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

Rodolphe Priam

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

R. Priam *

January 21, 2024

Abstract

It is proposed and compared new generalized estimators for small area with a Fay-Herriot model and a nonlinear transformation of the outcome within a finite interval. Different bias corrections and the estimation of the variance or mean squared error are presented. The methods are mainly related to an integral solution in closed-form via a 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. Other back-transformations discussed are a recent approach via an exact expectation from the Gaussian distribution, a direct expansion via a Taylor-Young serie, and a Gauss-Legendre quadrature. A transformation related to an extended log-transformation, when 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

*rpriam@gmail.com.

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.

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:

$$\begin{aligned} \mu_i &\sim N(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma_u^2) \\ z_i|\mu_i &\sim N(\mu_i, \sigma_{e_i}^2) \\ z_i &\sim N(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma_{e_i}^2 + \sigma_u^2)\,. \end{aligned}$$

Note that a limit with this notation is to not show the random effect explicitely.

Estimators

The best (or Bayes) estimator of μ_i is under squared error loss as follows, when one denotes the fraction $\gamma_i = \gamma_i(\sigma_u^2) = \sigma_{e_i}^2(\sigma_u^2 + \sigma_{e_i}^2)^{-1}$ for σ_u^2 given. Say the mean estimator from small area is obtained as:

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

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:

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

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 area estimation.

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,

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

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:

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

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. Here, PR, FH, ML and REML standands for Prasad-Rao, Fay-Herriot, Maximum Likelihood and Restricted (or residual, or reduced) Maximum Likelihood with respective estimators: $\hat{\sigma}^{*2}_{u,\mathrm{PR}}$, $\hat{\sigma}^{*2}_{u,\mathrm{FH}}$, $\hat{\sigma}^{*2}_{u,\mathrm{ML}}$, and $\hat{\sigma}^{*2}_{u,\mathrm{REML}}$. For the estimation of σ^2_u, the usual different approaches are as follows.

Name	Method (estimating equation)
PR	$\hat{\sigma}^{*2}_{u,\mathrm{PR}} = \dfrac{1}{m-p}\sum_{i=1}^{m}\left[(y_i - \mathbf{x}_i^T\hat{\boldsymbol{\beta}}_{LS})^2 - \sigma^2_{e_i}(1-\mathbf{x}\mathbf{x}_i)\right]$
FH	$\sum_{i=1}^{m}\dfrac{1}{\sigma^2_u+\sigma^2_{e_i}}\left[y_i - \mathbf{x}_i^T\tilde{\boldsymbol{\beta}}(\sigma^2_u)\right]^2 = m-p$
ML	$\sum_{i=1}^{m}\dfrac{\left[y_i - \mathbf{x}_i^T\tilde{\boldsymbol{\beta}}(\sigma^2_u)\right]^2}{(\sigma^2_u+\sigma^2_{e_i})^2} - \dfrac{1}{(\sigma^2_u+\sigma^2_{e_i})} = 0$
$REML$	$\sum_{i=1}^{m}\dfrac{\left[y_i - \mathbf{x}_i^T\tilde{\boldsymbol{\beta}}(\sigma^2_u)\right]^2 - (\sigma^2_u+\sigma^2_{e_i}) + \mathbf{x}_i^T\left(\sum_l \dfrac{1}{\sigma^2_u+\sigma^2_{e_l}}\mathbf{x}_l\mathbf{x}_l^T\right)^{-1}\mathbf{x}_i}{(\sigma^2_u+\sigma^2_{e_i})^2} = 0$

These different methods for the estimation of σ^2_u 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 $\hat{\sigma}^2_u$ 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 $\lambda = 0$,

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

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. Herein, a

Table 1: Generalized transformation for skewed and bounded outcomes

Name	a	b	c	$\kappa_\xi^{-1}(y)$	$\kappa_\xi(z)$
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	c	$\exp(z)$	$\frac{a+\exp(y)}{1+c\exp(y)}$
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	$\sinh(s_2 \sinh^{-1}(y) - s_1)$	$\sinh(\frac{\sinh^{-1}(z)+s_1}{s_2})$
Asinh-Exp	0	1	0	$\dfrac{e^y + \varrho(\lambda)e^{-y}}{2}$	$\log(z + \sqrt{z^2 + \varrho(\lambda)})$

transformation depending on a parameter is used in the experiments for positive and negative outcome but related to the log-transformation. This is a reparameterization of asinh with $\varrho(\lambda) \in [0;1]$ and $\varrho(0) = 0$ with for instance $\varrho(\lambda) = \tanh(\lambda^2/4)$ or $\varrho(\lambda) = 1/(1+\exp(\lambda))$, or eventually unbounded $\varrho(\lambda) = \lambda^2$, such that $z = \log(y + \sqrt{y^2 + \varrho(\lambda)})$ which is inverted with the function in Table 3 and called Asinh-Exp at the last row. The functions exponential and inverse hyperbolic sine are intensively studied, hence a mixture of both transformations may be convenient. Note that comparing the results from several concurrent transformations and choosing the best one is the best practice in order to get a model fitting as good as possible.

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.

Alternative nonlinearities

As seen before, the transformation of the outcome asks for a back-transformation and a bias correction. This may be more justified for an unit-level model where each observed outcome value of each unit is transformed. But also some dataset for this area-level model is obtained after a mean estimation from unit-level observations where the transformation may be applied before.

- 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.

- In the area-level model, the values for z_i may be obtained from a survey estimator and similarly for $\sigma_{e_i}^2$, such as instead of applying the transformation to z_i one may prefer alternative mean values in the whole area h_i where the unit-level observations are found:

$$z_i^\kappa = \sum_{j \in h_i} w_{ij} \kappa^{-1}(y_{ij})\,,$$

  instead of the more usual means which are implicitely involved in the next sections with z_{ij} keep unstransformed while this is z_i which is transformed. Otherwise, a back-transformation is also required and eventually approximated as if the transformation was applied to the mean directly, and out of the scope herein.

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.

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:

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

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^2_u 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,

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

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. Note that in [3] it has been also proposed for the log-transformation several estimators for the mean squared error which is unbiased at second-order. But except from numerical algorithms, a more general expression for any estimator is not available to our knowledge except in [17] whose mean estimator is discussed next section.

3 Bias correction for back-transformation via expectation

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. In this section, it is reviewed several way to deal with the expectation from a random variable in order to deduce related estimators. In the more general case, one writes $\hat{\vartheta}_{iy} = \kappa(\hat{\vartheta}_{iz})$ for the back-transformation which is corrected at a first stage as:

$$\begin{aligned} \hat{\vartheta}^{**}_{iy} &= \hat{E}_{\hat{\vartheta}_{iz}}[\kappa(\hat{\vartheta}_{iz})] \\ &= \int_{-\infty}^{+\infty} t \frac{1}{\sqrt{2\pi\hat{\sigma}^2_{iz}}} \exp\left(-\frac{(\kappa^{-1}(t)-\hat{\vartheta}_{iz})^2}{2\hat{\sigma}^2_{iz}}\right) \frac{d\kappa^{-1}(u)}{du}\bigg|_{u=t} dt \,. \end{aligned}$$

In order to avoid to need a new formula for each case, a generalized solution is developped in this subsection before numerical experiments for a second stage of bias correction. Note also that the method for bias correction leads often below to an expression which is similar to the one for the log-transformation when $\kappa = \exp$, with also an exponential and the variance of the estimator $\hat{\sigma}_{iz}$. If the integral w.r.t. the density function was proposed recently, one may prefer a direct expansion via a Taylor-Young serie, or even a numerical quadrature which allows a direct computation for the derivatives w.r.t. the parameters as required for instance in variance estimation. This also suggests the use of the empirical distribution with weights w_ℓ and Dirac distributions $\mathbf{1}_{\vartheta_\ell}$ instead of the asymptotic one with a Gaussian distribution $N(\vartheta_{iz}, \sigma^2_{iz})$ of mean ϑ_{iz} and variance σ^2_{iz} in the involved expectations. Such as two situations result from a choice for the approximation of the distribution of $\hat{\vartheta}_{iz}$ depending of the size of the available data sample and the chosen previously distribution for the linear model:

$$\hat{\vartheta}_{iz} \sim \begin{cases} N(\vartheta_{iz}, \sigma^2_{iz}) & \text{for asymptotic or exact distribution}\,, \\ \sum_\ell \alpha_\ell \mathbf{1}_{\omega_\ell} & \text{for empirical or sampled distribution}\,. \end{cases}$$

In this section, each one leads to very different estimators with different properties as expected: next subsections present several estimators for the back-transformation. The need for an approximation/alternative to the exact integral with a Gaussian distribution is explained by numerical reasons in simulation for running time reduction, or analytical reasons when the estimator is studied for its bias or its variance, and otherwise eventually by no trusting the parametric hypothesis.

3.1 Expectation from a Gaussian distribution

In this subsection, it is reviewed and proposed several method for computing the back-transformation, with usual methods and less usual ones. The resulting estimators are expected to behave differently for the variance.

Case of the exact estimator

Originally the estimator $\hat{\vartheta}^{**,int}_{iy,\mathrm{GEN}}$ in [17] has been kept as an integral without its usual second-order approximation. 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], but also in a general setting with analytical proofs for an estimator the mean square error unbiased at second-order. As examples, two cases are in closed-form while usually the integral needs to be approximated. In the case from the log-transformation, an usual exponential term for back-transformation is retrieved with the integral approach above in the conditional setting as one denotes that for the log-transformation the estimator $\hat{\vartheta}^{**}_{iy,\mathrm{LOG}}$. In the case from the transformation asinh-exp, the expression is for extending the sinh and logarithm transformation such that one may write that the estimator $\hat{\vartheta}^{**}_{iy,\mathrm{AE}}$. More formally, the expressions for the generic case and the two examples are as follows:

$$\begin{aligned} \hat{\vartheta}^{**,int}_{iy,\mathrm{GEN}} &= E_{\hat{\vartheta}_{iz}}(\kappa(\hat{\vartheta}_{iz})) \\ &\stackrel{\text{def}}{=\!=} \int_{-\infty}^{+\infty} \kappa(t) \frac{1}{\sqrt{2\pi\hat{\sigma}^2_{iz}}} \exp\left(-\frac{(t-\hat{\vartheta}_{iz})^2}{2\hat{\sigma}^2_{iz}}\right) dt \\ &\approx \left(1 + 0.5\frac{\ddot{\kappa}(\hat{\vartheta}_{iz})\hat{\sigma}^2_{iz}}{\kappa(\hat{\vartheta}_{iz})}\right) \kappa(\hat{\vartheta}_{iz})\,. \end{aligned}$$

$$\begin{aligned} \hat{\vartheta}^{**}_{iy,\mathrm{LOG}} &= \exp(0.5\hat{\sigma}^2_{iz})\, \exp(\hat{\vartheta}_{iz}) \\ \hat{\vartheta}^{**}_{iy,\mathrm{AE}} &= \exp(0.5\hat{\sigma}^2_{iz})\, \frac{\exp(\hat{\vartheta}_{iz}) + \varrho(\lambda)\exp(-\hat{\vartheta}_{iz})}{2}\,. \end{aligned}$$

This is the usual bias correction for the log-transformation. It is interesting to note the exponential term with a variance which appears in all these three cases even without an exponential function involved but the estimators above have different forms.

Case from numerical integrations

The most usual approach in order to approximate the integral for the expectation is via a quadrature[1]. For instance, this is written, considering given the weights α_ℓ and nodes ω_ℓ, as follows for the back-transformation:

$$\begin{aligned}\hat{\vartheta}^{**,qud}_{iy,\mathrm{GEN}} &\stackrel{\text{def}}{=\!=} \int_{-\infty}^{+\infty} \kappa(\hat{\vartheta}_{iz} + \sqrt{2}\hat{\sigma}_{zi}t) \sum_{\ell=1}^{\ell=L} \alpha_\ell \mathbf{1}_{\omega_\ell}(t)dt \\ &= \tfrac{1}{\sqrt{\pi}} \sum_{\ell=1}^{\ell=L} \alpha_\ell\, \kappa(\hat{\vartheta}_{iz} + \sqrt{2}\hat{\sigma}_{zi}\omega_\ell)\,.\end{aligned}$$

This is originally for numerical integration but this approximation is also used for variance estimation in the literature of mixed model such as in psychometrics or in model inference. It may be worth to compare this approximation with the ones presented above in the context of small area estimation. On the contrary to most of the approaches herein, the hypothesis of a gaussian distribution is not required when nodes and weights are available. For instance a boostrap of the statistic $\hat{\vartheta}_{iz}$ leads to an empirical distribution which may be approached by a mixture of Dirac distributions leading to such expression, but this seems not studied in the literature and out of the scope herein. This allows several possible estimates with the following weights and nodes:

$$\{(\omega_\ell, \alpha_\ell), 1 \leq \ell \leq L\} = \begin{cases} \text{from a standard quadrature for gaussian distribution}\,, \\ \text{from an empirical boostrap for empirical distribution}\,, \\ \text{from a Monte-Carlo sampling for a given distribution}\,. \end{cases}$$

Most of the transformations are for positive values and eventually a bounded interval: this can be handled via a Monte-Carlo (or quasi one) sampling despite the slow rate of convergece $O(n^{-0.5})$. Otherwise, for real values in finite intervals, the Gauss-Legendre and Gauss-Chebyshev are relevant[2] while for unbounded positive real values, Gauss-Laguerre is available. Note that an adaptative variant is likelily to better perform and improve the results, for instance an hybrid approah may find intervals for a stratified sampling from an adaptive Simpson or Trapezoidal rule before the stratified Monte-Carlo using the same strata. Despite that the function κ is monotone, an usual sampling from the density function f_ν did not performed well with about 10^7 generated samples in a first experiment. The Monte-Carlo may be slow in practice when implemented with an interpreted computer language, hence a Gauss-Legendre rule is preferred herein. when one denote $c = 0.5(b-a)$, and $d = 0.5(a+b)$ for the usual change of variable, the previous expression above becomes:

$$\begin{aligned}\hat{\vartheta}^{**,qgl}_{iy,\mathrm{GEN}} &\doteq \frac{c}{\sqrt{2\pi\hat{\sigma}^2_{iz}}} \int_{-1}^{1} \kappa\,(ct+d)\, \exp\left(-\tfrac{1}{2\hat{\sigma}^2_{iz}}\left(ct+d-\hat{\vartheta}_{iz}\right)^2\right) dt \\ &\stackrel{\text{def}}{=\!=} \frac{c}{\sqrt{2\pi\hat{\sigma}^2_{iz}}} \sum_{\ell=1}^{\ell=L} \alpha_\ell\, \kappa\,(c\omega_\ell+d)\, \exp\left(-\tfrac{1}{2\hat{\sigma}^2_{iz}}\left(c\omega_\ell+d-\hat{\vartheta}_{iz}\right)^2\right)\,.\end{aligned}$$

[1] `https://en.wikipedia.org/wiki/Gauss-Hermite_quadrature`

[2] `https://en.wikipedia.org/wiki/Gaussian_quadrature`

Finally, this covers the three possible cases of intervals $]-\infty;+\infty[$,$[0;+\infty[$,and $[a;b]$ for the back-transformation when the infinite bounds are truncated to large reals. The literature insures that the method is very accurate, when denoting an unknown value $t_{-1}^{+1} \in [-1;1]$ it has been shown that as found in any textbook on integral approximation which equivalent for K big from the Stirling formula to:

$$\hat{\vartheta}_{iy,\mathrm{GEN}}^{**,int} - \hat{\vartheta}_{iy,\mathrm{GEN}}^{**,qgl} \sim \left(\frac{e}{4K}\right)^{2K} \frac{\kappa^{(2K)}(t_{-1}^{+1})}{\sqrt{4\pi K^3}}\,.$$

With this estimator, more flexibility allows to avoid the truncated integral such as in a previous paragraph before, and it is also fast to compute hence it is more convenient for several reasons.

3.2 Expectation from a Gaussian distribution with sums of exp

Generally for any transformations with good properties, an approximated closed-form expression of the integral for bias-transformation may be computed analytically from a sum of exponential decaying (or increasing) functions. The back-transformation performed by an expectation may be computed with the expression just below 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 $\alpha_\ell = (\alpha_{1,\kappa}, \cdots, \alpha_{\ell,\kappa}, \alpha_{L,\kappa})$ and $\omega_\ell = (\omega_{1,\kappa}, \cdots, \omega_{\ell,\kappa}, \cdots, \omega_{L,\kappa})$ herein, thus:

$$\widehat{\kappa}(u) = \sum_{\ell} \alpha_{\ell,\kappa} e^{-u\omega_{\ell,\kappa}}\,.$$

This kind of functional approximations was developped with rigourous mathematical proofs of existence in the 1970's and 1980's by Braess, Kammler and McGlinn see [18, 19] and also [20] for earlier and later related researches, while the first algorithms for fitting are based on Prony's work in the 1790's, see [21].

This approximation leads to a general expression for the expectation with a Gaussian density function. Thus, it is obtained a bias-corrected estimator for a back-transformation κ when an enough sharp approximation with a sum of exponentials is available for the function κ supposed monotone for insuring its existence. Hence for instance, it is supposed:

$$\max_{u\in I} |\kappa(u) - \widehat{\kappa}(u)| \leq \epsilon_\kappa\,,$$

for a small constant, typically smaller than 10^{-3} for instance for intervals I enough large for computing accurately the integrals in this section. This allows to replace the true function for back-transformation κ by its approximation with a sum without changing the values of the integral except for a neglectable amount, because (by the mean value theorem for instance) and with $\int_I f_\nu \leq 1$, one may have informally:

$$\left|\int_I \widehat{\kappa} f_\nu - \int_I \kappa f_\nu\right| \leq \epsilon_\kappa\,,$$

when f_ν is a density function as herein. The resulting estimators are explained next paragraphs. Two cases are proposed, the one with the full infinite real line and the one with a bounded and finite interval.

Cases of the estimators with exponential form and expansion form

When denoting $K = 2K'$ without loss of generality, and $\kappa^{(k)}(\vartheta) = \frac{d^k\kappa(u)}{du^k}|_{u=\vartheta}$, the induced expression for a new estimator approximating the previous exact integral is thus as follows:

$$\begin{aligned}
\hat{\vartheta}^{**,sum}_{iy,\mathrm{GEN}} &\doteq \int_{-\infty}^{+\infty} \kappa(t)\frac{1}{\sqrt{2\pi\hat{\sigma}^2_{iz}}} \exp\left(-\frac{(t-\hat{\vartheta}_{iz})^2}{2\hat{\sigma}^2_{iz}}\right) dt \\
&\underline{\underline{\mathrm{def}}} \sum_{\ell=1}^{\ell=L} \frac{\alpha_{\ell,\kappa}}{\sqrt{2\pi\hat{\sigma}^2_{iz}}} \int_{-\infty}^{+\infty} \exp\left(-\frac{(t-\hat{\vartheta}_{iz})^2+2\hat{\sigma}^2_{iz}\omega_{\ell,\kappa}\, t}{2\hat{\sigma}^2_{iz}}\right) dt \\
&= \sum_{\ell=1}^{\ell=L} \alpha_{\ell,\kappa} \exp\left(-\omega_{\ell,\kappa}\hat{\vartheta}_{iz}\right) \exp\left(0.5\omega^2_{\ell,\kappa}\hat{\sigma}^2_{iz}\right) \\
&= \sum_{\ell=1}^{\ell=L} \alpha_{\ell,\kappa} \exp\left(-\omega_{\ell,\kappa}\hat{\vartheta}_{iz}\right) \left\{1 + \sum_{k=1}^{k=K'} \tfrac{1}{k!}(0.5\omega^2_{\ell,\kappa}\hat{\sigma}^2_{iz})^k\right\} + o(\hat{\sigma}^K_{iz}) \\
&\doteq \underbrace{\kappa(\hat{\vartheta}_{iz}) + \sum_{k=1}^{k=K'} \frac{1}{(2k)!}\,\kappa^{(2k)}(\hat{\vartheta}_{iz})(2k-1)!!\,\hat{\sigma}^{2k}_{iz}}_{\hat{\vartheta}^{**,sgm}_{iy,\mathrm{GEN}}} + o(\hat{\sigma}^K_{iz})
\end{aligned}$$

At the last row, the estimator written as an expansion is a second approximation to the integral estimator. It is recognized the central moments from the normal (or gaussian) distribution for powers even as they are zero otherwise. 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. 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. Note that because this is the exponential function, and considering the gaussian, one may suppose 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, with a truncated serie thanks to the remainder term. Thus one may write that approximatively whe nκ and its derivatives have same bounds for the approximation with a sum,

$$\begin{aligned}
\left|\hat{\vartheta}^{**,int}_{iy,\mathrm{GEN}} - \hat{\vartheta}^{**,sum}_{iy,\mathrm{GEN}}\right| &\leq \epsilon_\kappa \\
\left|\hat{\vartheta}^{**,int}_{iy,\mathrm{GEN}} - \hat{\vartheta}^{**,sgm}_{iy,\mathrm{GEN}}\right| &\leq \epsilon_\kappa \left(1 + \widehat{\kappa}(\hat{\vartheta}_{iz})\frac{(\omega^2_{*,\kappa}\hat{\sigma}^2_{iz})^{K'+1}\exp\left(0.5\omega^2_{*,\kappa}\hat{\sigma}^2_{iz}\right)}{2^{K'+1}(K'+1)!}\right).
\end{aligned}$$

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 terms, achieving an accurate approximation of the integral. As an expansion holds also for a direct Taylor-Young serie of κ under the integral sum, the remainder term is rewritten next subsection, as a complement. This means that one just need to upper bound a derivative of the back-transformation function in order to decide of the order for the sum, otherwise, this may be done numerically by comparing the integral with the sum for an increasing number of terms in order to obtain the final estimator.

Case of the generic transformation with finite bounds

Let denote $\Phi(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-t^2} dt$ for the cumulative distribution function for the Gaussian distribution, for an eventual a factor renormalizing the gaussian distribution, it follows when there exists for κ a relevant approximation as a sum of exponential functions,

$$\begin{aligned}
\hat{\vartheta}_{iy,\text{GEN}}^{**,[sum]} &\doteq \int_a^b \kappa(t) \frac{1}{\sqrt{2\pi\hat{\sigma}_{iz}^2}} \exp\left(-\frac{(t-\hat{\vartheta}_{iz})^2}{2\hat{\sigma}_{iz}^2}\right) dt \\
&\stackrel{\text{def}}{=\!=} \sum_{\ell=1}^{\ell=L} \frac{\alpha_{\ell,\kappa}}{\sqrt{2\pi\hat{\sigma}_{iz}^2}} \int_a^b \exp\left(-\frac{(t-\hat{\vartheta}_{iz})^2 + 2\hat{\sigma}_{iz}^2 \omega_{\ell,\kappa}\, t}{2\hat{\sigma}_{iz}^2}\right) dt \\
&= \sum_{\ell=1}^{\ell=L} \alpha_{\ell,\kappa}\, e^{-\omega_{\ell,\kappa}(\vartheta_{iz} - 0.5\sigma_{iz}^2 \omega_{\ell,\kappa})} \left[\Phi\left(\frac{b-\hat{\vartheta}_{iz}+\hat{\sigma}_{iz}^2\omega_{\ell,\kappa}}{\hat{\sigma}_{iz}}\right) - \Phi\left(\frac{a-\hat{\vartheta}_{iz}+\hat{\sigma}_{iz}^2\omega_{\ell,\kappa}}{\hat{\sigma}_{iz}}\right)\right].
\end{aligned}$$

This leads to a closed-formed expression which may be further bias corrected either from a numerical algorithm either an analytical approach as next section. The difference between the true integral and the approximation just above remains smaller or equal to the constant ϵ_κ. A limit to the solution for an interval with finite bound(s) is that it does not share anymore the exponential form of the unbounded case above. This may be solved by adding an approximation of the cumulative function Φ with a sum of exponential function for the interval $[a;b]$ such that one may write $\widehat{\Phi}(u) = \sum_m \alpha_{m,\Phi} e^{-u\omega_{m,\Phi}}$, leading a new approximation with a double sum. This is a closed-form expression which is expected to be very near the original function, such that the usual computation with log-normal distribution are still available, after linearizing the nonlinear function of the parameters within the exponential function.

Two examples of this case are: i) the inverse sine transformation with two finite bounds from the first quadrant of the unit circle in trigonometric for back-transformation to proportion and ii) the Box-cox transformation with one finite one, say 0, and one infinite for positive estimators to be back-transformed. For the first one, it has been suggested recentlya 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 with $\kappa = \sin^2$ as follows $\exp(-2\hat{\sigma}_{iz}^2)\sin^2(\hat{\vartheta}_{iz}) + 0.5\,(1-\exp(-2\hat{\sigma}_{iz}^2))$ for $a=-\infty$ and $b=\infty$. Here, the closed-form solution is found after classical trigonometry followed by a classical calculus for integrals. But the relevant estimator has finite bounds because a proportion in $[0;1]$ such that $a=0$ and $b=\pi/2$, which is further studied in the experimental part. Note that also in [17] the bounds were not given in this communication, but herein, they are required in order to write cautiously the estimators before the experimental section.

Three different types of resulting estimators

The three types met above are listed in Table 2 for three back-transformations, such that they share a similar expression including a multiplicative bias correction with the usual expectation after just a two-order expansion. The first one is a direct back-transformation without bias-correction, the second one is a back-transformation with bias-correction from the expectation with the Gaussian density function, the third one back-transformation with same bias-correction but after a second-order approximation of the function.

Table 2: Three generalized estimators for back-transformations (bc).

	No bc	bc by integration	bc by integration+ 2^{th}-order approximation
Exponential form	$\sum_{\ell=1}^{\ell=L} \alpha_{\ell,\kappa} e^{-\omega_{\ell,\kappa}\vartheta_{iz}}$	$\sum_{\ell=1}^{\ell=L} \alpha_{\ell,\kappa} e^{-\omega_{\ell,\kappa}(\vartheta_{iz}-0.5\sigma^2_{iz}\omega_{\ell,\kappa})}$	$\sum_{\ell=1}^{\ell=L} \alpha_{\ell,\kappa}(1+0.5\sigma^2_{iz}\omega^2_{\ell,\kappa})e^{-\omega_{\ell,\kappa}\vartheta_{iz}}$
Expansion form			$\left(1+0.5\frac{\ddot{\kappa}(\vartheta_{iz})}{\kappa(\vartheta_{iz})}\sigma^{(2)}_{iz}\right)\kappa(\vartheta_{iz})$

These estimators are written with exponential functions as met in a log-transformation, but there is a cost to this simpler expression in comparison to deal directly with the usual nonlinear function. The parameters α_ℓ and ω_ℓ in the sum of exponentials are written implicitely (see also the discussion next after) as a function of the initial parameters from the linear model fitted before returning to the original scale, because the sum is fitted with the nonlinear transformation.

3.3 Expectation from Taylor-Young serie

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 [22]). 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,

$$E_{\hat{\vartheta}_{iz}}[\kappa(\hat{\vartheta}_{iz})] \;=\; \kappa(\vartheta_{iz}) + \sum_{k=1}^{k=K} \tfrac{1}{k!}\,\kappa^{(k)}(\vartheta_{iz})E_{\hat{\vartheta}_{iz}}[(\hat{\vartheta}_{iz}-\vartheta_{iz})^k] + o(E_{\hat{\vartheta}_{iz}}[(\vartheta_{iz}-\hat{\vartheta}_{iz})^K])\,.$$

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

$$\begin{aligned}\hat{\vartheta}^{**,tly}_{iy,\mathrm{GEN}} &\doteq \hat{E}_{\hat{\vartheta}_{iz}}[\kappa(\hat{\vartheta}_{iz})] \\ &\stackrel{\text{def}}{=\!=} \kappa(\hat{\vartheta}_{iz}) + \sum_{k=1}^{k=K} \tfrac{1}{k!}\,\kappa^{(k)}(\hat{\vartheta}_{iz})\hat{E}_{\hat{\vartheta}_{iz}}[(\hat{\vartheta}_{iz}-\vartheta_{iz})^k]\end{aligned}$$

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 [23] instead. This allows several possible estimates such that:

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

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. The error is also known, as equal to approximatively,

$$\hat{\vartheta}^{**,int}_{iy,\mathrm{GEN}} - \hat{\vartheta}^{**,lty}_{iy,\mathrm{GEN}} = \frac{\kappa^{(K+1)}(t^{+1}_{-1})\hat{E}_{\hat{\vartheta}_{iz}}[(\hat{\vartheta}_{iz} - \vartheta_{iz})^{K+1}]}{(K+1)!}\,.$$

A limit of this approximation is asking either a full range for the estimator instead of an interval or at least a small variable $\hat{\sigma}^2_{iz}$, otherwise the expansion may not be accurate. 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 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,

$$\hat{\vartheta}^{**,tly}_{iy,\mathrm{GEN}} \;=\; \hat{h}_{iy}\,\kappa(\hat{\vartheta}_{iz}) \quad \text{where} \quad \hat{h}_{iy} \;\approx\; 1 + 0.5\frac{\ddot{\kappa}(\hat{\vartheta}_{iz})}{\kappa(\hat{\vartheta}_{iz})}\hat{\sigma}^2_{iz} \doteq \exp\left(0.5\frac{\ddot{\kappa}(\hat{\vartheta}_{iz})}{\kappa(\hat{\vartheta}_{iz})}\hat{\sigma}^2_{iz}\right)\,.$$

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 [24] 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.

3.4 Discussion

The different estimators defined in this section may be classified in order to better see their difference and equivalence w.r.t. the use of a gaussian distribution, parametric boostrapping and any eventual other hypothesis of extension: some have a gaussian hypothesis and some other ones not. A more synthetic way to understand these estimators is the approximation of the integral for back-transformation. When κ is the function for the back-transformation of the estimator ν and f_ν is its distribution, one may write:

$$\int \kappa f_\nu \quad \text{approximated by} \quad \text{or} \int \hat{\kappa} f_\nu \text{ or} \int \kappa \hat{f}_\nu \text{ or} \int \hat{\kappa}\hat{f}_\nu\,,$$

where $\hat{\kappa}$ and $\hat{f}_\nu$ are approximations of the true functions κ and f_ν respectively. Note that there this is also the integration which may be approximated via adaptive quadrature for instance, this is included by approaching the continuous density function by a discrete one.

In summary one or two functions need to be approximated in the estimators above. For the approximation of the back-transformation, one may consider the usual Taylor-Young expansion while it was proposed herein the sum of exponential functions in order to retrieve the nice properties met with a log-transformation and log-normal distribution. The sum of exponential function is also related to generative function for random variables, such that is brings also a tool for approximating the integral directly. The estimator with an exponential form is compact and share good properties with the an exponential back-transformation but it may ask for implicit function theorem [25] for the derivatives even if this dependence may be ignored at first. The approximation of the density function f_ν was shortly adressed by an eventual (parametric) boostraping which leads to an empirical distribution, but this is mostly a basic numerical quadrature which is involved nextafter. Note that this may be included in a numerical procedure but such approach is expected to be computer intensive in practice, hence asking for a powerful computer that is not always available. Note also that other expansions from the literature for distribution approximation are not discussed neither herein.

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. The new 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 and the existing expressions for bias-corrected estimators in the current literature.

4 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.

4.1 Multiplicative correction for a generalized estimator

According to the revisit of the existing approaches for bias corrections from the previous section, a generalized estimator needs an additional bias-correction from the sampling error. It is denoted ϕ_κ a constant equal to 0 or 1 (or varying between 0 and $\phi_{max} \geq 1$ for curve drawing and diagnostic purpose) in order to retrieve the estimator without a first bias correction. With $\hat{\sigma}_{iz}^{(k)} = \hat{E}_{\hat{\vartheta}_{iz}}[(\hat{\vartheta}_{iz} - \vartheta_{iz})^k]$, when following the previous section and considering eventually a parameter λ for the transformation, the estimators of interest are defined as follows:

$$\begin{aligned}
\hat{\vartheta}_{iy}^{**,e} &= \kappa(\hat{\vartheta}_{iz}) + \phi_\kappa \sum_{k=1}^{k=K} \tfrac{1}{k!}\, \kappa^{(k)}(\hat{\vartheta}_{iz})\, \hat{\sigma}_{iz}^{(k)} \\
\hat{\vartheta}_{iy}^{**,[q]} &= \tfrac{c}{\sqrt{2\pi\hat{\sigma}_{iz}^2}} \sum_{\ell=1}^{\ell=L} \alpha_\ell\, \kappa\left(c\omega_\ell + d\right)\, \exp\left(-\tfrac{1}{2\hat{\sigma}_{iz}^2}\left(c\omega_\ell + d - \hat{\vartheta}_{iz}\right)^2\right) \\
\hat{\vartheta}_{iy}^{**,[s]} &= \sum_{\ell=1}^{\ell=L} \alpha_{\ell,\kappa}\, e^{\left(0.5\omega_{\ell,\kappa}^2 \hat{\sigma}_{iz}^2 - \hat{\vartheta}_{iz}\omega_{\ell,\kappa}\right)} \left\{(1-\phi_\kappa)\left[\Phi\left(\tfrac{b - \hat{\vartheta}_{iz} + \hat{\sigma}_{iz}^2 \omega_{\ell,\kappa}}{\hat{\sigma}_{iz}}\right) - \Phi\left(\tfrac{a - \hat{\vartheta}_{iz} + \hat{\sigma}_{iz}^2 \omega_{\ell,\kappa}}{\hat{\sigma}_{iz}}\right)\right]\right\}.
\end{aligned}$$

The first one extends the usual estimator often considered in the literature for the log-transformation but with higher-order derivatives when the variance is larger than one. The second one is with an approximation of the Gaussian density function, it allows as a baseline a comparison hence a validation for the numerical results. The third one is with the exponential form with the function approximated instead of the Gaussian density function. These three generalized estimators are in stake next after for a validation of their expectations and a comparison of their variances.

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 in the Fay-Herriot model. Each transformation function would ask for a new analytical solution on the contrary to a generalized estimator. Using a factorisation form would lead to very related [26] estimators and requires the hypothesis of small variance.

Analytical expression for the multiplicative factor

According to the literature [22, 27], two ways for bias corrections are to construct from the expectation estimator $\hat{E}[\vartheta_{iy}^{**}]$, a new estimator: either the additive one $2\hat{\vartheta}_{iy}^{**} - \hat{E}[\hat{\vartheta}_{iy}^{**}]$ or either the multiplicative one $\hat{\vartheta}_{iy}^{**}\hat{\vartheta}_{iy}^{**}/\hat{E}[\hat{\vartheta}_{iy}^{**}]$ which are equivalent at first orders. The usual computational algorithm for an estimation of the expectation is available via the standard parametric boostrap often used in small area publications, hence a numerical comparison with the analytical result is required here for experiments. A perspective to the reader here may be to plug-in estimators from a bias correction for the linear model from the transformed outcome [28] or via boostran too in order to improve further the results. Closed-form expressions allow a faster access to the result and also provide a better understanding of the methods. The additive bias correction is obtained directly from the available expression below, hence this is the multiplicative bias correction, extending the case of the log-transformation in [4], which is treated here without loss of generality:

$$\hat{\vartheta}_{iy}^{*} = \hat{\rho}_i \hat{\vartheta}_{iy}^{**} \quad \text{with} \quad \hat{\rho}_i = \frac{\hat{\vartheta}_{iy}^{**}}{\hat{E}[\hat{\vartheta}_{iy}^{**}]} \, .$$

With $\phi_\rho, \phi_g, \phi_\lambda \in \{0, 1\}$, the expression for the corrective term $\rho_{i,\text{GEN}}$ comes from the vector of bias $m^{-1}[\psi_\beta, \psi_{\sigma_u^2}, \psi_{\lambda^2}]^T = m^{-1}\psi$, the gradient vector $g_{**} = [\partial[\vartheta_{iy}^{**}]/\partial\beta^T, \partial[\vartheta_{iy}^{**}]/\partial\sigma_u^2, \partial[\vartheta_{iy}^{**}]/\partial\lambda^T]^T$ with the first order derivatives, the hessian matrix H^{**} with the second-order derivatives and the covariance matrix as found below.

It is involved for $\hat{\theta} = (\hat{\beta}^T, \hat{\sigma}_u^2, \hat{\lambda}^T)^T$ for the estimation and θ for the true unknown parameter, while $V_{\hat{\beta}} = \text{var}(\hat{\beta})$, $V_{\hat{\sigma}_u^2} = \text{var}(\hat{\sigma}_u^2)$, $V_{\hat{\lambda}} = \text{var}(\hat{\lambda})$, $V_{\hat{\beta}\hat{\sigma}_u^2} = \text{cov}(\hat{\beta}, \hat{\sigma}_u^2)$, $V_{\hat{\beta}\hat{\lambda}} = \text{cov}(\hat{\beta}, \hat{\lambda})$, $V_{\hat{\lambda}\hat{\sigma}_u^2} = \text{cov}(\hat{\lambda}, \hat{\sigma}_u^2)$ which are aggregated in $V_{\hat{\theta}}$. It is also supposed an hypothesis of bias $m^{-1}\psi$ and of distribution (see for instance Lemma 2.1 from [17]) for the estimators from the parameters:

$$\sqrt{m}\left(\hat{\theta} - \theta\right) \underset{+\infty}{\longrightarrow} N\left(0, V_{\hat{\theta}}\right) \quad \text{as the hypothesis of distribution of the parameters estimator.}$$

Eventually a parameter λ is estimated such that for the Box-Cox transformation but its influence and the one for the bias from the linear estimation are both neglected herein for the analytical variance estimation when ϕ_g and ϕ_λ are both equal to zero.

General expression with a direct second-order approximation within expectation

A multiplicative term for bias correction may be written from the expectation as follows,

$$\begin{aligned}
\frac{1}{\vartheta_{iy}^{**}}E[\vartheta_{iy}^{**}] &= 1+\frac{\phi_\rho}{\vartheta_{iy}^{**}\,m}\left\{g_{**}^T\psi+\frac{1}{2}Tr\left[\begin{pmatrix} H_\beta^{**} & H_{\beta\sigma_u^2}^{**} & H_{\beta\lambda}^{**} \\ H_{\sigma_u^2\beta}^{**} & H_{\sigma_u^2}^{**} & H_{\sigma_u^2\lambda}^{**} \\ H_{\lambda\beta^T}^{**} & H_{\lambda\sigma_u^2}^{**} & H_{\lambda^2}^{**}\end{pmatrix}\begin{pmatrix} V_{\hat\beta} & V_{\hat\beta\hat\sigma_u^2} & V_{\hat\beta\hat\lambda} \\ V_{\hat\sigma_u^2\hat\beta} & V_{\hat\sigma_u^2} & V_{\hat\sigma_u^2\hat\lambda} \\ V_{\hat\lambda\hat\beta} & V_{\hat\lambda\hat\sigma_u^2} & V_{\hat\lambda}\end{pmatrix}\right]+o(1)\right\} \\
&\doteq 1+\frac{\phi_g}{m\,\vartheta_{iy}^{**}}\,g_{**}^T\psi+\frac{0.5\phi_\rho}{m\,\vartheta_{iy}^{**}}\left\{\begin{array}{l} Tr[H_\beta^{**}V_{\hat\beta}]+Tr[H_{\sigma_u^2}^{**}V_{\hat\sigma_u^2}]+\phi_\lambda Tr[H_{\sigma_u^2\lambda}^{**}V_{\hat\lambda\hat\sigma_u^2}] \\ +\phi_\lambda\,Tr[H_{\lambda^2}^{**}V_{\hat\lambda}]+\,\phi_\lambda\,Tr[H_{\beta\lambda}^{**}H_{\lambda\beta^T}^{**}] \\ +\phi_\lambda\,Tr[H_{\lambda\beta^T}^{**}V_{\hat\beta\hat\lambda}]\,+\,\phi_\lambda\,Tr[H_{\lambda\sigma_u^2}^{**}V_{\hat\sigma_u^2\hat\lambda}]\end{array}\right\} \\
&\doteq 1+\frac{0.5\phi_\rho}{m\,\vartheta_{iy}^{**}}\left[Tr\left(H_\beta^{**}V_{\hat\beta}\right)+\left(H_{\sigma_u^2}^{**}V_{\hat\sigma_u^2}\right)\right].
\end{aligned}$$

Here the correlations between the variance component and the regression coefficients are supposed zero as in maximum likelihood, $o(1)$ comes from the behavior of the third order moments of $\hat\theta$ and from Hölder's inequality as for instance in [17]. It is also denoted $Tr()$ for the matricial operator named the trace. It would be interesting as an alternative to check the behavior of such estimator with a concentrated likelihood in order to remove the part of the hessian relative to $\hat\lambda$ differently. The derivatives for the hessian matrix are given at the appendix section for the studied estimators. Let also have the derivatives for the linear estimator $\vartheta_{iz} = x_i^T\beta + \gamma_i(y_i - x_i^T\beta)$ at the appendix section. 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. Note that an alternative bias correction for the log-transformation was proposed in [4] and requires to deal with exponential functions, which becomes possible with the proposed approximation with a sum in $\hat\vartheta_{iy}^{**,[s]}$ after the second approximation for the cumulative function and the term in factor. According to the available experiments for the log-transformation such estimator may be not better hence is not studied further. Some other alternatives such as multivariate quadrature for instance are likely to provide a better estimator for smallers values of m but are more difficult to deal with the variance computation, hence are left as a perspective. This underlines why (parametric) boostrap is an usual method for such bias correction while requiring intensive numerical resampling procedures.

Retrieving the case of the log-transformation and extending to other one

Table 3: Examples of inverse transformation functions and their derivatives.

	$\kappa(\vartheta_{iz})$	$\dot\kappa(\vartheta_{iz})$	$\ddot\kappa(\vartheta_{iz})$	$\left.\frac{d^k\kappa(u)}{du^k}\right\rvert_{\vartheta_{iz}}$
Exponential	$\exp(\vartheta_{iz})$	$\exp(\vartheta_{iz})$	$\exp(\vartheta_{iz})$	$\exp(\vartheta_{iz})$
Box-Cox	$(1+\lambda\vartheta_{iz})^{\frac{1}{\lambda}}$	$(1+\lambda\vartheta_{iz})^{\frac{1-\lambda}{\lambda}}$	$(1-\lambda)(1+\lambda\vartheta_{iz})^{\frac{1-2\lambda}{\lambda}}$	$\prod_{\ell=1}^{k-1}(1-\ell\lambda)(1+\lambda\vartheta_{iz})^{\frac{1-k\lambda}{\lambda}}$
Asinh-Exp	$\frac{e^{\vartheta_{iz}}+\varrho(\lambda)e^{-\vartheta_{iz}}}{2}$	$\frac{e^{\vartheta_{iz}}-\varrho(\lambda)e^{-\vartheta_{iz}}}{2}$	$\frac{e^{\vartheta_{iz}}+\varrho(\lambda)e^{-\vartheta_{iz}}}{2}$	$\frac{e^{\vartheta_{iz}}+(-1)^k\varrho(\lambda)e^{-\vartheta_{iz}}}{2}$

The derivatives at first order $\dot\kappa(\vartheta_{iz})$ and second order $\ddot\kappa(\vartheta_{iz})$ for the involved inverse transformation functions $\kappa(\vartheta_{iz})$ are in the Table 3. It is retrieved the pionneer research for the log-transformation from [4] which is extended herein to other transformations like its usual generalisation for positive outcomes. Note that this is not discussed further but the derivarives for the inverted Box-cox function look not upper bounded for some values hence one may check this issue for the estimator written as an expansion, as this hypothesis has not been introduced when the remainder term has been upper bounded in the previous section. The problem with the Box-Cox transformation is to ask for positive values, which is not always true in practice, hence alternative ones [29] have been proposed in the literature.

4.2 Variance estimation for the generalized estimators

The proposed estimators from the previous sections are nearly equal to the integral approach but with an additional bias correction: its simpler expression allows a direct computation of the variance as explained next subsection. 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 [30] 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 [31] is implemented. See also [32] for further discussion and justification of the method for another estimator. 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 [33] 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.

Method

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.

Implementation with the proposed estimators

In the case of the small area, the hessian matrix are known from the literature. The derivatives of the estimators are found as follows for λ while extends to β and σ_u^2 by changing the related derivatives.

$$\begin{array}{lcl} \frac{\partial}{\partial\lambda}[\hat{\vartheta}_{iy}^*] & = & \frac{\partial}{\partial\lambda}[\rho_i\hat{\vartheta}_{iy}^{**}] \\ & = & \frac{\partial}{\partial\lambda}[\rho_i]\hat{\vartheta}_{iy}^{**} + \rho_i\frac{\partial}{\partial\lambda}[\hat{\vartheta}_{iy}^{**}] \end{array}$$

The related derivatives for $\hat{\vartheta}_{iy}^{**}$, the Hessian matrix and the traces are given at the appendix section. For the derivative of the multiplicative terms one may get that,

$$\frac{\partial}{\partial\lambda}[\rho_i] \quad = \quad \frac{0.5\phi_\rho}{\vartheta_{iy}^{**}}\left[-\frac{\frac{\partial}{\partial\lambda}[\vartheta_{iy}^{**}]}{\vartheta_{iy}^{**}}\left\{Tr\left(H_\beta^{**}V_{\hat{\beta}}\right) + \left(H_{\sigma_u^2}^{**}V_{\hat{\sigma}_u^2}\right)\right\} + \frac{\partial}{\partial\lambda}[Tr\left(H_\beta^{**}V_{\hat{\beta}}\right) + \left(H_{\sigma_u^2}^{**}V_{\hat{\sigma}_u^2}\right)]\right]\rho_i^2$$

Here the derivatives are required for β and σ_u^2, while the one for λ may be not involved for some transformation without parameters or when their influence is neglected.

5 Experiments

In this section, the estimators are compared with artificial and real data via simulations in order to validate the analytical expressions.

5.1 Experimental settings

The proposed estimators from the previous sections are nearly equal to the integral approach but with an additional bias correction: the simpler expression allows a direct computation of some analytical quantities. An analytical expressions as proposed herein may use an approximation which may be not always relevant. Numerical algorithms for bias correction and mean squared error estimation are fortunately available from the literature and may be compared to the analytical results. In small area estimation, it is often computed the mean squared error in order to know if the estimator is enough accurate but the bias correction of the mean estimator itself was almost never adressed before recently.
For the Gauss-Legendre quadrature, the nodes and weights for the numerical quadrature are computed numerically with the function available from the r package *pracma*[3] with $L = 500$ such that the resulting sum was nearly equal to the integral.

5.2 Assessing the estimators for back-transformation

Estimator with the exponential form for the inverse sine transformation

The estimator for back-transformation is defined as follows:

$$\hat{\vartheta}^{**}_{i,\mathrm{SIN}\frac{1}{2}} \overset{\text{def}}{=\!=} \int_0^{\frac{\pi}{2}} \sin^2(t) \frac{1}{\sqrt{2\pi\hat{\sigma}^2_{iz}}} \exp\left(-\frac{(t-\hat{\vartheta}_{iz})^2}{2\hat{\sigma}^2_{iz}}\right) dt\,.$$

When the integral is restricted to the relevant domain $[0; \pi/2]$, note that for the function $u \to \sin^2(u)$, a set of exponential functions and weights were found for $L = 3$ with the r package pracma as follows, with $\alpha_0 = -1.979235023$ for an additional constant, in Table 4. Note that a small noise (of variances 0.005 and 0.01) was added to the true 30 points selected after equal spacing for the abcisses between 0 and $\pi/2$, while several runs has been required in order to select several initial values while the one chosen were $-0.1, -0.5, -0.6$ as the starting points for the "starting values for the exponentials alone" according to the documentation. The obtained parameters are $\alpha_1 = 5.978669435\,10^{-6}$, $\alpha_2 = -8.758313137\,10^{3}$, $\alpha_3 = 8.760298018\,10^{3}$, $\omega_1 = 6.630547823$, $\omega_2 = -1.319349208$, $\omega_3 = -1.319031305$ as exemple of numerical values after fitting $\sin^2$ to a sum of exponential functions. The approximation for this function with a sum of exponential functions is accurate (around 10^{-2} for some integrals with some model parameters) while allowing the proposed expressions for back-transformation:

$$\widehat{\sin^2(u)} = \alpha_0 + \alpha_1 e^{\omega_1 u} + \alpha_2 e^{\omega_2 u} + \alpha_3 e^{\omega_3 u} \quad \text{for } u \in \left[0; \frac{\pi}{2}\right]\,.$$

There may exist numerical alternatives to improve the obtained parameters, such sums may be difficult to estimate with the available algorithms currently. More precise values for the weights and nodes may be also better for improving the precision. The resulting estimators are as follows in Table 4 for intrue, intsum, estsum, estqgl and estsig with same definition than above and range

[3]`https://cran.r-project.org/web/packages/pracma/index.html`

values for $\hat{\sigma}_{iz}$ and $\hat{\vartheta}_{iz}$, the estimate before back-transformation and its standard-error estimate as given in the table.

Estimator with the exponential form for the Box-Cox transformation

The estimator for back-transformation is defined as follows:

$$\hat{\vartheta}^{**}_{i,\mathrm{BC}} \quad \stackrel{\text{def}}{=\!=} \quad \int_0^{+\infty} (1+\lambda t)^{\frac{1}{\lambda}} \frac{1}{\sqrt{2\pi\hat{\sigma}^2_{iz}}} \exp\left(-\frac{(t-\hat{\vartheta}_{iz})^2}{2\hat{\sigma}^2_{iz}}\right) dt\,.$$

First of all, the infinite bound is replaced by a finite, one, here chosen in order to get two related integrals equal for a small relative difference, say 10^{-3}. It is obtained thus $a = 0$ and $b = 7$ with for instance $\lambda = 0.5$ in this example. The corresponding weights and abcissa for the sum of exponentials are equal to $\alpha_0 = 7.464383579$, $\alpha_1 = -7.826695199\,10^0$, $\alpha_2 = -2.880858092\,10^4$, $\alpha_3 = 2.880792708\,10^4$, $\alpha_4 = 1.405707821\,10^{-2}$, $\omega_1 = 0.1890400637$, $\omega_2 = -1.3633583448$, $\omega_3 = -1.3633745088$, $\omega_4 = 0.6852717213$. This is written as follows:

$$\widehat{(1+0.5u)^{\frac{1}{0.5}}} = \alpha_0 + \alpha_1 e^{\omega_1 u} + \alpha_2 e^{\omega_2 u} + \alpha_3 e^{\omega_3 u} + \alpha_4 e^{\omega_4 u} \qquad \text{for } u \in [0;7]\ .$$

The resulting estimators are as follows in Table 5 for intrue, intsum, estsum, estqgl and estsig with same definition than above and same range values for $\hat{\sigma}_{iz}$ and $\hat{\vartheta}_{iz}$, the estimate before back-transformation and its standard-error estimate.

Estimator with the exponential form for the asinh-exp transformation

The estimator for back-transformation is defined as follows:

$$\begin{aligned}\hat{\vartheta}^{**}_{i,\mathrm{AE}} \quad &\stackrel{\text{def}}{=\!=} \quad \int_0^{+\infty} \frac{e^y + \varrho(\lambda)e^{-y}}{2} \frac{1}{\sqrt{2\pi\hat{\sigma}^2_{iz}}} \exp\left(-\frac{(t-\hat{\vartheta}_{iz})^2}{2\hat{\sigma}^2_{iz}}\right) dt \\ &= \quad \exp(0.5\hat{\sigma}^2_{iz})\, \frac{\exp(\hat{\vartheta}_{iz}) + \varrho(\lambda)\exp(-\hat{\vartheta}_{iz})}{2}\,.\end{aligned}$$

First of all, the infinite bound is replaced by a finite, one, here chosen in order to get two related integrals equal for a small relative difference, say 10^{-3}, with thus $a = -7$ and $b = 7$ and $\lambda = 0.5$. The corresponding weights and abcissa for the sum of exponentials are equal to $\alpha_0 = 0$, $\alpha_1 = 0.5$, $\alpha_2 = 0.5\varrho(0.5)$, $\omega_1 = 1$, $\omega_2 = -1$. This is written as follows, but the approximation is here exact:

$$\widehat{\frac{\exp(u) + \varrho(\lambda)\exp(-u)}{2}} = \alpha_1 e^{\omega_1 u} + \alpha_2 e^{\omega_2 u} \qquad \text{for } u \in [-7;7]\ .$$

The resulting estimators are as follows in Table 6 for intrue, intsum, estsum, estqgl and estsig with same definition than above and same range values for $\hat{\sigma}_{iz}$ and $\hat{\vartheta}_{iz}$, the estimate before back-transformation and its standard-error estimate. An exact solution allows to check further the integral before the estimators for small area.

Summary

According to the examples in this subsection, the approximations with a sum of exponentual function is relevant for a bias or less of 10^{-2} because current software are able to provide a solution after a few tries. Better approximations may be constructed with less but ask more advanced developpements currently. Other approximation are also relevant, in particular numerical integration but ask to be carefull to the choice of the number of points, which is more wanted for small values of $\hat{\sigma}_{iz}$ and $\hat{\vartheta}_i^{**}$ such as adaptive estimators are appealing. The expansion with the standard-deviation is relevant only in a few cases when the standard deviation is small or all the real line in the range of the transformation as expected.

6 Conclusion

Herein, it is proposed for regression models with in particular 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. The usual second-order approximation is replaced by two alternatives, say an higher-order expansion or a sum of exponentials, in order to better approximate the expectation involved for back-transformation. On the contrary to alternative approximations such as the Laplace method [34] for instance, the approach herein is expected to lead to a very small error after the approximation while providing simpler analytical expressions. Algorithms for fitting such sums are of first importance in order to improve the method, eventually by adding constraints, higher order polynomial terms or more dimensions as this is not addressed currently in the literature.

This also suggests that some known results for the log-normal distribution may be extended to a nonlinear case with these new approximations. Perspectives for the reader are an extension to the unit-level models and the multivariate area-level models. 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.

Appendix A: Tables for comparing the approximations

Table 4: Exemple of numerical values for $\hat{\vartheta}^{**}_{i,\mathrm{SIN}\frac{1}{2}}$ after fitting $\sin^2$.

$\hat{\sigma}_{iz}$	$\hat{\vartheta}^{**}_i$	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3
0.05	intrue	0.089	0.153	0.231	0.320	0.415	0.515	0.613	0.707	0.793	0.867	0.926
	intsum	0.083	0.152	0.234	0.323	0.416	0.512	0.607	0.699	0.787	0.867	0.935
	estsum	0.083	0.152	0.234	0.323	0.416	0.512	0.607	0.699	0.787	0.867	0.935
	estqgl	0.089	0.153	0.231	0.320	0.415	0.515	0.613	0.707	0.793	0.867	0.926
	estsig	0.089	0.153	0.231	0.320	0.415	0.515	0.613	0.707	0.793	0.867	0.926
0.10	intrue.1	0.096	0.159	0.235	0.322	0.417	0.514	0.611	0.704	0.788	0.861	0.917
	intsum	0.089	0.157	0.237	0.325	0.417	0.512	0.606	0.697	0.784	0.862	0.924
	estsum	0.089	0.157	0.237	0.325	0.417	0.512	0.606	0.697	0.784	0.862	0.924
	estqgl	0.096	0.159	0.235	0.322	0.417	0.514	0.611	0.704	0.788	0.861	0.917
	estsig	0.096	0.159	0.235	0.322	0.417	0.514	0.611	0.704	0.788	0.861	0.920
0.20	intrue	0.118	0.178	0.251	0.333	0.422	0.513	0.604	0.690	0.762	0.809	0.809
	intsum	0.113	0.175	0.249	0.332	0.420	0.511	0.601	0.687	0.761	0.810	0.812
	estsum	0.113	0.175	0.249	0.332	0.420	0.511	0.601	0.687	0.761	0.810	0.812
	estqgl	0.118	0.178	0.251	0.333	0.422	0.513	0.604	0.690	0.762	0.809	0.809
	estsig	0.119	0.178	0.251	0.333	0.422	0.513	0.605	0.692	0.772	0.840	0.896
0.25	intrue	0.133	0.192	0.261	0.340	0.425	0.512	0.597	0.673	0.730	0.758	0.743
	intsum	0.128	0.188	0.259	0.338	0.423	0.510	0.594	0.671	0.730	0.759	0.745
	estsum	0.128	0.188	0.259	0.338	0.423	0.510	0.594	0.671	0.730	0.759	0.745
	estqgl	0.133	0.192	0.261	0.340	0.425	0.512	0.597	0.673	0.730	0.758	0.743
	estsig	0.136	0.193	0.262	0.340	0.425	0.513	0.600	0.684	0.760	0.825	0.878
0.50	intrue	0.208	0.255	0.305	0.355	0.404	0.448	0.484	0.510	0.523	0.522	0.506
	intsum	0.205	0.252	0.302	0.353	0.402	0.446	0.483	0.509	0.523	0.522	0.507
	estsum	0.205	0.252	0.302	0.353	0.402	0.446	0.483	0.509	0.523	0.522	0.507
	estqgl	0.208	0.255	0.305	0.355	0.404	0.448	0.484	0.510	0.523	0.522	0.506
	estsig	0.250	0.289	0.336	0.390	0.448	0.509	0.569	0.626	0.678	0.724	0.760
1.00	intrue	0.221	0.237	0.252	0.265	0.276	0.285	0.292	0.296	0.298	0.297	0.293
	intsum	0.220	0.236	0.250	0.263	0.275	0.284	0.291	0.296	0.297	0.296	0.293
	estsum	0.220	0.236	0.250	0.263	0.275	0.284	0.291	0.296	0.297	0.296	0.293
	estqgl	0.221	0.237	0.252	0.265	0.276	0.285	0.292	0.296	0.298	0.297	0.293
	estsig	0.444	0.453	0.463	0.475	0.488	0.502	0.515	0.528	0.540	0.550	0.558
1.20	intrue	0.204	0.214	0.223	0.232	0.239	0.244	0.248	0.251	0.252	0.251	0.249
	intsum	0.203	0.213	0.222	0.231	0.238	0.243	0.248	0.250	0.251	0.251	0.249
	estsum	0.203	0.213	0.222	0.231	0.238	0.243	0.248	0.250	0.251	0.251	0.249
	estqgl	0.204	0.214	0.223	0.232	0.239	0.244	0.248	0.251	0.252	0.251	0.249
	estsig	0.477	0.480	0.485	0.490	0.495	0.501	0.506	0.512	0.517	0.521	0.524

Table 5: Exemple of numerical values for $\hat{\vartheta}^{**}_{i,\mathrm{BC}\frac{1}{2}}$ after fitting $(1+0.5u)^{\frac{1}{0.5}}$.

$\hat{\sigma}_{iz}$	$\hat{\vartheta}^{**}_i$	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3
0.05	intrue	1.323	1.441	1.563	1.691	1.823	1.961	2.103	2.251	2.403	2.561	2.723
	intsum	1.329	1.446	1.567	1.694	1.825	1.961	2.103	2.250	2.402	2.560	2.723
	estsum	1.329	1.446	1.567	1.694	1.825	1.961	2.103	2.250	2.402	2.560	2.723
	estqgl	1.323	1.441	1.563	1.691	1.823	1.961	2.103	2.251	2.403	2.561	2.723
	estsig	1.324	1.442	1.564	1.692	1.825	1.962	2.105	2.253	2.406	2.563	2.726
0.10	intrue	1.324	1.442	1.565	1.692	1.825	1.962	2.105	2.253	2.405	2.562	2.725
	intsum	1.330	1.448	1.569	1.696	1.827	1.963	2.105	2.252	2.404	2.562	2.725
	estsum	1.330	1.448	1.569	1.696	1.827	1.963	2.105	2.252	2.404	2.562	2.725
	estqgl	1.324	1.442	1.565	1.692	1.825	1.962	2.105	2.253	2.405	2.563	2.725
	estsig	1.329	1.447	1.570	1.698	1.832	1.970	2.113	2.261	2.415	2.573	2.736
0.20	intrue	1.271	1.429	1.567	1.699	1.832	1.970	2.112	2.260	2.412	2.570	2.732
	intsum	1.276	1.434	1.571	1.702	1.834	1.971	2.113	2.260	2.412	2.570	2.733
	estsum	1.276	1.434	1.571	1.702	1.834	1.971	2.113	2.260	2.412	2.570	2.733
	estqgl	1.271	1.429	1.567	1.699	1.832	1.970	2.112	2.260	2.412	2.570	2.732
	estsig	1.349	1.469	1.594	1.724	1.859	2.000	2.145	2.295	2.451	2.612	2.777
0.25	intrue	1.236	1.406	1.557	1.698	1.836	1.975	2.118	2.266	2.418	2.576	2.738
	intsum	1.241	1.411	1.561	1.701	1.838	1.976	2.119	2.266	2.418	2.576	2.739
	estsum	1.241	1.411	1.561	1.701	1.838	1.976	2.119	2.266	2.418	2.576	2.739
	estqgl	1.236	1.406	1.557	1.698	1.836	1.975	2.118	2.266	2.418	2.576	2.738
	estsig	1.364	1.486	1.612	1.744	1.880	2.022	2.169	2.321	2.479	2.641	2.809
0.50	intrue	1.184	1.344	1.503	1.662	1.821	1.978	2.136	2.294	2.453	2.616	2.781
	intsum	1.187	1.346	1.506	1.665	1.823	1.980	2.137	2.295	2.455	2.617	2.783
	estsum	1.187	1.346	1.506	1.665	1.823	1.980	2.137	2.295	2.455	2.617	2.783
	estqgl	1.184	1.344	1.503	1.662	1.821	1.978	2.136	2.294	2.453	2.616	2.781
	estsig	1.499	1.632	1.771	1.915	2.065	2.221	2.382	2.550	2.722	2.901	3.085
1.00	intrue	1.382	1.513	1.649	1.791	1.938	2.089	2.245	2.406	2.570	2.739	2.912
	intsum	1.384	1.515	1.652	1.794	1.940	2.092	2.248	2.409	2.574	2.743	2.916
	estsum	1.384	1.515	1.652	1.794	1.940	2.092	2.248	2.409	2.574	2.743	2.916
	estqgl	1.382	1.513	1.649	1.791	1.938	2.089	2.245	2.406	2.570	2.739	2.912
	estsig	2.180	2.374	2.576	2.786	3.005	3.231	3.466	3.710	3.961	4.221	4.489
1.20	intrue	1.506	1.633	1.766	1.903	2.046	2.194	2.347	2.505	2.667	2.834	3.006
	intsum	1.508	1.635	1.768	1.906	2.049	2.197	2.350	2.508	2.671	2.838	3.010
	estsum	1.508	1.635	1.768	1.906	2.049	2.197	2.350	2.508	2.671	2.838	3.010
	estqgl	1.506	1.633	1.766	1.903	2.046	2.194	2.347	2.505	2.667	2.834	3.006
	estsig	2.717	2.958	3.210	3.472	3.744	4.027	4.319	4.622	4.936	5.259	5.593

Table 6: Exemple of numerical values for $\hat{\vartheta}^{**}_{i,\mathrm{AE}\frac{1}{2}}$ from the exact sum $(\exp(u) + \varrho(0.5)\exp(-u))/2$.

$\hat{\sigma}_{iz}$	$\hat{\vartheta}^{**}_{i}$	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3
0.05	intrue	0.699	0.768	0.844	0.929	1.024	1.128	1.244	1.372	1.514	1.672	1.845
	intsum	0.699	0.768	0.844	0.929	1.024	1.128	1.244	1.372	1.514	1.672	1.845
	estsum	0.699	0.768	0.844	0.929	1.024	1.128	1.244	1.372	1.514	1.672	1.845
	estqgl	0.699	0.768	0.844	0.929	1.024	1.128	1.244	1.372	1.514	1.672	1.845
	estsig	0.699	0.768	0.844	0.929	1.024	1.128	1.244	1.372	1.514	1.672	1.845
0.10	intrue	0.702	0.771	0.848	0.933	1.027	1.132	1.249	1.377	1.520	1.678	1.852
	intsum	0.702	0.771	0.848	0.933	1.027	1.132	1.249	1.377	1.520	1.678	1.852
	estsum	0.702	0.771	0.848	0.933	1.027	1.132	1.249	1.377	1.520	1.678	1.852
	estqgl	0.702	0.771	0.848	0.933	1.027	1.132	1.249	1.377	1.520	1.678	1.852
	estsig	0.702	0.771	0.848	0.933	1.027	1.132	1.249	1.377	1.520	1.678	1.852
0.20	intrue	0.712	0.782	0.860	0.947	1.043	1.150	1.268	1.398	1.543	1.703	1.880
	intsum	0.712	0.782	0.860	0.947	1.043	1.150	1.268	1.398	1.543	1.703	1.880
	estsum	0.712	0.782	0.860	0.947	1.043	1.150	1.268	1.398	1.543	1.703	1.880
	estqgl	0.712	0.782	0.860	0.947	1.043	1.150	1.268	1.398	1.543	1.703	1.880
	estsig	0.712	0.782	0.860	0.947	1.043	1.150	1.268	1.398	1.543	1.703	1.880
0.25	intrue	0.720	0.791	0.870	0.958	1.055	1.163	1.282	1.414	1.560	1.722	1.902
	intsum	0.720	0.791	0.870	0.958	1.055	1.163	1.282	1.414	1.560	1.722	1.902
	estsum	0.720	0.791	0.870	0.958	1.055	1.163	1.282	1.414	1.560	1.722	1.902
	estqgl	0.720	0.791	0.870	0.958	1.055	1.163	1.282	1.414	1.560	1.722	1.902
	estsig	0.720	0.791	0.870	0.958	1.055	1.163	1.282	1.414	1.560	1.722	1.902
0.50	intrue	0.791	0.869	0.956	1.052	1.159	1.277	1.408	1.553	1.714	1.892	2.089
	intsum	0.791	0.869	0.956	1.052	1.159	1.277	1.408	1.553	1.714	1.892	2.089
	estsum	0.791	0.869	0.956	1.052	1.159	1.277	1.408	1.553	1.714	1.892	2.089
	estqgl	0.791	0.869	0.956	1.052	1.159	1.277	1.408	1.553	1.714	1.892	2.089
	estsig	0.791	0.869	0.956	1.052	1.159	1.277	1.408	1.553	1.714	1.892	2.089
1.00	intrue	1.151	1.264	1.390	1.530	1.686	1.858	2.049	2.260	2.494	2.752	3.039
	intsum	1.151	1.264	1.390	1.530	1.686	1.858	2.049	2.260	2.494	2.752	3.039
	estsum	1.151	1.264	1.390	1.530	1.686	1.858	2.049	2.260	2.494	2.752	3.039
	estqgl	1.151	1.264	1.390	1.530	1.686	1.858	2.049	2.260	2.494	2.752	3.039
	estsig	1.151	1.264	1.390	1.530	1.686	1.858	2.049	2.260	2.494	2.752	3.039
1.20	intrue	1.434	1.575	1.732	1.907	2.100	2.315	2.552	2.816	3.107	3.429	3.786
	intsum	1.434	1.575	1.732	1.907	2.100	2.315	2.552	2.816	3.107	3.429	3.786
	estsum	1.434	1.575	1.732	1.907	2.100	2.315	2.552	2.816	3.107	3.429	3.786
	estqgl	1.434	1.575	1.732	1.907	2.100	2.315	2.552	2.816	3.107	3.429	3.786
	estsig	1.434	1.575	1.732	1.907	2.100	2.315	2.553	2.816	3.107	3.430	3.787

References

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

[2] M. J. Gurka, L. J. Edwards, K. E. Muller, and L. 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] E. V. Slud and T. 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] H. Chandra, K. Aditya, and S. 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] C. L. Bayes, J. L. Bazán, and M. 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] E. Lesaffre, D. Rizopoulos, and R. Tsonaka, "The logistic transform for bounded outcome scores," *Biostatistics*, vol. 8, no. 1, pp. 72–85, 04 2006.

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

[8] R. E. Fay and R. 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] G. S. Datta, T. Kubokawa, I. 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] N. Rojas-Perilla, S. Pannier, T. Schmid, and N. 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] S. Hadam, N. Würz, A.-K. Kreutzmann, and T. Schmid, "Estimating regional unemployment with mobile network data for functional urban areas in germany," *Statistical Methods & Applications*, Sep 2023.

[12] J. C. Pinheiro and D. M. Bates, *Mixed-Effects Models in S and S-PLUS*. New York: Springer, 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] S. X. Meng and S. 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] J. Bertrand, E. Comets, M. Chenel, and F. 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] T. T. Nguyen and F. 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] S. Sugasawa and T. Kubokawa, "Transforming response values in small area prediction," *Computational Statistics & Data Analysis*, vol. 114, pp. 47–60, 2017.

[18] D. W. Kammler, "Approximation with sums of exponentials in $L_p[0, +\infty)$," *Journal of Approximation Theory*, vol. 16, no. 4, pp. 384–408, 1976.

[19] R. 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.

[20] D. W. Kammler and R. J. McGlinn, "A bibliography for approximation with exponential sums," *Journal of Computational and Applied Mathematics*, vol. 4, pp. 167–173, 1978.

[21] R. Prony, "Essai experimental et analytique," *J. Ecole Polytechnique*, vol. 1, pp. 24–76, 1795.

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

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

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

[25] O. de Oliveira, "The Implicit and Inverse Function Theorems: Easy Proofs," *Real Analysis Exchange*, vol. 39, no. 1, pp. 207 – 218, 2013.

[26] W. S. Zeng and S. Z. Tang, "Bias correction in logarithmic regression and comparison with weighted regression for nonlinear models," *Nature Precedings*, Dec 2011.

[27] P. Hall and T. 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.

[28] X. Lin and N. 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.

[29] A. C. Tsai, M. Liou, M. Simak, and P. E. Cheng, "On hyperbolic transformations to normality," *Computational Statistics & Data Analysis*, vol. 115, pp. 250–266, 2017.

[30] 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.

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

[32] D. Shu, J. G. Young, S. Toh, and R. Wang, "Variance estimation in inverse probability weighted cox models," *Biometrics*, vol. 77, no. 3, pp. 1101–1117, 2021.

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

[34] R. L. Strawderman, "Higher-order asymptotic approximation: Laplace, saddlepoint, and related methods," *Journal of the American Statistical Association*, vol. 95, no. 452, pp. 1358–1364, 2000.