Data-driven Transformations in Small Area Estimation

Natalia Rojas-Perilla[*], Sören Pannier[*], Timo Schmid[*], and Nikos Tzavidis[**]

[*]Institute of Statistics and Econometrics, Freie Universität Berlin, Berlin, Germany
[**]Department of Social Statistics and Demography, Southampton Statistical Sciences Research Institute, University of Southampton, UK

Abstract

Small area models typically depend on the validity of model assumptions. For example, a commonly used version of the Empirical Best Predictor relies on the Gaussian assumptions of the error terms of the linear mixed regression model, a feature rarely observed in applications with real data. The present paper proposes to tackle the potential lack of validity of the model assumptions by using data-driven scaled transformations as opposed to ad-hoc chosen transformations. Different types of transformations are explored, the estimation of the transformation parameters is studied in detail under the linear mixed regression model and transformations are used in small area prediction of linear and non-linear parameters. The use of scaled transformations is crucial as it allows for fitting the linear mixed regression model with standard software and hence it simplifies the work of the data analyst. Mean squared error estimation that accounts for the uncertainty due to the estimation of the transformation parameters is explored using parametric and semi-parametric (wild) bootstrap. The proposed methods are illustrated using real survey and census data for estimating income deprivation parameters for municipalities in the Mexican state of Guerrero. Simulation studies and the results from the application show that using carefully selected, data-driven transformations can improve small area estimation.

Keywords: Random effects; bootstrap; adaptive transformations; maximum likelihood estimation; poverty mapping

1 Introduction

Model-based methods for small area estimation (SAE) are now widely used in practice for producing reliable estimates of linear and non-linear indicators for areas/domains with small sample sizes. Examples of indicators that are estimated by using model-based methods include poverty (income deprivation) and inequality measures such as the head count ratio, the poverty gap and the income quintile share ratio. Two popular small area methods in this case are the empirical best predictor (EBP), proposed by Molina and Rao (2010) and the World Bank method, proposed by Elbers et al. (2003). Both approaches are based on the use of unit-level linear mixed regression models. Although estimation of complex indicators can be also implemented with area-level linear mixed regression models (Fabrizi and Trivisano, 2016; Schmid et al., 2017), in this paper we focus on unit-level linear mixed regression models. In the original paper, Molina and Rao (2010) assumed that the error terms of the linear mixed regression model follow a Gaussian distribution. In case the model error terms significantly deviate from normality, the EBP estimator can be biased. What are the options available to the data analyst when the normality assumptions are

not met? One option is to formulate the EBP under alternative and more flexible parametric assumptions. Graf et al. (2018) study an EBP method under the generalized beta distribution of the second kind (GB2), whereas Diallo and Rao (2014) propose the use of skewed-normal distributions in applications with income data. One complication with using the EBP under alternative parametric distributions is that new tools for estimation must be developed and training for the data analyst is needed. In addition, misspecification of the model assumptions is still possible. Another option when the Gaussian assumptions are not satisfied is to use a methodology that minimizes the use of parametric assumptions. For instance, Elbers and van der Weide (2014) proposed an EBP method based on normal mixture models. With this method the distribution of the error terms is described by normal mixtures. Weidenhammer et al. (2014) recently proposed a method that aims at estimating the quantiles of the empirical distribution function of the data. The estimation of the quantiles is facilitated by a nested error regression model using the asymmetric Laplace distribution for the unit-level error terms as a working assumption. The estimation of the random effects can be made completely non-parametric by using a discrete mixture proposed by Marino et al. (2016). Another option, and the one we study in this paper, is to find an appropriate transformation such that the model assumptions (in this paper the Gaussian assumptions of the EBP method) hold. The aim is to find transformations that (a) are data-driven and optimal according to some criterion and (b) can be implemented by using standard software. To the best of our knowledge, the use and choice of transformations in SAE has not been extensively studied or it has been studied in fairly ad-hoc manner. Elbers et al. (2003) and Molina and Rao (2010) suggested the use of logarithmic-type transformations for income data. However, are such transformations the most appropriate choice? Can alternative transformations offer improved estimation? In order to answer these research questions, the paper investigates data-driven transformations for small area estimation.

The choice of transformations when modelling income-type outcomes presents different challenges. Transformations should be suitable for dealing with unimodal, leptokurtic and positively skewed data that may include zero and negative values. Besides the logarithmic transformation and its modifications (e.g. the log-shift transformation) a popular family of data-driven transformations that includes the logarithmic one as a special case is the Box-Cox family (Box and Cox, 1964). Since the Box-Cox transformation is not defined for negative values, when negative values are present, the data must be shifted to the positive range. Another difficulty with the use of the Box-Cox transformation is the truncation on the transformation parameter described later in Section 4. A solution to this problem can be offered by using of the dual power transformation. Although extensive literature on the use of transformations exists, see for example, John and Draper (1980), Bickel and Doksum (1981) and Yeo and Johnson (2000) among others. In this paper we focus on three types of transformations, namely log-shift, Box-Cox and dual power transformations.

In addition to selecting the type of transformation, estimating the transformation parameter adds another layer of complexity. To the best of our knowledge the use of transformations in recent applications of SAE has employed visual residual diagnostics for finding a suitable transformation parameter. In this paper we propose a structured, data-driven approach for estimating the transformation parameter. In particular, we introduce maximum likelihood and residual maximum likelihood methods for estimating the transformation parameter under the linear mixed regression model following Gurka et al. (2006). Alternative estimation approaches based on the minimization of distances (Cramér, 1928; Chakravarti and Laha, 1967) and on the minimization of the skewness (Carroll and Ruppert, 1987) are also discussed.

At this point we should emphasise some of the differences between the present paper and the paper published by Tzavidis et al. (2018). The latter paper proposes a general framework for the production

of small area statistics including measuring uncertainty. Broadly speaking, the proposed framework is based on three stages, namely specification of the problem, analysis of the data and adaptation of the model, and method evaluation. The paper focuses on practical aspects of the SAE process and not on proposing new methodology. The target audience includes practitioners using small area estimation methods for example, colleagues in National Statistical Institutes. The use of transformations, as a parsimonious approach to adapting the model, is mentioned in the paper but the methodological details are not derived. In contrast, the present paper focuses on developing new and generally applicable methodology that underpins the use of data-driven transformations in SAE and applies the methodology to real data problems. In particular, the current paper proposes the use of the empirical best predictor (EBP) (Molina and Rao, 2010) with data-driven transformations estimated with likelihood-based methods. The paper focuses on scaled transformations that allow the use of standard software for small area estimation. As we mentioned above, the focus is on the use of the log-shift, Box-Cox, and dual power transformations and the mathematical derivations for developing scaled transformations are presented. Illustrating how to derive scaled transformations for these three transformation types will allow researchers to use similar developments for other families of transformations. In addition, in the present paper we propose two bootstrap schemes (parametric- and wild-type) for estimating the MSE under data-driven transformations and extend these to capture the additional uncertainty due to the estimation of the transformation parameter. The wild bootstrap scheme can been viewed as an insurance policy in case there are some "mild" departures from Normality after using transformations. Finally, the present paper includes results from model-based simulation studies that are necessary for comparing the performance of data-driven transformations against the use of fixed, ad-hoc transformations. Emphasis is given to the estimation of poverty and inequality indicators due to their important socio-economic relevance and policy impact. We further study whether the impact of departures from Gaussian assumptions is different depending on the target of estimation. For instance, departures from normality may have lesser impact on estimates of median income compared to estimates indicators that are more sensitive in the data distribution. The use of model-based simulations was one of the method evaluation approaches recommended in the paper by Tzavidis et al. (2018).

The rest of the paper is structured as follows. The EBP approach is introduced in Section 2. Section 3 presents the survey data we use in this paper and makes the case, via the use of residual diagnostics, for using transformations. In Section 4 selected transformations are introduced and extended for their use with model-based SAE methods under the linear mixed regression model. This section includes the theoretical details about the choice of an appropriate scale and estimation of the transformation parameter. MSE estimation is discussed in Section 5. In Section 6 the proposed methods are applied to data from Guerrero in Mexico for estimating a range of deprivation and inequality indicators and corresponding estimates of uncertainty. In Section 7 the proposed methods are further evaluated by realistic - for income data - model-based simulations. Section 8 summarizes the main findings and outlines further research.

2 The Empirical Best Prediction (EBP) method

Let U denote a finite population of size N partitioned into D areas or domains (representing the small areas) $U_1, U_2, \ldots, U_D$ of sizes $N_1, \ldots, N_D$, where $i = 1, \ldots, D$ refers to the ith area. Let y_{ij} be the target variable defined for the jth individual belonging to the ith area, with $j = 1, \ldots, N_i$. Denote by $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_p)^T$ the design matrix containing p explanatory variables and define by s as the

set of sample units, with s_i the in-sample units in area i and by r be the set of non-sampled units, with r_i the out-of-sample units in area i. Let n_i denote the sample size in area i with $n = \sum_{i=1}^{D} n_i$. Hence, we define by $\mathbf{y}_i$ a vector with population elements of the target outcome for area i partitioned as $\mathbf{y}_i^T = \left(\mathbf{y}_{is}^T, \mathbf{y}_{ir}^T \right)$, where $\mathbf{y}_{is}$ and $\mathbf{y}_{ir}$ denote the sample elements s and the out-of-sample elements r in area i respectively. Let us now describe in more detail the EBP approach by Molina and Rao (2010), which is the methodology we focus on in this paper. Under this approach census predictions of the target outcome are generated by using the conditional predictive distribution of the out-of-sample data given the sample data. The point of departure is the standard parametric unit-level linear mixed regression model, which is also known as the unit-level nested error regression model. This is defined by Battese et al. (1988) as:

$$ y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + u_i + e_{ij}, \quad u_i \overset{iid}{\sim} N(0, \sigma_u^2) \quad \text{and} \quad e_{ij} \overset{iid}{\sim} N(0, \sigma_e^2), \tag{1} $$

where u_i, the area-specific random effects, and e_{ij}, the unit-level error, are assumed to be independent. Assuming normality for the unit-level error and the area-specific random effects, the conditional distribution of the out-of-sample data given the sample data is also normal. A Monte Carlo approach is used to obtain a numerically efficient approximation to the expected value of this conditional distribution as follows:

1. Use the sample data to obtain $\hat{\boldsymbol{\beta}}, \hat{\sigma}_u^2, \hat{\sigma}_e^2$ and the weighting factors $\hat{\gamma}_i = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \frac{\hat{\sigma}_e^2}{n_i}}$.

2. For $l = 1, \ldots, L$:

   (a) Generate $v_i^{(l)} \overset{iid}{\sim} N\left(0, \hat{\sigma}_u^2(1 - \hat{\gamma}_i)\right)$ and $e_{ij}^{(l)} \overset{iid}{\sim} N\left(0, \hat{\sigma}_e^2\right)$ and obtain a pseudo-population of the target variable by:
   $$ y_{ij}^{(l)} = \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}} + \hat{u}_i + v_i^{(l)} + e_{ij}^{(l)}, $$
   where the predicted random effect $\hat{u}_i$ is defined as $\hat{u}_i = E(u_i | \mathbf{y}_{is})$.

   (b) Calculate the indicator of interest $I_i^{(l)}$ in each area.

3. Finally, take the mean over the L Monte Carlo runs in each area to obtain a point estimate of the indicator of interest:
   $$ \hat{I}_i^{EBP} = \frac{1}{L} \sum_{l=1}^{L} I_i^{(l)}. $$

As is common in real applications, some areas are out-of-sample. For those areas, we cannot estimate an area-specific random effect, and hence the corresponding area-specific random effect is set equal to zero. Synthetic values of the outcome for the out-of-sample areas are then generated under the linear mixed regression model as follows:

$$ y_{ij}^{(l)} = \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}} + u_i^{(l)} + e_{ij}^{(l)}, $$

with $u_i^{(l)} \overset{iid}{\sim} N\left(0, \hat{\sigma}_u^2\right)$ and $e_{ij}^{(l)} \overset{iid}{\sim} N\left(0, \hat{\sigma}_e^2\right)$. Finally, a parametric bootstrap - under the assumed model - is used for the MSE estimation. This is discussed in some detail in Section 5. Assuming normality for the error terms is a convenient assumption as allows the conditional distribution of $\mathbf{y}_r | \mathbf{y}_s$ to be derived. However, in applications that involve modelling an income-type outcome, as is the case in this paper, assuming normality is unrealistic. If our primary target is to develop a methodology that can easily be used in practice, finding appropriate data transformations is important.

3 The Guerrero case study: Data source and initial analysis

In this section, we describe the data sources used in the application and provide a motivation for the use of transformations. The case study was carried out by using the open-source software R (R Core Team, 2017) and R packages. The data we use in this paper come from the Mexican state of Guerrero, one of the 32 states in Mexico. The state Guerrero is considered by the World Bank to be one of the states - next to the State of Mexico investigated by Tzavidis et al. (2018) - mostly contributing to income inequality in Mexico (Bedoya et al., 2013). Additionally, according to the United Nations Development Programme (UNDP), Guerrero has one of the highest rates of poverty and lack of infrastructural development (Tortajada, 2006). According to the general social development law in Mexico, the National Institute of Statistics and Geography (INEGI) has to provide relevant official statistics at the national, state and municipal-levels. Furthermore, the Social Development Law (SDL) in Mexico establishes that the National Council for the Evaluation of Social Development Policy (CONEVAL) should measure poverty at state level every two years and at municipal level every five years. For carrying out the analysis the statistical and geographical information was provided by INEGI through the Household Income and Expenditure Survey (ENIGH) 2010 and the National Population and Housing Census of 2010. Looking in more detail at the data available and their geographic coverage, Guerrero comprises 81 administrative divisions, known as municipalities. From the 81 municipalities 40 municipalities with 1611 households are in-sample (in the sample of the ENIGH survey) and the remaining 41 municipalities are out-of-sample. For the in-sample municipalities the maximum sample size in a municipality is 511, the minimum is 9 and the median is 24 households. Note that more than 30% of the sample is from a single municipality, the capital (Chilpancingo de los Bravo).

The survey and census data include a large number of socio-demographic variables, which are common and are measured similarly in both data sources. The total household per capita income (*ictpc*, measured in pesos) is a variable recorded for households and is available in the survey but not in the census. We used this variable as a proxy that best approximates the living standard in Guerrero and as the outcome variable in our models. Socio-economic variables available for the households both in the survey and census data are used as explanatory variables. The underlying linear mixed regression model (1) of the EBP has two levels, households and municipalities. The variables available in the survey and census data, which are identified by using Bayesian information criterion (BIC) as good predictors of *ictpc*, are described in Table 1. From now on, the working model is assumed to be known and fixed.

Table 1: Description of the explanatory variables used in the working model

Determinant	Variable
Occupation	1) Indicator if the head of household and the spouse are employed
	2) Type of household occupation
	3) Total number of employees older than 14 years in a household
	4) Percentage of employees older than 14 years in a household
Sources of income	5) Indicator of a household receiving remittances
Socioeconomic level	6) Availability of assets in the household
	7) Total number of goods in the household
Education	8) Average standardized years of schooling (by age and sex) within the household relative to the population

The next step after the identification of a possible set of covariates is assessing the predictive power of the model. Nakagawa and Schielzeth (2013) propose the use of two coefficients of determination suitable for linear mixed regression models: (a) the marginal R_m^2, which is a measure for the variance

explained by fixed effects and (b) the conditional R_c^2, which measures the variance explained by both, the fixed and random effects. Without using any transformation, these measures are both around 35% and the corresponding intraclass correlation (ICC) under the model is 0.027.

In order to explore the validity of the Gaussian assumptions underlying the linear mixed regression model, it is common practice to perform normality tests and some residual diagnostics. The p-values of the Shapiro-Wilk (S-W) test statistic are equal to $2.2 \cdot 10^{-16}$ for the household-level and 0.197 for the municipal-level. These results indicate that the null hypothesis of normality for the household-level is rejected. As normality tests like Shapiro-Wilk have some problems we also present some visual approaches in addition. Figure 1 presents the Normal probability quantile-quantile (Q-Q) plots for household-level and municipal-level residuals. As expected, in the case of using the non-transformed *ictpc* variable, the shape of the Q-Q plots is clearly different from what would be expected under normality. In addition, the analysis of skewness and kurtosis for both error terms is also informative. The skewness and kurtosis for a Normal distribution are equal to zero and three, respectively. The skewness and kurtosis of the household-level are equal to 6.338 and 75.483, and for the municipal-level equal to 0.448 and 3.250. These results indicate severe departures - especially for the household-level - from the Gaussian assumptions when modelling the non-transformed income.
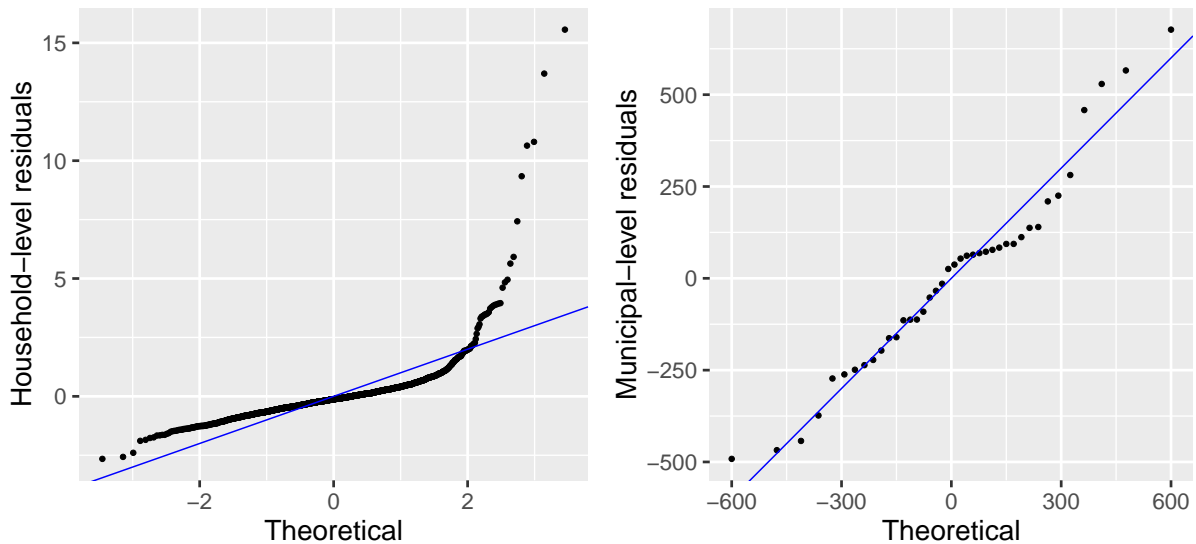

Figure 1: Q-Q plots of the household- and municipal-level error terms

4 Use of transformations

In order to get closer to normality, it is common to use a one-to-one transformation $T(y_{ij}) = y_{ij}^*$ of the target variable. The application of the natural logarithmic transformation, which is a popular choice for income data, leads in many cases from right-skewed to more symmetric distributions. This approach is followed by the paper by Molina and Martín (2018), in which the logarithmic transformation is applied to an income-type variable for meeting the assumptions of the model proposed by Battese et al. (1988). In particular, the paper by Molina and Martín (2018) proposes analytic MSE estimators and develops bias correction terms necessary when using estimating small area averages using a logarithmic transformation under the linear mixed regression model. The logarithmic transformation is frequently used for dealing with non-normality due to its simplicity. However, can an alternative transformation with data-driven

parameter(s) λ, $T_\lambda(y_{ij}) = y^*_{ij}(\lambda)$, offer small area estimates with improved precision?

The structure of the section is as follows. In Section 4.1 we introduce the EBP approach with data-driven transformations. In Section 4.2 we propose likelihood-based approaches for estimating the transformation parameter, λ, in general and discuss three particular subcases - the log-shift, Box-Cox and dual power transformations - in detail. Finally, in Section 4.3 we discuss alternative to likelihood-based approaches for estimating the transformation parameter.

4.1 EBP under transformations

In order to apply the EBP method by using transformations, the linear mixed regression model is re-defined as follows:

$$y^*_{ij}(\lambda) = \mathbf{x}^T_{ij}\boldsymbol{\beta} + u_i + e_{ij}, \quad u_i \overset{iid}{\sim} N(0, \sigma^2_u) \text{ and } e_{ij} \overset{iid}{\sim} N(0, \sigma^2_e). \tag{2}$$

The EBP approach under transformations can be re-written as follows:

1. Select a transformation and obtain $T_\lambda(y_{ij}) = y^*_{ij}(\lambda)$.

2. Use the transformed sample data to obtain $\hat{\boldsymbol{\beta}}, \hat{\sigma}^2_u, \hat{\sigma}^2_e$ and calculate the weighting factors, $\hat{\gamma}_i = \frac{\hat{\sigma}^2_u}{\hat{\sigma}^2_u + \frac{\hat{\sigma}^2_e}{n_i}}$.

3. For $l = 1, \ldots, L$:

    (a) Generate $v^{(l)}_i \overset{iid}{\sim} N(0, \hat{\sigma}^2_u(1 - \hat{\gamma}_i))$ and $e^{(l)}_{ij} \overset{iid}{\sim} N(0, \hat{\sigma}^2_e)$ and obtain a pseudo-population of the target variable by:
    $$y_{ij}{}^{*(l)} = \mathbf{x}^T_{ij}\hat{\boldsymbol{\beta}} + \hat{u}_i + v^{(l)}_i + e^{(l)}_{ij}.$$

    (b) Back-transform $y_{ij}^{*(l)}$ to the original scale $y^{(l)}_{ij} = T^{-1}_\lambda\left(y_{ij}^{*(l)}\right)$.

    (c) Calculate the indicator of interest $I^{(l)}_i$ in each area.

4. Finally, take the mean over the L Monte Carlo generations in each area to obtain an estimate of the indicator of interest:
$$\hat{I}^{EBP}_i = \frac{1}{L}\sum_{l=1}^{L} I^{(l)}_i.$$

4.2 Likelihood-based approach for estimating λ

For estimating the transformation parameter λ, the linear mixed regression model defined in (2) is used. Assume that the transformed vectors $\boldsymbol{y}^*_i$ are independent and normally distributed for some unknown λ,

$$\boldsymbol{y}^*_i(\lambda) \sim N(\boldsymbol{\mu}_i, \mathbf{V}_i) \text{ for } i = 1, \ldots, D,$$

where

$$\boldsymbol{\mu}_i = \mathbf{X}_i\boldsymbol{\beta} \quad \text{and} \quad \mathbf{V}_i = \sigma^2_u \mathbf{1}_{N_i}\mathbf{1}'_{N_i} + \sigma^2_e \mathbf{I}_{N_i},$$

with $\mathbf{1}_{N_i}$ a column vector of ones of size N_i and $\boldsymbol{I}_{N_i}$ the $N_i \times N_i$ identity matrix, the vector of unknown model parameters is $\boldsymbol{\theta}^T = (\boldsymbol{\beta}, \sigma^2_u, \sigma^2_e, \lambda)$. The log-likelihood function under the model is defined as

follows:

$$
l_{\mathrm{ML}}(\boldsymbol{y}^*, \lambda | \boldsymbol{\theta}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{D}\log|\boldsymbol{V}_i|
$$

$$
- \frac{1}{2}\sum_{i=1}^{D}[\boldsymbol{y}_i^*(\lambda) - \boldsymbol{X_i}\hat{\boldsymbol{\beta}}]^T \boldsymbol{V}_i^{-1}[\boldsymbol{y}_i^*(\lambda) - \boldsymbol{X_i}\hat{\boldsymbol{\beta}}].
$$

The log-likelihood function in relation to the original observations is obtained by multiplying the normal density by the log of the Jacobian of the transformation from $\boldsymbol{y}_i$ to $\boldsymbol{y}_i^*(\lambda)$. The Jacobian $J(\lambda, \boldsymbol{y})$ is defined as $\prod_{i=1}^{D}\prod_{j=1}^{n_i}\left|\frac{dy_{ij}^*(\lambda)}{dy_{ij}}\right|$ and is incorporated as follows:

$$
l_{\mathrm{ML}}(\boldsymbol{y}, \lambda | \boldsymbol{\theta}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{D}\log|\boldsymbol{V}_i|
$$

$$
- \frac{1}{2}\sum_{i=1}^{D}[\boldsymbol{y}_i^*(\lambda) - \boldsymbol{X_i}\hat{\boldsymbol{\beta}}]^T \boldsymbol{V}_i^{-1}[\boldsymbol{y}_i^*(\lambda) - \boldsymbol{X_i}\hat{\boldsymbol{\beta}}] + \log J(\lambda, \boldsymbol{y}).
$$

The maximization of $l_{\mathrm{ML}}(\boldsymbol{y}, \lambda | \boldsymbol{\theta})$ produces maximum likelihood (ML) estimates of the unknown parameters $\boldsymbol{\theta}$. However, in the theory of linear mixed regression models, when interest focuses on accurate estimators of the variance components, restricted maximum likelihood (REML) theory is recommended (Verbeke and Molenberghs, 2000). The REML is defined as follows:

$$
l_{\mathrm{REML}}(\boldsymbol{y}, \lambda | \boldsymbol{\theta}) = -\frac{n-p}{2}\log(2\pi) + \frac{1}{2}\log\left|\sum_{i=1}^{D}\boldsymbol{X}_i^T\boldsymbol{X}_i\right| - \frac{1}{2}\sum_{i=1}^{D}\log|\boldsymbol{V}_i|
$$

$$
- \frac{1}{2}\log\left|\sum_{i=1}^{D}\boldsymbol{X}_i^T\boldsymbol{V}_i^{-1}\boldsymbol{X}_i\right|
$$

$$
- \frac{1}{2}\sum_{i=1}^{D}[\boldsymbol{y}_i^*(\lambda) - \boldsymbol{X_i}\hat{\boldsymbol{\beta}}]^T \boldsymbol{V}_i^{-1}[\boldsymbol{y}_i^*(\lambda) - \boldsymbol{X_i}\hat{\boldsymbol{\beta}}] + \log J(\lambda, \boldsymbol{y}). \tag{3}
$$

The use of the scaled version of a selected transformation, defined by $\frac{y_{ij}^*(\lambda)}{J(\lambda, \boldsymbol{y})^{\frac{1}{n}}} = z_{ij}^*(\lambda)$, is crucial for estimating the transformation parameter λ under the REML approach presented above. The Jacobian of such a scaled transformation is equal to 1. This means, the scale of the likelihood is preserved independently of the transformation and its parameter λ. Therefore, values of the log-likelihood function - under differently transformed $y_{ij}^*(\lambda)$ - can be directly compared and the log-likelihood function simplifies to the log-likelihood function of the linear mixed regression model. As a result, standard software for fitting this model can be used to estimate the transformation parameter λ. Eventhough using scaled transformations aids the implementation of the methods in practice, appropriate scaling factors must be developed depending on the type of transformation used.

Although the theory is applicable to data-driven transformations in general, we focus on the three types of transformations we presented in Section 1, namely the log-shift, Box-Cox and dual power transformations. Additionally, we use the frequently applied logarithmic transformation as a benchmark. This transformation is defined by $y_{ij}^* = \log(y_{ij} + s)$, where s denotes a fixed parameter such that $y_{ij} + s > 0$. The log-shift transformation (Yang, 1995), presented below, extends the logarithmic transformation by

including the data-driven transformation parameter, $\lambda \geq s$, which needs to be estimated:

$$y_{ij}^*(\lambda) = \log(y_{ij} + \lambda).$$

When $\lambda = s$, the logarithmic transformation is obtained. The Box-Cox transformation (Box and Cox, 1964) is defined as follows:

$$y_{ij}^*(\lambda) = \begin{cases} \frac{(y_{ij}+s)^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0, \\ \log(y_{ij} + s) & \text{if } \lambda = 0. \end{cases}$$

where s denotes a fixed parameter such that $y_{ij} + s > 0$. If $\lambda = 0$, the logarithmic transformation is then a special case and if $\lambda = 1$, the data are only shifted. One difficulty with the Box-Cox type transformations is the long-standing truncation, i.e. $y_{ij}^*(\lambda)$ is bounded, from below by $\frac{1}{\lambda}$ if $\lambda > 0$ and from above by $\frac{-1}{\lambda}$ if $\lambda < 0$. This is the key motivation for the third type of transformation. The dual power transformation, introduced by Yang (2006), is defined as follows:

$$y_{ij}^*(\lambda) = \begin{cases} \frac{(y_{ij}+s)^\lambda - (y_{ij}+s)^{-\lambda}}{2\lambda} & \text{if } \lambda > 0, \\ \log(y_{ij} + s) & \text{if } \lambda = 0, \end{cases}$$

where s is defined as in the case of the Box-Cox transformation.

The corresponding Jacobian used in (3) and scaled versions of the log-shift, Box-Cox and dual power transformations are presented in Table 2. For more details we refer to the developments in Appendix A.

Table 2: Jacobian and scaled data-driven transformations for log-shift, Box-Cox and dual

Transformation	Jacobian J	Scaled transformation $z_{ij}^*(\lambda)$
Log-Shift	$\prod_{i=1}^D \prod_{j=1}^{n_i} (y_{ij} + \lambda)^{-1}$	$J^{\frac{-1}{n}} \log(y_{ij} + \lambda)$
Box-Cox	$\prod_{i=1}^D \prod_{j=1}^{n_i} y_{ij}^{\lambda-1}$	$J^{\frac{-1}{n}} \frac{(y_{ij}+s)^\lambda - 1}{\lambda}, \quad \text{if } \lambda \neq 0$ $J^{\frac{-1}{n}} \log(y_{ij} + s) \quad \text{if } \lambda = 0$
Dual	$\frac{\prod_{i=1}^D \prod_{j=1}^{n_i} \left((y_{ij}+s)^{\lambda-1} + (y_{ij}+s)^{-\lambda-1} \right)}{2}$	$J^{\frac{-1}{n}} \frac{(y_{ij}+s)^\lambda - (y_{ij}+s)^{-\lambda}}{2\lambda}, \quad \text{if } \lambda \neq 0$ $J^{\frac{-1}{n}} \log(y_{ij} + s) \quad \text{if } \lambda = 0$

4.3 Alternative approaches for estimating λ

The ML and REML approaches introduced in 4.2 rely on parametric assumptions that may be influenced by outliers in the data. The kurtosis and skewness are crucial features for defining the shape of a distribution and a proximity measure can be minimized in order to find a transformation parameter under which the empirical distribution of residuals has skewness and kurtosis as close as possible to zero and three respectively. In general, skewness is considered more important than kurtosis, therefore, minimizing the skewness is an approach already considered in the literature (Royston and Lambert, 2011) for linear models as follows:

$$\hat{\lambda}_{\text{skew}} = \underset{\lambda}{\arg\min} |S_{e_\lambda}|,$$

where S_{e_λ} is the skewness and $\sigma^2_{e_\lambda}$ denotes the variance of the unit-level error terms. Note that the index λ is used to emphasize that the skewness and the variance parameters depend on the transformation parameter. In the context of linear mixed regression models, an additional problem arises as there are two independent error terms to be considered. We propose a pooled skewness approach that uses a weight w to ensure that the larger the error term variance $\sigma^2_{e_\lambda}$ is, the more weight its skewness will have in the minimization. Let S_{u_λ} be the skewness and $\sigma^2_{u_\lambda}$ be the variance of the area-specific random effects u_i of the linear mixed regression model. The estimation criteria in the pooled skewness approach is defined as follows:

$$\hat{\lambda}_{\text{poolskew}} = \underset{\lambda}{\arg\min} \left(w|S_{e_\lambda}| + (1-w)|S_{u_\lambda}| \right),$$

$$\text{where } w = \frac{\hat{\sigma}^2_{e_\lambda}}{\hat{\sigma}^2_{u_\lambda} + \hat{\sigma}^2_{e_\lambda}}.$$

Considering only the skewness may ignore other properties of the distribution. Hence, a measure describing the distance between two distribution functions is another alternative. Two distance measures, the Kolmogorov-Smirnov (KS) and the Cramér-von Mises (CvM) are used,

$$\hat{\lambda}_{\text{KS}} = \underset{\lambda}{\arg\min} \ \sup \ |F_n(\cdot) - \Phi(\cdot)|,$$

$$\hat{\lambda}_{\text{CvM}} = \underset{\lambda}{\arg\min} \ \int_{-\infty}^{\infty} [F_n(\cdot) - \Phi(\cdot)]^2 \, \phi(\cdot),$$

where $F_n(\cdot)$ is the empirical cumulative distribution function estimated by using the normalized residuals, $\Phi(\cdot)$ is the distribution function of a standard normal distribution and $\phi(\cdot)$ its density. The impact of using alternative approaches for estimating λ is studied in a model-based simulation study in Section 7.3.

5 MSE estimation under transformations

Estimating the MSE of small area estimates is a challenging problem. In the case of the EBP Molina and Rao (2010) propose a parametric bootstrap procedure following González-Manteiga et al. (2008). In this section we propose two bootstrap schemes for estimating the MSE under transformations. These bootstrap MSE estimators are extended to capture the additional uncertainty due to the estimation of the transformation parameter λ. The difference between the two bootstrap schemes is the mechanism used for generating the bootstrap population. In particular, the first bootstrap generates bootstrap realisations of the random effects and unit-level error terms parametrically. In contrast, the second one is a semi-parametric wild bootstrap which aims to protect against departures from the assumptions of the model in particular, those of the unit-level error term.

The steps of the proposed parametric bootstrap are as follows:

1. For $b = 1, ..., B$

   (a) Using the sample estimates, $\hat{\boldsymbol{\beta}}, \hat{\sigma}^2_u, \hat{\sigma}^2_e, \hat{\lambda}$, generate $u_i^{(b)} \overset{iid}{\sim} N\left(0, \hat{\sigma}^2_u\right)$ and $e_{ij}^{(b)} \overset{iid}{\sim} N\left(0, \hat{\sigma}^2_e\right)$ and simulate a bootstrap super-population $y_{ij}^{*(b)} = \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}} + u_i^{(b)} + e_{ij}^{(b)}$.

   (b) Back-transform $y_{ij}^{*(b)}$ to the original scale $y_{ij}^{(b)} = T_\lambda^{-1}\left(y_{ij}^{*(b)}\right)$ and compute the population value of the indicator of interest $I_{i,b}$.

(c) Extract the bootstrap sample in $y_{ij}{}^{(b)}$ and perform the EBP method, as described in Section 4.1. Note, as the back-transformed sample data are used, the transformation parameter λ is re-estimated in each bootstrap replication b.

(d) Obtain $\hat{I}_{i,b}^{EBP}$.

2. $\widehat{MSE}\left(\hat{I}_i^{EBP}\right) = B^{-1}\sum_{b=1}^B \left(\hat{I}_{i,b}^{EBP} - I_{i,b}\right)^2.$

As mentioned before, the proposed parametric bootstrap allows for the additional uncertainty due to the estimation of the transformation parameter. Although the use of an optimal transformation may reduce the deviation from normality, there may still be departures from normality especially in the tails of the distribution of the unit-level error term. To overcome this problem, we propose a semi-parametric bootstrap that relies on the normality of the random effects but generates the unit-level error terms by using the empirical distribution of suitably scaled unit-level residuals. The proposed wild bootstrap scheme is described below:

1. Fit the model 1 using an appropriate transformation $T(y_{ij}) = y_{ij}^*$ and obtain $\hat{\boldsymbol{\beta}}, \hat{\sigma}_u^2, \hat{\sigma}_e^2, \hat{\lambda}$.

2. Calculate the sample residuals by $\hat{e}_{ij} = y_{ij} - \mathbf{x}_{ij}^T\hat{\boldsymbol{\beta}} - \hat{u}_i$.

3. Scale and center the residuals using $\hat{\sigma}_e$. The scaled and centered residuals are denoted by $\hat{\epsilon}_{ij}$.

4. For $b = 1, ..., B$

   (a) Generate $u_i^{(b)} \overset{iid}{\sim} N(0, \hat{\sigma}_u^2)$.

   (b) Calculate the linear predictor $\eta_{ij}^{(b)}$ by $\eta_{ij}^{(b)} = \mathbf{x}_{ij}^T\hat{\boldsymbol{\beta}} + u_i^{(b)}$.

   (c) Match $\eta_{ij}^{(b)}$ with the set of estimated linear predictors $\{\hat{\eta}_k | \eta \in n\}$ from the sample by using

   $$\min_{k \in n} \left| \eta_{ij}^{(b)} - \hat{\eta}_k \right|$$

   and define $\tilde{k}$ as the corresponding index.

   (d) Generate weights w from a distribution satisfying the conditions in Feng et al. (2011) where w is a simple two-point mass distribution with probabilities 0.5 at $w = 1$ and $w = -1$, respectively.

   (e) Calculate the bootstrap population as $y_{ij}^{*(b)} = \mathbf{x}_{ij}^T\hat{\boldsymbol{\beta}} + u_i^{(b)} + w_k|\hat{\epsilon}_{\tilde{k}}^{(b)}|$.

   (f) Back-transform $T(y_{ij}^{*(b)})$ to the original scale and compute the population value $I_{i,b}$.

   (g) Extract the bootstrap sample in $y_{ij}{}^{(b)}$ and use the EBP method, as described in Section 4.

   (h) Obtain $\hat{I}_{i,b}^{EBP}$.

5. $\widehat{MSE}_{wild}\left(\hat{I}_i^{EBP}\right) = B^{-1}\sum_{b=1}^B \left(\hat{I}_{i,b}^{EBP} - I_{i,b}\right)^2.$

The performance of both MSE estimators is compared in a model-based simulation study in Section 7.

6 The Guerrero case study: Application of data-driven transformations

The benefits of using the proposed EBP approach with data-driven transformations for estimating deprivation and inequality indicators are illustrated in an application using the household data from the

ENIGH survey 2010 and the National Population and Housing Census 2010 we introduced in Section 3. The aim is to estimate the head count ratio (HCR) and the poverty gap (PGAP) as well as the income quintile share ratio (QSR) for the 81 municipalities in Guerrero. As the ENIGH survey and the census contains only contain information on household level we estimate the poverty and inequality indicators for households and not individuals.

The indicators HCR and PGAP are special cases of the Foster-Greer-Thorbecke (FGT) indicators (Foster et al., 1984) and they depend on a poverty line t which is equal to 0.6 times the median of the target variable. The FGT index of type α for an area i is defined by

$$F_i(\alpha, t) = \frac{1}{N_i} \sum_{j=1}^{N_i} \left(\frac{t - y_{ij}}{t} \right)^\alpha \mathbb{I}(y_{ij} \leq t), \quad \text{for } \alpha = 0, 1, 2,$$

where $\mathbb{I}(\cdot)$ denotes an indicator function which returns 1 if $(\cdot)$ holds and 0 otherwise. When $\alpha = 0$, $F_i(\alpha, t)$ is the HCR and represents the proportion of the households whose income is below the poverty line t. Taking $\alpha = 1$, $F_i(\alpha, t)$ defines the PGAP which is a measure of poverty intensity and quantifies the degree, to which the average income of people living under the poverty line differs from the poverty line. In addition to the two deprivation indicators, we investigate inequality by the QSR defined by

$$\text{QSR}_i = \frac{\sum_{j=1}^{N_i} \mathbb{I}(y_{ij} \geq \mathbf{y}_{0.8}) \, y_{ij}}{\sum_{j=1}^{N_i} \mathbb{I}(y_{ij} \leq \mathbf{y}_{0.2}) \, y_{ij}},$$

where $\mathbf{y}_{0.8}$ and $\mathbf{y}_{0.2}$, denote the 80% and 20% quantiles of the target variable respectively. The QSR is a widely used inequality indicator due to its simplicity and straightforward interpretation (Eurostat, 2004). The estimation of the QSR is challenging characterized by a large variability also in large samples. However, we have decided to report the estimated QSR in this paper only for illustrative purposes. In particular, we are interested in showing the increasing importance of the model assumptions when the target parameter depends of the tails of the data distribution and the importance of using data transformation parameters.

Before focusing on the state of Guerrero, we briefly illustrate the need for data-driven transformations in different states in Mexico. Figure 2 represents the estimated data-driven Box-Cox transformation parameters $\hat{\lambda}$ (by REML) for each state in Mexico. These estimates vary between 0.13 and 0.37, showing the adaptive feature of data-driven transformations for each state in Mexico. Furthermore, we observe that a fixed logarithmic transformation is not suitable for any of the states.

6.1 Model checking and residual diagnostics

In Section 3 we show that the model assumptions of the working model in the state of Guerrero are not met. We now discuss the use of the proposed data-driven transformations for adapting the working model. In particular, we focus on the three data-driven transformations presented in Section 4.2, denoted by *Log-Shift*, *Box-Cox* and *Dual* power transformations and their comparison to (a) a model that use a logarithmic transformation (*Log*) and (b) a model that uses the untransformed income variable (*No*).

To start with, Figure 3 provides a graphical representation of the REML maximization for the transformation parameter λ for log-shift, Box-Cox and dual power transformations in the state of Guerrero. In this case the optimal λs are approximately equal to 156.44, 0.20 and 0.23, respectively (cf. Table 3).

In order to analyze whether the use of transformations improves the predictive power of the model, Table 3 reports the percentage of variability explained for each model and its corresponding ICC. As
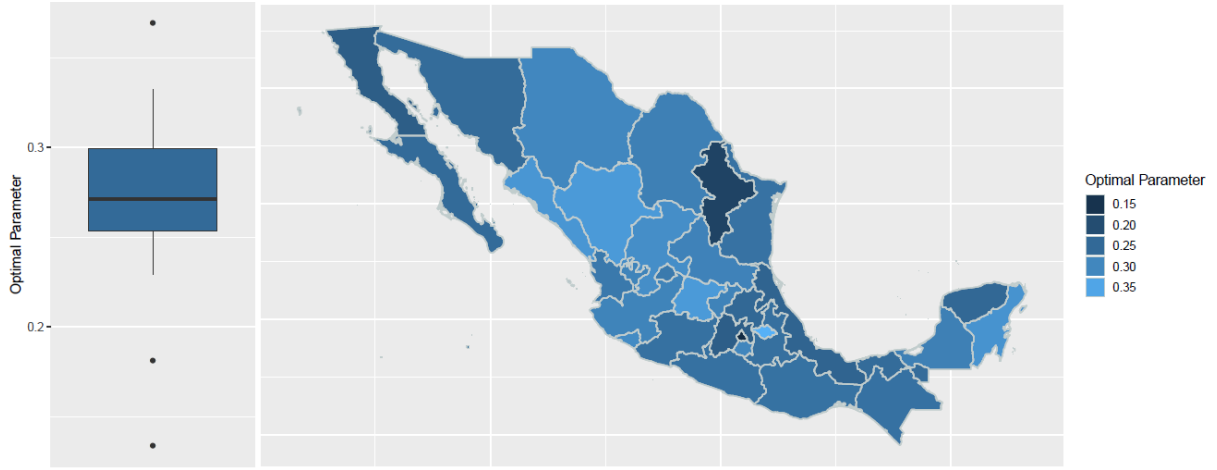
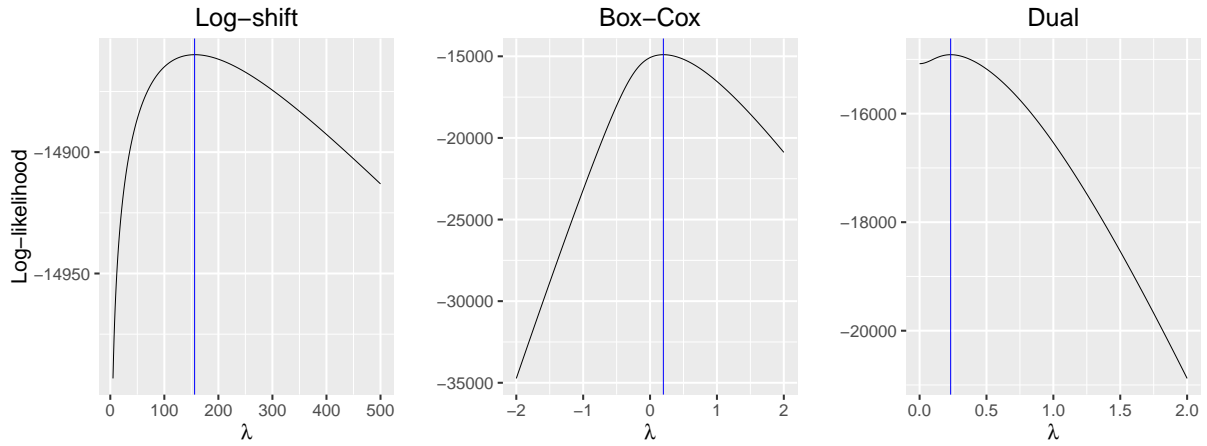Figure 2: Estimated transformation parameters of the Box-Cox transformation in the different states of Mexico

Figure 3: Optimal transformation parameter λs for the working model under the log-shift, Box-Cox and dual power transformations in the state of Guerrero

the ICC is larger than 0 in all cases, there appears to be unexplained between area variability and hence the use of the mixed model may be appropriate. Using the untransformed $ictpc$ outcome leads to a marginal (R_m^2) and conditional (R_c^2) coefficients of determination of 0.35 and 0.37, respectively. The use of a logarithmic transformation improves the predictive power of the model in terms of the conditional R_c^2 and the marginal R_m^2. However, it can clearly be noted that the use of data-driven transformations increases the predictive power of the model.

A detailed analysis of the Gaussian assumptions of the working models corresponding to each transformation is now carried out. The results summarizing the skewness, kurtosis and S-W normality tests are presented in Table 4 and the Q-Q plots are presented in Figure 4. It should be noted, that at municipal-level, all three data-driven transformations perform similarly and yield good approximations to the normal distribution. In contrast, the household-level residuals show clear departures from normality, especially under the model with a fixed logarithmic transformation and without a transformation. The picture considerably improves for the data-driven transformations. The log-shift, Box-Cox and dual power transformations lead to very similar results in terms of skewness and kurtosis. We note that the log-shift transformation performs slightly better in terms of kurtosis and skewness compared to the Box-Cox and dual power transformation. These findings are supported by the Q-Q plots displayed in Figure 4.

Table 3: R_m^2, R_c^2, λ_s, and ICC for the working model under the different transformations

	R_m^2	R_c^2	λ	ICC
No	0.351	0.368	-	0.027
Log	0.361	0.458	-	0.151
Log-Shift	0.460	0.522	156.443	0.114
Box-Cox	0.454	0.513	0.199	0.108
Dual	0.454	0.512	0.232	0.106

The data-driven transformations lead to similar Q-Q plots with more symmetrical and less extreme tails compared to the fixed log transformation. Overall, it appears that the proposed data-driven transformations improve the predictive power of the model and clearly give better approximations to the underlying model assumptions of the linear mixed regression model compared to the use of a fixed logarithmic transformation.

Table 4: Skewness, kurtosis and values of the S-W p-values for the municipal- and household-level error terms of the working models for EBP under the different transformations

	Household-level residuals			Municipal-level residuals		
Transformation	Skewness	Kurtosis	p-value	Skewness	Kurtosis	p-value
No	6.338	75.483	0.000	0.448	3.250	0.197
Log	-2.046	16.986	0.000	-1.491	7.059	0.001
Log-Shift	-0.024	4.143	0.000	-0.276	3.485	0.893
Box-Cox	-0.055	6.085	0.000	-0.389	3.861	0.662
Dual	-0.045	6.542	0.000	-0.387	3.889	0.657

6.2 Deprivation and inequality indicators for municipalities in Guerrero

Based on the analysis in Section 6.1, estimates for the deprivation and inequality indicators presented in Section 2 are calculated by using the EBP method under the three data-driven transformations and the fixed logarithmic transformation. MSE estimation is implemented with the wild bootstrap we introduced in Section 5 with $B = 500$ bootstrap replications.

Table 5 shows summaries over municipalities of point estimates and root MSEs (RMSEs) under the different transformations. In addition we provide a detailed comparison between EBP methods under the different transformations and the direct estimator with corresponding coefficients of variation (CVs) as part of the supplementary material. We observe that the estimates based on the EBP with data-driven transformations are more efficient on average (in terms of RMSE) than the corresponding estimates based on a fixed logarithmic transformation. The effect is especially pronounced for indicators that rely on the tail of the distribution like the QSR. Furthermore, the use of data-driven transformations also has an effect on the point estimates of the indicators. For the HCR and PGAP, the results obtained under the three data-driven transformations are similar to each other, main differences are noticeable when just the fixed logarithmic transformation and no transformation are applied. For instance, the EBP estimates under the model with a logarithmic transformation are on average 5% higher compared to the EBP estimates with data-driven transformations for HCR (see Table 1 in the supplementary material). In addition, the distribution (over municipalities) of the point estimates obtained under EBP with data-driven transformations appear to be closer to the distribution of the direct estimates than the distribution
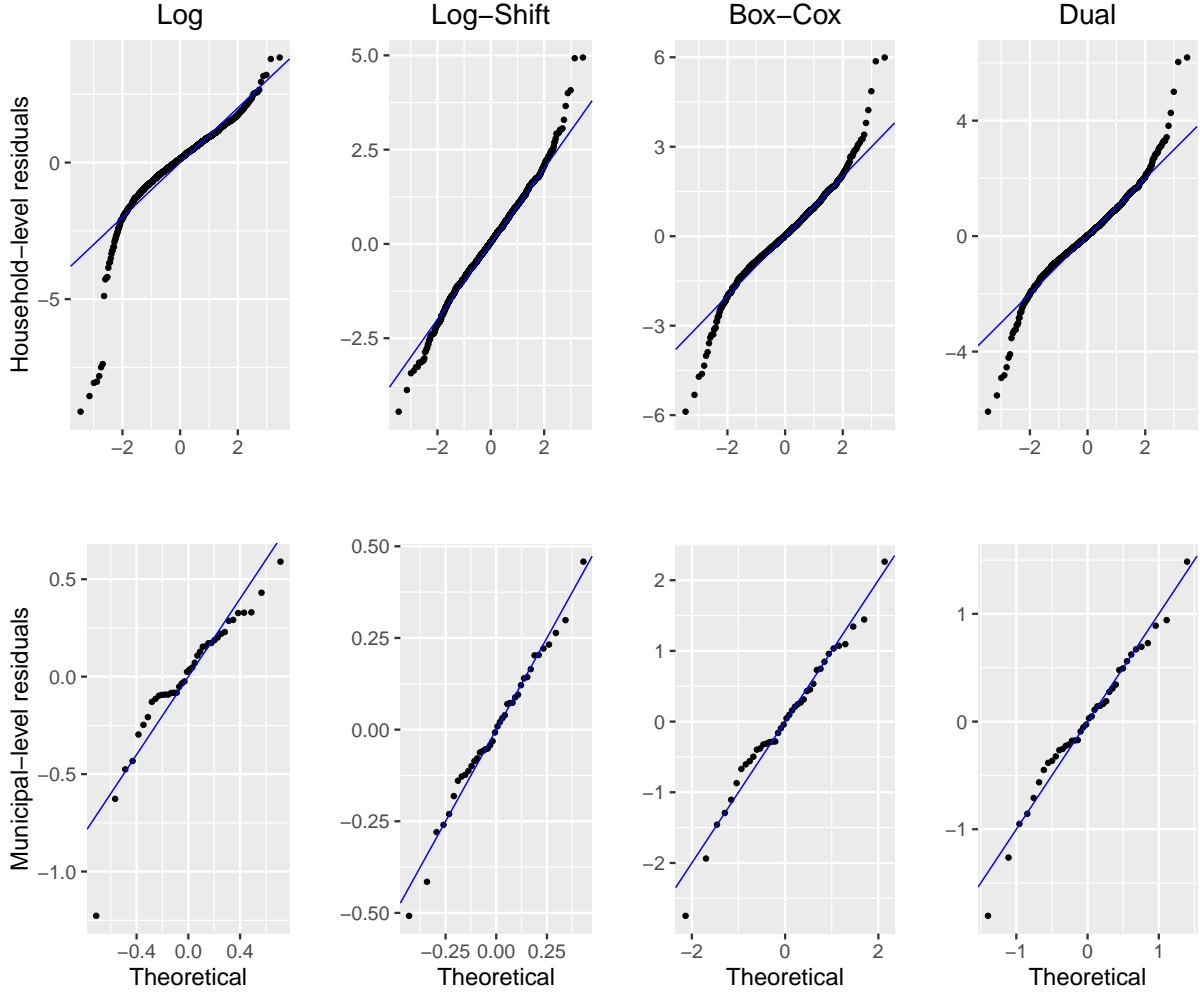
Figure 4: Q-Q-plots for the Pearson household-level (upper panels) and municipal-level (lower panels) residuals of the working model for EBP under the different transformations

of the EBP under a fixed logarithmic transformation.

Having assessed the estimates from a statistical perspective, we investigate the results in the context of the spatial distribution of poverty and inequality in the state of Guerrero. Figure 5 presents the point estimates of HCR, PGAP and QSR at municipal-level. As the point estimates based on the three data-driven transformations are almost identical, we only show the results for the EBP with the log-shift transformation. We observe clear regional differences between the municipalities. Having a closer look to the coastal area in the south-west of Guerrero, where the largest city Acapulco is located, we observe lower levels of poverty (HCR and PGAP) and inequality (QSR) compared to other parts of the state. The coastline to the Pacific Ocean is wealthier due to several tourist destinations like Acapulco, Ixtapa and Zihuatanejo. In contrast, there is also a clear deprivation hotspot in the eastern part of the state Guerrero (e.g. municipalities: Cochoapa el Grande, Metlatnoc and Atlamajalcingo del Monte) with high poverty and inequality rates. These municipalities are home to indigenous populations living in isolated mountain areas.

7 Model-based simulation study

In this section, we present results from a model-based simulation study that aims to evaluate the perfor-mance of the proposed methods. In Section 7.1 we analyze the behaviour of the data-driven transforma-

Table 5: Summaries of point estimates and corresponding RMSEs over municipalities in Guerrero

| Point Estimation	HCR		PGAP		QSR	
Transformation	Mean	Median	Mean	Median	Mean	Median
Log	0.48	0.49	0.24	0.23	18.03	17.82
Log-Shift	0.44	0.44	0.21	0.20	15.56	14.39
Box-Cox	0.44	0.44	0.22	0.22	16.78	16.39
Dual	0.44	0.44	0.22	0.22	17.38	16.98
RMSE	HCR		PGAP		QSR	
	Mean	Median	Mean	Median	Mean	Median
Log	0.11	0.13	0.07	0.08	32.19	27.54
Log-Shift	0.09	0.09	0.06	0.05	4.68	1.91
Box-Cox	0.09	0.10	0.06	0.06	2.98	2.95
Dual	0.09	0.10	0.06	0.06	2.85	2.76
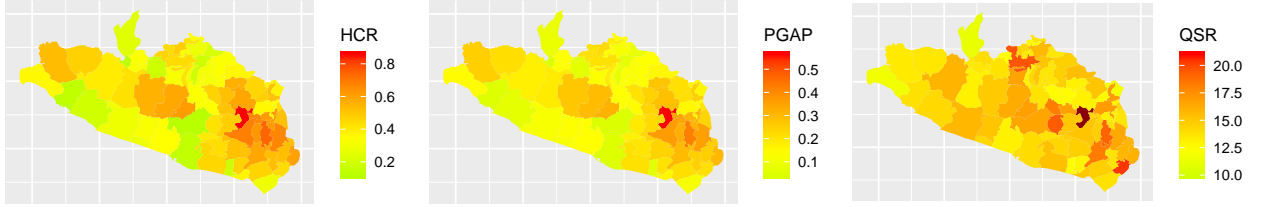
Figure 5: Maps of the HCR, PGAP and QSR in Guerrero for the EBP method under the log-shift transformation at municipal-level

tion parameter under four scenarios for the distributions of the area and unit-level error terms. In Section 7.2 we investigate the ability of the proposed methods to provide more precise small area estimates than the EBP with a fixed logarithmic transformation or without a transformation and assess the performance of the proposed MSE estimators. Finally, in Section 7.3 we evaluate the methods for estimating the transformation parameter. In addition we have also conducted a design-based simulation study with a variable available in the census data that is highly correlated with the target variable (*ictpc*) in the application. The results are provided as part of the supplementary material.

We generate finite populations U of size $N = 10000$, partitioned into $D = 50$ areas $U_1, U_2, \ldots, U_D$ of sizes $N_i = 200$. The samples are selected by a stratified random sampling with strata defined by the 50 small areas. This leads to a sample size of $n = \sum_{i=1}^{D} n_i = 921$ whereby the area-specific sample sizes n_i vary between 8 and 29. We chose the sample sizes mainly because of two reasons. First, we want to assess the data-driven transformations under extreme but realistic cases. Second, the sample sizes are similar in the case study.

Four scenarios, denoted by *Normal*, *Log-scale*, *Pareto* and *GB2*, are considered. Details about the data generating mechanisms of the different scenarios are provided in Table 6. Under scenario *Normal*, data are generated by using Normal distributions for the random effects and unit-level errors. Under the second scenario random effects and unit-level errors are generated under a log-normal distribution such that a fixed logarithmic transformation is suitable. Scenarios *Pareto* and *GB2* are settings that attempt to replicate realistic situations for income data. In particular, these distributions mimic the features of income-based variables, namely unimodal, leptokurtic, and highly skewed data influenced by outliers. Random effects are generated by using a Normal distribution and unit-level error terms are generated under a *Pareto* and *GB2* scenario respectively. Each setting was repeated independently $M = 500$

16

times. We focus on the three data-driven transformations, namely log-shift, Box-Cox, and dual power transformations, and compare these to the case of a fixed logarithmic transformation and the case of using untransformed data.

Table 6: Model-based simulation settings for the analysis of the MSE

Scenario	Model	x_{ij}	z_{ij}	μ_i	u_i	e_{ij}
Normal	$4500 - 400x_{ij} + u_i + e_{ij}$	$N(\mu_i, 3)$	-	$U[-3, 3]$	$N\left(0, 500^2\right)$	$N(0, 1000^2)$
Log-scale	$\exp(10 - x_{ij} - 0.5z_{ij} + u_i + e_{ij})$	$N(\mu_i, 2)$	$N(0, 1)$	$U[2, 3]$	$N\left(0, 0.4^2\right)$	$N\left(0, 0.8^2\right)$
Pareto	$12000 - 400x_{ij} + u_i + e_{ij} - \bar{e}$	$N(\mu_i, 7.5)$	-	$U[-3, 3]$	$N\left(0, 500^2\right)$	$\sqrt{2}\text{Pareto}(3, 2000^2)$
GB2	$8000 - 400x_{ij} + u_i + e_{ij} - \bar{e}$	$N(\mu_i, 5)$	-	$U[-1, 1]$	$N\left(0, 500^2\right)$	$GB2(2.5, 1700, 18, 1.46)$

7.1 Behaviour of the data-driven transformation parameters

Figure 6 shows box plots of the estimated transformation parameters λ for the log-shift, Box-Cox and dual power transformations (over $M = 500$ replications) under the four simulation settings. The data-driven transformation parameters are estimated by REML. Under the *Normal* setting the parameters of the Box-Cox and dual power transformations are close to one indicating that no transformation is needed. In the *Log-scale* scenario, the data was generated in such a way that normality may be achieved by applying the logarithmic transformation. In this case the log-shift transformation parameter is close to zero and the same holds for the parameters of Box-Cox and dual power transformations. For the other two scenarios (*Pareto* and *GB2*), the data-driven parameters are between 0.25 and 0.5, so neither using a logarithmic transformation nor ignoring the need for a transformation is appropriate. Overall, the results indicate that the data-driven transformations behave as expected in the four scenarios and adapt to the shapes of the data distributions.
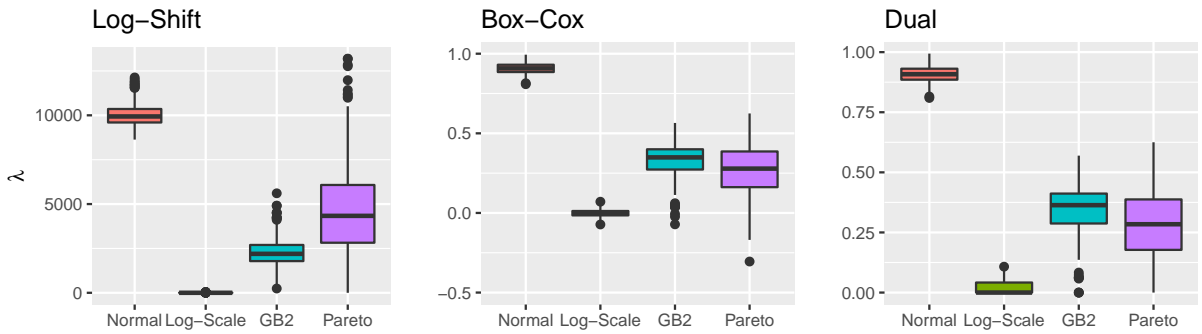

Figure 6: Estimated transformation parameters for the log-shift, Box-Cox and dual power transformations under the different settings.

7.2 Performance of the EBP under data-driven transformations

In this section we compare the performance of the proposed methods to the case of (a) fixed logarithmic transformation and (b) no transformation. We then assess the performance of the MSE estimators. Five estimators of small area deprivation and inequality indicators (HCR, PGAP and QSR) are evaluated. The EBP and the corresponding MSE estimators are implemented using $L = 100$ and $B = 500$. The following quality measures averaged over Monte-Carlo replications M are used to assess the performance of a

small area estimator in area i:

$$\text{RMSE}\left(\hat{I}_i^{\text{method}}\right) = \left[\frac{1}{M}\sum_{m=1}^{M}\left(\hat{I}_i^{\text{method}(m)} - I_i^{(m)}\right)^2\right]^{1/2},$$

$$\text{Bias}\left(\hat{I}_i^{\text{method}}\right) = \frac{1}{M}\sum_{m=1}^{M}\left(\hat{I}_i^{\text{method}(m)} - I_i^{(m)}\right),$$

where $\hat{I}_i^{\text{method}}$ denotes the estimated indicator in area i based on any of the five methods under consideration and I_i denotes the corresponding true value in area i. Table 7 presents the results split by the four scenarios. It shows median and mean values of RMSE and bias averaged over small areas. Under the *Normal* scenario the EBP without transformation is the gold, but the EBP with data-driven transformations (log-shift, Box-Cox and dual power) perform similarly in terms of RMSE and bias. The same picture emerges in the *Log-scale* scenario where the EBP with a logarithmic transformation is the gold standard, but again the EBP with data-driven transformations perform well both in terms of RMSE and bias. These results confirm our expectations that the EBP with data-driven transformations adapt to the shape of the data distribution. Under the *GB2* and *Pareto* scenarios we notice that the EBP with a fixed transformation or without transformation is inferior to the EBP with data-driven transformations both in terms of RMSE and Bias. The differences are especially pronounced for QSR which is very sensitive to the tails of the distribution. Furthermore, the estimates based on data-driven transformations are almost unbiased or have a small bias. A closer look at the data-driven transformations indicates that EBP with a log-shift transformation performs slightly better than the EBP with Box-Cox and dual power transformations under the *GB2* and *Pareto* scenarios. Overall, it appears that the proposed EBP method with data-driven transformations adapts to the underlying distribution of the data, and hence improves the precision of small area estimates.

We now turn our attention to the performance of the MSE estimators. We denote by *parametric* and *wild* the proposed parametric bootstrap and proposed semi-parametric wild bootstrap respectively. The aim of this part is twofold. Firstly, we assess the performance of the two proposed MSE estimators we introduced in Section 5. Secondly, we investigate the ability of the wild bootstrap to protect against departures from the assumptions of the unit-level error term. Starting with the first aim, Table 8 reports the results for the two MSE estimators and presents the mean and median values of relative RMSE and relative bias -over Monte-Carlo replications and areas- of the EBP with Box-Cox transformation. For calculating the RMSE and relative bias we treat the empirical MSE (over Monte-Carlo replications) as the true MSE. The results for the EBP with a log-shift transformation and dual power transformation are very similar and although they are omitted they are available on request from the authors.

We note that, on average, the proposed *parametric* and *wild* bootstrap approaches for the EBP with a Box-Cox transformation have small positive relative bias (HCR and PGAP indicators) in the *Normal* and *Log-scale* settings. However, *parametric* bootstrap shows some underestimation in the case of QSR. In this latter case *wild* bootstrap appears to be associated with smaller relative bias. For the *Normal* and *Log-scale* scenarios *parametric* bootstrap also has smaller relative RMSE than *wild* bootstrap. Nevertheless, *wild* bootstrap provides reasonable results for HCR and PGAP and reduces the underestimation for QSR. When the distributional assumptions are not met, as in the (*GB2* and *Pareto*) scenarios, *parametric* tends to clearly underestimate the MSE (except for the HCR in the *Pareto* scenario). Although *wild* bootstrap does not completely eliminate this bias, it greatly reduces it and provides more stable MSE estimates in terms of relative RMSE. These results indicate that departures from the model assumptions -even after

using data transformations- can impact MSE estimation with parametric methods. The problem is more pronounced when estimating parameters that depend on the tails of the distribution as is the case with the QSR. In those cases, the use of semi-parametric bootstrap, at least as a supplementary MSE estimation method, can offer some protection against misspecification.

7.3 Impact of alternative estimation methods for λ

In this last section we explore the use of non-parametric alternatives to the REML approach for estimating data-driven transformation parameters (see Section 4.3). Here, we study five estimation methods. These are the REML approach, the minimization of the skewness (*Skew*) and the pooled skewness (*poolSkew*), and the distance-based criteria Kolomogorov-Smirnov (*KS*) and Cramér-von Mises (*CvM*) we introduced in Section 4.3.

The five methods estimate transformation parameters close to the theoretically correct ones, in the scenarios those are known. For instance, in the Log-scale scenario, the estimated transformation parameters under the different estimation methods are shown in Figure 7 and Table 9. We observe that although the five methods provide similar estimates of λ, the REML method has smaller variability. In our model-based simulations we further studied the impact of the estimation method of the transformation parameter on point and MSE estimation and we conclude that this only marginally influences the quality of small area estimates. These results are available from the authors upon request.

Overall, these results suggest that for the scenarios we considered in this paper the method used to estimate the transformation parameter does not have a noticeable impact on small area estimation and REML appears to be the most stable method.
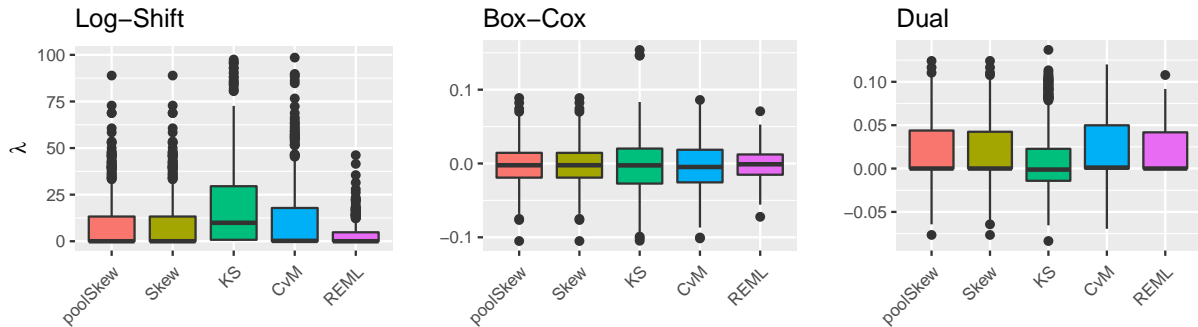

Figure 7: Box-plots of estimated transformation parameters for the log-scale scenario using different estimation methods

Table 7: Summaries of estimated RMSEs and Bias over the model-based settings

	Indicator	HCR		PGAP		QSR	
	Estimator	Median	Mean	Median	Mean	Median	Mean
Normal							
RMSE	No	0.0338	0.0357	0.0136	0.0154	0.3259	1.2765
	Log-Shift	0.0344	0.0363	0.0155	0.0175	0.3898	0.6710
	Box-Cox	0.0343	0.0358	0.0134	0.0156	0.3348	1.1178
	Dual	0.0343	0.0358	0.0134	0.0156	0.3346	0.5797
BIAS	No	0.0000	0.0007	0.0002	0.0009	0.0049	0.0899
	Log-Shift	0.0029	0.0039	-0.0067	-0.0076	-0.1000	-0.2190
	Box-Cox	0.0016	0.0027	-0.0021	-0.0025	-0.0396	-0.0807
	Dual	0.0016	0.0027	-0.0021	-0.0024	-0.0458	-0.1193
Log-Scale							
RMSE	Log	0.0583	0.0605	0.0358	0.0367	4.9100	4.8969
	Log-Shift	0.0583	0.0605	0.0358	0.0367	4.9024	4.8985
	Box-Cox	0.0581	0.0604	0.0358	0.0367	4.9731	4.9717
	Dual	0.0584	0.0605	0.0359	0.0367	4.9025	4.9093
BIAS	Log	-0.0011	-0.0009	-0.0007	-0.0003	0.0394	0.1143
	Log-Shift	-0.0020	-0.0017	-0.0011	-0.0007	-0.0873	-0.0072
	Box-Cox	-0.0009	-0.0006	-0.0008	-0.0004	0.1499	0.2106
	Dual	-0.0024	-0.0021	-0.0009	-0.0005	-0.1610	-0.0992
GB2							
RMSE	No	0.0650	0.0656	0.0552	0.0552	17.7364	32.0686
	Log	0.0912	0.0908	0.0272	0.0270	1.8979	1.9002
	Log-Shift	0.0418	0.0415	0.0127	0.0132	0.4286	0.4411
	Box-Cox	0.0471	0.0469	0.0136	0.0139	0.4708	0.4753
	Dual	0.0472	0.0470	0.0137	0.0140	0.4715	0.4760
BIAS	No	0.0471	0.0477	0.0481	0.0479	1.8355	2.0825
	Log	0.0746	0.0747	0.0169	0.0169	1.4718	1.4692
	Log-Shift	0.0176	0.0179	-0.0008	-0.0013	0.0546	0.0523
	Box-Cox	0.0274	0.0274	0.0035	0.0031	0.1780	0.1721
	Dual	0.0275	0.0274	0.0037	0.0034	0.1800	0.1747
Pareto							
RMSE	No	0.0448	0.0444	0.0622	0.0613	1.6814	3.6057
	Log	0.0304	0.0306	0.0082	0.0084	0.3887	0.3994
	Log-Shift	0.0185	0.0196	0.0060	0.0063	0.1661	0.1779
	Box-Cox	0.0192	0.0202	0.0059	0.0062	0.1786	0.1901
	Dual	0.0192	0.0203	0.0059	0.0062	0.1782	0.1902
BIAS	No	0.0277	0.0287	0.0166	0.0160	0.3173	0.3132
	Log	0.0086	0.0081	-0.0030	-0.0037	0.2068	0.2034
	Log-Shift	0.0003	-0.0001	-0.0034	-0.0041	0.0305	0.0300
	Box-Cox	0.0030	0.0026	-0.0031	-0.0037	0.0525	0.0530
	Dual	0.0030	0.0027	-0.0031	-0.0037	0.0522	0.0530

Table 8: Performance of MSE estimators in model-based simulations: EBP with Box-Cox transformation

Indicator	Estimator	HCR Median	HCR Mean	PGAP Median	PGAP Mean	QSR Median	QSR Mean
Normal							
rel. RMSE[%]	Parametric	8.30	9.22	9.15	9.47	15.25	21.23
	Wild	14.57	14.77	14.21	14.61	17.46	20.93
rel. Bias[%]	Parametric	6.64	7.27	-1.17	-0.12	-7.72	-12.61
	Wild	8.05	8.04	2.17	3.23	-1.01	-1.46
Log-Scale							
rel. RMSE[%]	Parametric	11.14	12.00	19.19	19.57	19.10	19.75
	Wild	16.82	17.00	22.70	22.95	25.34	25.62
rel. Bias[%]	Parametric	6.10	6.29	5.70	6.36	7.91	7.92
	Wild	7.69	7.82	7.34	7.39	6.58	6.78
GB2							
rel. RMSE[%]	Parametric	21.71	21.86	20.89	20.57	43.75	43.58
	Wild	19.01	19.39	14.76	15.12	26.21	27.23
rel. Bias[%]	Parametric	-20.04	-19.74	-16.88	-15.92	-42.90	-42.74
	Wild	-14.59	-14.64	-5.45	-5.75	-21.72	-22.53
Pareto							
rel. RMSE[%]	Parametric	11.31	12.60	35.60	34.78	50.04	51.63
	Wild	26.18	28.44	23.58	26.04	28.60	33.40
rel. Bias[%]	Parametric	2.43	3.38	-33.82	-31.16	-49.51	-51.06
	Wild	19.21	21.37	-8.28	-3.28	-23.02	-26.79

Table 9: Mean and median of estimated transformation parameters under the log-scale scenario using different estimation methods

	Log-Shift Mean	Log-Shift Median	Box-Cox Mean	Box-Cox Median	Dual Mean	Dual Median
poolSkew	9.381	0.000	-0.002	-0.002	0.016	0.000
Skew	9.381	0.000	-0.002	-0.002	0.015	0.000
KS	23.906	10.816	-0.003	-0.003	0.009	-0.001
CvM	11.954	0.211	-0.004	-0.005	0.025	0.001
REML	3.349	0.000	-0.002	-0.001	0.021	0.000

8 Conclusions and future research directions

In this paper we investigate data-driven transformations for small area estimation. In particular, we propose an EBP approach with data-driven transformations estimated with likelihood-based methods. The use of scaled transformations (conditional on the Jacobian) allows for the use of standard software for fitting the linear mixed regression model. Three types of transformations are discussed log-shift, Box-Cox and dual power transformations. We further explore the use of parametric and semi-parametric wild bootstrap for MSE estimation that also captures the uncertainty from estimating the data driven transformation parameter. Semi-parametric bootstrap is used for protecting against departures from the model assumptions. Model-based simulations demonstrate the ability of the proposed EBP method to adapt to the shape of the data distribution and hence provide more efficient estimates than a fixed logarithmic transformation or the case where no transformation is used. Although the paper focuses on the EBP the proposed methods are applicable to other small area estimators for example, the ELL approach (Elbers et al., 2003). The methods proposed in this paper can be implemented by using the R Package **emdi** (Kreutzmann et al., 2019). The package supports the user by estimating and mapping regionally disaggregated indicators. Although this package already includes the logarithm and Box-Cox transformations, some research effort should be shifted towards the development of relevant software which includes in more detail the use of data-driven transformations in the SAE context.

Further research can investigate the use of multiparameter transformation families. This may allow for better control of higher moments and hence better adaptation to the distribution of the data. Since likelihood-based approaches might be influenced by outliers, it would be also interesting to investigate robust estimation methods. Model selection with data driven transformations presents additional challenges. Finding a good working model depends on the method of transformation. In this paper we first find a working model and keep this fixed when considering different data-driven transformations. However, this may not offer the best approach to model selection. Approaches that simultaneously consider both steps for linear regression models have been proposed (Laud and Ibrahim, 1995; Hoeting and Ibrahim, 1998; Hoeting et al., 2002). Extending these approaches to the case of linear mixed models is an open research problem. Finally, comparing the EBP with data-driven transformations to EBP approaches with alternative parametric assumptions (Diallo and Rao, 2014; Graf et al., 2018) is empirical work that remains open.

A Derivation of scaled transformations

In this appendix we derive the Jacobian and the corresponding scaling factors presented in Table 2 for the log-shift, Box-Cox, and dual power transformations.

A.1 Log-shift transformation

Let $J(\lambda, \boldsymbol{y})$ be the Jacobian of the log-shift transformation from $\boldsymbol{y}_i$ to $\boldsymbol{y}_i^*(\lambda)$, defined as:

$$J(\lambda, \boldsymbol{y}) = \prod_{i=1}^{D} \prod_{j=1}^{n_i} \left| \frac{dy_{ij}^*(\lambda)}{dy_{ij}} \right|$$

$$= \prod_{i=1}^{D} \prod_{j=1}^{n_i} (y_{ij} + \lambda)^{-1}.$$

The log-likelihood function in (3) can be rewritten as follows:

$$
\begin{aligned}
l_{\text{REML}}(\boldsymbol{y}, \lambda | \boldsymbol{\theta}) &= -\frac{n-p}{2} \log(2\pi) + \frac{1}{2} \log \left| \sum_{i=1}^{D} \boldsymbol{X}_i^T \boldsymbol{X}_i \right| - \frac{1}{2} \sum_{i=1}^{D} \log |\boldsymbol{V}_i| \\
&\quad - \frac{1}{2} \log \left| \sum_{i=1}^{D} \boldsymbol{X}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{X}_i \right| \\
&\quad - \frac{1}{2} \sum_{i=1}^{D} [\boldsymbol{y}_i^*(\lambda) - \boldsymbol{X}_i \hat{\boldsymbol{\beta}}]^T \boldsymbol{V}_i^{-1} [\boldsymbol{y}_i^*(\lambda) - \boldsymbol{X}_i \hat{\boldsymbol{\beta}}] - n \log \underbrace{\left(\prod_{i=1}^{D} \prod_{j=1}^{n_i} (y_{ij} + \lambda) \right)^{\frac{1}{n}}}_{=\bar{y}_\lambda}.
\end{aligned}
$$

In order to obtain the scaled log-shift transformation, $z_{ij}^*(\lambda)$, the denominator of the term $\frac{y_{ij}^*(\lambda)}{J(\lambda, \boldsymbol{y})^{1/n}}$ is given by:

$$1/J(\lambda, \boldsymbol{y})^{\frac{1}{n}} = J(\lambda, \boldsymbol{y})^{-\frac{1}{n}} = \left[\prod_{i=1}^{D} \prod_{j=1}^{n_i} (y_{ij} + \lambda)^{-1} \right]^{-\frac{1}{n}}$$

$$= \bar{y}_\lambda.$$

Therefore, the scaled log-shift transformation is defined as follows:

$$z_{ij}^*(\lambda) = \frac{y_{ij}^*(\lambda)}{J(\lambda, \boldsymbol{y})^{1/n}} = \bar{y}_\lambda \, \log(y_{ij} + \lambda)$$

for $y_{ij} > -\lambda$.

A.2 Box-Cox transformation

Let $J(\lambda, \boldsymbol{y})$ be the Jacobian of the Box-Cox transformation from $\boldsymbol{y}_i$ to $\boldsymbol{y}_i^*(\lambda)$, defined as:

$$J(\lambda, \boldsymbol{y}) = \prod_{i=1}^{D} \prod_{j=1}^{n_i} \left| \frac{dy_{ij}^*(\lambda)}{dy_{ij}} \right|$$

$$= \prod_{i=1}^{D} \prod_{j=1}^{n_i} (y_{ij} + s)^{\lambda - 1}.$$

The log-likelihood function in (3) can be rewritten as follows:

$$
\begin{aligned}
l_{\text{REML}}(\boldsymbol{y}, \lambda | \boldsymbol{\theta}) &= -\frac{n-p}{2} \log(2\pi) + \frac{1}{2} \log \left| \sum_{i=1}^{D} \boldsymbol{X}_i^T \boldsymbol{X}_i \right| - \frac{1}{2} \sum_{i=1}^{D} \log |\boldsymbol{V}_i| \\
&\quad - \frac{1}{2} \log \left| \sum_{i=1}^{D} \boldsymbol{X}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{X}_i \right| \\
&\quad - \frac{1}{2} \sum_{i=1}^{D} [\boldsymbol{y}_i^*(\lambda) - \boldsymbol{X}_i \hat{\boldsymbol{\beta}}]^T \boldsymbol{V}_i^{-1} [\boldsymbol{y}_i^*(\lambda) - \boldsymbol{X}_i \hat{\boldsymbol{\beta}}] + n(\lambda-1) \log \underbrace{\left(\prod_{i=1}^{D} \prod_{j=1}^{n_i} (y_{ij} + s) \right)^{\frac{1}{n}}}_{=\bar{y}}.
\end{aligned}
$$

In order to obtain the scaled transformation of the Box-Cox family, $z_{ij}^*(\lambda)$, the denominator of the term $\frac{y_{ij}^*(\lambda)}{J(\lambda, \boldsymbol{y})^{1/n}}$ is given by:

$$
1/J(\lambda, \boldsymbol{y})^{\frac{1}{n}} = J(\lambda, \boldsymbol{y})^{-\frac{1}{n}} = \left[\prod_{i=1}^{D} \prod_{j=1}^{n_i} (y_{ij} + s)^{\lambda-1} \right]^{-\frac{1}{n}}
$$

$$
= \bar{y}^{-(\lambda-1)}.
$$

Therefore, the scaled Box-Cox transformation is defined as follows:

$$
z_{ij}^*(\lambda) = \frac{y_{ij}^*(\lambda)}{J(\lambda, \boldsymbol{y})^{1/n}} = \begin{cases} \frac{(y_{ij}+s)^{\lambda}-1}{\bar{y}^{\lambda-1}\lambda}, & \lambda \neq 0, \\ \bar{y} \log(y_{ij} + s), & \lambda = 0, \end{cases}
$$

for $y_{ij} > -s$.

A.3 Dual power transformation

Let $J(\lambda, \boldsymbol{y})$ be the Jacobian of the dual power transformation from $\boldsymbol{y}_i$ to $\boldsymbol{y}_i^*(\lambda)$, defined as:

$$
\begin{aligned}
J(\lambda, \boldsymbol{y}) &= \prod_{i=1}^{D} \prod_{j=1}^{n_i} \left| \frac{dy_{ij}^*(\lambda)}{dy_{ij}} \right| \\
&= \prod_{i=1}^{D} \prod_{j=1}^{n_i} \frac{(y_{ij} + s)^{\lambda-1} + (y_{ij} + s)^{-\lambda-1}}{2}.
\end{aligned}
$$

The log-likelihood function in (3) can be rewritten as follows:

$$
\begin{aligned}
l_{\text{REML}}(\boldsymbol{y}, \lambda | \boldsymbol{\theta}) &= -\frac{n-p}{2}\log(2\pi) + \frac{1}{2}\log\left|\sum_{i=1}^{D}\boldsymbol{X}_i^T\boldsymbol{X}_i\right| - \frac{1}{2}\sum_{i=1}^{D}\log|\boldsymbol{V}_i| \\
&\quad - \frac{1}{2}\log\left|\sum_{i=1}^{D}\boldsymbol{X}_i^T\boldsymbol{V}_i^{-1}\boldsymbol{X}_i\right| \\
&\quad - \frac{1}{2}\sum_{i=1}^{D}[\boldsymbol{y}_i^*(\lambda) - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}]^T\boldsymbol{V}_i^{-1}[\boldsymbol{y}_i^*(\lambda) - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}] \\
&\quad + n\log\underbrace{\left(\prod_{i=1}^{D}\prod_{j=1}^{n_i}\frac{(y_{ij}+s)^{\lambda-1} + (y_{ij}+s)^{-\lambda-1}}{2}\right)^{\frac{1}{n}}}_{=\bar{y}_\lambda}.
\end{aligned}
$$

In order to obtain the scaled dual transformation, $z_{ij}^*(\lambda)$, the denominator of the term $\frac{y_{ij}^*(\lambda)}{J(\lambda,\boldsymbol{y})^{1/n}}$ is given by:

$$
\begin{aligned}
1/J(\lambda, \boldsymbol{y})^{1/n} = J(\lambda, \boldsymbol{y})^{-\frac{1}{n}} &= \left[\prod_{i=1}^{D}\prod_{j=1}^{n_i}\frac{(y_{ij}+s)^{\lambda-1} + (y_{ij}+s)^{-\lambda-1}}{2}\right]^{-\frac{1}{n}} \\
&= \bar{y}_\lambda^{-1}.
\end{aligned}
$$

Therefore, the scaled dual transformation is defined as follows:

$$
z_{ij}^*(\lambda) = \frac{y_{ij}^*(\lambda)}{J(\lambda, \boldsymbol{y})^{1/n}} = \begin{cases} \bar{y}_\lambda^{-1}\frac{(y_{ij}+s)^\lambda - (y_{ij}+s)^{-\lambda}}{2\lambda} & \text{if } \lambda > 0; \\ \bar{y}_\lambda^{-1}\log(y_{ij}+s) & \text{if } \lambda = 0, \end{cases}
$$

for $y_{ij} > -s$.

References

Battese, G. E., Harter, R. M. and Fuller, W. A. (1988) An error component model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, **83**, 28–36.

Bedoya, H., Freije, S., Vila, L., Echeverria, G., Biller, D., Grandolini, G. M., Albisetti, R., Quintrell, E. and Vish, R. (2013) Country partnership strategy for the united mexican states (2014-2019). *Tech. rep.*, World Bank Group, Washington DC.

Bickel, P. J. and Doksum, K. A. (1981) An analysis of transformations revisited. *Journal of the American Statistical Association*, **76**, 296 – 311.

Box, G. E. P. and Cox, D. R. (1964) An analysis of transformations. *Journal of the Royal Statistical Society: Series B*, **26**, 211–252.

Carroll, R. J. and Ruppert, D. (1987) Diagnostics and robust estimation when transforming the regression model and the response. *Technometrics*, **29**, 287–299.

Chakravarti, I. M. and Laha, R. G. (1967) Handbook of methods of applied statistics. In *Handbook of methods of applied statistics*. John Wiley & Sons.

Cramér, H. (1928) On the composition of elementary errors: First paper: Mathematical deductions. *Scandinavian Actuarial Journal*, **1928**, 13–74.

Diallo, M. S. and Rao, J. N. K. (2014) Small area estimation of complex parameters under unit-level models with skew-normal errors. *JSM 2014*, Survey Research Methods Section.

Elbers, C., Lanjouw, J. and Lanjouw, P. (2003) Micro-level estimation of poverty and inequality. *Econometrica*, **71**, 355–364.

Elbers, C. and van der Weide, R. (2014) Estimation of normal mixtures in a nested error model with an application to small area estimation of poverty and inequality. Policy Research working paper.

Eurostat (2004) Common cross-sectional eu indicators based on eu-silc; the gender pay gap. *Unit D-2: Living conditions and social protection, Directorate D: Single Market, Employment and Social statistics, Eurostat, Luxembourg*.

Fabrizi, E. and Trivisano, C. (2016) Small area estimation of the gini concentration coefficient. *Computational Statistics & Data Analysis*, **99**, 223–234.

Feng, X., He, X. and Hu, J. (2011) Wild bootstrap for quantile regression. *Biometrika*, **98**, 995–999.

Foster, J., Greer, J. and Thorbecke, E. (1984) A class of decomposable poverty measures. *Econometrica*, **52**, 761–766.

González-Manteiga, W., Lombardía, M., Molina, I., Morales, D. and Santamaría, L. (2008) Bootstrap mean squared error of a small-area eblup. *Journal of Statistical Computation and Simulation*, **78**, 443–462.

Graf, M., Marín, J. M. and Molina, I. (2018) A generalized mixed model for skewed distributions applied to small area estimation. *TEST*, forthcoming.

Gurka, M. J., Edwards, L. J., Muller, K. E. and Kupper, L. L. (2006) Extending the Box-Cox transformation to the linear mixed model. *Journal of the Royal Statistical Society: Series A*, **169**, 273–288.

Hoeting, J. A. and Ibrahim, J. G. (1998) Bayesian predictive simultaneous variable and transformation selection in the linear model. *Computational Statistics & Data Analysis*, **28**, 87–103.

Hoeting, J. A., Raftery, A. E. and Madigan, D. (2002) Bayesian variable and transformation selection in linear regression. *Journal of Computational and Graphical Statistics*, **3**, 485–507.

John, J. A. and Draper, N. R. (1980) An alternative family of transformations. *Journal of the Royal Statistical Society: Series C*, **29**, 190–197.

Kreutzmann, A.-K., Pannier, S., Rojas-Perilla, N., Schmid, T., Templ, M. and Tzavidis, N. (2019) **emdi**: estimating and mapping disaggregated indicators. *Journal of Statistical Software*, forthcoming.

Laud, P. W. and Ibrahim, J. G. (1995) Predictive model selection. *Journal of the Royal Statistical Society: Series B*, **57**, 247–262.

Marino, M. F., Tzavidis, N. and Alfo, M. (2016) Mixed hidden Markov quantile regression models for longitudinal data with possibly incomplete sequences. *Statistical Methods in Medical Research*, forthcoming.

Molina, I. and Martín, N. (2018) Empirical best prediction under a nested error model with log transformation. *The Annals of Statistics*, **46**, 1961–1993.

Molina, I. and Rao, J. N. K. (2010) Small area estimation of poverty indicators. *The Canadian Journal of Statistics*, **38**, 369–385.

Nakagawa, S. and Schielzeth, H. (2013) A general and simple method for obtaining r2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, **4**, 133–142.

R Core Team (2017) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: `\url{http://www.R-project.org/}`.

Royston, P. and Lambert, P. C. (2011) *Flexible parametric survival analysis using Stata: Beyond the Cox model*. StataCorp LP.

Schmid, T., Bruckschen, F., Salvati, N. and Zbiranski, T. (2017) Constructing sociodemographic indicators for national statistical institutes by using mobile phone data: estimating literacy rates in Senegal. *Journal of the Royal Statistical Society: Series A*, **180**, 1163–1190.

Tortajada, C. (2006) Who has access to water case study of mexico city metropolitan area human development report 2006. *Tech. rep.*, United Nations Development Programme.

Tzavidis, N., Hernandez, A. L., Zhang, L. C., Schmid, T. and Rojas-Perilla, N. (2018) From start to finish: A framework for the production of small area official statistics. *Journal of the Royal Statistical Society: Series A*, **181**, 927–979.

Verbeke, G. and Molenberghs, G. (2000) *Linear mixed models for longitudinal data*, vol. 1. Springer Series in Statistics.

Weidenhammer, B., Tzavidis, N., Schmid, T. and Salvati, N. (2014) Domain prediction for counts using microsimulation via quantiles. In *Small Area Estimation 2014 Conference*. Poznan, Poland.

Yang, L. (1995) *Transformation-density estimation*. Ph.D. thesis, University of North Carolina, Chapel Hill.

Yang, Z. (2006) A modified family of power transformations. *Economics Letters*, **92**, 14–19.

Yeo, I.-K. and Johnson, R. A. (2000) A new family of power transformations to improve normality or symmetry. *Biometrika*, **87**, 954–959.