-corrected estimator, such as the existing ones which are related to [3] for the unit-level model and also the one just described above, a first relevant estimator is this sample version from the direct expectation of the back-transformation. According to the first order expansion, the method with an integral is nearly equivalent to the usual method when the noise is small otherwise a higer expansion may be required.

Instead of looking directly for the multiplicative factor ρ as introduced in [3], one may write before, when keeping only the second-order terms,

$$\begin{aligned} \hat{\vartheta}^{**,tly}_{iy,\text{GEN}} &= \hat{E}_{\hat{\vartheta}_{iz}}[\kappa(\hat{\vartheta}_{iz})] \quad \text{where} \quad & \hat{h}_{iy} &= \hat{h}_{iy}(\hat{\beta},\hat{\sigma}^2_u) \\ &= \hat{h}_{iy}\,\kappa(\hat{\vartheta}_{iz}) & &= 1+0.5\frac{\ddot{\kappa}(\hat{\vartheta}_{iz})}{\kappa(\hat{\vartheta}_{iz})}\hat{\sigma}^2_{iz} \doteq \exp\left(0.5\frac{\ddot{\kappa}(\hat{\vartheta}_{iz})}{\kappa(\hat{\vartheta}_{iz})}\hat{\sigma}^2_{iz}\right). \end{aligned}$$

It is denoted the first and second order derivative of the function for the back-transformation, say respectively $\dot{\kappa} = \frac{d\kappa(u)}{d_u}$ and $\ddot{\kappa} = \frac{d^2\kappa(u)}{d^2u}$ jointly with partial derivarives w.r.t. the parameters. Note that with $(1+u)^{+1} \approx e^{+u}$, it may be preferred the exponential form of the multiplicative term. But a direct way is with an additional term, with the form $\kappa(\hat{\vartheta}_{iz}) + \hat{h}_{iy}(\hat{\beta},\hat{\sigma}^2_u)$ which may be more convenient later as in the next section. The form of the estimator insures that it remains positive, it is for instance as found in [21] in the case of the Box-Cox transformation, $\ddot{\kappa}(\hat{\vartheta}_{iz})/\kappa(\hat{\vartheta}_{iz})$ equal to $(1-\lambda)\kappa^{-\lambda/2}(\hat{\vartheta}_{iz})$, hence a similar expression but less general.

Expectation from Laplace-like approximation and first moments

Another quadratic approximation leads directly to a multiplicative correction, when this is the logarithm of the back-transformation, $\kappa_\ell = \log\kappa$, which is involved. Here, it is denoted $\hat{a}_{i,z} = (1-\hat{\sigma}^2_{iz}\ddot{\kappa}_\ell(\hat{\vartheta}_{i,z})))/2\hat{\sigma}^2_{iz}$, $\hat{b}_{i,z} = -(\hat{\vartheta}_{i,z}+\hat{\sigma}^2_{iz}\dot{\kappa}_\ell(\hat{\vartheta}_{i,z})+\hat{\sigma}^2_{iz}[-\ddot{\kappa}_\ell(\hat{\vartheta}_{i,z})]\hat{\vartheta}_{i,z})/\hat{\sigma}^2_{iz}$, $\hat{c}_{i,z} = (\hat{\vartheta}^2_{i,z}+2\hat{\sigma}^2_{iz}\dot{\kappa}_\ell(\hat{\vartheta}_{i,z})\hat{\vartheta}_{i,z}+\hat{\sigma}^2_{iz}[-\ddot{\kappa}_\ell(\hat{\vartheta}_{i,z})]\hat{\vartheta}^2_{i,z})/2\hat{\sigma}^2_{iz}$ in order to retrieve the usual formula for the integral of $\exp(-(at^2+bt+c))$ from the literature on the gaussian density function. This leads to:

$$\begin{aligned} \hat{\vartheta}^{**,lpc}_{iy,\text{GEN}} &= E_{\hat{\vartheta}_{iz}}[\exp(\kappa_\ell(\hat{\vartheta}_{iz}))] \\ &\doteq E_{\hat{\vartheta}_{iz}}\left[\exp\left(\kappa_\ell(\vartheta_{i,z}) + \dot{\kappa}_\ell(\hat{\vartheta}_{i,z})(\vartheta_{iz}-\hat{\vartheta}_{i,z}) + 0.5\ddot{\kappa}_\ell(\hat{\vartheta}_{i,z})(\vartheta_{iz}-\hat{\vartheta}_{i,z})^2\right)\right] \\ &\doteq \frac{\exp\left(\frac{\hat{b}^2_{iz}}{4\hat{a}_{iz}} - \hat{c}_{iz}\right)}{\sqrt{1-\hat{\sigma}^2_{iz}\ddot{\kappa}_\ell(\hat{\vartheta}_{i,z})}}\kappa(\hat{\vartheta}_{i,z})\,. \end{aligned}$$

The derivatives for the logarithm of the back-transformation are found from the untransformed function, say: $\dot{\kappa}_\ell = \dot{\kappa}/\kappa$ and $\ddot{\kappa}_\ell = \ddot{\kappa}/\kappa - \dot{\kappa}^2/\kappa^2$. As the logarithm is more linear this approximation is expected to be sligthly better than the one above, but the expression suggests the condition $\hat{\sigma}^2_{iz}\ddot{\kappa}_\ell(\hat{\vartheta}_{i,z}) < 1$, hence not considered further herein.

Three different types of resulting estimators

For the linearization in order to get the multiplicative term for $\kappa(\vartheta_{iz})$, one may write the new expression for the derivatives of $\vartheta_{iy} = \kappa(\vartheta_{iz})$, as follows with an usual expectation after a two-order expansion for the three cases which are suggested by the previous section. The three types met above are listed in Table 2 for three back-transformations, such that they share a similar expression multiplicative bias correction.

Table 2: Three generalized back-transformations for a small area-level model.

	Type I	Type II	Type III
ρ_i	$\rho^{(2,\kappa)}_{i,\text{GEN}}$	$\rho^{(2,h)}_{i,\text{GEN}}$	$\rho^{(2,h\kappa)}_{i,\text{GEN}}$
ϑ^{**}_{iy}	$\kappa(\vartheta_{iz})$	$h(\vartheta_{iz}, \beta, \sigma^2_u)$	$h_{iy}\kappa(\vartheta_{iz})$
Exponential form:	$\sum_{\ell=1}^{\ell=L} \alpha_{\ell,\kappa} e^{-\omega_{\ell,\kappa}\vartheta_{iz}}$	$\sum_{\ell=1}^{\ell=L} \alpha_{\ell,\kappa} e^{-\omega_{\ell,\kappa}(\vartheta_{iz}-0.5\sigma^2_{iz}\omega_{\ell,\kappa})}$	$\sum_{\ell=1}^{\ell=L} \alpha_{\ell,\kappa}(1+0.5\sigma^2_{iz}\omega^2_{\ell,\kappa})e^{-\omega_{\ell,\kappa}\vartheta_{iz}}$

The two estimators above have a similar behavior for a small variance. An additional bias correction is introduced at the next section with the final definition for a bias corrected back-transformed mean estimator from the involved small area-level model:

$$\hat{\vartheta}^*_{iy} \;\; = \;\; \hat{\rho}_i\hat{\vartheta}^{**}_{iy}\,.$$

Herein $\hat{\vartheta}^{**}_{iy}$ is from the multiplicative-corrected estimator $\hat{\vartheta}^{**,tly}_{iy,\text{GEN}}$ just above in order to compare with the one without correction. Hence, $\hat{\rho}_i$ is a term different from before because of the additional term $\hat{h}_{iy}$, or $\kappa(\hat{\vartheta}_{iz})$ needs to be replaced implicitely by $\hat{h}_{iy}\kappa(\hat{\vartheta}_{iz})$ in order to keep the same definition. This estimator includes explicitely a first correction from the back-transformation transformation itself before the one from the sampling error which comes as another source of bias. This is in correspondence with the log-transformation extended to the Box-Cox transformation in order to retrieve their respective expressions of their bias-corrected estimators from the current literature. Hence to define $\hat{h}_{iy}$, one has an exponential function and one has a first order approximation of the exponential function while both include a weighted variance term.

3 Bias correction after the back-transformation

When the transformation is inverted a bias appears as explained in the literature [3, 4, 17]. The Fay-Herriot method or area level method is defined directly for the area instead of the units inside the area. This section proposed a general approach for bias correction after a non linear transformation by extending a recent existing work for log-transformation and also by discussing the variance of the resulting back-transformed estimators. For the log transformation, the authors in [4] have suggested two approaches but the first one did not perform well hence is not considered herein. The second one (see also [22] for such multiplicative bias correcting method) is extended to any estimator after back-transformaton as proposed in the previous section. This allows to retrieve the log-transformation as a particular case, before discussing the variance estimation.

3.1 Multiplicative correction for a generalized estimator

For the Box-cox transformation and the logistic-bounded Box-Cox transformation, the former bias correction needs to be updated because the involved function is different. For the more general case, the back-transformation leads to a function with a general expression:

$$\hat{\vartheta}_{iy} \quad = \quad \kappa(\hat{\vartheta}_{iz})\,.$$

In order to avoid to need a new formula for each case, a generalized solution is developped in this subsection before numerical experiments. Note also that the method for bias correction leads to a nearly similar expression than for the log transformation, which is expected.

According to the revisit from above of the existing approaches for bias corrections, a generalized estimator needs to be bias-corrected from the sampling error. Let denote $\hat{\sigma}_{iz}^{(k)} = \hat{E}_{\hat{\vartheta}_{iz}}[(\hat{\vartheta}_{iz} - \vartheta_{iz})^k]$, $\kappa^{(k)}(\vartheta) = \frac{d^k\kappa(u)}{du^k}|_{u=\vartheta}$ and ϕ_κ a constant equal to 0 or 1 (or varying between 0 and $\phi_{max} > 1$ for curve drawing and diagnostic purpose) in order to retrieve the estimator without a first bias correction. The proposed estimator is defined as follows:

$$\hat{\vartheta}_{iy}^{**,a} \quad = \quad \kappa(\vartheta_{iz}) + \phi_\kappa \sum_{k=1}^{k=K} \frac{1}{k!}\,\kappa^{(k)}(\hat{\vartheta}_{iz})\,\hat{\sigma}_{iz}^{(k)}\,.$$

According to the available experiments in [4] the direct correction of the bias from a second order approximation of the expectation from the model parameters reduces the bias. Each transformation function would ask for a new analytical solution on the contrary to a generalized estimator. Using a factorisation would lead to very related [23] estimators and requires the hypothesis of small variance. This generalized estimator $\hat{\vartheta}_{iy}^{**,a}$ is in stake next after.

Analytical expression for the multiplicative factor

With $\phi_\rho \in \{0,1\}$, the expression for the corrective term ρ comes from the derivative as found below.

$$\begin{aligned}
\rho_{i,\mathrm{GEN}} \quad &= \quad \frac{1}{1 + \frac{0.5\phi_\rho}{\vartheta_{iy}^{**}} Tr\left[\begin{pmatrix} \frac{\partial^2[\vartheta_{iy}^{**}]}{\partial\beta\partial\beta^T} & \frac{\partial^2[\vartheta_{iy}^{**}]}{\partial\beta\partial\sigma_u^2} & \frac{\partial^2[\vartheta_{iy}^{**}]}{\partial\beta\partial\lambda} \\ \frac{\partial^2[\vartheta_{iy}^{**}]}{\partial\sigma_u^2\partial\beta^T} & \frac{\partial^2[\vartheta_{iy}^{**}]}{\partial^2\sigma_u^2} & \frac{\partial^2[\vartheta_{iy}^{**}]}{\partial\sigma_u^2\partial\lambda} \\ \frac{\partial^2[\vartheta_{iy}^{**}]}{\partial\lambda\partial\beta^T} & \frac{\partial^2[\vartheta_{iy}^{**}]}{\partial\lambda\partial\sigma_u^2} & \frac{\partial^2[\vartheta_{iy}^{**}]}{\partial^2\lambda^2} \end{pmatrix}\begin{pmatrix} V_{\hat{\beta}} & & \\ & V_{\hat{\sigma}_u^2} & \\ & & V_{\hat{\lambda}} \end{pmatrix}\right]} \\
&\approx \quad \frac{1}{1 + \frac{0.5\phi_\rho}{\vartheta_{iy}^{**}}\left[Tr\left(\frac{\partial^2[\vartheta_{iy}^{**}]}{\partial\beta\partial\beta^T} V_{\hat{\beta}}\right) + \left(\frac{\partial^2[\vartheta_{iy}^{**}]}{\partial\sigma_u^2} V_{\hat{\sigma}_u^2}\right)\right]}\,.
\end{aligned}$$

Here eventually a parameter λ is estimated such that for the Box-Cox transformation but its influence is neglected herein. It is also denoted $Tr()$ for the matricial operator named the trace. The derivatives for the linear estimator $\vartheta_{iz} = x_i^T\beta + \gamma_i(y_i - x_i^T\beta)$ may be written as follows:

$$\begin{aligned}
\frac{\partial[\vartheta_{iz}]}{\partial\beta} \quad &= \quad (1-\gamma_i)x_i \\
\frac{\partial[\vartheta_{iz}]}{\partial\sigma_u^2} \quad &= \quad \tau_i^{-1}(1-\gamma_i)(y_i - x_i^T\beta) \\
\frac{\partial^2[\vartheta_{iz}]}{\partial\sigma_u^2} \quad &= \quad -2\tau_i^{-2}(1-\gamma_i)(y_i - x_i^T\beta)\,.
\end{aligned}$$

For the two estimators above, the derivatives may be computed as follows.

$$
\begin{aligned}
\frac{\partial^2[\vartheta_{iy}]}{\partial\beta\partial\beta^T} &= \kappa^{(2)}(\vartheta_{iz})\frac{\partial[\vartheta_{iz}]}{\partial\beta}\frac{\partial[\vartheta_{iz}]}{\partial\beta^T} \\
\frac{\partial^2[\vartheta_{iy}]}{\partial\sigma_u^2} &= \kappa^{(2)}(\vartheta_{iz})\left(\frac{\partial[\vartheta_{iz}]}{\partial\sigma_u^2}\right)^2 + \kappa^{(1)}(\vartheta_{iz})\frac{\partial^2[\vartheta_{iz}]}{\partial^2\sigma_u^2} \\
\frac{\partial^2[\vartheta_{iy}^{**,a}]}{\partial\beta\partial\beta^T} &= \left\{\kappa^{(2)}(\vartheta_{iz}) + \phi_\kappa\sum_{k=1}^{k=K}\frac{1}{k!}\sigma_{iz}^{(k)}\kappa^{(\mathrm{k+1})}(\vartheta_{iz})\right\}\frac{\partial[\vartheta_{iz}]}{\partial\beta}\frac{\partial[\vartheta_{iz}]}{\partial\beta^T} \\
\frac{\partial^2[\vartheta_{iy}^{**,a}]}{\partial^2\sigma_u^2} &= \frac{\partial^2[\kappa(\vartheta_{iz})]}{\partial^2\sigma_u^2} + \phi_\kappa\sum_{k=1}^{k=K}\frac{1}{k!}\left\{\sigma_{iz}^{(k)}\frac{\partial^2[\kappa^{(\mathrm{k})}(\vartheta_{iz})]}{\partial^2\sigma_u^2} + 2\frac{\partial[\kappa^{(\mathrm{k})}(\vartheta_{iz})]}{\partial\sigma_u^2}\frac{\partial[\sigma_{iz}^{(k)}]}{\partial\sigma_u^2} + \kappa^{(\mathrm{k})}(\vartheta_{iz})\frac{\partial^2[\sigma_{iz}^{(k)}]}{\partial^2\sigma_u^2}\right\}
\end{aligned}
$$

The derivatives for ϑ_{iz} and $\sigma_{iz}^{(k)}$ w.r.t. β and σ_u^2 are replaced in order to achieve the computations.A perspective to the reader here may be to plug-in estimators from a bias correction for the linear model from the transformed outcome [24] in order to improve further the results.

Retrieving the case of the log-transformation

In the case of the logarithm transformation, the back-transformation leads to:

$$
\begin{aligned}
\hat{\vartheta}_{iy} &= \kappa(\hat{\vartheta}_{iz}) \\
&= \exp(\hat{\vartheta}_{iz})\,.
\end{aligned}
$$

When $G_i = 2(y_i - x_i^T\beta) + \sigma_{e_i}^2$, in [4] it was proven:

$$
\begin{aligned}
\frac{\partial^2[\vartheta_{iy}]}{\partial\beta\partial\beta^T} &= Tr\left(A_i V_{\hat{\beta}}\right)\vartheta_{iy} \text{ with } A_i = (1-\gamma_i)^2 x_i^T x_i \\
\frac{\partial^2[\vartheta_{iy}]}{\partial\sigma_u^2} &= \left(B_i V_{\hat{\sigma}_u}\right)\vartheta_{iy} \text{ with } B_i = \frac{\sigma_{e_i}^2}{\tau_i^3}G_i\left(\frac{\sigma_{e_i}^2}{4\tau_i}G_i - 1\right)\,.
\end{aligned}
$$

With the approximation $1+u \approx e^u$, this leads to a second multiplicative term for the bias correction:

$$
\rho_{i,\mathrm{LOG}}^{(2)} = \exp\left[-0.5\left\{Tr\left(A_i V_{\hat{\beta}}\right) + B_i V_{\hat{\sigma}_u}\right\}\right] \qquad \text{for the Log case}\,.
$$

The same expression is expected to be found with the proposed generalized expression just above. This justifies the serie truncated at K terms instead of just the two first ones, because otherwise an hypothesis of small variance would be wanted here, which is not anymore required. Note that for the Box-Cox transformation, the k-st derivative is equal to $\prod_{\ell=1}^{k-1}(1-\ell\lambda)(1+\lambda\vartheta_{iz})^{\frac{1-k\lambda}{\lambda}}$ instead of just $\exp(\vartheta_{iz})$.
In summary the proposed estimator is nearly equal to the integral approach but with an additional multiplicative bias correction: its simpler expression allows a direct computation of the variance as explained next subsection.

3.2 Variance estimation for the generalized estimators

As a statistics depending on a sample, the proposed estimators have an expectation and a variance. In small area estimation, it is often computed the mean squared error in order to know if the estimator is enough precise. The more usual method is via an unbiased mse [25] introduced by Prasad and Rao, here left as a perspective, as bias is the main concern herein. In this subsection, a recent approach by Deville via linearizations and influence functions [26] is implemented. See also [27] for further discussion and justification of the method. To follow this idea, one has to focus on the estimating equations which are written as sums from the observed data vectors. An additional parameter may be estimated for the transformation in this case. Also empirical likelihood [28] is often shown to perform better via weighting, an additional estimating equation for the unknown lagrange multiplier coming from the estimation of the weights is required in this case. Such that it may be written when λ denotes the additional parameters,

$$\hat{\vartheta}_i = g_i(\hat{\beta}, \hat{\sigma}_u^2, \hat{\lambda})\,.$$

With $\theta = (\beta^T, \sigma_u^2, , \lambda)$ the following approximation may be involved.

$$\hat{\vartheta}_i - \vartheta_i \doteq \left.\frac{\partial g_i}{\partial \beta}\right|_{\theta}^{T} (\hat{\beta} - \hat{\beta}) + \left.\frac{\partial g_i}{\partial \sigma_u^2}\right|_{\theta}^{T} (\hat{\sigma}_u^2 - \sigma_u^2) + \left.\frac{\partial g_i}{\partial \lambda}\right|_{\theta}^{T} (\hat{\lambda} - \lambda)\,.$$

For λ, β and σ_u^2, it is supposed the two following estimating equations, available from the numerical algorithms for the inference,

$$\begin{array}{lcl} \frac{1}{m}\sum_j^m \phi_{j,\beta}(\hat{\beta}, \hat{\sigma}_u^2, \hat{\lambda}) & = & 0 \\ \frac{1}{m}\sum_j^m \phi_{j,\sigma_u^2}(\hat{\beta}, \hat{\sigma}_u^2, \hat{\lambda}) & = & 0 \\ \frac{1}{m}\sum_j^m \phi_{j,\lambda}(\hat{\beta}, \hat{\sigma}_u^2, \hat{\lambda}) & = & 0\,. \end{array}$$

Similarly, with $H_\beta = E[\frac{1}{m} \left.\frac{\partial \sum_j \phi_{j,\beta}}{\partial \beta}\right|_{\theta}^{T}]$, $H_{\sigma_u^2} = E[\frac{1}{m} \left.\frac{\partial \sum_j \phi_{j,\sigma_u^2}}{\partial \sigma_u^2}\right|_{\theta}^{T}]$ and $H_\lambda = E[\frac{1}{m} \left.\frac{\partial \sum_j \phi_{j,\lambda}}{\partial \lambda}\right|_{\theta}^{T}]$ a linearization leads to:

$$\begin{array}{lcl} \sum_j^m \phi_{j,\beta}(\beta, \sigma_u^2, \lambda) + H_\beta(\hat{\beta} - \hat{\beta}) & \doteq & 0 \\ \sum_j^m \phi_{j,\sigma_u^2}(\beta, \sigma_u^2, \lambda) + H_{\sigma_u^2}(\hat{\sigma}_u^2 - \sigma_u^2) & \doteq & 0 \\ \sum_j^m \phi_{j,\lambda}(\beta, \sigma_u^2, \lambda) + H_\lambda(\hat{\lambda} - \lambda) & \doteq & 0\,. \end{array}$$

This may induce the following linearization:

$$\hat{\vartheta}_i - \vartheta_i \doteq \sum_j^m I_{j,i}^{lin}\,,$$

where,

$$I_{j,i}^{lin} = \left\{ \left.\frac{\partial g_i}{\partial \beta}\right|_{\theta}^{T} H_\beta^{-1}\phi_{j,\beta}(\beta, \sigma_u^2) + \left.\frac{\partial g_i}{\partial \sigma_u^2}\right|_{\theta}^{T} H_{\sigma_u^2}^{-1}\phi_{j,\sigma_u^2}(\beta, \sigma_u^2) + \left.\frac{\partial g_i}{\partial \lambda}\right|_{\theta}^{T} H_\lambda^{-1}\phi_{j,\lambda}(\beta, \lambda) \right\}\,.$$

Finally, an estimation of the variance may be obtained as the following sum,

$$\hat{V}_{\hat{\vartheta}_i}^{lin} = \frac{1}{m(m-1)} \sum_{j=1}^m (\hat{I}_{j,i}^{lin} - \bar{I}_i^{lin})(\hat{I}_{j,i}^{lin} - \bar{I}_i^{lin})^T \quad \text{with} \quad \bar{I}_i^{lin} = \frac{1}{m}\sum_{j=1}^m \hat{I}_{j,i}^{lin}.$$

Note that this approach allows to compare with a resampling, with typically the parametric boostrap which is well suited for a model with parameters by making the hypothesis that the chosen model is relevant.

4 Conclusion

Herein, it is suggested, for a small area-level model, several multiplicative (and additive) bias corrections for the Box-Cox transformation with a generalized setting which may be relevant for other nonlinear transformations.
Perspectives for the reader are an extension to the unit-level model and the multivariate area-level model. Computational bias corrections are required in order to compare with the analytical solutions. An analytical expression or a computational algorithm for the mean squared error after such bias correction is also wanted. Also the parameter estimator may be corrected before the back-transformation for a better result with a smaller number of areas. Another direction may be to correct the bias from the nonlinear estimator without back-transformation and to compute the resulting mean squared error with an analytical or a computational method.

References

[1] Jiming Jiang and P. Lahiri, "Mixed model prediction and small area estimation," *TEST*, vol. 15, no. 1, pp. 1–96, Jun 2006.

[2] Matthew J. Gurka, Lloyd J. Edwards, Keith E. Muller, and Lawrence L. Kupper, "Extending the Box-Cox Transformation to the Linear Mixed Model," *Journal of the Royal Statistical Society Series A: Statistics in Society*, vol. 169, no. 2, pp. 273–288, 10 2005.

[3] Eric V. Slud and Tapabrata Maiti, "Mean-Squared Error Estimation in Transformed Fay-Herriot Models," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 68, no. 2, pp. 239–257, 03 2006.

[4] Hukum Chandra, Kaustav Aditya, and Sushil Kumar, "Small-area estimation under a log-transformed area-level model," *Journal of Statistical Theory and Practice*, vol. 12, no. 3, pp. 497–505, Sep 2018.

[5] Cristian L. Bayes, Jorge L. Bazán, and Mário de Castro, "A quantile parametric mixed regression model for bounded response variables," *Statistics and Its Interface*, vol. 10, no. 3, pp. 483–493, 2017.

[6] Emmanuel Lesaffre, Dimitris Rizopoulos, and Roula Tsonaka, "The logistic transform for bounded outcome scores," *Biostatistics*, vol. 8, no. 1, pp. 72–85, 04 2006.

[7] Anthony C. Atkinson, Marco Riani, and Aldo Corbellini, "The Box-Cox Transformation: Review and Extensions," *Statistical Science*, vol. 36, no. 2, pp. 239 – 255, 2021.

[8] Robert E. Fay and Roger A. Herriot, "Estimates of income for small places: An application of james-stein procedures to census data," *Journal of the American Statistical Association*, vol. 74, no. 366, pp. 269–277, 1979.

[9] Gauri Sankar Datta, Tatsuya Kubokawa, Isabel Molina, and J. N. K. Rao, "Estimation of mean squared error of model-based small area estimators," *TEST*, vol. 20, no. 2, pp. 367–388, Aug 2011.

[10] Natalia Rojas-Perilla, Sören Pannier, Timo Schmid, and Nikos Tzavidis, "Data-Driven Transformations in Small Area Estimation," *Journal of the Royal Statistical Society Series A: Statistics in Society*, vol. 183, no. 1, pp. 121–148, 07 2019.

[11] Sandra Hadam, Nora Würz, Ann-Kristin Kreutzmann, and Timo Schmid, "Estimating regional unemployment with mobile network data for functional urban areas in germany," *Statistical Methods & Applications*, Sep 2023.

[12] José C. Pinheiro and Douglas M. Bates, *Mixed-Effects Models in S and S-PLUS*, Springer, New York, 2000.

[13] E. Kuhn and M. Lavielle, "Maximum likelihood estimation in nonlinear mixed effects models," *Computational Statistics & Data Analysis*, vol. 49, no. 4, pp. 1020–1038, 2005.

[14] Shawn X. Meng and Shongming Huang, "Improved Calibration of Nonlinear Mixed-Effects Models Demonstrated on a Height Growth Function," *Forest Science*, vol. 55, no. 3, pp. 238–248, 06 2009.

[15] Julie Bertrand, Emmanuelle Comets, Marylore Chenel, and France Mentré, "Some alternatives to asymptotic tests for the analysis of pharmacogenetic data using nonlinear mixed effects models," *Biometrics*, vol. 68, no. 1, pp. 146–155, 2012.

[16] Thu Thuy Nguyen and France Mentré, "Evaluation of the fisher information matrix in nonlinear mixed effect models using adaptive gaussian quadrature," *Computational Statistics & Data Analysis*, vol. 80, pp. 57–69, 2014.

[17] Shonosuke Sugasawa and Tatsuya Kubokawa, "Transforming response values in small area prediction," *Computational Statistics & Data Analysis*, vol. 114, pp. 47–60, 2017.

[18] Robert J McGlinn, "Uniform approximation of completely monotone functions by exponential sums," *Journal of Mathematical Analysis and Applications*, vol. 65, no. 1, pp. 211–218, 1978.

[19] James G. MacKinnon and Anthony A. Smith, "Approximate bias correction in econometrics," *Journal of Econometrics*, vol. 85, no. 2, pp. 205–230, 1998.

[20] Tatsuya Kubokawa and Bui Nagashima, "Parametric bootstrap methods for bias correction in linear mixed models," *Journal of Multivariate Analysis*, vol. 106, pp. 1–16, 2012.

[21] Xiaojun Pu and Michael Tiefelsdorf, *A Variance-Stabilizing Transformation to Mitigate Biased Variogram Estimation in Heterogeneous Surfaces with Clustered Samples*, pp. 271–280, 01 2017.

[22] Peter Hall and Tapabrata Maiti, "On parametric bootstrap methods for small area prediction," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 68, no. 2, pp. 221–238, 2006.

[23] Wei Sheng Zeng and Shou Zheng Tang, "Bias correction in logarithmic regression and comparison with weighted regression for nonlinear models," *Nature Precedings*, Dec 2011.

[24] Xihong Lin and Norman E. Breslow, "Bias correction in generalized linear mixed models with multiple components of dispersion," *Journal of the American Statistical Association*, vol. 91, no. 435, pp. 1007–1016, 1996.

[25] N. G. N. Prasad and J. N. K. Rao, "The estimation of the mean squared error of small-area estimators," *Journal of the American Statistical Association*, vol. 85, no. 409, pp. 163–171, 2023/10/01/ 1990.

[26] Deville J.-C, "Variance estimation for complex statistics and estimators: Linearization and residual techniques," *Survey Methodology*, vol. 25, no. 2, pp. 193–203, December 1999.

[27] Di Shu, Jessica G. Young, Sengwee Toh, and Rui Wang, "Variance estimation in inverse probability weighted cox models," *Biometrics*, vol. 77, no. 3, pp. 1101–1117, 2021.

[28] Jin Qin and Jerry Lawless, "Empirical likelihood and general estimating equations," *Ann. Statist.*, vol. 22, no. 1, pp. 300–325, 1994.