We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,700
Open access books available

182,000
International authors and editors

195M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS

BOOK CITATION INDEX

INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Chapter

On the Selection of Power Transformation Parameters in Regression Analysis

Haithem Taha Mohammed Ali and Azad Adil Shareef

Abstract

In multiple linear regression, there are several classical methods used to estimate the parameters of power transformation models that are used to transform the response variable. Traditionally, these parameters can be estimated using either Maximum Likelihood Estimation or Bayesian methods in conjunction with the other model parameters. In this chapter, attention has been paid to four indicators of the efficiency and reliability of the regressive modeling, and study the possibility of considering them as decision rules through which the optimal power parameter can be chosen. The indicators are the coefficient of determination and p-value of the general linear F-test statistic. Also, the p-value of Shapiro-Wilk test (SWT) statistic for the residual's normality of the estimated linear regression of the transformed response vector and the estimated nonlinear regression of the original response vector resulting from the back transform of the power Transformation model. Real data were used and a computational algorithm was proposed to estimate the optimal power parameter. The authors concluded that the multiplicity of indicators does not lead to obtaining an optimal single value for the power parameter, but this multiplicity may be useful in fortifying the decision-making ability.

Keywords: Box-Cox transformation, multiple linear regression, Shapiro-Wilk test, general linear F-test statistic, Maximum likelihood estimation

1. Introduction

It is known that when some conditions of the statistical analysis are not met in the linear regression inputs, this means that the outputs of the statistical inference will be unreliable. The most important two conditions that must be fulfilled in the estimated linear regression model are the normality of residuals and constancy of its variances and it is the most violating condition as well [1]. Also, the unfulfilled of these conditions means that the estimated response mean function has no straight line shapes in its relationships with the explanatory variables. In this regard as well the lack of conditions becomes evident in complicated nonlinear models when the residuals in the original model are additive [2]. Therefore, the data transformation tools to linearity "Especially those that belong to the power transformation (PT) family" have been

used to greatly enhance the utility of statistical modeling and obtain a better fit as a general goal. That is, the main goal of data transformation is to prepare it to be compatible with the requirements of statistical inference tools [3]. In short, the confirming conditions for the best estimate of linear regression model are (i) the transformed response should be normally distributed with constant variance for each value of the predictor variables [4] or (ii) have more closing on a better fit to normality [5].

A large body of literature provides various suggestions and developments about the uses of PT for continuous variables in regression models, whether for the dependent variable, independent variables, or both. In this regard, two main research directions can be distinguished; the first is concerned with various proposals and strategies for developing the mathematical functions of PT models to address more complexities in data patterns. "For example, see [6–11]". While the second direction, which will be focused on in this chapter was concerned with the selecting methods of optimal power parameters in different PT families and datasets, "For example, see [8, 12–18]". There are many methods used to estimate the power parameters in Multiple Linear Regression (MLR). Traditionally, these parameters can be estimated using either Maximum Likelihood Estimation (MLE) or Bayesian methods in conjunction with the other model parameters [13]. It is also known that MLE is very sensitive to outliers [8]. Therefore, in addition to the traditional estimation methods, there are some other proposed methods based on the indicators of statistical modeling efficiency. These indicators were used as decision rules to choose the optimal value of the PT parameter [14, 15]. In general, multiplicity of criteria used for a particular dataset does not lead to a single value or at least a closed feasible region for a power parameter. Also, the values of the power parameters differ according to the transformation models.

Outside of the traditional methods, Bartlett's method was to choose a transformation based on the minimizing some measure of the heterogeneity of variance [16]. Tukey, 1949 [17] used the efficiency indicators of ANOVA such as minimization of the F-test value for non-additivity, minimization of the F ratio for interaction versus error, and maximization of the F ratio for treatments versus error [18]. Anscombe, 1961 and Anscombe and Tukey 1963 indicated how a certain function of the residuals can be providing us with a certain insight into the PT model [19]. While some other authors went on to propose algorithms for power parameter selection using the goodness of fit tests of the normality transformed data [12, 20] and coefficient of determination of the estimated linear regression of transformed response [15, 21, 22].

The chapter was divided into four sections. The second section included a short review of the PT models. Third section included the application and the computational algorithm. While the fourth section included the conclusions.

2. Power transformation: Short review

Finney, 1947 [23] assumed the following simple family of PT to transform both sides of the Dose-Response regression $Y = \eta(x, \beta) + \varepsilon$,

$$\psi(y, x) = \begin{cases} \left(y^{\lambda_1}, x^{\lambda_2}\right) & \lambda \neq 0 \\ (\ln y, \ln x) & \lambda = 0 \end{cases} \tag{1}$$

to form a monotonic simple linear regression $E(\psi(y)) = \eta(\psi(x), \beta))$ for the nonlinear relationship of the positive response Y given the positive dose X. λ_1 and λ_2 are the power parameters that can be estimated from the data.

Tukey, in 1957, developed another simple family of PT to accommodate negative y's by assuming [24],

$$\psi(y) = \begin{cases} (y+a)^\lambda & \lambda \neq 0 \\ ln\ (y+a) & \lambda = 0 \end{cases} \tag{2}$$

where the value of a can be chosen such that $(y + a) > 0$. In general, it is assumed that for each λ, $\psi(y)$ is a monotonic function of y over the admissible range [13].

Considering the common family of Box-Cox transformation (BCT) [13], it is possible to propose the following generalized approach,

$$\psi(y) = \begin{cases} \dfrac{(y+a)^\lambda - b}{\lambda \{gm(y+a)\}^{\lambda-1}} & \lambda \neq 0 \\ gm(y+a)\,ln\ (y+a) & \lambda = 0 \end{cases} \tag{3}$$

where a and b are constant quantities and a is chosen so that $(y + a) > 0$. $gm(y + a)$ represents the geometric mean of the shifted response $(y + a)$. Eq. (3) of BCT family hold for $(y + a) > 0$ and for $y > -a$. A number of PT models have been derived from this family; the following PT is equivalent to the simple version of Finney transformation Eq. (1) when $a = 0, b = 1$ and $gm(y + a) = 1$,

$$\psi(y) = \begin{cases} \dfrac{y^\lambda - 1}{\lambda} & \lambda \neq 0 \\ ln\ y & \lambda = 0 \end{cases} \tag{4}$$

As for the following PT, it is an extended form of Eq. (4) when $a \neq 0, b = 1$ and $gm(y + a) = 1$ and equivalent to Tukey transformation according to Eq. (2) since the analysis of variance is unchanged by a linear transformation [24],

$$\psi(y) = \begin{cases} \dfrac{(y+a)^\lambda - 1}{\lambda} & \lambda \neq 0 \\ ln\ (y+a) & \lambda = 0 \end{cases} \tag{5}$$

While as for $a = 0, b = 1$ can get a PT model equivalent to Eq. (4),

$$\psi(y) = \begin{cases} \dfrac{y^\lambda - 1}{\lambda \{gm(y)\}^{\lambda-1}} & \lambda \neq 0 \\ \{gm(y)\}\,ln\ y & \lambda = 0 \end{cases} \tag{6}$$

finally, as for $a \neq 1, b = 1$, can get the following PT model,

$$\psi(y) = \begin{cases} \dfrac{(y+a)^\lambda - 1}{\lambda \{gm(y+a)\}^{\lambda-1}} & \lambda \neq 0 \\ \{gm(y+a)\}\,ln\ (y+a) & \lambda = 0 \end{cases} \tag{7}$$

The main three properties of the PT family are: the first, is the continuity at λ go to zero, consider the BCT according to (Eq. 4), by the use of L'Hospital's Rule, it can be shown that, $\lim_{\lambda \to 0} (y^\lambda - 1)/\lambda = \ln y$. The second property is the concavity of the transformation function $\psi(y)$ that leads to obtaining a non-linear regression model for the original data after performing a back transformation of the transformed data model. While the third property is flexibility, as the transformation by power is suitable for dealing with a lot of data structures, and is also suitable for achieving a number of goals.

In BCT family models, if the transformation parameter was negative, the order of the variable would be reversed. That is, when Y is increasing, $\psi(y)$ is decreasing for $\lambda < 0$. So, Tukey, 1977 proposed the following model to maintain the order of the transformed variable [24],

$$\psi(y) = \begin{cases} y^\lambda & \lambda > 0 \\ \ln y & \lambda = 0 \\ -(y^\lambda) & \lambda < 0 \end{cases} \tag{8}$$

BCT, according to Eq. (4) and Eq. (6), is applicable and restricted to positive data. So, Yeo and Johnson, 2000 [25] generalized BCT to include negative and positive values in datasets. They used a smoothness condition to combine the transformations for positive and negative observations, obtaining a one-parameter transformation family. For $Y \in R$, Yeo-Jonson Transformation (YJT) is given by,

$$\Psi(y) = \begin{cases} \left((y + 1)^\lambda - 1\right)/\lambda & \lambda \neq 0 \text{ and } y \geq 0 \\ Ln(y + 1) & \lambda = 0 \text{ and } y \geq 0 \\ -\left((-y + 1)^{2-\lambda} - 1\right)/(2 - \lambda) & \lambda \neq 2 \text{ and } y < 0 \\ Ln(-y + 1) & \lambda = 2 \text{ and } y < 0 \end{cases} \tag{9}$$

Three properties of YJT namely [26]; (i) For $Y \geq 0$, then $\Psi(y) \geq 0$, and for $Y < 0$, then $\Psi(y) < 0$. (ii) $\Psi(y)$ is continuous at $\lambda \to 0$ and $\lambda \to 2$. (iii) $\Psi(y)$ is convex with $\lambda > 1$, and concave with $\lambda < 1$.

In MLR, for all previous PT families, the optimal power parameter $\lambda^* = 1$ confirm the linearity of the regression relationship and no transformation is required, $\lambda^* < 1$ refers to the fact that the regression relationship of the original data is not linear due to the skew of the response distribution towards the right and vice versa for $\lambda^* > 1$ [8].

The main idea of the use of PT models in data processing is based on the assumption that the transformed response variable in MLR follows a normal distribution. As a result, the original response follows an unknown and somewhat complex Probability Density Function (PDF) in the exponential family. In the sense that the response transformation changes the shape of data and its original unit of measure [27]. Thus, the optimal power parameter and other model parameters are estimated for the transformed data by the common estimation methods. In the end, the back-transformation will represent the fitted nonlinear regression model of the original data. Mathematically, for the univariate $Y > 0$, based on the main assumption; $Y^{(\lambda)} = \psi(y) \sim N(\mu, \sigma^2)$, the PDF of the univariate $Y > 0$ is given by $f_Y(y; \lambda, \mu, \sigma^2) = f_{Y^{(\lambda)}}(\psi(y); \lambda, \mu, \sigma^2).J(Y, \lambda)$, where, $J(Y, \lambda) = |d\psi(y)/dy|$ is the Jacobian factor to transform $(Y_1, \ldots, Y_n) \to (\psi(Y_1, \lambda), \ldots, \psi(Y_n, \lambda))$.

Consider the MLR model $\mathbf{Y}^{(\lambda)} = \mathbf{X\beta} + \boldsymbol{\varepsilon}$, where $\mathbf{Y}^{(\lambda)} = \psi(\mathbf{y})$ represents the *(nx1)* column vector of transformed values of response variable vector $\mathbf{Y}$. $\mathbf{X}$ is the $nx(p+1))$ known information matrix. $\boldsymbol{\beta}$ is the *(p + 1)x1* unknown parameters vector and $\boldsymbol{\varepsilon}$ is the *(nx1)* column vector of residuals and distributed according to the normal distribution with mean equal to *(nx1)* zero vector and identity variances matrix equal to $\sigma^2 \mathbf{I}_n$. Also, based on the main assumption; $\psi(\mathbf{y}) \sim N(\mathbf{X\beta}, \sigma^2 \mathbf{I}_n)$, the joint PDF of response variable vector $\mathbf{Y}$, is given by the following likelihood function,

$$L\left(\lambda, \boldsymbol{\beta}, \sigma^2 | \mathbf{y}, \mathbf{X}\right) = f_{\mathbf{Y}}(\mathbf{y}) = \left(2\Pi\sigma^2\right)^{-n/2} . exp\left\{\frac{-\left(\mathbf{Y}^{(\lambda)} - \mathbf{X\beta}\right)^T \left(\mathbf{Y}^{(\lambda)} - \mathbf{X\beta}\right)}{2\sigma^2}\right\} . J(\mathbf{Y}, \lambda)$$

(10)

Where $J(\mathbf{Y}, \lambda) = \prod_{i=1}^{n}\left|dy_i^{(\lambda)}/dy_i\right|$. Applying the method of MLE for Eq. (10) and solving $\partial Ln\ L/\partial\boldsymbol{\beta} = 0$ and $\partial Ln\ L/\partial\sigma^2 = 0$, we get the following estimates for each value of λ,

$$\hat{\boldsymbol{\beta}}(\lambda) = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Y}^{(\lambda)}$$

(11)

$$\sigma^2(\lambda) = (1/n)\left(\mathbf{Y}^{(\lambda)}\right)^T H\left(\mathbf{Y}^{(\lambda)}\right)$$

(12)

Where $H = \mathbf{I} - \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T$. Substituting the estimates $\hat{\boldsymbol{\beta}}(\lambda)$ and $\sigma^2(\lambda)$ in the logarithm of likelihood function Eq. (10) gives what might be called the Box-Cox objective function after ignoring the constant term,

$$L(\lambda, \mathbf{y}) = -(n/2)\ log\ \sigma^2(\lambda) + log\ J(\mathbf{Y}, \lambda)$$

(13)

Note that the likelihood for a given λ is inversely proportional to the sum of the squared residuals $SS_{res}(\lambda)$ of the regression $\psi(\mathbf{y})$ on $\mathbf{X}$. The likelihood function is maximized when $SS_{res}(\lambda)$ value is minimized. The value of the power parameter λ is optimal when $L(\lambda, \mathbf{y})$ is at its maximum.

3. An application and computational algorithm

We consider a real economic dataset that includes a set of five explanatory variables affecting the Current Account of the Republic of Iraq in the period 2004–2020 (**Table 1**). The dataset has been obtained from Iraqi Central Bank and is also available at https://cbiraq.org/. R program was used to analyze the data.

Evident from (**Figure 1**) that there are three outliers among the values of the response variable, which are the values y_9, y_{10} and y_{11}. Also, regarding BCT and the conditions for its implementation, the response positively constraint is not fulfilled due to the presence of some negative values. Therefore, the estimating of MLR for these data would be risky, and the diagnostic and inference tools might give misleading results. So, there is a certain and definite need to conduct some mathematical preparations to shift the data to another space.

Years	Current Account Y	Deficit/Surplus in general budget X_1	GDP at Current Prices X_2	Oil Revenues X_3	Other Revenues X_4	Public Expenditures X_5
2004	−5,796,516	865,248	53,235,358	28639.1	4343.6	32117.5
2005	5,048,118	14,127,715	73,533,598	33627.2	6875.7	26375.1
2006	18,521,580	10,248,866	95,587,954	41076.2	7987.1	38076.8
2007	33,161,857	15,568,219	111,455,813	44646.1	9953.4	39031.2
2008	42,020,417	20,848,807	157,026,061	70,124	10128.1	59403.3
2009	−493,311	2,642,328	130,642,187	43309.2	11900.1	52,567
2010	1,453,244	44,022	162,064,566	59,794	10384.2	70134.2
2011	29,228,742	30,049,726	217,327,107	98090.2	10717.2	78757.6
2012	378,788,640	14,677,649	254,225,490	109772.1	10045.1	105139.5
2013	430,082,730	−5,360,605	273,587,529	112894.3	945.7	119127.5
2014	224,949,984	−8,086,894	266,420,384	97072.4	8537.4	113473.5
2015	−4,377,124	−39,277,264	199,715,699	51312.6	15157.6	70397.5
2016	46,126,504	−12,658,167	203,869,832	44,267	10142.2	67067.4
2017	93,634,588	1,932,057	225,995,179	65071.9	12350.2	75490.1
2018	−11,244,618	−12,514,516	226,455,132	95619.8	10,950	80873.1
2019	27,714,354	−4,156,528	276,157,867	99216.3	8350.6	111723.5
2020	−1,582,698	−12,882,754	219,768,798	54448.5	8751.1	76082.4

Table 1.
The current account and some explanatory variables of Republic of Iraq for the period 2004–2020 "Million IQD".
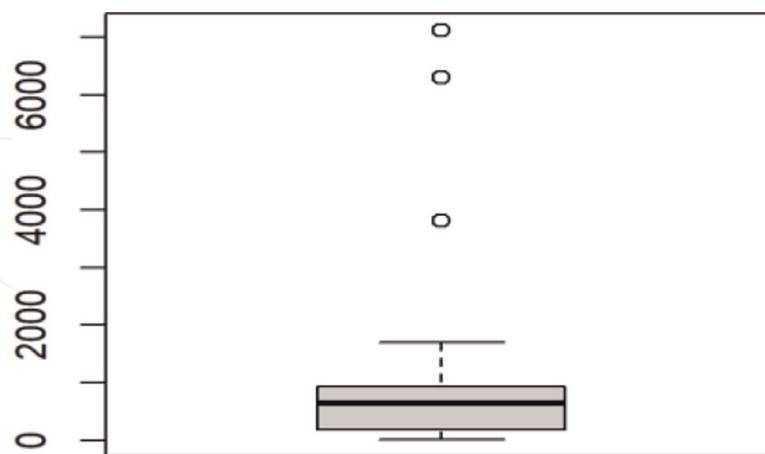

Figure 1.
Box plot of response variable values.

So, the following MLR model was chosen, which addresses the presence of negative values in the data and might have some robustness to get past the implications of having outliers,

$$Z^{(\lambda)} = U\beta + \varepsilon \qquad (14)$$

$Z^{(\lambda)}$ represents the (17×1) column vector of transformed values of Simple Index Numbers (SIN) of the original response variable vector Y. $Z^{(\lambda)}$ is defined according to the following simplified version of BCT family,

$$Z^{(\lambda)} = \begin{cases} \dfrac{z^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \ln z & \lambda = 0 \end{cases} \tag{15}$$

And the ith-value in the Z's vector is defined to the following SIN with considering the first year as a base year,

$$z_i = \frac{y_{i+1} + a}{y_1 + a}(100) \tag{16}$$

a is a constant to shift the location of the response vector to positive space where it is chosen to ensure the BCT's constraint $(Y + a) > 0$. U is the (17x5) known information matrix of the SIN considering the first year as a base year of the explanatory variables, where,

$$u_{ik} = \frac{x_{(i+1)k}}{x_{1k}}(100) \tag{17}$$

and $u_{1k} = 100\%$ For $k = 2, 3, 4, 5$. While the SIN for the first explanatory variable is defined as,

$$u_{i1} = \frac{x_{(i+1)1} + b_k}{x_{1k} + b_k}(100) \tag{18}$$

where $u_{11} = 100\%$ and b_k is a constant to shift the location of the explanatory variables to positive values where it is chosen so that $(X + a) > 0$. β is the (6x1) unknown parameters vector and ε is the (17×1) column vector of residuals and distributed according to the normal distribution with mean equal to (17×1) zero vector and identity variances matrix equal to $\sigma^2 I_n$.

Finally, the nonlinear multiple regression model for the original data regression Z given X is derived from the following back-transform of BCT,

$$Z = \begin{cases} \left(\lambda Z^{(\lambda)} + 1\right)^{1/\lambda} & \lambda \neq 0 \\ \exp\left(Z^{(\lambda)}\right) & \lambda = 0 \end{cases} \tag{19}$$

Thus, we can have obtained the estimated multiple nonlinear regression model for the original data regression from the estimated MLR of transformed data,

$$\widehat{Z} = \begin{cases} \left(\lambda\, U\widehat{\beta} + 1\right)^{1/\lambda} & \lambda \neq 0 \\ \exp\left(U\widehat{\beta}\right) & \lambda = 0 \end{cases} \tag{20}$$

A number of modeling efficiency indicators are included in our search algorithm to obtain optimal power parameter λ^*. The first is the traditional MLE. The second,

third, and fourth are the coefficient of determination (CoD), p-value of SWT statistic for the residual's normality, and p-value of the general linear F-test statistic of the estimated linear regression of the transformed response vector. The fifth is the p-value of SWT statistic for the residual's normality of the estimated nonlinear regression of the original response vector resulting from the back-transform of BCT. The proposed computational algorithm is as follows:

Step 1: Transform the original response vector $\boldsymbol{Y}$ to SIN's vector $\boldsymbol{Z}$ according to Eq. (16) of vector elements and the original information matrix $\boldsymbol{X}$ to SIN's matrix $\boldsymbol{U}$ according to Eq. (17) of matrix elements.

Step 2: Choose a set of candidate values for the power parameter. For example, fix $\lambda \in \Lambda$, where $\Lambda = \{-2, -1.9, \dots, 0, \dots, 1.9, 2\}$. Λ can be expanded to an acceptable range from which we can obtain a convex curve for MLE, and the same applies to CoD. Also, obtaining a minimum value of the p-value of general linear F-test statistic within Λ can be an indicator of acceptance of the candidate range.

Step 3: Transform the SIN's vector $\boldsymbol{Z}$ to $\psi(\boldsymbol{Z})$ using the simple version of the BCT family according to Eq. (15) by the first candidate λ^* in Λ.

Step 4: Estimate the parameters $\hat{\boldsymbol{\beta}}(\lambda^*)$ and $\sigma^2(\lambda^*)$ of MLR of $\boldsymbol{Z}^{(\lambda)}$ given $\boldsymbol{X}$ according to Eq. (14) using Eq. (11) and Eq. (12).

Step 5: Estimate log-likelihood function $L(\lambda^*, \boldsymbol{z})$ according to Eq. (13). Calculate CoD, a p-value of SWT statistics to test the residual vector normality, and p-value of the general linear F-test statistic.

Step 6: Estimate the multiple nonlinear regression model for the original data regression using Eq. (20).

Step 7: Calculate the p-value of SWT of the residual vector normality of the estimated multiple nonlinear regression model of the original data.

Step 8: Repeat all the steps from 3 to 7 for all values of $\lambda \in \Lambda$.

The tables below show the results of applying the computational algorithm. **Table 2** shows the optimal values of λ against each indicator in its optimal state. **Table 3** shows the estimates of power parameters according to the five indicators for all $\Lambda = \{-3, -2.9, \dots, 0, \dots, 2.9, 3\}$.

Based on the results of p-values of general linear F-test statistic for all $\lambda \in \Lambda$ in (**Table 3**), we conclude that the full estimated models "whether for the non-Linear multiple regression models when $\hat{\lambda} \neq 1$ or MLR in which $\hat{\lambda} = 1$" are appropriate for the data. It is also clear that the residuals are close to normality shape for transformed data models except in the case Ln Z based on the indicator of the p-value of SWT of residuals normality (**Table 2**).

Indicators	Value	Optimal λ
MLE	-22.7	0
CoD	0.69	-0.5
p-value of SWT of Residuals Normality -Transformed data	0.99	3
p-value of SWT of Residuals Normality -Back Transformed data	0.89	2
p-value of F-test statistics	0.01	-0.5

Table 2.
The optimal values of λ against each indicator in its optimal state.

$\widehat{\lambda}$	MLE	CoD	p-value of SWT of Residuals Normality		p-value of F-test statistics
			Transformed data	Back Transformed data	
1	−39.8	0.63	0.94	0.84	0.03
(−3.0, −1.2)	(−161.6, −61.5)	0.68	(0.62, 0.67)	0.00	(0.01, 0.02)
(−1.1, −0.5)	(−58.5, −30.2)	0.69	(0.72, 0.84)	0.00	0.01
(−0.4, −0.3)	(−28.4, −25.6)	0.68	(0.58, 0.70)	0.00	(0.01, 0.02)
−0.2	−24.5	0.67	0.24	0.00	0.02
−0.1	−23.1	0.66	0.10	0.00	0.02
(0, 0.1)	(−22.8, −22.7)	0.65	(0.04, 0.05)	0.00	(0.02, 0.03)
(0.2, 0.5)	(−28.0, −23.1)	0.64	(0.11, 0.76)	(0.01, 0.11)	0.03
(0.6, 1.7)	(−63.1, −29.2)	0.63	(0.71, 0.96)	(0.20, 0.89)	0.03
(1.8, 2.3)	(−84.9, −65.2)	0.62	(0.68, 0.85)	(0.79, 0.89)	(0.03, 0.04
(2.4, 3.0)	(−110.3, −87.2)	0.61	(0.89, 0.99)	(0.46, 0.75)	0.04

Table 3.
Estimates of the power parameter according to the five indicators for all $\lambda \in \Lambda$.
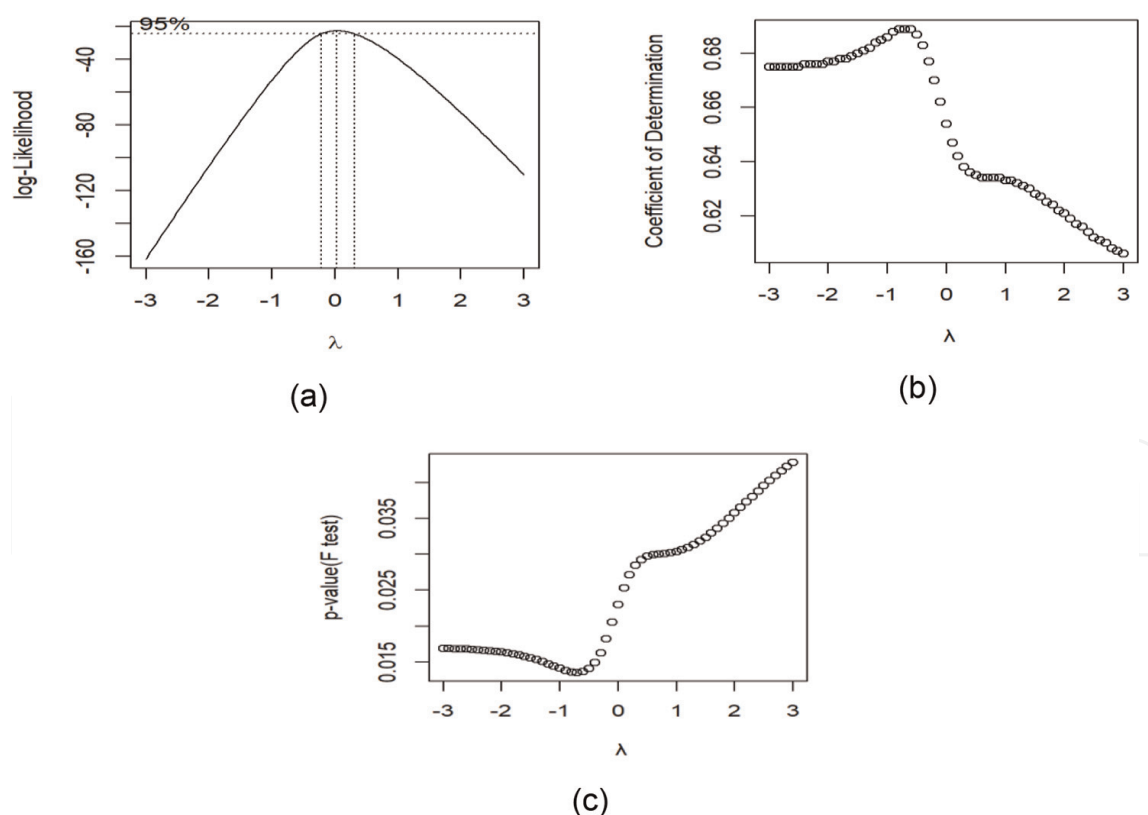

(a)

(b)

(c)

Figure 2.
For all $\lambda \in \Lambda$ (a) The Log-likelihood curve (b) The CoD estimates (c) p-values of general linear F-test statistic.

As for the MLE, the highest point corresponds to the value of the parameter when it is close to zero, (**Figure 2(a)**). That is, the optimal transformation is *Ln y*. On the other hand, according to the p-value of SWT of residuals normality, it is quite clear

that the residuals are abnormal. Therefore, it can be said that the results of general linear F-test statistics are not reliable.

As we mentioned in the article, the value of the optimal λ varies according to the different methods and indicators of estimation. Confirmation of that, the results of the optimal case for two of the five indicators led to obtaining identical values for the optimal power parameter at $\widehat{\lambda} = -0.5$ which are CoD (**Figure 2(b)**) and the p-value of general linear F-test statistic (**Figure 2(c)**).

4. Conclusions

The use of power transformation models to transform the response variable in regression relationships is, in fact, a way to create a nonlinear model for the data when the requirements of linear regression analysis are not met. In the sense that the statistical modeling operations of the transformed data are more like an intermediate station, the statistical analysis does not succeed unless the operations in this station are accurate and meet the requirements of the model construction. Therefore, there are many indicators of the success of statistical analysis, depending on the multiplicity of its reliability conditions. In this regard and when using PT models there are many methods for selecting the optimal power parameters. Two common directions can be identified: the first is the use of well-known estimation methods such as the method of MLE. The second is the use of some efficiency criteria in regression modeling as a decision rule for estimating the power parameter. We conclude that the multiplicity of criteria for selecting the power parameter does not mean that it can lead to a single value. However, the multiplicity of decision rules can contribute to providing features for optimal solutions and support the decision to choose the optimal power parameter.

Author details

Haithem Taha Mohammed Ali[1,2]* and Azad Adil Shareef[3]

1 Department of Economic Sciences, University of Zakho, Kurdistan Region, Iraq

2 Department of Economics, Nawroz University, Kurdistan Region, Iraq

3 Department of Statistics, University of Duhok, Kurdistan Region, Iraq

*Address all correspondence to: haithem.taha@uoz.edu.krd

IntechOpen

References

[1] Chatterjee S, Price B. Regression Analysis by Example. New York: John Wiley and Sons, Inc.; 1977. pp. 19-22

[2] Cook RD, Weisberg S. Diagnostics for heteroscedasticity in regression. Biometrika. 1983;**70**(1):1-10. DOI: 10.1093/biomet/70.1.1

[3] O'Hara RB, Kotze DJ. Do not log-transform count data. Methods in Ecology and Evolution. 2010;**1**:118-122. DOI: 10.1111/j.2041-210X.2010.00021.x

[4] van Albada SJ, Robinson PA. Transformation of arbitrary distributions to the normal distribution with application to EEG test-retest reliability. Journal of Neuroscience Methods. 2007;**161**(2):205-211. DOI: 10.1016/j.jneumeth.2006.11.004

[5] Box GEP, Cox DR. An analysis of transformations revisited, rebutted. Journal of the American Statistical Association. 1982;**77**(377):209-210

[6] Klein Entink RH, van der Linden WJ, Fox JPA. Box–Cox normal model for response times. British Journal of Mathematical and Statistical Psychology. 2009;**62**(Pt 3):621-640

[7] Fischer C. Comparing the logarithmic transformation and the box-cox transformation for individual tree basal area increment models. Forest Science. 2016;**62**(3):297-306. DOI: 10.5849/forsci.15-135

[8] Raymaekers J, Rousseeuw PJ. Transforming Variables to Central Normality. Machine Learning. 2021. DOI: 10.1007/s10994-021-05960-5

[9] Ferrari SLP, Fumes G. Box-Cox symmetric distributions and applications to nutritional data. AStA Advances in Statistical Analaysis. 2017;**101**:321-344. DOI: 10.1007/s10182-017-0291-6

[10] Yeo IK, Johnson RA. A new family of power transformations to improve normality or symmetry. Biometrika. 2000;**87**(4):954-959

[11] Vélez JI, Correa JC, Marmolejo-Ramos F. A new approach of the Box Cox transformation. Frontiers in Applied Mathmatics and Statistics. 2015;**1**(12):1-10. DOI: 10.3389/fams.2015.00012

[12] Asar Ö, Ilk O, Dag O. Estimating Box-Cox power transformation parameter via goodness of fit tests. Communications in Statistics - Simulation and Computation. 2017;**46**(1):91-105. DOI: 10.1080/03610918.2014.957839

[13] Box GEP, Cox DR. An Analysis of Transformations. Journal of the Royal Statistical Society. Series B (Methodological). 1964;**26**(2):211-252. DOI: 10.1111/j.2517-6161.1964.tb00553.x

[14] Alyousif HT, Abduahad FN. Develop a nonlinear model for the conditional expectation of the Bayesian probability distribution (Gamma – Gamma). Al-Nahrain Journal of Science. 2018;**17**(2):205-212 Available from: https://anjs.edu.iq/index.php/anjs/article/view/462/408

[15] Al-Saffar A, Mohammed Ali HT. Using power transformations in response surface methodology. In: 2022 International Conference on Computer Science and Software Engineering (CSASE), Iraq: IEEE; 2022. pp. 374-379. DOI: 10.1109/CSASE51777.2022.9759781

[16] Tukey JW. Dyadic ANOVA, an analysis of variance for vectors. Human Biology. 1950;**21**:65-110

[17] Box GEP, Tidwell PW. Transformation of the independent variables. Technometrics. 1962;**4**:531-550

[18] Tukey JW. One degree of freedom for non-additivity. Biometrics. 1949; **5**(3):232-242

[19] Velez JI, Marmolejo RF. A new approach to the Box-Cox transformation. Frontiers in Applied Mathematics and Statistics. 2015;**1**(12):1-10. DOI: 10.3389/fams.2015.00012

[20] Chen G, Lockhart RA, Stephens MA. Box–Cox transformations in linear models: large sample theory and tests of normality. Canadian Journal of Statistics. 2002;**30**(2):1-59. DOI: 10.2307/3315946

[21] Draper NR, Smith H. Applied Regression Analysis. NY: John Wiley and Sons Inc.; 1981

[22] Atkinson AC, Riani M, Corbellini A. The box-cox transformation: Review and extensions. Statistical Science. 2021;**36** (2):239-255. DOI: 10.1214/20-STS778

[23] Finney DJ. The Principles of Biological Assay. Supplement to the Journal of the Royal Statistical Society. 1947;**9**(1):46-81. DOI: 10.2307/2983571

[24] Tukey JW. On the comparative anatomy of Transformations. The Annals of Mathematical Statistics. 1957; **28**(3):602-632

[25] Yeo IK, Johnson RA. A new family of power transformations to improve normality or symmetry. Biometrika. 2000;**56**(I):87-90. DOI: 10.1093/biomet/ 87.4.954

[26] Samira, S. Exact Box-Cox analysis. Electronic thesis and dissertation repository. 2018. Available from: https:// ir.lib.uwo.ca/etd/5308

[27] Cook RD, Weisberg S. Residuals and Influence in Regression. New York: Chapman and Hall; 1982