

Unbiased Fitting versus Functional Inversion in Square-Root Models

D. Sarala¹ and Martin L. William²

¹Department of Statistics, Presidency College
Chennai 600005, India
Email : mercy.sarala@gmail.com

²Department of Statistics, Loyola College
Chennai 600034, India
Email : martin_lw@yahoo.com

ABSTRACT

The question of fitting appropriate theoretical values for the response variable from a regression model is a classical one. When a model is constructed by making a transformation on the dependent variable to achieve stable variance and normality, the fitting of theoretical values for the 'original' dependent variable is generally obtained by a simple functional inversion of the transformation. However, these theoretical values are not unbiased and nothing is known about their properties. A functional inversion of the transformation appears functionally proper, but it is statistically inappropriate. In this paper, we address the question of obtaining unbiased theoretical value of an observation on the original (untransformed) dependent variable that has been subjected to the square-root transformation for model building. A comparison of the 'functional inversion fitting' approach with the proposed 'statistical fitting' (unbiased fitting) approach is carried out under a squared-error risk function and the latter is shown to be the preferred one with reduced mean-square risk. Extension to fitting of theoretical values simultaneously for multiple observations under a matrix mean-square risk is also presented.

Keywords: Functional Inversion; Square-root Transformation; Statistical Fit

2000 Mathematics Subject Classification: 62F99, 62J99

1. INTRODUCTION

In fitting of linear regression models, the constant variance condition is required to apply the ordinary least squares technique. Normality of observations is required to perform tests and to obtain confidence intervals. Transformation is sometimes resorted to in order to bring about constant variance / normality of the transformed response variable. We refer to the classical paper of Box and Cox (1964) for analytical discussion on transformations of variables. An early work on prediction of untransformed observations from a model built by power transformation of the dependent variable is a paper by Carroll and Ruppert (1981). Duan (1983) proposed the notion of a 'smearing estimate', a nonparametric estimate of the expected response in the untransformed scale, after fitting a linear model on a transformed scale. Taylor (1985) discussed the properties of the measure of location of the original variable obtained by applying the functional inversion to the measure of location of the Box-Cox power transformed variable. In another paper, Taylor (1986) discussed an approximate method to estimate the mean of a dependent variable after a linear model is fitted to the Box-Cox power transformed variable. A review of various techniques for forming point and interval predictions in regression models with Box-Cox transformed variables is found in Collins (1991). Goncalves and Meddahi (2011) considered a new class of transformations based on the Box-Cox transformation.

A recent work on a new family of power transformations to deal with modelling situations where the dependent variable can assume non-positive values is a paper by Yeo and Johnson (2000). Freeman and Modarres (2006) develop the exact expectation and variance of the inverse Box-Cox transformation by Chebyshev-Hermite polynomials. In all these works, the focus is on the method of fitting the expected (theoretical) values of the 'original' (untransformed) dependent variable by a simple functional inversion of the theoretical fit values obtained for the transformed version of the dependent variable. Hitherto, there has been no effort at alternative ways of generating appropriate theoretical fit values of the untransformed dependent variable in regression modelling. It is well known that square-root transformation is generally recommended when the response variable is a Poisson count. However, in recent years, other approaches like GLM have been followed by several authors in modeling count responses. We refer to Astuti *et. al.* (2013) for an interesting work on a Generalized Poisson Regression model for mortality data.

In this paper, we address the question of fitting appropriate theoretical values of the original dependent variable which are founded on strong statistical considerations enjoying lower average error for the situation where the square-root transformation is applied to build the linear regression model under ordinary least squares approach. The conventional process of doing a 'functional inversion' (squaring) to get theoretical values of the original (untransformed) response variable is statistically unsatisfactory as these are biased. We attempt to obtain different theoretical fit values of the dependent variable which are not only unbiased but also more efficient than those fitted by functional inversion.

This paper is structured as follows. Section 2 provides the background of the research problem being addressed, gives a brief about the functionally-inverted fit for any observation on the original dependent variable and specifies the mean-square error criterion for comparing the performances (efficiencies) of different theoretical fits. In Section 3, a new unbiased fit is proposed and the comparison of the mean-square risk of this statistical fit with that of the functionally-inverted fit is carried out. The superiority of the newly proposed fit is established. In Section 4, we give an extension of this approach for simultaneous fitting of multiple observations with respect to the criterion of reducing a matrix mean-square risk. Section 5 reports the results of our simulation study which has been carried out for models with two regressors based on 200 observations each and several values of the regression parameters and the error variance. This is followed by some remarks in the concluding section.

2. BACKGROUND OF THE PROBLEM

Consider a response variable Y which does not satisfy the 'constant variance' / 'normality' conditions but satisfies these requirements when subjected to square-root transformation. Let $X_1, X_2, \dots, X_k$ be the regressors and suppose that a linear regression model is built with $\sqrt{Y}$ as the dependent variable using OLS principle. That is, we build the following linear model using a dataset of 'n' observations:

$$Y_i^* = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + U_i, i = 1, 2, \dots, n \quad (2.1)$$

where $Y_i^* = \sqrt{Y_i}$. The situation we consider here is one where this model satisfies the conditions of constant variance and normality of errors. That is, $E(U_i) = 0$, $\text{Var}(U_i) = \sigma^2$ for $i = 1, 2, \dots, n$ and $\text{Cov}(U_i, U_j) = 0$ for $i \neq j$.

If $\hat{Y}_i$ is any fit for the observation Y_i , the mean-square risk of $\hat{Y}_i$ is given by

$$M(\hat{Y}_i) = E [\hat{Y}_i - Y_i]^2 \quad (2.2)$$

The mean-square risk is a performance measure that is suited to compare biased fits and unbiased fits as well. The OLS approach gives the estimates for the regression parameters and using these estimates, the following theoretical fit of Y_i^* is usually considered:

$$\hat{Y}_i^* = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \dots + \hat{\beta}_k X_{ki} \quad (2.3)$$

For fitting the original (untransformed) response variable Y_i , the following fit is taken according to the convention of inverting a transformation:

$$\hat{Y}_{i(oid)} = (\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \dots + \hat{\beta}_k X_{ki})^2 \quad (2.4)$$

As the transformation involves the square-root operation, the functional inversion involves the squaring operation. The above $\hat{Y}_{i(oid)}$ is a biased fit for Y_i . While biased fits do perform well at times, our objective is to provide an unbiased fit and compare it with $\hat{Y}_{i(oid)}$ for their mean squared-error risks.

3. A NEW STATISTICAL FIT FOR A SINGLE OBSERVATION AND COMPARISON WITH FUNCTIONAL INVERSION FIT

With the notation introduced in the previous section, we have

$$\begin{aligned} E(Y_i) &= E\{Y_i^{*2}\} = \{E(Y_i^*)\}^2 + \text{Var}(Y_i^*) \\ &= \{\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}\}^2 + \sigma^2 \\ &= \{C_i^T \beta\}^2 + \sigma^2 \end{aligned}$$

where $C_i = (1, X_{1i}, \dots, X_{ki})^T$, the i^{th} row of the data matrix X , and $\beta = (\beta_0, \beta_1, \dots, \beta_k)^T$

Now, $E(\hat{Y}_{i(oid)}) = \{C_i^T \beta\}^2 + \sigma^2 C_i^T (X^T X)^{-1} C_i \neq E(Y_i)$ showing that $\hat{Y}_{i(oid)}$ is a biased fit for Y_i .

While it is true that biased estimates sometimes do perform well over unbiased estimates, we take the view that the fit value be unbiased. Towards this, we propose the following unbiased fit for Y_i :

$$\begin{aligned} \tilde{Y}_{i(new)} &= (\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \dots + \hat{\beta}_k X_{ki})^2 + \hat{\sigma}^2 (1 - C_i^T (X^T X)^{-1} C_i) \\ &= \hat{Y}_{i(oid)} + \hat{\sigma}^2 (1 - h_{ii}) \end{aligned} \quad (3.1)$$

where $h_{ii} = C_i^T (X^T X)^{-1} C_i = (i, i)^{th}$ element of the 'Hat' matrix $X (X^T X)^{-1} X^T$ and $\hat{\sigma}^2 = \text{Mean-Square Error of the model built with } Y_i^* \text{ as the dependent variable.}$

We know that $\hat{\sigma}^2$ is an unbiased estimate of the error variance σ^2 of the model (2.1). It is easily seen that $E(\tilde{Y}_{i(new)}) = E(Y_i)$ showing that $\tilde{Y}_{i(new)}$ is an unbiased fit for Y_i .

We shall call $\hat{Y}_{i(oid)}$ the functionally inverted fit and $\tilde{Y}_{i(new)}$ the statistical fit.

The mean-square risks (expected loss or error) involved in using $\hat{Y}_{i(oid)}$ and $\tilde{Y}_{i(new)}$ as theoretical fits for Y_i are given by $M(\hat{Y}_{i(oid)}) = E [\hat{Y}_{i(oid)} - Y_i]^2$ and $M(\tilde{Y}_{i(new)}) = E [\tilde{Y}_{i(new)} - Y_i]^2$. In the sequel,

we show that $\tilde{Y}_{i(new)}$ is preferable over $\hat{Y}_{i(old)}$ as theoretical fit of Y_i in terms of reduced mean-square risk.

Theorem 3.1: $M(\tilde{Y}_{i(new)}) \leq M(\hat{Y}_{i(old)})$

Proof: We have the relation

$$\begin{aligned} M(\tilde{Y}_{i(new)}) &= M(\hat{Y}_{i(old)}) + 2(1 - h_{ii}) E[\hat{\sigma}^2 (\hat{Y}_{i(old)} - Y_i)] + (1 - h_{ii})^2 E[\hat{\sigma}^4] \\ &= M(\hat{Y}_{i(old)}) + 2(1 - h_{ii}) E[\hat{\sigma}^2 (\hat{Y}_i^{*2} - Y_i^{*2})] + (1 - h_{ii})^2 E[\hat{\sigma}^4] \\ &= M(\hat{Y}_{i(old)}) - 2(1 - h_{ii}) E[\hat{\sigma}^2 (Y_i^* - \hat{Y}_i^*)^2] + 2\hat{\sigma}^2 \hat{Y}_i^* (Y_i^* - \hat{Y}_i^*) + (1 - h_{ii})^2 E[\hat{\sigma}^4] \\ &= M(\hat{Y}_{i(old)}) - 2(1 - h_{ii}) E[\hat{\sigma}^2 (Y_i^* - \hat{Y}_i^*)^2] - 4(1 - h_{ii}) E[\hat{\sigma}^2 \hat{Y}_i^* (Y_i^* - \hat{Y}_i^*)] + (1 - h_{ii})^2 E[\hat{\sigma}^4] \\ &= M(\hat{Y}_{i(old)}) - 2(1 - h_{ii}) E[\hat{\sigma}^2 e_i^{*2}] - 4(1 - h_{ii}) E[\hat{\sigma}^2 \hat{Y}_i^* e_i^*] + (1 - h_{ii})^2 E[\hat{\sigma}^4] \end{aligned} \quad (3.2)$$

where $e_i^* = Y_i^* - \hat{Y}_i^*$ is the residual for the i^{th} observation obtained from the model built with Y^* as the dependent variable.

Denote $E[\hat{Y}_i^* e_i^* | \hat{\sigma}^2] = g(\hat{\sigma}^2)$. Then $E[g(\hat{\sigma}^2)] = E[\hat{Y}_i^* e_i^*] = 0$ by a property of residuals in a linear model. Using the fact that the family of distributions of $\hat{\sigma}^2$ is the chi-square family which is complete, we see that $g(\hat{\sigma}^2) = 0$.

$$\text{Now, } E[\hat{\sigma}^2 \hat{Y}_i^* e_i^*] = E[\hat{\sigma}^2 E(\hat{Y}_i^* e_i^* | \hat{\sigma}^2)] = E[\hat{\sigma}^2 g(\hat{\sigma}^2)] = 0 \quad (3.3)$$

Considering $E[e_i^{*2} - \hat{\sigma}^2(1 - h_{ii})] = \sigma^2(1 - h_{ii}) - \sigma^2(1 - h_{ii}) = 0$, we get

$$E[E[e_i^{*2} - \hat{\sigma}^2(1 - h_{ii}) | \hat{\sigma}^2]] = 0$$

which, on using the completeness of the family of distributions of $\hat{\sigma}^2$, yields

$$E[e_i^{*2} - \hat{\sigma}^2(1 - h_{ii}) | \hat{\sigma}^2] = 0$$

Thus, we get $E[e_i^{*2} | \hat{\sigma}^2] = \hat{\sigma}^2(1 - h_{ii})$ so that $E[\hat{\sigma}^2 e_i^{*2} | \hat{\sigma}^2] = \hat{\sigma}^4(1 - h_{ii})$ and hence

$$E[\hat{\sigma}^2 e_i^{*2}] = E[E[\hat{\sigma}^2 e_i^{*2} | \hat{\sigma}^2]] = (1 - h_{ii}) E[\hat{\sigma}^4] \quad (3.4)$$

Using (3.3) and (3.4) in (3.2), we get

$$\begin{aligned} M(\tilde{Y}_{i(new)}) &= M(\hat{Y}_{i(old)}) - 2(1 - h_{ii})^2 E[\hat{\sigma}^4] + (1 - h_{ii})^2 E[\hat{\sigma}^4] \\ &= M(\hat{Y}_{i(old)}) - (1 - h_{ii})^2 E[\hat{\sigma}^4] \\ &\leq M(\hat{Y}_{i(old)}) \end{aligned}$$

This shows that for the i^{th} observation of the original response variable Y_i , the statistical fit $\tilde{Y}_{i(new)}$ proposed in (3.1) is superior to the traditional functionally inverted fit $\hat{Y}_{i(old)}$ given in (2.3).

4. SIMULTANEOUS FITTING OF MULTIPLE OBSERVATIONS

In this section we present the case of obtaining theoretical fit for multiple observations; i.e. for a subset of the available observations. Let the subset of 'm' observations for which theoretical fit are to be obtained be denoted as Y_S . If $\hat{Y}_S$ be an m-dimensional vector fit of Y_S , then the matrix mean-square risk of $\hat{Y}_S$ is given by

$$M[\hat{Y}_S] = E[(\hat{Y}_S - Y_S) \cdot (\hat{Y}_S - Y_S)^T]$$

Let us suppose that the subset 's' contains the observation labels $i_1, i_2, \dots, i_m$. That is, the vector of observations to be fitted is $\mathbf{Y}_S = (Y_{i_1}, Y_{i_2}, \dots, Y_{i_m})^T$. We shall denote the vector of square-root transformed observations as $\mathbf{Y}_S^* = (Y_{i_1}^*, Y_{i_2}^*, \dots, Y_{i_m}^*)^T$.

According to the convention of functional inversion, the following is generally considered as the theoretical fit for $\mathbf{Y}_S$:

$$\hat{\mathbf{Y}}_{S(oid)} = (\hat{Y}_{i_1(oid)}, \hat{Y}_{i_2(oid)}, \dots, \hat{Y}_{i_m(oid)})^T$$

where we refer to (2.4) for $\hat{Y}_{i(oid)}$.

Note that $\hat{\mathbf{Y}}_{S(oid)} = \hat{\mathbf{Y}}_S^* \circ \hat{\mathbf{Y}}_S^*$ where the symbol $\circ$ denotes 'Hadamard' product of vectors and

$$\hat{\mathbf{Y}}_S^* = (\hat{Y}_{i_1}^*, \hat{Y}_{i_2}^*, \dots, \hat{Y}_{i_m}^*)^T$$

where we refer to (2.3) for $\hat{Y}_i^*$. Denoting $\mathbf{Y}_S^* - \hat{\mathbf{Y}}_S^* = \mathbf{e}_S^*$ we now propose the following statistical fit which is unbiased:

$$\tilde{\mathbf{Y}}_{S(new)} = (\tilde{Y}_{i_1(new)}, \tilde{Y}_{i_2(new)}, \dots, \tilde{Y}_{i_m(new)})^T$$

where we refer to (3.1) for $\tilde{Y}_{i(new)}$.

Denoting $\mathbf{1}_S = (1, 1, \dots, 1)^T$, $\mathbf{0}_S = (0, 0, \dots, 0)^T$, both $m \times 1$ vectors, and $\mathbf{h}_S = (h_{i_1 i_1}, h_{i_2 i_2}, \dots, h_{i_m i_m})^T$, we note that

$$\tilde{\mathbf{Y}}_{S(new)} = \hat{\mathbf{Y}}_{S(oid)} + \hat{\sigma}^2 (\mathbf{1}_S - \mathbf{h}_S) \quad (4.1)$$

We shall now establish that $\tilde{\mathbf{Y}}_{S(new)}$ is preferable to $\hat{\mathbf{Y}}_{S(oid)}$ as a theoretical fit for the vector $\mathbf{Y}_S$, by showing that $\mathbf{M} [\tilde{\mathbf{Y}}_{S(new)}] \leq \mathbf{M} [\hat{\mathbf{Y}}_{S(oid)}]$ where the statement $A \leq B$ is to be understood as 'B - A is a non-negative definite matrix'.

Theorem 4.1: $\mathbf{M} [\tilde{\mathbf{Y}}_{S(new)}] \leq \mathbf{M} [\hat{\mathbf{Y}}_{S(oid)}]$

$$\begin{aligned} \text{Proof: } \mathbf{M} [\tilde{\mathbf{Y}}_{S(new)}] &= E[(\tilde{\mathbf{Y}}_{S(new)} - \mathbf{Y}_S) \cdot (\tilde{\mathbf{Y}}_{S(new)} - \mathbf{Y}_S)^T] \\ &= E[(\hat{\mathbf{Y}}_{S(oid)} + \hat{\sigma}^2 (\mathbf{1}_S - \mathbf{h}_S) - \mathbf{Y}_S) \cdot (\hat{\mathbf{Y}}_{S(oid)} + \hat{\sigma}^2 (\mathbf{1}_S - \mathbf{h}_S) - \mathbf{Y}_S)^T] \\ &= \mathbf{M} [\hat{\mathbf{Y}}_{S(oid)}] + (\mathbf{1}_S - \mathbf{h}_S) E [\hat{\sigma}^2 \cdot (\hat{\mathbf{Y}}_{S(oid)} - \mathbf{Y}_S)^T] \\ &\quad + E[(\hat{\sigma}^2 (\hat{\mathbf{Y}}_{S(oid)} - \mathbf{Y}_S)) (\mathbf{1}_S - \mathbf{h}_S)^T] + E[\hat{\sigma}^4] (\mathbf{1}_S - \mathbf{h}_S)(\mathbf{1}_S - \mathbf{h}_S)^T \end{aligned} \quad (4.2)$$

$$\text{Consider } E[(\hat{\sigma}^2 (\hat{\mathbf{Y}}_{S(oid)} - \mathbf{Y}_S))] = E[(\hat{\sigma}^2 (\hat{\mathbf{Y}}_S^* \circ \hat{\mathbf{Y}}_S^* - \mathbf{Y}_S^* \circ \mathbf{Y}_S^*))]$$

$$\begin{aligned} &= -E[\hat{\sigma}^2 ((\mathbf{Y}_S^* - \hat{\mathbf{Y}}_S^*) \circ (\mathbf{Y}_S^* - \hat{\mathbf{Y}}_S^*)) + 2\hat{\sigma}^2 \hat{\mathbf{Y}}_S^* \circ (\mathbf{Y}_S^* - \hat{\mathbf{Y}}_S^*)] \\ &= -E[\hat{\sigma}^2 e_S^* \circ e_S^*] - 2E[\hat{\sigma}^2 \hat{\mathbf{Y}}_S^* \circ e_S^*] \\ &= -E[\hat{\sigma}^2 e_S^* \circ e_S^*] - 2E[\hat{\sigma}^2 E[\hat{\mathbf{Y}}_S^* \circ e_S^* | \hat{\sigma}^2]] \end{aligned} \quad (4.3)$$

Denote $E[\hat{\mathbf{Y}}_S^* \circ e_S^* | \hat{\sigma}^2] = G_S(\hat{\sigma}^2)$. Then $E[G_S(\hat{\sigma}^2)] = E[\hat{\mathbf{Y}}_S^* \circ e_S^*] = \mathbf{0}_S$ by a property of residuals.

Using the fact that the family of distributions of $\hat{\sigma}^2$ is complete, we see that $G_S(\hat{\sigma}^2) = \mathbf{0}_S$ and so,

$$E[\hat{\sigma}^2 E[\hat{\mathbf{Y}}_S^* \circ e_S^* | \hat{\sigma}^2]] = \mathbf{0}_S \quad (4.4)$$

Considering $E[e_S^* \circ e_S^* - \hat{\sigma}^2 (\mathbf{1}_S - \mathbf{h}_S)] = \sigma^2 (\mathbf{1}_S - \mathbf{h}_S) - \sigma^2 (\mathbf{1}_S - \mathbf{h}_S) = \mathbf{0}_S$, we get

$$E [E [e_S^* \circ e_S^* - \hat{\sigma}^2 (\mathbf{1}_S - \mathbf{h}_S) | \hat{\sigma}^2]] = \mathbf{0}_S$$

which on using the completeness of the family of distributions of $\hat{\sigma}^2$, yields

$$E [e_S^* \circ e_S^* - \hat{\sigma}^2 (\mathbf{1}_S - \mathbf{h}_S) | \hat{\sigma}^2] = \mathbf{0}_S$$

This gives $E [e_S^* \circ e_S^* | \hat{\sigma}^2] = \hat{\sigma}^2 (\mathbf{1}_S - \mathbf{h}_S)$ so that $E [\hat{\sigma}^2 e_S^* \circ e_S^* | \hat{\sigma}^2] = \hat{\sigma}^4 (\mathbf{1}_S - \mathbf{h}_S)$ and hence

$$E [\hat{\sigma}^2 e_S^* \circ e_S^*] = E [E [\hat{\sigma}^2 e_S^* \circ e_S^* | \hat{\sigma}^2]] = E [\hat{\sigma}^4] (\mathbf{1}_S - \mathbf{h}_S) \quad (4.5)$$

Using (4.4) and (4.5) in (4.3) gives

$$E[(\hat{\sigma}^2 (\hat{\mathbf{Y}}_{S(old)} - \mathbf{Y}_S))] = -E[\hat{\sigma}^4] (\mathbf{1}_S - \mathbf{h}_S) \quad (4.6)$$

which on using in (4.2) gives

$$\begin{aligned} \mathbf{M} [\tilde{\mathbf{Y}}_{S(new)}] &= \mathbf{M} [\hat{\mathbf{Y}}_{S(old)}] - (\mathbf{1}_S - \mathbf{h}_S) E [\hat{\sigma}^4] (\mathbf{1}_S - \mathbf{h}_S)^T \\ &\quad - E [\hat{\sigma}^4] (\mathbf{1}_S - \mathbf{h}_S) (\mathbf{1}_S - \mathbf{h}_S)^T + E [\hat{\sigma}^4] (\mathbf{1}_S - \mathbf{h}_S) (\mathbf{1}_S - \mathbf{h}_S)^T \\ &= \mathbf{M} [\hat{\mathbf{Y}}_{S(old)}] - E [\hat{\sigma}^4] (\mathbf{1}_S - \mathbf{h}_S) (\mathbf{1}_S - \mathbf{h}_S)^T \end{aligned}$$

$$\text{Therefore, } \mathbf{M} [\hat{\mathbf{Y}}_{S(old)}] - \mathbf{M} [\tilde{\mathbf{Y}}_{S(new)}] = E [\hat{\sigma}^4] (\mathbf{1}_S - \mathbf{h}_S) (\mathbf{1}_S - \mathbf{h}_S)^T$$

where the matrix on the right-hand side is clearly a non-negative definite matrix.

Hence, $\mathbf{M} [\tilde{\mathbf{Y}}_{S(new)}] \leq \mathbf{M} [\hat{\mathbf{Y}}_{S(old)}]$ as required.

This shows that for the subset of observations $\mathbf{Y}_S = (Y_{i_1}, Y_{i_2}, \dots, Y_{i_m})^T$ of the original response variable, the statistical fit $\tilde{\mathbf{Y}}_{S(new)}$ proposed in (4.1) is superior to the traditional fit $\hat{\mathbf{Y}}_{S(old)}$ that is provided by the functional inversion approach.

5. SIMULATION STUDY

To strengthen the theoretical result of Section 3, we carry out a simulation study to establish the superior performance of the new 'statistical' fit over the conventional functional inversion for fitting the theoretical values to the response variable.

We have simulated 100 datasets each containing 200 observations on two randomly generated regressor variables (say, X_1 and X_2) and an independently generated normal error term, by varying the regression parameters and the error variance over a 'good' range of values. We have considered a range of positive values from 0 to 3.5 for the intercept parameter β_0 and the slope parameters β_1 and β_2 and a range of positive values from 0 to 16 for the error variance σ^2 and generated the 'observations' on the dependent variable Y . Specifically, we have generated $Y^* (= \sqrt{Y})$ through the equations

$$Y_i^* = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + U_i$$

where the error terms U_i are independently generated from $N(0, \sigma^2)$.

A small sample of 20 records from the 200 generated in one of the 100 simulated datasets (with $\beta_0 = 0.5$, $\beta_1 = 1.0$, $\beta_2 = 1.5$ and $\sigma^2 = 1.0404$) is presented in Table 1. The OLS procedure for estimation of the parameters gives the estimates: $\hat{\beta}_0 = 0.8884$, $\hat{\beta}_1 = 0.476$, $\hat{\beta}_2 = 1.4133$. The fitted values for Y_i^* are obtained as $\hat{Y}_i^* = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i}$.

The fitted values of the 'original' response variable 'Y' by the conventional functional inversion approach [that is, $\hat{Y}_{i(old)} = \hat{Y}_i^{*2}$] and the newly proposed statistical fit [$\tilde{Y}_{i(new)}$] are reported in the following Table 1.

Table 1: A sample of records and theoretical values fitted by the two methods

X1	X2	U-Error	Yi*	Yi	(Yi*)sq	Yi~ new
0.23	0.82	-0.2436	1.7164	2.9460	4.6519	5.4821
0.35	0.49	-0.1450	1.4400	2.0736	3.0539	3.8919
0.61	0.02	0.0397	1.1797	1.3916	1.4569	2.2838
0.16	0.87	1.1309	3.0959	9.5846	4.8144	5.6403
0.71	0.33	1.3935	3.0985	9.6008	2.8655	3.7024
0.09	0.04	0.2546	0.9046	0.8183	0.9757	1.7927
0.96	0.54	-0.9524	1.3176	1.7360	4.4462	5.2775
0.32	0.62	0.9121	2.6621	7.0865	3.6749	4.5117
0.76	0.87	-1.0372	1.5278	2.3341	6.1494	6.9806
0.88	0.83	-0.0939	2.5311	6.4063	6.1523	6.9816
0.14	0.34	0.3097	1.4597	2.1306	2.0609	2.8906
0.34	0.2	0.4716	1.6116	2.5972	1.7767	2.6091
0.89	0.94	0.4160	3.2160	10.3425	6.9728	7.7975
0.89	0.44	0.2200	2.2700	5.1530	3.7401	4.5739
0.19	0.56	0.0098	1.5398	2.3708	3.1340	3.9674
0.37	0.97	0.4361	2.7611	7.6235	5.9315	6.7594
0.6	0.34	-0.7201	0.8899	0.7919	2.7375	3.5758
0.06	0.17	-0.5086	0.3064	0.0939	1.3392	2.1603
0.03	0.54	0.8116	2.1516	4.6294	2.7752	3.6013
0.48	0.43	-1.4054	0.2196	0.0482	2.9743	3.8138

For each of the 100 samples generated, we compute the Average Square-error for both the conventional functional inversion approach and the newly suggested statistical approach of fitting the original (untransformed) response variable. The 'mean square errors' for both methods of fitting obtained for a subset of 10 of the 100 simulated datasets are summarized in Table 2 below. It is found from the table that, on an average, the new statistical fitting approach gives closer fit than the conventional functional inversion with respect to the average squared error criterion. We observe that in 76 of the 100 simulated samples, the statistical-fitting approach outperforms the functional inversion approach, thereby providing an empirical support to the theoretical results established in Section 3 and Section 4.

Table 2: Comparison of residuals from the two methods for ten of the simulated datasets

Dataset #	1	2	3	4	5	6	7	8	9	10
Sq.res.old	16.771	26.823	39.866	28.017	149.753	134.161	109.239	122.431	277.356	263.546
Sq.res.new	16.074	25.825	38.062	26.745	140.633	125.572	102.455	114.856	261.399	249.602

6. CONCLUDING REMARKS

The present paper gives a new approach to obtain unbiased theoretical fits to the observations on a response variable whose square-root transformation is used for model-building. The square-root transformation is necessitated in situations where the original response variable does not satisfy the equal variance and normality requirements while the square-root transformation of the dependent variable satisfies these conditions. An instance for this is a response variable which is a Poisson count. It is noted that Poisson count is not the only response variable that calls for square-root transformation to be applied for building a linear regression model. The Box-Cox approach to build a model after transforming the variable may be initially invoked and if the square-root transformation is recommended (even approximately) as the suitable one, the theory presented in this paper is applicable. From the theory developed in this paper, we find that the statistical approach of obtaining unbiased fit for the 'original' response variable is a champion over the conventional mathematical approach of taking functional inversion.

The related problem of prediction of new (unknown / future) responses on the original response variable from a square-root model corresponding to a set of values of the predictor / explanatory variables not observed in the available data set is being addressed separately and will be communicated in another article.

7. REFERENCES

- Astuti, E. T., Budiantara, I. N., Sunaryo, S. and Dokhi, M. 2013. Statistical Modeling for Mortality Data using Local Generalized Poisson Regression Model, *International Journal of Applied Mathematics and Statistics*, **33(3)**, 92-101.
- Box, G.E.P. and Cox, D.R., 1964, An Analysis of transformations. *Journal of the Royal Statistical Society, Series B*, **26(2)**, 211-252
- Carroll, R. J., and Ruppert, D. 1981. On prediction and the power transformation family. *Biometrika*, **68(3)**, 609-615
- Collins, S., 1991, Prediction Techniques for Box-Cox Regression Models. *Journal of Business and Economic Statistics*, **9(3)**, 267-277
- Goncalves, S., 2011, And Meddahi, N. Box-Cox transforms for realized volatility, *Journal of Econometrics*, **160**, 129-144
- Duan, N., 1983, Smearing Estimate: A nonparametric retransformation method. *Journal of the American Statistical Association*, **78(383)**, 605-610
- Freeman, J. And Modarres, R. 2006, Inverse Box-Cox: The power normal distribution. *Statistics and Probability Letters*, **76**, 764-772
- Taylor, J.M.G., 1985, Measures of location of skew distributions obtained through Box-Cox transformations. *Journal of the American Statistical Association*, **80(390)**, 427-432
- Taylor, J.M.G., 1986, The retransformed mean after a fitted power transformation. *Journal of the American Statistical Association*, **81(393)**, 114-118
- Yeo, I. W. And Johnson, R. A., 2000, A new family of power transformations to improve Normality or symmetry. *Biometrika*, **87(4)**, 954-959.