

PAPER • OPEN ACCESS

A simulation study with log, Box-Cox, and dual-power transformation on handling curvilinear relationship in small area estimation

To cite this article: D K Bodro *et al* 2019 *IOP Conf. Ser.: Earth Environ. Sci.* **299** 012029

View the [article online](#) for updates and enhancements.

A simulation study with log, Box-Cox, and dual-power transformation on handling curvilinear relationship in small area estimation

D K Bodro^a, B Sartono^a and K Sadik^{a*}

^aDepartment of Statistics, Institut Pertanian Bogor, Bogor, Indonesia

*E-mail: kusmansadik@gmail.com

Abstract. A standard small area estimation method may fail to produce reasonable estimates when the normality assumption is not met or the relationship between the interest parameter and the auxiliary variables is not linear. A logarithm transformation has been widely used to help this issue and works well for some cases. However, it may not be generally valid so that some transformation such as Box-Cox (BC) and dual power (DP). This paper discusses a simulation study on how BC and DP could overcome circumstances where those aforementioned problems are there in the data. Several different forms of the relationship were studied and the study revealed that BC and DP transformation are the recommended methods because they produced smaller Mean Absolute Percentage Error (MAPE) values than ones without transformation or using logarithm transformation. Even if DP does not consistently provide the lowest value, but technically, DP can overcome truncated problems that occur in BC. The findings of this work indicate that this new transformation, DP, is proposed to be the best transformation to overcome the abnormality of an interesting variable because this transformation produces a monotonic function that has a domain of positive real numbers ($\mathbb{R}_+$) and range of whole real numbers ($\mathbb{R}$).

1. Introduction

The linear mixed models (LMM) with both random and fixed effects have been studied from both theoretical and applied aspects in the literature. As special normal LMM, the Fay-Herriot model has been used in small area estimation (SAE). Small area estimation is a method used to estimate parameters in a small area by utilizing information from outside the area, from within the area itself and from outside the survey and an area is said to be small-area if the sample taken in the area is not sufficient to estimate directly with accurate results [1].

The Fay-Herriot model that is applied in estimating small areas generally needs to be transformed so that the estimation obtained produces a good estimator. Transformation is needed generally to meet data distribution assumptions. For interested variables that are positive data types such as community income data, the standard transformation is a log, but this transformation is not always appropriate to use because it depends on the distribution of the interested variable. For interested variables that are positive data types such as income data, the standard transformation is a log, but this transformation is not always appropriate to use. Another conventional transformation commonly used is the Box-Cox (BC) transformation. However, this transformation has a truncated problem, so that the BC transformation does not match the normal assumption. In practice, this truncation effect may be negligible as claimed by many researchers, but technically it causes difficulties in deriving the



asymptotic properties of the estimated model. Sugasawa and Kubokawa [2] argued that another drawback of BC transformation was that the estimator of maximum likelihood (ML) for λ was inconsistent, so the EBLUP method used in SAE did not produce the best predictor.

The dual power transformation (DP) is proposed to overcome the problem of truncated problems in BC transformation [3]. The primary motivation of this new transformation is to remove the limitations on BC transformation and maintain non-negative traits for interesting variables. This transformation is a modification of the BC model. The application of these two types of transformation was carried out by Yang [3] which was applied to the regression model and gave better empirical results for dual power transformation in terms of the goodness of the model. Sugasawa and Kubokawa [2] applied a DP transformation to the SAE model for the interested variable that skewness to the right, this study resulted in an estimation value of four parameters in the SAE model which was able to increase the value of its parameters. This study suggests leaving the log transformation method in handling skewed data and switching to using other transformation methods that can provide better results in estimating small area estimation models. The research study will examine and evaluate the parametric transformation method in the Fay-Herriot model by applying log-normal transformations, Box-Cox transformations and dual power on the interesting variable that has a non-linear pattern and compare and evaluate the application of log transformation, Box-Cox (BC) and dual power (DP).

Materials

We investigated the performance of transformed interested variable on Fay-Herriot models through simulation experiments. The Fay-Herriot model is defined as follows [4]:

$$h(y_i, \lambda) = x_i^T \beta + v_i + e_i \quad (1)$$

Let i be the index of the number of areas ($1, 2, \dots, m$), y_i be a vector of original interested variable, $h(y_i, \lambda)$ be a vector of transformed observations, λ be a power transformation, β be regression coefficients and x_i^T be columns contain the values of the explanatory variable and generated from normal distribution. Let the random effect v_i is generated from $v_i \sim N(0, \tau^2)$, and the sampling error e_i is generated from $N(0, \sigma_i^2)$ with $\sigma_i^2 = 0.1, 0.2, 0.3, 0.4, 0.5$. Stage of generating data simulation;

1. Determine the parameter values of the Fay-Herriot model:

- transformation parameter (initial lambda) : $0 < \lambda < 2$
- regression coefficients : $\beta_0 = 0.5, \beta_1 = 1$
- random effect variance: $\tau^2 = 0.16$
- the error variance : $0 < \sigma_i^2 < 0.5$

each one value σ_i^2 represents $m/5$ sequential area until all areas as much as $m = 50$ are represented.

2. Generates explanatory variable $x_i \sim N(50, 15)$ and random effects $v_i \sim N(0, \tau^2)$,
3. Generates y vectors that have linear and non-linear patterns using the model equation as follows:

$$y = \{\lambda(\beta_0 + \beta_1 x_i + v_i + e_i) + \sqrt{\lambda^2[\beta_0 + \beta_1 x_i + v_i + e_i]^2 + 1}\}^{1/\lambda} \quad (2)$$

lambda used is 0.1; 0.3 and 0.6 to produce non-linear patterned data. The equation of the above model with three different lambda values will be generated with three scenarios, namely y_i^N, y_i^t and y_i^g . These three interesting variable sets are used to evaluate the DP transformation method in different interested variable conditions. All interested variables used must have an error with expectation value equal to 0 ($E[e] = 0$) which will be described as follows:

- y_i^N that has an error term $e_i \sim N(0, \sigma_i^2)$
- y_i^t that has error term $e_i \sim t(3/\sigma_i^2)$, $E[e] = 0$ when $v > 0$
- y_i^g that has error term $e_i \sim r(\alpha, \beta)$ where $\alpha = 1.5, \beta = 1$, $E[e] = \frac{\alpha}{\beta}$ to get the expectation

value equal to 0 then $E[e] = \frac{\alpha}{\beta} - \frac{\alpha}{\beta} = 0$ so that the error term is $e_i \sim r(1.5, 1) - \frac{1.5}{1}$

4. For each vector, y will be used as an interested variable on the Fay-Herriot model either not transformed $h(y)$ or transformed by log, BC and DP given by:

$$h^{\log}(y, \lambda) = \log(y_i) = \beta_0 + \beta_1 x_i + v_i + e_i \quad (3)$$

$$h^{BC}(y, \lambda) = \frac{y_i^\lambda - 1}{\lambda} = \beta_0 + \beta_1 x_i + v_i + e_i \quad (4)$$

$$h^{DP}(y, \lambda) = \frac{y_i^\lambda - y_i^{-\lambda}}{2\lambda} = \beta_0 + \beta_1 x_i + v_i + e_i \quad (5)$$

5. Stages 2 through 4 will be raised 100 times to estimate 4 parameters, there are $\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\tau}^2$, and $\widehat{\lambda}$ for each transformation. From the 100 datasets, MAPE is obtained from the error size of the actual interested variable (y) with its estimation ($\widehat{y}$).
6. Evaluate the results of the application of two transformations by comparing the MAPE obtained.

2. Methods

In the section, we present the methods of the works as following:

2.1. Parametric transformed Fay-Herriot models

Let $h(y, \lambda)$ be a monotone transformation from $\mathbb{R}_+$ to $\mathbb{R}$ for positive y . It is noted that the transformation involves unknown parameter λ . It is assumed that positive data $y_1, y_2, \dots, y_m$ are available, where y_1 is an area-level data like a sample mean for the i th small area. For $i = 1, 2, \dots, m$ assume that the transformed data $h(y_i, \lambda)$ has a linear mixed model given by model in (1).

In this paper, we want to consider a class of the transformations $h(y, \lambda)$ so that the maximum likelihood estimator (MLE) of λ is consistent. In order for transformation of the interested variable to satisfy normality assumptions, there are three assumptions that must be satisfied:

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

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

$$h_y(y, \lambda) = \frac{\partial h(y, \lambda)}{\partial y}, h_\lambda(y, \lambda) = \frac{\partial h(y, \lambda)}{\partial \lambda}, h_{\lambda\lambda}(y, \lambda) = \frac{\partial^2 h(y, \lambda)}{\partial \lambda \partial \lambda'},$$

$$h_{y\lambda}(y, \lambda) = \frac{\partial^2 h(y, \lambda)}{\partial y \partial \lambda}, h_{y\lambda\lambda}(y, \lambda) = \frac{\partial^3 h(y, \lambda)}{\partial y \partial \lambda \partial \lambda'}$$

(A.3) Transformation function $h(y, \lambda)$ satisfied the integrability conditions given by

$$E[\{h(y, \lambda) - \mu\}, h_\lambda(y, \lambda)] = O(1), E[h_\lambda(y, \lambda)] = O(1),$$

$$E[h_{\lambda\lambda}(y, \lambda)] = O(1), E\left[\frac{d}{d\lambda} \frac{h_{y\lambda}(y, \lambda)}{h_y(y, \lambda)}\right] = O(1)$$

where $h(y, \lambda)$ is normally distributed.

When $h(y, \lambda) = y$, the above model refers to the model developed by Fay and Herriot [3], while Slud and Maiti [5] used this model with logarithmic transformation on the interested variable so that $h(y, \lambda) = \log(y)$. Another alternative method for parametric transformation is Box-Cox (BC) transformation but this transformation does not satisfy the assumption (A.1), so that it does not conform to the assumption of normality and makes the parameter estimation λ to be inconsistent. As for the dual-power (DP) transformation, the three assumptions above are satisfied. Thus the DP transformation is recommended to be used for the transformation of interested variable [2].

2.2. Estimation of parameters

We derive consistent estimators of the parameters β, τ^2 , and λ in the model (1). We first provide estimators $\widehat{\tau}^2(\lambda)$ and $\widehat{\beta}(\lambda)$ of τ^2 and β , respectively, when λ is fixed. we next derive an estimator $\widehat{\lambda}$ by solving an equation for estimating λ , and then we get estimators $\widehat{\tau}^2(\widehat{\lambda})$ and $\widehat{\beta}(\widehat{\lambda})$ by plugging in the estimator $\widehat{\lambda}$. We begin by estimating β and τ^2 when λ is given. In this case, the conventional procedures given in the literature for the Fay-Herriot model can be inherited to the transformed model. Thus, for given τ^2 and λ , the MLE of β is given by

$$\widehat{\beta}(\tau^2, \lambda) = \left\{ \sum_{i=1}^m (\tau^2 + \sigma_i^2)^{-1} \mathbf{x}_i \mathbf{x}_i' \right\}^{-1} \sum_{i=1}^m (\tau^2 + \sigma_i^2)^{-1} \mathbf{x}_i h(y_i, \lambda) \quad (6)$$

Concerning the estimation of τ^2 given λ , we consider a class of estimators $\widehat{\tau^2}(\lambda)$ satisfying the following assumption:

$$(A.4) \quad \widehat{\tau^2}(\lambda) = \tau^2 + \mathbf{O}_p(m^{-\frac{1}{2}})$$

$$(A.5) \quad \frac{\partial \widehat{\tau^2}(\lambda)}{\partial \lambda} = \mathbf{O}_p(1),$$

$$(A.6) \quad \frac{\partial \widehat{\tau^2}(\lambda)}{\partial \lambda} - E\left[\frac{\partial \widehat{\tau^2}(\lambda)}{\partial \lambda}\right] = \mathbf{O}_p\left(m^{-\frac{1}{2}}\right)$$

Assumption (A.4) implies that the estimator $\widehat{\tau^2}(\lambda)$ is consistent. Assumption (A.5) and (A.6) will be used for approximating prediction errors of EBLUP. Let us define $\hat{\beta}(\lambda)$ by

$$\hat{\beta}(\lambda) = \hat{\beta}(\widehat{\tau^2}(\lambda), \lambda) \quad (7)$$

Which is provide by substituting $\widehat{\tau^2}(\lambda)$ into $\hat{\beta}(\tau^2, \lambda)$ in (6). Asymptotic properties of $\hat{\beta}(\lambda)$ can be investigated under the following standard conditions on σ_i^2 and x_i . The following are assumed for σ_i^2 and x_i :

$$(A.7) \quad m^{-1} \sum_{i=1}^m x_i x_i' \text{ converges to a positive definite matrix as } m \rightarrow \infty$$

$$(A.8) \quad \text{There exist constants } \underline{\sigma}^2 \text{ and } \bar{\sigma}^2 \text{ such that } \underline{\sigma}^2 \leq \sigma_i^2 \leq \bar{\sigma}^2 \text{ for } i = 1, 2, \dots, m \text{ and } \underline{\sigma}^2 \text{ and } \bar{\sigma}^2 \text{ are positive constants independent of } m.$$

Since $\hat{\beta}(\tau^2, \lambda) \sim N_p(\beta, \{\sum_{i=1}^m (\tau^2 + \sigma_i^2)^{-1} x_i x_i'\}^{-1})$, it is clear that $\hat{\beta}(\tau^2, \lambda)$ is consistent and $\hat{\beta}(\tau^2, \lambda) - \beta = \mathbf{O}_p\left(m^{-\frac{1}{2}}\right)$ under assumption (A.7) and (A.8). The MLE of τ^2 is obtained as a solution of the equation

$$\sum_{i=1}^m (\tau^2 + \sigma_i^2)^{-2} \{h(y_i, \lambda) - x_i' \hat{\beta}(\tau^2, \lambda)\}^2 = \sum_{i=1}^m (\tau^2 + \sigma_i^2)^{-1} \quad (8)$$

We provide a consistent estimator of the transformation parameter λ . For estimating λ , we use the log-likelihood function, which is expressed as

$$L(\lambda, \tau^2, \beta) \propto -\frac{1}{2} \sum_{i=1}^m \log(\tau^2 + \sigma_i^2) - \frac{1}{2} \sum_{i=1}^m \frac{\{h(y_i, \lambda) - x_i' \beta\}^2}{\tau^2 + \sigma_i^2} + \sum_{i=1}^m \log h_y(y_i, \lambda) \quad (9)$$

The derivate with respect to λ is written as

$$\begin{aligned} F(\lambda, \tau^2, \beta) &\equiv \frac{\partial L(\lambda, \tau^2, \beta)}{\partial \lambda} \\ &= \sum_{i=1}^m \frac{h_{y\lambda}(y_i, \lambda)}{h_y(y_i, \lambda)} - \sum_{i=1}^m (\tau^2 + \sigma_i^2)^{-1} \{h(y_i, \lambda) - x_i' \beta\} h_{\lambda}(y_i, \lambda) \end{aligned} \quad (10)$$

Thus, we suggest estimator $\hat{\lambda}$ as a solution to the equation:

$$F(\hat{\lambda}, \widehat{\tau^2}(\hat{\lambda}), \hat{\beta}(\hat{\lambda})) = 0$$

where $\widehat{\tau^2}(\lambda)$ is the maximum likelihood estimator of τ^2 , the estimator results from the equation above are ML estimators of λ which implies that the estimation is consistent and in accordance with the asymptotic normality [2].

2.3. Empirical Best Linear Unbiased Predictor (EBLUP)

Best linear unbiased predictor (BLUP) is a parameter estimation that minimizes mean squared error (MSE) among other bias parameters of the estimation class [6]. BLUP is produced with the assumption that the variance component is known. But in practice the variance component is unknown. Therefore, it is necessary to estimate the various components through sample data. The EBLUP method substitutes this unknown component with its estimator. Henderson [7] shows that replacing the components of variance within a BLUP with its estimator can lead to bias. But Kackar and Harville [8] show that two approaches by estimating the components of variance then use them to

predict and predict fixed parameters and random components will be able to produce estimators that are not biased.

The Fay-Herriot model is defined as follows $h(y_i, \lambda) = x_i^T \beta + z_i^T v_i$. Consider the problem of predicting $\theta_i = x_i^T \beta + z_i^T v_i$, the best predictor of θ_i is given by

$$\hat{\theta}_i^{BP} = \hat{\theta}_i(y_i | \beta, \tau^2, \lambda) = x_i^T \beta + \frac{\tau^2}{\tau^2 + \sigma_i^2} \{h(y_i, \lambda) - x_i^T \beta\} \quad (11)$$

with $\text{MSE}(\hat{\theta}_i^{BP}) = \text{Var}(\hat{\theta}_i) = \frac{\tau^2 \sigma_i^2}{\tau^2 + \sigma_i^2} = g_{1i}(\tau^2)$. Since β is unknown, substituting $\hat{\beta}(\tau^2, \lambda)$ in (6), into

$$\hat{\theta}_i^{BLUP} = \hat{\theta}_i(y_i | \tau^2, \lambda) = x_i^T \hat{\beta}(\tau^2, \lambda) + \left(\frac{\tau^2}{\tau^2 + \sigma_i^2}\right) (h(y_i, \lambda) - x_i^T \hat{\beta}(\tau^2, \lambda)) \quad (12)$$

with $\text{MSE}(\hat{\theta}_i^{BLUP}) = g_{1i}(\tau^2) + g_{2i}(\tau^2)$, $g_{1i}(\tau^2) = \frac{\tau^2 \sigma_i^2}{\tau^2 + \sigma_i^2}$, $g_{2i}(\tau^2) = \frac{\sigma_i^2}{(\tau^2 + \sigma_i^2)^2 [x_i^T (x^T V^{-1} x) x_i]}$.

Which is the best linear unbiased predictor (BLUP) as a function of $h(y_i, \lambda)$. For the parameters β and τ^2 , we use the estimators $\hat{\beta}(\hat{\lambda})$ and $\hat{\tau}^2(\hat{\lambda})$ into BLUP, we get empirical best linear unbiased predictor (EBLUP)

$$\hat{\theta}_i^{EBLUP} = x_i^T \hat{\beta}(\hat{\lambda}) + \left(\frac{\hat{\tau}^2(\hat{\lambda})}{\hat{\tau}^2(\hat{\lambda}) + \sigma_i^2}\right) \{h(y_i, \hat{\lambda}) - x_i^T \hat{\beta}(\hat{\lambda})\} \quad (13)$$

with $\text{MSE}(\hat{\theta}_i^{EBLUP}) = g_{1i}(\hat{\tau}^2) + g_{2i}(\hat{\tau}^2) + 2g_{3i}(\hat{\tau}^2)$. where $g_{3i}(\tau^2) = \frac{\sigma_i^2 \text{Var}(\hat{\tau}^2)}{2(\tau^2 + \sigma_i^2)^2}$. The addition of g_{2i} and g_{3i} from $\text{MSE}(\hat{\theta}_i^{BP})$ to $\text{MSE}(\hat{\theta}_i^{BLUP})$ and $\text{MSE}(\hat{\theta}_i^{EBLUP})$ to correct uncertainty due to estimation of β and τ^2 .

2.4. Mean Absolute Percentage Error (MAPE)

According to Montgomery *et. Al.* [9], the error which includes the standard statistical is the mean error, mean absolute error, and mean square error. Error sizes that include relative size are mean percentage error and mean absolute percentage error (MAPE). In this paper, the error size used is the mean absolute percentage error (MAPE), with the following equation:

$$\text{MAPE} = \frac{1}{m} \sum_{t=1}^m \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100 \quad (14)$$

where y_t is the actual value and $\hat{y}_t$ is the estimation value. MAPE indicates an error is estimating a value compared to its actual value.

3. Results and Discussion

This study carried out three transformations on positive value interested variable, namely transformation Log, BC, and DP. In most of the economic application, however, the data are nonnegative. Three transformations were applied to see the performance of transformation to normalize nonnegative data. Log and BC transformation are commonly used transformations, whereas DP transformation is the development of BC transformation. Removing the bound in the BC transformation while at the same time preserving the non-negativeness of y is the key motivation of the new transformation. In BC and DP transformations there is a parameter λ which is the power of transformation, whereas for Log transformation does not have a lambda value for its transformation parameters.

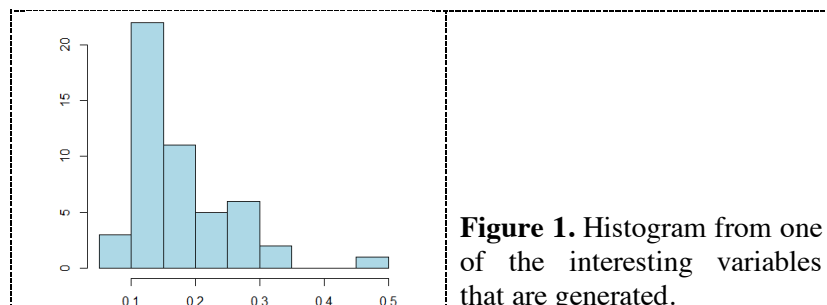


Figure 1. Histogram from one of the interesting variables that are generated.

The interesting variable distribution pattern that will be used to prove the truncated problem in BC transformation will be shown in Figure 1. The interesting variable is generated by following the y^N with a value of $\lambda = 0.6$ which produce a nonlinear data pattern and the skewness of data to the right.

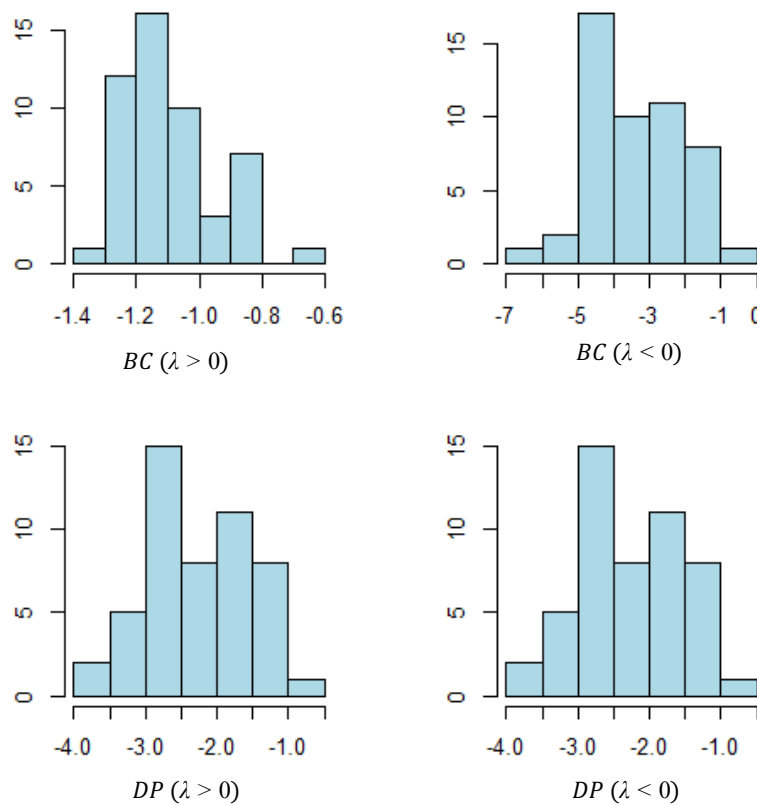


Figure 2. Histogram $h^{BC}(y, \lambda)$ and $h^{DP}(y, \lambda)$.

Figure 2 presents the results of $h^{BC}(y, \lambda)$ and $h^{DP}(y, \lambda)$ when $\lambda > 0$ ($\lambda = 0.6$) dan $\lambda < 0$ ($\lambda = -0.6$). From the histogram, it can be seen that the interested variable that originally skewness to the right occurred in a change approaching the symmetric form (close to the normal distribution) by transforming DP and BC. DP results show more symmetric data distribution than BC at $\lambda > 0$ or $\lambda < 0$. Table 1 shows the resulting skewness and kurtosis values $h^{BC}(y, \lambda)$ and $h^{DP}(y, \lambda)$ for positive or negative λ . If the distribution of data is normal then the skewness and kurtosis values will approach 0. The skewness and kurtosis values on DP are closest to 0 compared to the BC. This shows that the DP transformation can produce a data set that can change the interested variable which was originally not normally distributed into data that is closely normal distribution as shown in Figure 2 and supported by skewness and kurtosis values in Table 1.

Table 1. Skewness and kurtosis for $h^{BC}(y, \lambda)$ dan $h^{DP}(y, \lambda)$.

$h(y, \lambda)$	Skewness	Kurtosis
BC ($\lambda > 0$)	0.95	0.81
BC ($\lambda < 0$)	-0.06	-0.63
DP ($\lambda > 0$)	0.04	-0.62
DP ($\lambda < 0$)	0.04	-0.62

Illustration for truncated problem handling is presented in Figure 3 and Figure 4 which uses the values $\lambda = 0.6$ and $\lambda = -0.6$, so that it produces a value of $1/\lambda = 1.6667$ and $-1/\lambda = -1.6667$. Truncated problems in BC transformations can be seen in the domain $h^{BC}(y, \lambda)$ for $\lambda > 0$ and $\lambda < 0$, which is $h^{BC}(y, \lambda)$ always above $-1/\lambda$ when $\lambda > 0$ while in when $\lambda < 0$ domain $h^{BC}(y, \lambda)$ will be below $1/\lambda$. interested variable is generated so that it can be located in the area of $y > 1$ and $0 < y < 1$. When $\lambda > 0$, the main difference between the two transformations happens at the part where y takes value from 0 to 1: BC maps $[0, 1]$ to $[-1/\lambda, 0]$, whereas the dual power transformation maps it to $(-\infty, 0]$. Both functions map the $y > 1$ into $[0, \infty)$ with the curve of DP lying below that of the BC. When $\lambda < 0$, BC is able to translate the domain $y [0, 1]$ part into $[-\infty, 0)$ but maps the $y > 1$ part into $[0, 1/\lambda]$. The DP is symmetric in λ , and hence a negative λ gives the same function as a positive one.

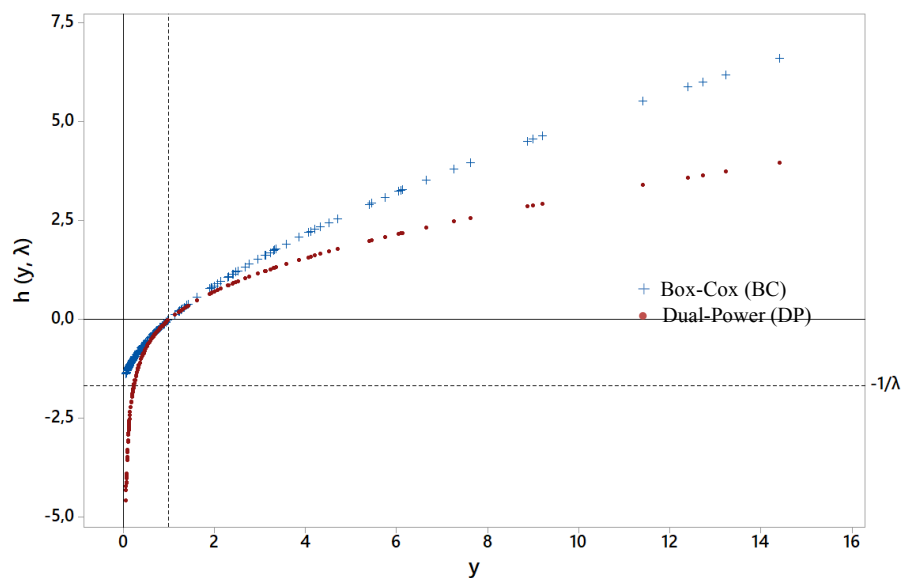


Figure 3. Scatterplot interested variable with $h^{BC}(y, \lambda)$ and $h^{DP}(y, \lambda)$ when $\lambda > 0$.

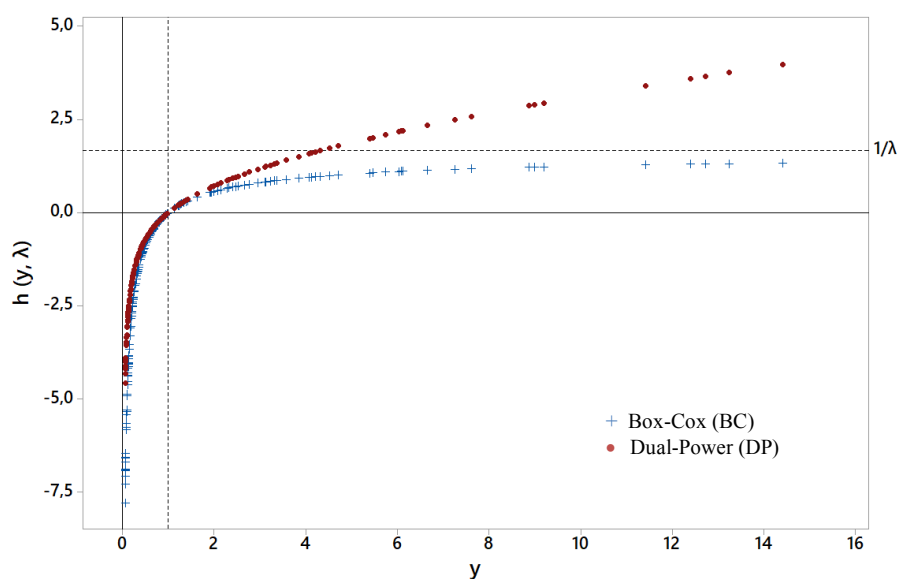


Figure 4. Scatterplot interested variable with $h^{BC}(y, \lambda)$ and $h^{DP}(y, \lambda)$ when $\lambda < 0$.

The data generated in the simulation experiment consisted of three data, namely are y^N , y^t and y^g . The scenario was tested with the Fay-Herriot model without transformation and transformation. Transformation is namely Log, BC, and DP. The data simulation consist of a non-linear interested variable. Non-linear interested variable is generated when data is patterned dual-power with lambda 0.1, 0.3 and 0.6. Figure 5 shows plot illustrations when using interested variable generation dual power with lambda which is close to 0 will result in a non-linear interested variable, whereas when using lambda value close to 1 it produces interested variable that is close to linear.

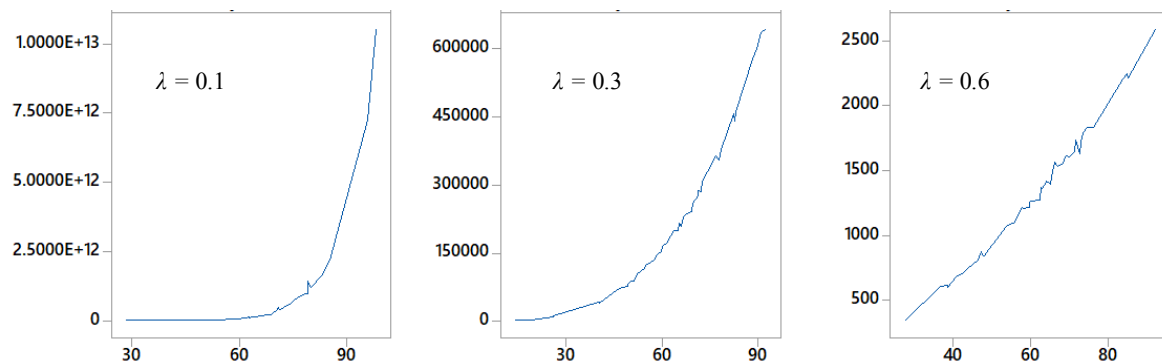


Figure 5. Plot x and interested variable.

Table 2 shows the results of calculating the average value of MAPE for all four methods in the three interested variables. When the y is not linear ($\lambda < 1$), the MAPE results obtained in the Fay-Herriot model without transformation are very large compared to the model that has been transformed. Even though the Log method is capable of correcting the interested variable estimation by producing a smaller MAPE than the model without transformation, the BC and DP methods give smaller MAPE compared to the Log method. This shows that the BC and DP methods can correct the estimation of interested variable well compared to the Log transformation method. Both of these transformations compete with each other to produce the smallest average MAPE value. From the data in Table 2 it can be seen that the results of MAPE for DP transformation are not always consistently smaller than the BC transformation. Even if it does not consistently produce the smallest value, but technically, DP can overcome truncated problems that occur in BC. So that this new transformation is proposed to be the best transformation to overcome the abnormality of an interested variable. Because this transformation produces a monotonic function that has a domain of positive real numbers ($\mathbb{R}_+$) and range of whole real numbers ($\mathbb{R}$).

Table 2. The average value of MAPE

λ	Interested variable	Without Transformation	Log	BC	DP
0.1	y^N	4.177319 E+8	754.4541	10.0015	10.0511
	y^t	3.800183 E+8	8022.1000	30.4603	36.4165
	y^g	7.006713 E+7	835.6814	17.7954	17.6675
0.3	y^N	593233	19.8421	6.9795	5.9217
	y^t	9067505	49.2529	11.2855	10.4276
	y^g	4141.252	20.4223	8.4711	8.2263
0.6	y^N	7.081	7.3029	2.1543	2.1478
	y^t	7.781	9.0590	4.2723	4.0294
	y^g	8.212	7.9797	3.7037	3.3806

References

- [1] Rao J N K 2003 *Small Area Estimation* John Willey&Sons Inc New Jersey
- [2] Sugasawa A and Kubokawa T 2015 Parametric transformed Fay–Herriot model for small area estimation *Journal of Multivariate Analysis*. **139** pp 295-311
- [3] Yang Z 2002 A Modified Family of Power Transformations *Econom Lett* **92** pp 14–19
- [4] Fay R E and Herriot R A 1979 Estimation of Income for Small Places: An Application of James-Stein Procedures to Census Data *Journal of the American Statistical Associations* **74** pp 269-277
- [5] Slud E V and Maiti T 2006 Mean-squared error estimation in transformed Fay-Herriot models. *Journal of the Royal Statistical Society* **68** pp 239-257
- [6] Kurnia A and Notodiputro K A 2006 Penerapan Metode Jackknife dalam Pendugaan Area Kecil *Forum Statistika dan Komputasi* **11(1)** pp 12-16
- [7] Henderson C R 1975 Best linear unbiased estimation and prediction under selection model *Biometrics* **31** pp 423-447
- [8] Kackar R N and Harville D A 1981 Unbiased for two-stage estimation and prediction procedure for mixed linear models *Communications in Statistics Theory and Methods* **A10** pp 1249-1261
- [9] Montgomery D C, Jennings C L, Kulahci M 2008 *Introduction to Time Series Analysis and Forecasting* John Wiley and Sons Inc Canada