國 立 交 通 大 學

統 計 學 研 究 所

碩 士 論 文

Likelihood Inference under the Transformed Truncated Normal Mode Regression Model

研 究 生：沈彥廷

指導教授：陳志榮 博士

中 華 民 國 九 十 九 年 六 月

變換截常態眾數迴歸模型的概似推論

Likelihood Inference under the Transformed Truncated Normal Mode Regression Model

研 究 生：沈彥廷　　Student：Yen-Tin Sen

指導教授：陳志榮　博士　　Advisor：Dr. Chih-Rung Chen

國 立 交 通 大 學
統 計 學 研 究 所
碩 士 論 文

A Thesis
Submitted to Institute of Statistics
College of Science
National Chiao Tung University
In Partial Fulfillment of the Requirements
for the Degree of Master
in
Statistics
June 2010
Hsinchu, Taiwan

中 華 民 國 九 十 九 年 六 月

變換截常態眾數迴歸模型的概似推論

學生：沈彥廷　　　　　　　　指導教授：陳志榮 博士

國立交通大學統計學研究所

摘要

當一組資料經過變換以後，其值域有可能不包含某些實數；在此情況下，變換後的資料不可能滿足傳統的常態假設。因此我們提出一個變換截常態眾數迴歸模型及其概似推論，然後應用到兩組實際的資料中，並與傳統的常態假設做比較。最後比較變換截常態眾數、平均數和中位數三種不同迴歸模型的計算複雜性。

Likelihood Inference under the Transformed Truncated Normal Mode Regression Model

Student: Yen-Tin Sen Advisor: Dr. Chih-Rung Chen

Institute of Statistics

National Chiao Tung University

Abstract

In the paper, the likelihood inference under the transformed truncated normal mode regression model is proposed when the range of the transformation is possibly different from the whole real line. The proposed methodology is then applied to two real data sets in Box and Cox (1964), where the truncated normality assumption is compared with the conventional normality assumption. Finally, the proposed model is compared with the transformed truncated normal mean and median regression models via the computational complexity.

誌　謝

我最需要感謝的，是陳志榮老師的指導，有老師的幫忙才能順利完成這篇論文，而且老師對於其學生付出的心力，遠比想像中還要多很多，從老師身上可以感覺出對研究的熱忱，除了學到解決問題的能力之外，更了解處理事情應該要細心、謹慎；會以老師為榜樣，在未來的日子，懷抱熱情並以求知的精神走下去。感謝兩年來所上老師們的教導，同時也要謝謝所上行政人員，在剛來統研所這個新環境時的幫助，還有同學們在課業上互相學習、競爭，這對我助益良多。很高興能在統研所和老師、同學一起度過兩年。

最後，感謝父母給於生活上的安定，讓我可以專心於學業上，不必為其他事操煩，有父母的付出、老師的教導、同學們的關愛讓我能順順利利完成學業。

沈 彥 廷 謹至于

國 立 交 通 大 學 統 計 學 研 究 所

中華民國九十九年六月

Contents

1 Introduction **1**

2 Transformed Truncated Normal Mode Regression Model **4**

2.1 Transformed Truncated Normal Mode Regression Model 4

2.2 Maximum Likelihood Estimation 4

2.3 Hypothesis Testing and Confidence Regions 7

2.4 Prediction Region of Future Observations 8

3 Two Real Data Sets **11**

3.1 A Biological Experiment Using a 3×4 Factorial Design 11

3.2 A Textile Experiment Using a Single Replicate of a 3^3 Design 13

4 Conclusions and Discussion **16**

List of Tables

1 Survival times (1 unit = 10 hours) of animals in a 3×4 factorial experiment. 29
2 MLEs under the false normality assumption and the truncated normality assumption, respectively, for Example 3.1. 30
3 MLEs without interaction under the false normality assumption and the truncated normality assumption, respectively, for Example 3.1. 31
4 MLEs with $\lambda = -1$ under the false normality assumption and the truncated normality assumption, respectively, for Example 3.1. 32
5 MLEs without interaction and with $\lambda = -1$ under the false normality assumption and the truncated normality assumption, respectively, for Example 3.1. 33
6 Cycles to failure of worsted yarn: 3^3 factorial experiment without replication. 34
7 MLEs under the false normality assumption and the truncated normality assumption, respectively, for Example 3.2. 35
8 MLEs without quadratic terms under the false normality assumption and the truncated normality assumption , respectively, for Example 3.2. 36
9 MLEs with $\lambda = 0$ under the false normality assumption and the truncated normality assumption , respectively, for Example 3.2. 37
10 MLEs with $\lambda = 0$ and without quadratic terms under the false normality assumption and the truncated normality assumption, respectively, for Example 3.2. 38

List of Figures

1 Some different modified power transformations. 39

2 (a) Residual plot against fitted values for the original data under the two-way ANOVA effects model for Example 3.1. (b) Residual plot against fitted values for the transformed data under the Box-Cox transformed mode regression model for Example 3.1. 40

3 (a) Normal probability plot under the false normality assumption for Example 3.1. (b) Normal probability plot under the truncated normality assumption for Example 3.1. 40

4 (a) Normal probability plot under the false normality assumption without interactions for Example 3.1. (b) Normal probability plot under the truncated normality assumption without interactions for Example 3.1. . . 41

5 (a) Normal probability plot under the false normality assumption with $\lambda = -1$ for Example 3.1. (b) Normal probability plot under the truncated normality assumption with $\lambda = -1$ for Example 3.1. 41

6 (a) Normal probability plot under the false normality assumption without interactions and with $\lambda = -1$ for Example 3.1. (b) Normal probability plot under the truncated normality assumption without interactions and with $\lambda = -1$ for Example 3.1. 42

7 (a) Residual plot against fitted values for the original data under the quadratic regression model for Example 3.2. (b) Residual plot against fitted values for the transformed data under the Box-Cox transformed mode regression model for Example 3.2. 42

8 (a) Normal probability plot under the false normality assumption for Example 3.2. (b) Normal probability plot under the truncated normality assumption for Example 3.2. 43

9 (a) Normal probability plot under the false normality assumption without quadratic effects and interactions for Example 3.2. (b) Normal probability plot under the truncated normality assumption without quadratic effects and interactions for Example 3.2. 43

10 (a) Normal probability plot under the false normality assumption with $\lambda = 0$ for Example 3.2. (b) Normal probability plot under the truncated normality assumption with $\lambda = 0$ for Example 3.2. 44

11 (a) Normal probability plot under the false normality assumption without quadratic effects and interactions and with $\lambda = 0$ for Example 3.2. (b) Normal probability plot under the truncated normality assumption without quadratic effects and interactions and with $\lambda = 0$ for Example 3.2. 44

1 Introduction

The techniques for linear models are justified by assuming simplicity of systematic structure, constancy of error variances, normality of distributions, and independence of responses. In analyzing data which do not satisfy the traditional assumptions for linear models, Tukey (1957) suggested two alternatives: either a new analysis must be devised or the data must be transformed to satisfy the assumptions. If a satisfactory transformation can be found, it is usually easier to use the conventional techniques for linear models to analyze the transformed data than to develop a new method to analyze the original data.

It is common practice simply to assume the following normal regression model:

$$y_i = f(x_i; \beta) + \varepsilon_i \tag{1}$$

for $i = 1, \ldots, n$, where y_i is the response for subject i; x_i is a known covariate vector for subject i; β is an unknown finite-dimensional regression parameter vector; $f(\cdot; \beta)$ is a known regression function for each β, e.g., $f(x_i; \beta) = x_i^T \beta$ or $\exp\{x_i^T \beta\}$; and ε_is are *i.i.d.* $N(0, \sigma^2)$ errors with unknown positive standard deviation σ. Notice that all of the mean, median, and mode of y_i are the same as $f(x_i; \beta)$ for $i = 1, \ldots, n$.

When there exist heteroscedastic errors and/or departures from normality in the data, one possible approach is to transform the data. A widely used family of transformations to transform positive continuous data is the family of modified power transformations

$$u^{(\lambda)} \equiv \begin{cases} (u^\lambda - 1)/\lambda & \text{for } \lambda \neq 0, \\ \log(u) & \text{for } \lambda = 0. \end{cases} \tag{2}$$

Figure 1 shows some different modified power transformations.

In such situations, Box and Cox (1964) proposed the following Box-Cox transformed linear normal regression model:

$$y_i^{(\lambda)} = x_i^T \beta + \varepsilon_i \tag{3}$$

for $i = 1, \ldots, n$, where y_i has support $(0, \infty)$ and λ is an unknown real-valued transformation parameter.

When both heteroscedastic errors and departures from normality cannot be simultaneously removed in the data by any single transformation, Carroll and Ruppert (1988) proposed the following Box-Cox transformed heteroscedastic normal regression model:

$$y_i^{(\lambda)} = f(x_i; \beta) + \varepsilon_i \tag{4}$$

for $i = 1, \ldots, n$, where ε_is are independent errors distributed as $N(0, g^2(f(x_i;\beta), z_i;\gamma)\,\sigma^2)$ such that z_i is a known covariate vector for subject i, e.g., z_i is a known function of x_i; γ is an unknown finite-dimensional parameter vector; $g(\cdot,\cdot;\gamma)$ is a known positive function for each γ, e.g., $g(f(x_i;\beta), z_i;\gamma) = \exp\{f(x_i;\beta)\gamma_1 + z_i^T\gamma_2\}$ with $\gamma \equiv (\gamma_1, \gamma_2^T)^T$; and σ is an unknown positive scale parameter. In Carroll and Ruppert (1988), the constants $1/g^2(f(x_i;\beta), z_i;\gamma)$ are called the true weights. Notice that model (3) is a special case of model (4) when $f(x_i;\beta) = x_i^T\beta$ and $g(f(x_i;\beta), z_i;\gamma) = 1$ for $i = 1, \ldots, n$.

However, $y_i^{(\lambda)} \in (-1/\lambda, \infty)$ for $\lambda > 0$, $(-\infty, \infty)$ ($\equiv \mathcal{R}$) for $\lambda = 0$, and $(-\infty, -1/\lambda)$ for $\lambda < 0$, respectively. Thus, except for $\lambda = 0$, $y_i^{(\lambda)}$ cannot be normally distributed. Hence, Poirier (1978) modified model (3) to the following Box-Cox transformed linear truncated normal mode regression model:

$$y_i^{(\lambda)} = x_i^T\beta + \varepsilon_i \tag{5}$$

for $i = 1, \ldots, n$, where ε_is are independent errors distributed as either $N(0, \sigma^2)$ for $\lambda = 0$ or truncated $N(0, \sigma^2)$ for $\lambda \neq 0$ with unknown positive scale parameter σ. Notice that, for $i = 1, \ldots, n$, $x_i^T\beta$ is the mode of $y_i^{(\lambda)}$ when it is in the support of $y_i^{(\lambda)}$; however, it is generally neither the mean nor median of $y_i^{(\lambda)}$.

In Chen and Wang (2003), three widely used families of transformations with ranges possibly different from $\mathcal{R}$ are reviewed as follows:

Example 1.1 The family of shifted power transformations (Box and Cox, 1964)

$$h(u;\lambda) \equiv (u-a)^{(\lambda)} = \begin{cases} [(u-a)^\lambda - 1]/\lambda & \text{for } \lambda \neq 0, \\ \log(u-a) & \text{for } \lambda = 0, \end{cases} \tag{6}$$

can be used to transform data with known support (a, ∞), where $a \in \mathcal{R}$, e.g., $a = 0$. Then the range $h((a,\infty);\lambda)$ is $(-1/\lambda, \infty)$ for $\lambda > 0$, $\mathcal{R}$ for $\lambda = 0$, and $(-\infty, -1/\lambda)$ for $\lambda < 0$, respectively. Similarly, the family of transformations

$$h(u;\lambda) \equiv -(b-u)^{(\lambda)} = \begin{cases} [1-(b-u)^\lambda]/\lambda & \text{for } \lambda \neq 0, \\ -\log(b-u) & \text{for } \lambda = 0, \end{cases} \tag{7}$$

can be used to transform data with known support $(-\infty, b)$, where $b \in \mathcal{R}$, e.g., $b = 0$. Then the range $h((-\infty,b);\lambda)$ is $(-\infty, 1/\lambda)$ for $\lambda > 0$, $\mathcal{R}$ for $\lambda = 0$, and $(1/\lambda, \infty)$ for $\lambda < 0$, respectively.

Example 1.2 The family of shifted folded power transformations (Mosteller and

Tukey, 1977)

$$h(u;\lambda) \equiv (u-a)^{(\lambda)} - (b-u)^{(\lambda)} = \begin{cases} [(u-a)^{\lambda} - (b-u)^{\lambda}]/\lambda & \text{for } \lambda \neq 0, \\ \log[(u-a)/(b-u)] & \text{for } \lambda = 0, \end{cases} \tag{8}$$

can be used to transform data with known support (a, b), where $-\infty < a < b < \infty$, e.g., $(a, b) = (0, 1)$. Then the range $h((a, b); \lambda)$ is $(-(b-a)^{\lambda}/\lambda, (b-a)^{\lambda}/\lambda)$ for $\lambda > 0$ and $\mathcal{R}$ for $\lambda \leq 0$, respectively.

Example 1.3 The family of shifted modulus power transformations (John and Draper, 1980)

$$\begin{aligned} h(u;\lambda) &\equiv \operatorname{sgn}(u-\lambda_2)(|u-\lambda_2|+1)^{(\lambda_1)} \\ &= \begin{cases} \operatorname{sgn}(u-\lambda_2)[(|u-\lambda_2|+1)^{\lambda_1} - 1]/\lambda_1 & \text{for } \lambda_1 \neq 0, \\ \operatorname{sgn}(u-\lambda_2)\log(|u-\lambda_2|+1) & \text{for } \lambda_1 = 0, \end{cases} \end{aligned} \tag{9}$$

can be used to transform data with support $\mathcal{R}$, where $\lambda \equiv (\lambda_1, \lambda_2)^T$ and $\operatorname{sgn}(u) = 1$ for $u > 0$, 0 for $u = 0$, and -1 for $u < 0$, respectively. Then the range $h(\mathcal{R}; \lambda)$ is $\mathcal{R}$ for $\lambda_1 \geq 0$ and $(1/\lambda_1, -1/\lambda_1)$ for $\lambda_1 < 0$, respectively.

In order to cover such kinds of families in Examples 1–3, Chen and Wang (2003) modified model (4) to the following transformed truncated normal median regression model:

$$h(y_i; \lambda) = f(x_i; \beta) + \varepsilon_i \tag{10}$$

for $i = 1, \ldots, n$, where y_i has known support (a, b) $(\subset \mathcal{R})$, e.g., $(a, b) = (0, \infty)$, or $(0, 1)$, or $\mathcal{R}$; λ is an unknown finite-dimensional transformation vector; $h(\cdot; \lambda)$ is a known strictly increasing and differentiable real-valued function on (a, b), e.g., Examples 1.1–1.3; and ε_is are independent errors distributed as either $N(0, g^2(f(x_i; \beta), z_i; \gamma)\,\sigma^2)$ or truncated $N(\mu_i(\lambda, \beta, \sigma, \gamma), g^2(f(x_i; \beta), z_i; \gamma)\,\sigma^2)$ with median 0 for some $\mu_i(\lambda, \beta, \sigma, \gamma) \in \mathcal{R}$. Notice that, for $i = 1, \ldots, n$, $f(x_i; \beta)$ is the median of $h(y_i; \lambda)$; however, it is generally neither the mean nor mode of $h(y_i; \lambda)$.

In Section 2, the transformed truncated normal mode regression model is proposed to extend model (5) and then the corresponding likelihood inference is discussed thoroughly. In Section 3, the proposed methodology is applied to two real data sets in Box and Cox (1964). Finally, conclusions and discussion are given in Section 4.

2 Transformed Truncated Normal Mode Regression Model

In this section, the transformed truncated normal mode regression model is proposed to extend model (5) and then the corresponding likelihood inference is discussed thoroughly.

2.1 Transformed Truncated Normal Mode Regression Model

Consider the following transformed truncated normal mode regression model:

$$h(y_i; \lambda) = f(x_i; \beta) + \varepsilon_i \tag{11}$$

for $i = 1, \ldots, n$, where y_i is the response for subject i with known support (a, b) $(\subset \mathcal{R})$, e.g., $(0, \infty)$, or $(0, 1)$, or $\mathcal{R}$; λ is an unknown finite-dimensional transformation vector; $h(\cdot; \lambda)$ is a known strictly increasing and differentiable real-valued function on (a, b), e.g., Examples 1.1–1.3 in Section 1; x_i is a known covariate vector for subject i; β is an unknown finite-dimensional regression parameter vector; $f(\cdot; \beta)$ is a known regression function for each β, e.g., $f(x_i; \beta) = x_i^T \beta$ or $\exp\{x_i^T \beta\}$; and ε_is are independent errors distributed as either $N(0, g^2(f(x_i; \beta), z_i; \gamma)\, \sigma^2)$ or truncated $N(0, g^2(f(x_i; \beta), z_i; \gamma)\, \sigma^2)$ such that z_i is a known covariate vector for subject i, e.g., z_i is a known function of x_i; γ is an unknown finite-dimensional parameter vector; $g(\cdot, \cdot; \gamma)$ is a known positive function for each γ, e.g., $g(f(x_i; \beta), z_i; \gamma) = \exp\{f(x_i; \beta)\, \gamma_1 + z_i^T \gamma_2\}$ with $\gamma \equiv (\gamma_1, \gamma_2^T)^T$; and σ is an unknown positive scale parameter. Notice that, for $i = 1, \ldots, n$, $f(x_i; \beta)$ is the mode of $h(y_i; \lambda)$ when it is in the support of $h(y_i; \lambda)$; however, it is generally neither the mean nor median of $h(y_i; \lambda)$.

2.2 Maximum Likelihood Estimation

Let θ $(\equiv (\theta_1, \ldots, \theta_d)^T)$ denote the d-dimensional parameter vector $(\lambda^T, \beta^T, \sigma, \gamma^T)^T$ in the parameter space Θ, where Θ is a non-empty open subset of the d-dimensional Euclidean space $\mathcal{R}^d$. Let $\Phi(\cdot)$ denote the cumulative distribution function (c.d.f.) of $N(0, 1)$ and $\phi(\cdot)$ the probability density function (p.d.f.) of $N(0, 1)$. Set

$$e_i(u; \theta) \equiv \frac{h(u; \lambda) - f(x_i; \beta)}{g(f(x_i; \beta), z_i; \gamma)\, \sigma} \equiv \frac{h(u; \lambda) - f_i(\beta)}{g_i(\beta, \gamma)\, \sigma} \tag{12}$$

for $u \in [a,b]$, $\theta \in \Theta$, and $i = 1,\ldots,n$, where $h(a;\lambda) \equiv \lim_{u\downarrow a} h(u;\lambda)$ and $h(b;\lambda) \equiv \lim_{u\uparrow b} h(u;\lambda)$.

Under model (11), the p.d.f. of y_i is

$$p_i(y_i;\theta) = \frac{\phi(e_i(y_i;\theta))h'(y_i;\lambda)}{g_i(\beta,\gamma)\,\sigma\,[\Phi(e_i(b;\theta)) - \Phi(e_i(a;\theta))]} \cdot 1_{(a,b)}(y_i) \tag{13}$$

for $i = 1,\ldots,n$, where $h'(y_i;\lambda) \equiv \partial h(u;\lambda)/\partial u|_{u=y_i} \equiv h_i'(\lambda)$ and $1_{(a,b)}(y_i) = 1$ for $y_i \in (a,b)$ and 0 otherwise. Set $y \equiv (y_1,\ldots,y_n)^T$. Set $h_i(\lambda) \equiv h(y_i;\lambda)$, and $e_i(\theta) \equiv e_i(y_i;\theta)$ for $\theta \in \Theta$ and $i = 1,\ldots,n$. Then, given y, the likelihood function for θ is

$$\prod_{i=1}^{n} \frac{\phi(e_i(\theta))h_i'(\lambda)}{g_i(\beta,\gamma)\,\sigma\,[\Phi(e_i(b;\theta)) - \Phi(e_i(a;\theta))]} \equiv L(\theta) \tag{14}$$

and the log-likelihood function for θ is

$$\log[L(\theta)] \equiv \ell(\theta) \equiv \sum_{i=1}^{n} \ell_i(\theta), \tag{15}$$

where

$$\begin{aligned} \ell_i(\theta) \;=\; & \log[\phi(e_i(\theta))] + \log[h_i'(\lambda)] - \log[g_i(\beta,\gamma)] - \log(\sigma) \\ & - \log[\Phi(e_i(b;\theta)) - \Phi(e_i(a;\theta))]. \end{aligned} \tag{16}$$

Assume that

$$\frac{\partial}{\partial\theta}\int_a^b p_i(y_i;\theta)\,dy_i = \int_a^b \frac{\partial p_i(y_i;\theta)}{\partial\theta}\,dy_i \tag{17}$$

and

$$\frac{\partial^2}{\partial\theta\partial\theta^T}\int_a^b p_i(y_i;\theta)\,dy_i = \int_a^b \frac{\partial^2 p_i(y_i;\theta)}{\partial\theta\partial\theta^T}\,dy_i \tag{18}$$

for $\theta \in \Theta$ and $i = 1,\ldots,n$. Then the score function for θ is

$$\frac{\partial\ell(\theta)}{\partial\theta} = \sum_{i=1}^{n}\frac{\partial\ell_i(\theta)}{\partial\theta} \equiv \sum_{i=1}^{n} S_i(\theta) \equiv S(\theta), \tag{19}$$

the observed Fisher information for θ is

$$-\frac{\partial^2\ell(\theta)}{\partial\theta\partial\theta^T} = -\sum_{i=1}^{n}\frac{\partial^2\ell_i(\theta)}{\partial\theta\partial\theta^T} \equiv \sum_{i=1}^{n} J_i(\theta) \equiv J(\theta), \tag{20}$$

and the expected Fisher information for θ is

$$Cov_\theta\,(S(\theta)) = \sum_{i=1}^{n} Cov_\theta(S_i(\theta)) \equiv \sum_{i=1}^{n} I_i(\theta) \equiv I(\theta), \tag{21}$$

where both $S_i(\theta)$ and $J_i(\theta)$ are put in Appendices A and B, respectively. By equation (17), $E_\theta(S_i(\theta)) = 0_{d\times 1}$ for $\theta \in \Theta$ and $i = 1, \ldots, n$, where $0_{d\times 1}$ denotes the $d \times 1$ vector $(0, \ldots, 0)^T$. By equations (17) and (18), $E_\theta(J_i(\theta)) = I_i(\theta)$ for $\theta \in \Theta$ and $i = 1, \ldots, n$.

Assume that, given y, there exists a unique maximum likelihood estimate (MLE) $\hat{\theta}(y)$ ($\equiv \hat{\theta}$) of θ. Then $\hat{\theta}$ solves the score equation $S(\hat{\theta}) = 0_{d\times 1}$ for θ. One possible approach to evaluate $\hat{\theta}$ is as follows: First choose a good initial value $\hat{\theta}^{(0)}$ and then iterate the following equations

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} + M^{-1}(\hat{\theta}^{(k)})S(\hat{\theta}^{(k)}) \tag{22}$$

for $k = 0, 1, 2, \ldots$ until $||S(\hat{\theta}^{(k+1)})|| < \varepsilon$ for some small positive value ε, e.g., $\varepsilon = 10^{-3}$, where $||a|| \equiv (a^T a)^{1/2}$ for $a \in \mathcal{R}^d$. When $M(\hat{\theta}^{(k)}) = I(\hat{\theta}^{(k)})$ for $k = 0, 1, 2, \ldots$, it is called the Fisher scoring method. However, it will take too much time to evaluate $I(\hat{\theta}^{(k)})$s because there is generally no closed-form formula for each $I(\hat{\theta}^{(k)})$. When $M(\hat{\theta}^{(k)}) = J(\hat{\theta}^{(k)})$ for $k = 0, 1, 2, \ldots$, it is called the Newton-Raphson method. It is usually difficult to find a good initial value for the Newton-Raphson method specially when d is not a small positive integer. Moreover, it is not necessary that $\ell(\hat{\theta}^{(k+1)}) > \ell(\hat{\theta}^{(k)})$ for $k = 0, 1, 2, \ldots$. Thus, a modified Newton-Raphson method is suggested as follows: First choose a good initial value $\hat{\theta}^{(0)}$ and a fraction $\lambda \in (0, 1)$, e.g., $\lambda = 1/2$. When $\hat{\theta}^{(k)}$ is obtained and $S^T(\hat{\theta}^{(k)})J^{-1}(\hat{\theta}^{(k)})S(\hat{\theta}^{(k)}) \neq 0$ for some non-negative integer k, iterate the following equations

$$\hat{\theta}^{(k+1,j)} \equiv \hat{\theta}^{(k)} + \mathrm{sgn}(S^T(\hat{\theta}^{(k)})J^{-1}(\hat{\theta}^{(k)})S(\hat{\theta}^{(k)}))\, \lambda^j\, J^{-1}(\hat{\theta}^{(k)})S(\hat{\theta}^{(k)}) \tag{23}$$

for $j = 0, 1, 2, \ldots, m_k$, where m_k is the first j such that $\ell(\hat{\theta}^{(k+1,j)}) > \ell(\hat{\theta}^{(k)})$. Set $\hat{\theta}^{(k+1)} \equiv \hat{\theta}^{(k+1,m_k)}$ for $k = 0, 1, 2, \ldots$ until $||S(\hat{\theta}^{(k+1)})|| < \varepsilon$ for some small positive value ε, e.g., $\varepsilon = 10^{-3}$.

When $S^T(\hat{\theta}^{(k)})J^{-1}(\hat{\theta}^{(k)})S(\hat{\theta}^{(k)}) \neq 0$ for some non-negative integer k, it follows from the first-order Taylor expansion that

$$\begin{aligned}
& \ell(\hat{\theta}^{(k+1,j)}) \\
= \;& \ell(\hat{\theta}^{(k)}) + S^T(\hat{\theta}^{(k)}) \left[\mathrm{sgn}(S^T(\hat{\theta}^{(k)})J^{-1}(\hat{\theta}^{(k)})S(\hat{\theta}^{(k)}))\, \lambda^j J^{-1}(\hat{\theta}^{(k)})S(\hat{\theta}^{(k)})\right] + o(\lambda^j) \\
= \;& \ell(\hat{\theta}^{(k)}) + \lambda^j |S^T(\hat{\theta}^{(k)})J^{-1}(\hat{\theta}^{(k)})S(\hat{\theta}^{(k)})| + o(\lambda^j)
\end{aligned} \tag{24}$$

as $j \to \infty$, which implies that $\ell(\hat{\theta}^{(k+1,j)}) > \ell(\hat{\theta}^{(k)})$ for large j and thus m_k is well-defined.

Now consider the case where the sample size n tends to infinity. Assume that the following conditions hold:

(i) the minimum eigenvalue of $I(\theta)$ tends to infinity as $n \to \infty$;

(ii) $E_\theta(\max_{1\le i\le n} |\partial\ell_i(\theta)/\partial\theta_j|)/[Var_\theta(\partial\ell(\theta)/\partial\theta_j)]^{1/2} \to 0$ as $n \to \infty$ for $j = 1, \ldots, d$;

(iii) $I^{-1/2}(\theta)J(\theta)I^{-1/2}(\theta) \xrightarrow{p} I_d$ as $n \to \infty$, where I_d denotes the identity matrix of order d; and

(iv) $[\text{diag}\{I_{11}(\theta), \ldots, I_{dd}(\theta)\}]^{-1/2}I(\theta)[\text{diag}\{I_{11}(\theta), \ldots, I_{dd}(\theta)\}]^{-1/2} \to \Sigma(\theta)$ as $n \to \infty$, where $I_{jj}(\theta)$ denotes the jth diagonal element of $I(\theta)$ for $j = 1, \ldots, d$ and $\Sigma(\theta)$ is a positive definite covariance matrix.

Let $M(\theta)$ denote either $I(\theta)$ or $J(\theta)$. Then, by Theorem 1.80 of Prakasa Rao (1999),

$$M^{-1/2}(\theta)S(\theta) \xrightarrow{d} N_d(0_{d\times1}, I_d) \tag{25}$$

as $n \to \infty$, where $N_d(0_{d\times1}, I_d)$ denotes the d-variate normal distribution with mean vector $0_{d\times1}$ and covariance matrix I_d. Assume that

$$I^{-1/2}(\theta)\{S(\hat\theta) - [S(\theta) - J(\theta)(\hat\theta - \theta)]\} = o_p(1) \tag{26}$$

as $n \to \infty$. Then, by condition (iii) and equations (25) and (26),

$$M^{1/2}(\theta)(\hat\theta - \theta) = M^{-1/2}(\theta)S(\theta) + o_p(1) \xrightarrow{d} N_d(0_{d\times1}, I_d) \tag{27}$$

as $n \to \infty$. Thus, by condition (i) and equation (27), $\hat\theta$ is a weakly consistent estimator of θ. Assume that $I^{-1/2}(\theta)I(\hat\theta)I^{-1/2}(\theta) \xrightarrow{p} I_d$ and $J^{-1/2}(\theta)J(\hat\theta)J^{-1/2}(\theta) \xrightarrow{p} I_d$ as $n \to \infty$. Then, by equation (27),

$$M^{1/2}(\hat\theta)(\hat\theta - \theta) = M^{1/2}(\theta)(\hat\theta - \theta) + o_p(1) \xrightarrow{d} N_d(0_{d\times1}, I_d) \tag{28}$$

as $n \to \infty$.

2.3 Hypothesis Testing and Confidence Regions

In this subsection, let ω ($\equiv (\psi^T, \chi^T)^T \in \Omega \subset \mathcal{R}^d$) be a one-to-one reparameterization of θ such that $\det(\partial\theta/\partial\omega^T) \neq 0$ and $\partial^2\theta_j/\partial\chi\partial\chi^T$ is a continuous function of χ for $j = 1, \ldots, d$, where ψ is the d_0-dimensional parameter vector of interest and χ is a $(d - d_0)$-dimensional nuisance parameter vector with $d_0 \in \{1, \ldots, d\}$. Here χ does not exist when $d_0 = d$. Suppose that we are interested in testing the null hypothesis H_0: $\psi = \psi_0$ versus the alternative H_1: $\psi \neq \psi_0$.

Set $S_\psi(\chi) \equiv \partial\ell(\theta)/\partial\chi, I_\psi(\chi) \equiv Cov_\omega(S_\psi(\chi))$, and $J_\psi(\chi) \equiv -\partial S_\psi(\chi)/\partial\chi^T$ for $\omega \in \Omega$. Then $S_\psi(\chi) = [\partial\theta^T/\partial\chi]S(\theta), I_\psi(\chi) = [\partial\theta^T/\partial\chi]I(\theta)[\partial\theta/\partial\chi^T]$, and

$$J_\psi(\chi) = \frac{\partial\theta^T}{\partial\chi}J(\theta)\frac{\partial\theta}{\partial\chi^T} - \sum_{j=1}^{d}\frac{\partial^2\theta_j}{\partial\chi\partial\chi^T}S(\theta)_j \tag{29}$$

for $\omega \in \Omega$, where $S(\theta) \equiv (S(\theta)_1, \ldots, S(\theta)_d)^T$. Assume that, given y, there exists a unique MLE $\hat{\chi}_\psi(y)$ ($\equiv \hat{\chi}_\psi$) of χ for fixed ψ. Then $\hat{\chi}_\psi$ solves the score equation $S_\psi(\hat{\chi}_\psi) = 0_{(d-d_0)\times 1}$ for χ, where $0_{(d-d_0)\times 1}$ denotes the $(d-d_0)\times 1$ vector $(0, \ldots, 0)^T$.

Set $W(\psi) \equiv 2[\ell(\hat{\theta}) - \ell(\theta(\psi, \hat{\chi}_\psi))]$. Assume that $I_\psi^{-1/2}(\chi)J_\psi(\hat{\chi}_\psi)I_\psi^{-1/2}(\chi) \xrightarrow{p} I_{d-d_0}$,

$$I_\psi^{1/2}(\chi)(\hat{\chi}_\psi - \chi) = I_\psi^{-1/2}(\chi)S_\psi(\chi) + o_p(1), \tag{30}$$

$$\ell(\theta) = \ell(\hat{\theta}) + S^T(\hat{\theta})(\theta - \hat{\theta}) - \frac{1}{2}(\theta - \hat{\theta})^T J(\hat{\theta})(\theta - \hat{\theta}) + o_p(1), \tag{31}$$

and

$$\ell(\theta) = \ell(\theta(\psi, \hat{\chi}_\psi)) + S_\psi^T(\hat{\chi}_\psi)(\chi - \hat{\chi}_\psi) - \frac{1}{2}(\chi - \hat{\chi}_\psi)^T J_\psi(\hat{\chi}_\psi)(\chi - \hat{\chi}_\psi) + o_p(1) \tag{32}$$

as $n \to \infty$. Then, by equations (27) and (28),

$$\begin{aligned} & W(\psi) \\ = \; & S^T(\theta)I^{-1/2}(\theta)\left\{I_d - I^{1/2}(\theta)\frac{\partial\theta}{\partial\chi^T}\left[\frac{\partial\theta^T}{\partial\chi}I(\theta)\frac{\partial\theta}{\partial\chi^T}\right]^{-1}\frac{\partial\theta^T}{\partial\chi}I^{1/2}(\theta)\right\}I^{-1/2}(\theta)S(\theta) \\ & + o_p(1) \\ \xrightarrow{d} \; & \chi^2_{d_0} \end{aligned} \tag{33}$$

as $n \to \infty$.

Let $\alpha \in (0, 1)$ be fixed, e.g., $\alpha = 0.05$. The likelihood ratio test with asymptotic size α is to reject H_0: $\psi = \psi_0$ if and only if the likelihood ratio test statistic $W(\psi_0) > \chi^2_{\alpha,d_0}$, where χ^2_{α,d_0} denotes the upper α quantile of the χ^2 distribution with d_0 degrees of freedom. To evaluate $W(\psi_0)$, we need to evaluate $\hat{\chi}_{\psi_0}$. One possible approach to evaluate $\hat{\chi}_{\psi_0}$ is to utilize a modified Newton-Raphson method in Section 2.2. Therefore, $\{\psi_0\colon W(\psi_0) \le \chi^2_{\alpha,d_0}\}$ is an asymptotic size $1-\alpha$ confidence region for ψ.

2.4 Prediction Region of Future Observations

Suppose that

$$h(y_{n+j};\lambda) = f(x_{n+j};\beta) + \varepsilon_{n+j} \tag{34}$$

for $j = 1, \ldots, m$, where m is a known positive integer, y_{n+j} is the future observation for subject $n+j$ with support (a, b), x_{n+j} is a known covariate vector for subject $n+j$, and ε_{n+j} is an error distributed as either $N(0, g^2(f(x_{n+j};\beta), z_{n+j};\gamma)\,\sigma^2)$ or truncated $N(0, g^2(f(x_{n+j};\beta), z_{n+j};\gamma)\,\sigma^2)$ with known covariate vector z_{n+j}, and $\varepsilon_1, \ldots, \varepsilon_{n+m}$ are independent. For $\theta \in \Theta$, $u \in [a, b]$, and $j = 1, \ldots, m$, set

$$e_{n+j}(u;\theta) \equiv \frac{h(u;\lambda) - f(x_{n+j};\beta)}{g(f(x_{n+j};\beta), z_{n+j};\gamma)\,\sigma} \equiv \frac{h(u;\lambda) - f_{n+j}(\beta)}{g_{n+j}(\beta,\gamma)\,\sigma}. \tag{35}$$

Let $\alpha \in (0,1)$ be fixed, e.g., $\alpha = 0.05$. For $\theta \in \Theta$ and $j = 1, \ldots, m$, let $\Phi_{n+j}(\cdot;\theta)$ denote the c.d.f. of ε_{n+j} and $q_{n+j,\alpha}(\theta)$ the α quantile of y_{n+j}. Then

$$q_{n+j,\alpha}(\theta) = h^{-1}(f_{n+j}(\beta) + \Phi^{-1}_{n+j}(\alpha;\theta);\lambda) \tag{36}$$

with MLE $q_{n+j,\alpha}(\hat\theta)$ for $\theta \in \Theta$ and $j = 1, \ldots, m$, where

$$\begin{aligned}\Phi^{-1}_{n+j}(\alpha;\theta) &= g_{n+j}(\beta,\gamma)\,\sigma\,\Phi^{-1}((1-\alpha)\,\Phi(e_{n+j}(a;\theta)) + \alpha\,\Phi(e_{n+j}(b;\theta)))\\ &\equiv g_{n+j}(\beta,\gamma)\,\sigma\,\Phi^{-1}_{n+j}(\alpha, a, b;\theta).\end{aligned} \tag{37}$$

Assume that $q_{n+j,\alpha}(\theta)$ is a continuously differentiable function of θ with $\partial q_{n+j,\alpha}(\theta)/\partial\theta \neq 0_{d\times 1}$ for $\theta \in \Theta$ and $j = 1, \ldots, m$. Then

$$\begin{aligned}\frac{\partial q_{n+j,\alpha}(\theta)}{\partial\theta} &= \frac{\partial f_{n+j}(\beta)/\partial\theta + \sigma\,\Phi^{-1}_{n+j}(\alpha,a,b;\theta)\,\partial g_{n+j}(\beta,\gamma)/\partial\theta}{h'(q_{n+j,\alpha}(\theta);\lambda)}\\ &\quad+\frac{g_{n+j}(\beta,\gamma)\,\Phi^{-1}_{n+j}(\alpha,a,b;\theta)\,\partial\,\sigma/\partial\theta + g_{n+j}(\beta,\gamma)\,\sigma\,\partial\Phi^{-1}_{n+j}(\alpha,a,b;\theta)/\partial\theta}{h'(q_{n+j,\alpha}(\theta);\lambda)}\\ &\quad-\frac{\partial h(u;\lambda)/\partial\theta|_{u=q_{n+j,\alpha}(\theta)}}{h'(q_{n+j,\alpha}(\theta);\lambda)}\end{aligned} \tag{38}$$

for $\theta \in \Theta$ and $j = 1, \ldots, m$, where

$$\frac{\partial\Phi^{-1}_{n+j}(\alpha,a,b;\theta)}{\partial\theta} = \frac{(1-\alpha)\,\partial\Phi(e_{n+j}(a;\theta))/\partial\theta + \alpha\,\partial\Phi(e_{n+j}(b;\theta))/\partial\theta}{\phi(\Phi^{-1}_{n+j}(\alpha,a,b;\theta))} \tag{39}$$

with both $\partial\Phi(e_{n+j}(a;\theta))/\partial\theta$ and $\partial\Phi(e_{n+j}(b;\theta))/\partial\theta$ being evaluated by similar formulas in Appendix A.

By equations (27) and (28),

$$\left[\frac{\partial q_{n+j,\alpha}(\theta)}{\partial\theta^T} M^{-1}(\theta)\,\frac{\partial q_{n+j,\alpha}(\theta)}{\partial\theta}\right]^{-1/2}[q_{n+j,\alpha}(\hat\theta) - q_{n+j,\alpha}(\theta)] \xrightarrow{d} N(0,1) \tag{40}$$

and

$$\left[\left.\frac{\partial q_{n+j,\alpha}(\theta)}{\partial\theta^T}\right|_{\theta=\hat\theta} M^{-1}(\hat\theta)\left.\frac{\partial q_{n+j,\alpha}(\theta)}{\partial\theta}\right|_{\theta=\hat\theta}\right]^{-1/2}[q_{n+j,\alpha}(\hat\theta) - q_{n+j,\alpha}(\theta)] \xrightarrow{d} N(0,1) \tag{41}$$

as $n \to \infty$ for $\theta \in \Theta$ and $j = 1, \ldots, m$, where M denotes either I or J.

Set $\alpha_m \equiv [1 - (1-\alpha)^{1/m}]/2$. Since $y_{n+1}, \ldots, y_{n+m}$ are independent,

$$
\begin{aligned}
& P_\theta\left(\bigcap_{j=1}^{m}\{y_{n+j} \in [q_{n+j,\alpha_m}(\theta), q_{n+j,1-\alpha_m}(\theta)]\}\right) \\
= & \prod_{j=1}^{m} P_\theta\left(\{y_{n+j} \in [q_{n+j,\alpha_m}(\theta), q_{n+j,1-\alpha_m}(\theta)]\}\right) \;=\; (1 - 2\,\alpha_m)^m \;=\; 1-\alpha, \qquad (42)
\end{aligned}
$$

which implies that $[q_{n+1,\alpha_m}(\theta), q_{n+1,1-\alpha_m}(\theta)] \times \cdots \times [q_{n+m,\alpha_m}(\theta), q_{n+m,1-\alpha_m}(\theta)]$ is a size $1-\alpha$ prediction region for $(y_{n+1}, \ldots, y_{n+m})^T$ with MLE $[q_{n+1,\alpha_m}(\hat{\theta}), q_{n+1,1-\alpha_m}(\hat{\theta})] \times \cdots \times [q_{n+m,\alpha_m}(\hat{\theta}), q_{n+m,1-\alpha_m}(\hat{\theta})]$.

3 Two Real Data Sets

In this section, the proposed methodology is applied to two real data sets in Box and Cox (1964).

3.1 A Biological Experiment Using a 3×4 Factorial Design

Table 1 shows the survival times of animals in a 3×4 factorial experiment, the factors being A with three poisons and B with four treatments. Each combination of these two factors is replicated for four animals, the allocation to animals being completely randomized. The two-way analysis-of-variance (ANOVA) effects model is

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk} \tag{43}$$

for $i = 1, 2, 3$ and $j, k = 1, 2, 3, 4$, where y_{ijk} is the kth observation for the ith poison of factor A and the jth treatment of factor B; μ is the overall mean, τ_i is the main effect of the ith level of factor A, β_j is the main effect of the jth level of factor B, $(\tau\beta)_{ij}$ is the interaction between the ith level of factor A and the jth level of factor B, and ε_{ijk}s are *i.i.d.* $N(0, \sigma^2)$ errors with unknown positive standard deviation σ. Figure 2(a) shows the residual plot against fitted values for the original data under the two-way ANOVA effects model. It is seen that $Var(y_{ijk})$ increases as $E(y_{ijk})$ increases.

Now consider the following Box-Cox transformed truncated normal mode regression model:

$$y_{ijk}^{(\lambda)} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk} \tag{44}$$

for $i = 1, 2, 3$ and $j, k = 1, 2, 3, 4$, where each y_{ijk} has support $(0, \infty)$, λ is am unknown real-valued transformation parameter and ε_{ijk}s are independent errors distributed as either $N(0, \sigma^2)$ or truncated $N(0, \sigma^2)$ with unknown positive scale parameter σ. Figure 2(b) shows the residual plot against fitted values for the transformed data under the Box-Cox transformed truncated normal mode regression model.

One possible way to find a good initial value $\hat{\theta}^{(0)}$ in this case is put in Appendix C. How to plot the normal probability plot in this case is put in Appendix D. Table 2 shows the MLEs under the false normality assumption and under the truncated normality assumption, respectively. Figure 3 shows the normal probability plots under the false normality assumption and the truncated normality assumption, respectively.

First, test the null hypothesis H_0: $(\tau\beta)_{ij} = 0$ for all (i,j) versus the alternative H_1: $(\tau\beta)_{ij} \neq 0$ for some (i,j). Then the asymptotic p-value is 0.3168 and thus it fails to reject the null hypothesis H_0.

Table 3 shows the MLEs without interactions under the false normality assumption and the truncated normality assumption, respectively. Figure 4 shows the normal probability plots under the false normality assumption and the truncated normality assumption, respectively, without interactions.

Similarly, we are also interested in testing the null hypothesis H_0: $\lambda = -1$ and $(\tau\beta)_{ij} = 0$ for all (i,j) versus the alternative H_1: $\lambda \neq -1$ or $(\tau\beta)_{ij} \neq 0$ for some (i,j). The asymptotic p-value is 0.2829 under the truncated normality assumption, and thus it also fails to reject the null hypothesis H_0.

Table 4 shows the MLEs with $\lambda = -1$ under the false normality assumption and the truncated normality assumption, respectively. Figure 5 shows the normal probability plots under the false normality assumption and the truncated normality assumption, respectively, with $\lambda = -1$. Table 5 shows the MLEs under the false normality assumption and under the truncated normality assumption, respectively, without interactions and with $\lambda = -1$. Figure 6 shows the normal probability plots under the false normality assumption and the truncated normality assumption, respectively, without interactions and with $\lambda = -1$.

Suppose that

$$y^{(\lambda)}_{i_l j_l k_l} = \mu + \tau_{i_l} + \beta_{j_l} + \varepsilon_{i_l j_l k_l} \tag{45}$$

for $l = 1, \ldots, m$, where m is a positive integer, $y_{i_l j_l k_l}$ is the k_lth observation with the i_lth poisson of factor A and the j_lth treatment of factor B, all (i_l, j_l, k_l)s are different for $k_l \geq 5$, $\varepsilon_{i_l j_l k_l}$ is an error distributed as either $N(0, \sigma^2)$ or truncated $N(0, \sigma^2)$, and ε_{ijk}s and $\varepsilon_{i_l j_l k_l}$s are independent.

Let $\alpha \in (0,1)$ be fixed, e.g., 0.05. For $l = 1, \ldots, m$, let $\Phi_{i_l j_l k_l}(\cdot;\theta)$ denote the c.d.f. of $\varepsilon_{i_l j_l k_l}$ and $q_{i_l j_l k_l,\alpha}(\theta)$ the α quantile of $y_{i_l j_l k_l}$. Then

$$q_{i_l j_l k_l,\alpha}(\theta) = \left\{1 + \lambda\left[\mu + \tau_{i_l} + \beta_{j_l} + \Phi^{-1}_{i_l j_l k_l}(\alpha;\theta)\right]\right\}^{1/\lambda} \tag{46}$$

for $l = 1, \ldots, m$, where

$$\Phi^{-1}_{i_l j_l k_l}(\alpha;\theta) = \sigma\,\Phi^{-1}\left(\alpha\,\Phi\left(\frac{-1/\lambda - \mu - \tau_{i_l} - \beta_{j_l}}{\sigma}\right)\right). \tag{47}$$

Thus, $[q_{i_1j_1k_1,\alpha_m}(\theta), q_{i_1j_1k_1,1-\alpha_m}(\theta)] \times \cdots \times [q_{i_mj_mk_m,\alpha_m}(\theta), q_{i_mj_mk_m,1-\alpha_m}(\theta)]$ is a size $1-\alpha$ prediction region for $(y_{i_1j_1k_1}, \ldots, y_{i_mj_mk_m})^T$ with MLE $[q_{i_1j_1k_1,\alpha_m}(\hat{\theta}), q_{i_1j_1k_1,1-\alpha_m}(\hat{\theta})] \times \cdots \times [q_{i_mj_mk_m,\alpha_m}(\hat{\theta}), q_{i_mj_mk_m,1-\alpha_m}(\hat{\theta})]$, where $\alpha_m \equiv [1-(1-\alpha)^{1/m}]/2$.

3.2 A Textile Experiment Using a Single Replicate of a 3^3 Design

Table 6 shows the numbers of cycles to failure, y, obtained in a single replicate of a 3^3 factorial experiment in which the factors are

x_1: length of test specimen (250, 300, 350 mm),

x_2: amplitude of loading cycle (8, 9, 10 mm),

x_3: load (40, 45, 50 gm).

In Table 6, the levels of the x_1, x_2, and x_3 are coded as $-1, 0, 1$, respectively. Consider the following quadratic regression model:

$$y_i = \beta_0 + \sum_{j=1}^{3} \beta_j x_{ij} + \sum_{1 \le j \le k \le 3} \beta_{jk} x_{ij} x_{ik} + \varepsilon_i \tag{48}$$

for $i = 1, \ldots, 27$, where y_i is the response for $(x_1, x_2, x_3) = (x_{i1}, x_{i2}, x_{i3})$, β_0 is the intercept, β_js and β_{jk}s are regression coefficients, and ε_is are *i.i.d.* $N(0, \sigma^2)$ errors with unknown positive standard deviation σ. Figure 7(a) shows the residual plot against fitted values for the original data under the quadratic regression model. It is easily seen that there is an obvious pattern in Figure 7(a).

Now consider the following Box-Cox transformed truncated normal mode regression model:

$$y_i^{(\lambda)} = \beta_0 + \sum_{j=1}^{3} \beta_j x_{ij} + \sum_{1 \le j \le k \le 3} \beta_{jk} x_{ij} x_{ik} + \varepsilon_i \tag{49}$$

for $i = 1, \ldots, 27$, where y_i has support $(0, \infty)$ and ε_is are independent errors distributed as either $N(0, \sigma^2)$ or truncated $N(0, \sigma^2)$ with unknown positive standard deviation σ. Figure 7(b) is the residual plot against fitted values for the transformed data under the Box-Cox transformed truncated normal mode regression model.

Table 7 shows the MLEs under the false normality assumption and under the truncated normality assumption, respectively. It is seen that the MLEs under the false normality assumption are nearly the same as under the truncated normality assumption. Figure 8 shows the normal probability plots under the false normality assumption and the truncated normality assumption, respectively.

First, test the null hypothesis H_0: $\beta_{jk} = 0$ for all (j,k) versus the alternative H_1: $\beta_{jk} \neq 0$ for some (j,k). Then the asymptotic p-value is 0.3487 and thus it fails to reject the null hypothesis H_0.

Table 8 shows the MLEs under the false normality assumption and under the truncated normality assumption, respectively, without quadratic effects and interactions. Figure 9 shows the normal probability plots under the false normality assumption and the truncated normality assumption, respectively, without quadratic effects and interactions.

Similarly, we are also interested in testing the null hypothesis H_0: $\lambda = 0$ and $\beta_{jk} = 0$ for all (j,k) versus the alternative H_1: $\lambda \neq 0$ or $\beta_{jk} \neq 0$ for some (j,k). The asymptotic p-value is 0.4313 under the truncated normality assumption, and thus it also fails to reject the null hypothesis H_0.

Table 9 shows the MLEs with $\lambda = 0$ under the false normality assumption and the truncated normality assumption, respectively. Figure 10 shows the normal probability plots under the false normality assumption and the truncated normality assumption, respectively, with $\lambda = 0$. Table 10 shows the MLEs under the false normality assumption and the truncated normality assumption, respectively, without quadratic effects and interactions and with $\lambda = 0$. Figure 11 shows the normal probability plots under the false normality assumption and the truncated normality assumption, respectively, without quadratic effects and interactions and with $\lambda = 0$.

Suppose that

$$y_l^{(\lambda)} = \beta_0 + \sum_{j=1}^{3} \beta_j x_{lj} + \varepsilon_l \tag{50}$$

for $l = 27+1, \ldots, 27+m$, where m is a positive integer, y_l is the lth observation for $(x_1, x_2, x_3) = (x_{l1}, x_{l2}, x_{l3})$, ε_l is the $(l-27)$-th future error distributed as either $N(0,\sigma^2)$ or truncated $N(0,\sigma^2)$, and $\varepsilon_1, \ldots, \varepsilon_{27+m}$ are independent. Let $\alpha \in (0,1)$ be fixed, e.g., 0.05. For $l = 27+1, \ldots, 27+m$, let $\Phi_l(\cdot;\theta)$ denote the c.d.f. of ε_l and $q_{l,\alpha}(\theta)$ the α quantile of y_l. Then

$$q_{l,\alpha}(\theta) = \left\{ 1 + \lambda \left[\beta_0 + \sum_{j=1}^{3} \beta_j x_{lj} + \Phi_l^{-1}(\alpha;\theta) \right] \right\}^{1/\lambda} \tag{51}$$

for $l = 27+1, \ldots, 27+m$, where

$$\Phi_l^{-1}(\alpha;\theta) = \sigma\, \Phi^{-1} \left(\alpha\, \Phi \left(\frac{-1/\lambda - \beta_0 - \sum_{j=1}^{3} \beta_j x_{lj}}{\sigma} \right) \right). \tag{52}$$

Thus, $[q_{27+1,\alpha_m}(\theta), q_{27+1,1-\alpha_m}(\theta)] \times \cdots \times [q_{27+m,\alpha_m}(\theta), q_{27+m,1-\alpha_m}(\theta)]$ is a size $1-\alpha$ prediction region for $(y_{27+1}, \ldots, y_{27+m})^T$ with MLE $[q_{27+1,\alpha_m}(\hat{\theta}), q_{27+1,1-\alpha_m}(\hat{\theta})] \times \cdots \times [q_{27+m,\alpha_m}(\hat{\theta}), q_{27+m,1-\alpha_m}(\hat{\theta})]$, where $\alpha_m \equiv [1-(1-\alpha)^{1/m}]/2$.

4 Conclusions and Discussion

Now consider the following transformed truncated normal mean regression model:

$$h(y_i;\lambda) = f(x_i;\beta) + \varepsilon_i \tag{53}$$

for $i = 1, \ldots, n$, where y_i is the response for subject i with known support (a, b) $(\subset \mathcal{R})$; λ is an unknown finite-dimensional transformation vector; $h(\cdot;\lambda)$ is a known strictly increasing and differentiable real-valued function on (a, b); x_i is a known covariate vector for subject i; β is an unknown finite-dimensional regression parameter vector; $f(\cdot;\beta)$ is a known regression function for each β; and ε_is are independent errors distributed as either $N(0, g^2(f(x_i;\beta), z_i;\gamma)\,\sigma^2)$ or truncated $N(\mu_i(\theta), \sigma_i^2(\theta))$ such that $\mu_i(\theta)$ is an unknown mean parameter, $\sigma_i(\theta)$ is an unknown positive standard deviation parameter, z_i is a known covariate vector for subject i; γ is an unknown finite-dimensional parameter vector; $g(\cdot,\cdot;\gamma)$ is a known positive function for each γ; and σ is an unknown positive scale parameter. Notice that, for $i = 1, \ldots, n$, $f(x_i;\beta)$ is the mean of $h(y_i;\lambda)$ when it is in the support of $h(y_i;\lambda)$, and $g^2(f(x_i;\beta), z_i;\gamma)\sigma^2$ is the variance of $h(y_i;\lambda)$.

By Johnson and Kotz (1994), using well-known formulas for the truncated normal distributions, it can show that suppose $\varepsilon_i \sim N(\mu_i(\theta), \sigma_i(\theta))$ has a normal distribution and lies within the interval $\varepsilon_i \in (a_i, b_i)$. Set $a_i' = [a_i - \mu_i(\theta)]/\sigma_i(\theta)$, $b_i' = [b_i - \mu_i(\theta)]/\sigma_i(\theta)$, ε_i conditional on $a_i < \varepsilon_i < b_i$ has a truncated normal distribution with probability density function

$$f(\varepsilon_i; \mu_i(\theta), \sigma_i(\theta), a_i, b_i) = \frac{\phi((\varepsilon_i - \mu_i(\theta))/\sigma_i(\theta)}{\sigma_i(\theta)\,[\Phi(b_i') - \Phi(a_i')]}. \tag{54}$$

Then

$$E_\theta(\varepsilon_i|\{a_i < \varepsilon_i < b_i\}) = \mu_i(\theta) + \frac{\phi(a_i') - \phi(b_i')}{\Phi(b_i') - \Phi(a_i')}\,\sigma_i(\theta) \tag{55}$$

and

$$Var_\theta(\varepsilon_i|\{a_i < \varepsilon_i < b_i\}) = \sigma_i^2(\theta)\left[1 + \frac{a_i'\phi(a_i') - b_i'\phi(b_i')}{\Phi(b_i') - \Phi(a_i')} - \left\{\frac{\phi(a_i') - \phi(b_i')}{\Phi(b_i') - \Phi(a_i')}\right\}^2\right]. \tag{56}$$

Since simultaneously solving equations (55) and (56), it will take too much time to evaluate the MLEs and the corresponding likelihood inference.

Consider the following transformed truncated normal median regression model:

$$h(y_i;\lambda) = f(x_i;\beta) + \varepsilon_i \tag{57}$$

where ε_is are independent errors distributed as either $N(0, g^2(f(x_i;\beta), z_i;\gamma)\,\sigma^2)$ or truncated $N(\mu_i(\theta), \sigma_i^2(\theta))$, Notice that, for $i = 1, \ldots, n$, $f(x_i;\beta)$ is the median of $h(y_i;\lambda)$, and $g(f(x_i;\beta), z_i;\gamma)\,\sigma$ is the interquartile range of $h(y_i;\lambda)$.

One way to obtain $\mu_i(\theta)$s is to utilize the Newton-Raphson method, but $\mu_i(\theta)$ generally it has no closed-form to be evaluated directly. It will take too much time to evaluate $\mu_i(\theta)$.

In this paper, we propose the transformed truncated normal mode regression model. The important advantage of our model is that the MLEs are easy and fast to compute. In the proposed model, we utilize the MLEs and likelihood function to do hypothesis testing and statistic intervals, and we compare the MLEs under truncated normality assumption with the MLEs under false normality assumption.

Under the false normality assumption, the log-likelihood function for θ is

$$\log[L(\theta)] \equiv \ell(\theta) \equiv \sum_{i=1}^{n} \ell_i(\theta), \tag{58}$$

where

$$\ell_i(\theta) \;=\; \log[\phi(e_i(\theta))] + \log[h_i'(\lambda)] - \log[g_i(\beta,\gamma)] - \log(\sigma). \tag{59}$$

Then the score function for θ is

$$\frac{\partial \ell(\theta)}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial \ell_i(\theta)}{\partial \theta} \equiv \sum_{i=1}^{n} S_i(\theta) \equiv S(\theta). \tag{60}$$

We compare equations (59) and (60) with equations (16) and (19).

Consider the standard deviation is fixed, if the sample size is not large enough, the difference between the score function for θ under the false normality assumption and under the truncated normality assumption will be small. Hence, the MLEs under the false normality assumption are similiar with under the truncated normality assumption.

Consider the sample size is fixed, if the standard deviation is very small, $e_i(b;\theta)$ tends to be ∞ and $e_i(a;\theta)$ tends to be $-\infty$ generally. Thus, the difference between the score function for θ under the false normality assumption and under the truncated normality assumption will be small. Hence, the MLEs under the false normality assumption are similiar with under the truncated normality assumption.

In Tables 2-5, there is no significant differences between the MLEs under the false normality assumption and the truncated normality assumption. A possible reason is that the sample size in Example 3.1 is not large enough.

In Tables 6-10, the MLEs under the false normality assumption are nearly the same as under the truncated normality assumption. Some possible reasons are that λ and σ are closed to 0, and the sample size is also not large enough in Example 3.2.

When the range of the response transformation is possibly different from $\mathcal{R}$, the likelihood inference under the coventional normality assumption is inappropriate and thus should not be used. Therefore, when the range of the response transformation is possibly different from $\mathcal{R}$, we may assume that the proposed model holds and the likelihood inference under the proposed model in Section 2 can be used.

References

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

[2] Carroll, R. J. and Ruppert, D. (1988) *Transformation and Weighting in Regression*. Chapman and Hall, New York.

[3] Chen, C.-R. and Wang, L.-C. (2003) Likelihood inference under the general response transformation model with heteroscedastic errors. *Taiwanese Journal of Mathematics*, **7**, 261–273.

[4] John, J. A. and Draper, N. R. (1980) An alternative family of transformations. *Applied Statistics*, **29**, 190–197.

[5] Johnson, N. L. and Kotz, S. (1994) *Distributions in Statistics*: *Continuous Univariate Distributions*. John Wiley & Sons, New York.

[6] McLachlan, G. J. (2008) *The EM Algorithm and Extensions*. John Wiley & Sons, Hoboken.

[7] Mosteller, F. and Tukey, J. W. (1977) *Data Analysis and Linear Regression*. Addison-Wesley, Reading, Massachusetts.

[8] Poirier, D. J. (1978) The use of the Box–Cox transformation in limited dependent variable models. *Journal of the American Statistical Association*, **73**, 284–287.

[9] Prakasa Rao, B. L. S. (1999) *Semimartingales and Their Statistical Inference*. Chapman & Hall/CRC, Boca Raton.

[10] Tukey, J. W. (1957) On the comparative anatomy of transformations. *The Annals of Mathematical Statistics*, **28**, 602–632.

Appendix A

For $i = 1, \ldots, n$,

$$\begin{aligned} S_i(\theta) &= \frac{\phi'(e_i(\theta))[\partial e_i(\theta)/\partial\theta]}{\phi(e_i(\theta))} + \frac{\partial h_i'(\lambda)/\partial\theta}{h_i'(\lambda)} - \frac{\partial\sigma/\partial\theta}{\sigma} - \frac{\partial g(\beta,\gamma)/\partial\theta}{g(\beta,\gamma)} \\ &\quad - \frac{\partial\Phi(e_i(b;\theta))/\partial\theta - \partial\Phi(e_i(a;\theta))/\partial\theta}{\Phi(e_i(b;\theta)) - \Phi(e_i(a;\theta))}, \end{aligned}$$

where

$$\phi'(e_i(\theta)) = -e_i(\theta)\phi(e_i(\theta))$$

and for $e_i(u;\theta) \in \mathcal{R}$,

$$\begin{aligned} \frac{\partial e_i(u;\theta)}{\partial\lambda} &= \frac{\partial h(u;\lambda)/\partial\lambda}{\sigma\, g_i(\beta,\gamma)}, \\ \frac{\partial e_i(u;\theta)}{\partial\beta} &= -\frac{\partial f(x_i;\beta)/\partial\beta}{\sigma\, g_i(\beta,\gamma)} - \frac{h(u;\lambda) - f(x_i;\beta)}{\sigma\, g_i^2(\beta,\gamma)} \frac{\partial g_i(\beta,\gamma)}{\partial\beta}, \\ \frac{\partial e_i(u;\theta)}{\partial\sigma} &= -\frac{e_i(u;\theta)}{\sigma}, \\ \frac{\partial e_i(u;\theta)}{\partial\gamma} &= -\frac{e_i(u;\theta)}{g_i(\beta,\gamma)} \frac{\partial g_i(\beta,\gamma)}{\partial\gamma}, \\ \frac{\partial\Phi(e_i(u;\theta))}{\partial\theta} &= \phi(e_i(u;\theta)) \frac{\partial e_i(u;\theta)}{\partial\theta} 1_{\mathcal{R}}(e_i(u;\theta)). \end{aligned}$$

As an example, when $a = 0$, $b = \infty$, $h(u;\lambda) = u^{(\lambda)}$, $f_i(\beta) = x_i^T\beta$, and $g_i(\beta,\gamma) = 1$ for $u \in (0,\infty)$ and $i = 1, \ldots, n$,

$$S_i(\theta) = -e_i(\theta)\frac{\partial e_i(\theta)}{\partial\theta} + y_i^{1-\lambda}\frac{\partial h_i'(\lambda)}{\partial\theta} - \sigma^{-1}\frac{\partial\sigma}{\partial\theta} + \frac{\phi(e_i(0;\theta))}{1 - \Phi(e_i(0;\theta))}\frac{\partial e_i(0;\theta))}{\partial\theta},$$

where

$$\begin{aligned}
e_i(\theta) &= \frac{y_i^{(\lambda)} - x_i^T\beta}{\sigma}, \\
\frac{\partial e_i(\theta)}{\partial \lambda} &= \frac{\log(y_i)y_i^{\lambda} - y_i^{(\lambda)}}{\sigma\lambda}, \\
\frac{\partial e_i(\theta)}{\partial \beta} &= -\frac{x_i}{\sigma}, \\
\frac{\partial e_i(\theta)}{\partial \sigma} &= -\frac{e_i(\theta)}{\sigma}, \\
e_i(0;\theta) &= -\frac{1/\lambda + x_i^T\beta}{\sigma}, \\
\frac{\partial e_i(0;\theta)}{\partial \lambda} &= \frac{1}{\sigma\lambda^2}, \\
\frac{\partial e_i(0;\theta)}{\partial \beta} &= -\frac{x_i}{\sigma}, \\
\frac{\partial e_i(0;\theta)}{\partial \sigma} &= -\frac{e_i(0;\theta)}{\sigma}, \\
\frac{\partial \sigma}{\partial \sigma} &= 1, \\
h_i'(\lambda) &= y_i^{\lambda-1},
\end{aligned}$$

and

$$\frac{\partial h_i'(\lambda)}{\partial \lambda} = \log(y_i)y_i^{\lambda-1}.$$

Appendix B

For $i = 1, \ldots, n$,

$$\begin{aligned}
J_i(\theta) \;=\; & -\frac{\phi''(e_i(\theta))[\partial e_i(\theta)/\partial\theta][\partial e_i(\theta)/\partial\theta^T] + \phi'(e_i(\theta))[\partial^2 e_i(\theta)/\partial\theta\partial\theta^T]}{\phi(e_i(\theta))} \\
& + \frac{[\phi'(e_i(\theta))]^2[\partial e_i(\theta)/\partial\theta][\partial e_i(\theta)/\partial\theta^T]}{\phi^2(e_i(\theta))} - \frac{\partial^2 h_i'(\lambda)/\partial\theta\partial\theta^T}{h_i'(\lambda)} \\
& + \frac{[\partial h_i'(\lambda)/\partial\theta][\partial h_i'(\lambda)/\partial\theta^T]}{(h_i')^2(\lambda)} + \frac{\partial^2\sigma/\partial\theta\partial\theta^T}{\sigma} - \frac{[\partial\sigma/\partial\theta][\partial\sigma/\partial\theta^T]}{\sigma^2} \\
& + \frac{\partial^2 g(\beta,\gamma)/\partial\theta\partial\theta^T}{g(\beta,\gamma)} - \frac{[\partial g(\beta,\gamma)/\partial\theta][\partial g(\beta,\gamma)/\partial\theta^T]}{g^2(\beta,\gamma)} \\
& + \frac{\partial^2\Phi(e_i(b;\theta))/\partial\theta\partial\theta^T - \partial^2\Phi(e_i(a;\theta))/\partial\theta\partial\theta^T}{\Phi(e_i(b;\theta)) - \Phi(e_i(a;\theta))} \\
& - \frac{[\partial\Phi(e_i(b;\theta))/\partial\theta - \partial\Phi(e_i(a;\theta))/\partial\theta][\partial\Phi(e_i(b;\theta))/\partial\theta^T - \partial\Phi(e_i(a;\theta))/\partial\theta^T]}{[\Phi(e_i(b;\theta)) - \Phi(e_i(a;\theta))]^2},
\end{aligned}$$

where

$$\begin{aligned}
\phi'(e_i(\theta)) &= -e_i(\theta)\phi(e_i(\theta)), \\
\phi''(e_i(\theta)) &= -\phi(e_i(\theta)) + e_i^2(\theta)\phi(e_i(\theta))
\end{aligned}$$

and for $e_i(u;\theta) \in \mathcal{R}$,

$$
\begin{aligned}
\phi'(e_i(\theta)) &= -e_i(\theta)\phi(e_i(\theta)),\\
\phi''(e_i(\theta)) &= -\phi(e_i(\theta)) + e_i^2(\theta)\phi(e_i(\theta)),\\
\frac{\partial e_i(u;\theta)}{\partial\lambda} &= \frac{\partial h(u;\lambda)/\partial\lambda}{g_i(\beta,\gamma)\,\sigma},\\
\frac{\partial e_i(u;\theta)}{\partial\beta} &= -\frac{\partial f_i(\beta)/\partial\beta}{g_i(\beta,\gamma)\,\sigma} - \frac{h(u;\lambda)-f_i(\beta)}{g_i^2(\beta,\gamma)\,\sigma}\frac{\partial g_i(\beta,\gamma)}{\partial\beta},\\
\frac{\partial e_i(u;\theta)}{\partial\sigma} &= -\frac{e_i(u;\theta)}{\sigma},\\
\frac{\partial e_i(u;\theta)}{\partial\gamma} &= -\frac{e_i(u;\theta)}{g_i(\beta,\gamma)}\frac{\partial g_i(\beta,\gamma)}{\partial\gamma},\\
\frac{\partial^2 e_i(u;\theta)}{\partial\lambda\partial\lambda^T} &= \frac{\partial^2 h(u;\lambda)/\partial\lambda\partial\lambda^T}{g_i(\beta,\gamma)\,\sigma},\\
\frac{\partial^2 e_i(u;\theta)}{\partial\beta\partial\beta^T} &= -\frac{\partial^2 f_i(\beta)/\partial\beta\partial\beta^T}{g_i(\beta,\gamma)\,\sigma} + \frac{\partial f_i(\beta)/\partial\beta}{g_i^2(\beta,\gamma)\,\sigma}\frac{\partial g_i(\beta,\gamma)}{\partial\beta^T} + \frac{\partial f_i(\beta)/\partial\beta}{g_i^2(\beta,\gamma)\,\sigma}\frac{\partial g_i(\beta,\gamma)}{\partial\beta^T}\\
&\quad +2\,\frac{h(u;\lambda)-f_i(\beta)}{g_i^3(\beta,\gamma)\,\sigma}\frac{\partial g_i(\beta,\gamma)}{\partial\beta}\frac{\partial g_i(\beta,\gamma)}{\partial\beta^T} - \frac{h(u;\lambda)-f_i(\beta)}{g_i^2(\beta,\gamma)\,\sigma}\frac{\partial^2 g_i(\beta,\gamma)}{\partial\beta\partial\beta^T},\\
\frac{\partial^2 e_i(u;\theta)}{\partial\sigma^2} &= 2\frac{e_i(u;\theta)}{\sigma^2},\\
\frac{\partial^2 e_i(u;\theta)}{\partial\gamma\partial\gamma^T} &= 2\frac{e_i(u;\theta)}{g_i^2(\beta,\gamma)}\frac{\partial g_i(\beta,\gamma)}{\partial\gamma}\frac{\partial g_i(\beta,\gamma)}{\partial\gamma^T} - \frac{e_i(u;\theta)}{g_i(\beta,\gamma)}\frac{\partial^2 g_i(\beta,\gamma)}{\partial\gamma\partial\gamma^T},\\
\frac{\partial^2 e_i(u;\theta)}{\partial\lambda\partial\beta^T} &= -\frac{\partial h(u;\lambda)/\partial\lambda}{g_i^2(\beta,\gamma)\,\sigma}\frac{\partial g_i(\beta,\gamma)}{\partial\beta^T},\\
\frac{\partial^2 e_i(u;\theta)}{\partial\lambda\partial\sigma} &= -\frac{\partial h(u;\lambda)/\partial\lambda}{g_i(\beta,\gamma)\,\sigma^2},\\
\frac{\partial^2 e_i(u;\theta)}{\partial\lambda\partial\gamma^T} &= -\frac{\partial h(u;\lambda)/\partial\lambda}{g_i^2(\beta,\gamma)\sigma}\frac{\partial g_i(\beta,\gamma)}{\partial\gamma},\\
\frac{\partial^2 e_i(u;\theta)}{\partial\beta\partial\sigma} &= \frac{\partial f_i(\beta)/\partial\beta}{g_i(\beta,\gamma)\,\sigma^2} + \frac{h(u;\lambda)-f_i(\beta)}{g_i^2(\beta,\gamma)\,\sigma^2}\frac{\partial g_i(\beta,\gamma)}{\partial\beta},\\
\frac{\partial^2 e_i(u;\theta)}{\partial\beta\partial\gamma^T} &= \frac{\partial f_i(\beta)/\partial\beta}{g_i^2(\beta,\gamma)\,\sigma}\frac{\partial g_i(\beta,\gamma)}{\partial\gamma^T} + 2\frac{h(u;\lambda)-f_i(\beta)}{g_i^3(\beta,\gamma)\,\sigma}\frac{\partial g_i(\beta,\gamma)}{\partial\beta}\frac{\partial g_i(\beta,\gamma)}{\partial\gamma^T}\\
&\quad -\frac{h(u;\lambda)-f_i(\beta)}{g_i^2(\beta,\gamma)\,\sigma}\frac{\partial^2 g_i(\beta,\gamma)}{\partial\beta\partial\gamma^T},\\
\frac{\partial^2 e_i(u;\theta)}{\partial\sigma\partial\gamma} &= \frac{e_i(u;\theta)}{g_i(\beta,\gamma)\,\sigma}\frac{\partial g_i(\beta,\gamma)}{\partial\gamma},\\
\frac{\partial\Phi(e_i(u;\theta))}{\partial\theta} &= \phi(e_i(u;\theta))\frac{\partial e_i(u;\theta)}{\partial\theta}\,1_{\mathcal{R}}(e_i(u;\theta)),
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial^2\Phi(e_i(u;\theta))}{\partial\theta\partial\theta^T} &= -e_i(u;\theta)\phi(e_i(u;\theta))\frac{\partial e_i(u;\theta)}{\partial\theta}\frac{\partial e_i(u;\theta)}{\partial\theta^T}\,1_{\mathcal{R}}(e_i(u;\theta))\\
&\quad +\phi(e_i(u;\theta))\frac{\partial^2 e_i(u;\theta)}{\partial\theta\partial\theta^T}\,1_{\mathcal{R}}(e_i(u;\theta)).
\end{aligned}
$$

As an example, when $a=0$, $b=\infty$, $h(u;\lambda)=u^{(\lambda)}$, $f_i(\beta)=x_i^T\beta$, and $g_i(\beta,\gamma)=1$ for $u\in(0,\infty)$ and $i=1,\ldots,n$,

$$\begin{aligned}
J_i(\theta) &= [1-e_i^2(\theta)][\partial e_i(\theta)/\partial\theta][\partial e_i(\theta)/\partial\theta^T] + e_i(\theta)[\partial^2 e_i(\theta)/\partial\theta\partial\theta^T] \\
&\quad + e_i^2(\theta)[\partial e_i(\theta)/\partial\theta][\partial e_i(\theta)/\partial\theta^T] - y_i^{1-\lambda}\frac{\partial^2 h_i'(\lambda)}{\partial\theta\partial\theta^T} \\
&\quad + y_i^{2\,(1-\lambda)}\frac{\partial h_i'(\lambda)}{\partial\theta}\frac{\partial h_i'(\lambda)}{\partial\theta^T} + \sigma^{-1}\frac{\partial^2\sigma}{\partial\theta\partial\theta^T} - \sigma^{-2}\frac{\partial\sigma}{\partial\theta}\frac{\partial\sigma}{\partial\theta^T} \\
&\quad + \frac{e_i(0;\theta)\phi(e_i(0;\theta))}{1-\Phi(e_i(0;\theta))}\frac{\partial e_i(0;\theta)}{\partial\theta}\frac{\partial e_i(0;\theta)}{\partial\theta^T} - \frac{\phi(e_0(u;\theta))}{1-\Phi(e_i(0;\theta))}\frac{\partial^2 e_i(0;\theta)}{\partial\theta\partial\theta^T} \\
&\quad - \left[\frac{\phi(e_i(0;\theta))}{1-\Phi(e_i(0;\theta))}\right]^2\frac{\partial e_i(0;\theta)}{\partial\theta}\frac{\partial e_i(0;\theta)}{\partial\theta^T},
\end{aligned}$$

where

$$\begin{aligned}
e_i(\theta) &= \frac{y_i^{(\lambda)} - x_i^T\beta}{\sigma}, \\
\frac{\partial e_i(\theta)}{\partial\lambda} &= \frac{\log(y_i)y_i^{\lambda} - y_i^{(\lambda)}}{\sigma\,\lambda}, \\
\frac{\partial e_i(\theta)}{\partial\beta} &= -\frac{x_i}{\sigma}, \\
\frac{\partial e_i(\theta)}{\partial\sigma} &= -\frac{e_i(\theta)}{\sigma}, \\
\frac{\partial^2 e_i(\theta)}{\partial\lambda^2} &= \frac{[\log(y_i)]^2 y_i^{\lambda} - [\log(y_i)y_i^{(\lambda)} - y_i^{(\lambda)}]/\lambda}{\sigma\,\lambda} - \frac{\log(y_i)y_i^{\lambda} - y_i^{(\lambda)}}{\sigma\,\lambda^2}, \\
\frac{\partial^2 e_i(\theta)}{\partial\beta^2} &= 0, \\
\frac{\partial^2 e_i(\theta)}{\partial\sigma^2} &= 2\frac{e_i(\theta)}{\sigma^2}, \\
\frac{\partial^2 e_i(\theta)}{\partial\lambda\partial\beta} &= 0, \\
\frac{\partial^2 e_i(\theta)}{\partial\lambda\partial\sigma} &= -\frac{\log(y_i)y_i^{\lambda} - y_i^{(\lambda)}}{\sigma^2\lambda}, \\
\frac{\partial^2 e_i(\theta)}{\partial\beta\partial\sigma} &= \frac{x_i}{\sigma^2}, \\
e_i(0;\theta) &= -\frac{1/\lambda + x_i^T\beta}{\sigma},
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial e_i(0;\theta)}{\partial \lambda} &= \frac{1}{\sigma\,\lambda^2}, \\
\frac{\partial e_i(0;\theta)}{\partial \beta} &= -\frac{x_i}{\sigma}, \\
\frac{\partial e_i(0;\theta)}{\partial \sigma} &= -\frac{e_i(0;\theta)}{\sigma}. \\
\frac{\partial^2 e_i(0;\theta)}{\partial \lambda^2} &= -\frac{2}{\sigma\,\lambda^3}, \\
\frac{\partial^2 e_i(0;\theta)}{\partial \beta \partial \beta^T} &= 0, \\
\frac{\partial^2 e_i(0;\theta)}{\partial \sigma^2} &= 2\,\frac{e_i(0;\theta)}{\sigma^2}, \\
\frac{\partial^2 e_i(0;\theta)}{\partial \lambda \partial \beta} &= 0, \\
\frac{\partial^2 e_i(0;\theta)}{\partial \lambda \partial \sigma} &= -\frac{1}{\sigma^2\,\lambda^2}, \\
\frac{\partial^2 e_i(0;\theta)}{\partial \beta \partial \sigma} &= \frac{x_i}{\sigma^2}, \\
\frac{\partial \sigma}{\partial \sigma} &= 1, \\
h_i'(\lambda) &= y_i^{\lambda-1}, \\
\frac{\partial h_i'(\lambda)}{\partial \lambda} &= \log(y_i) y_i^{\lambda-1},
\end{aligned}
$$

and

$$
\frac{\partial^2 h_i'(\lambda)}{\partial \lambda^2} = [\log(y_i)]^2 y_i^{\lambda-1}.
$$

Appendix C

Consider the following Box-Cox transformed truncated normal two-way ANOVA model:

$$y_{ijk}^{(\lambda)} = \mu_{ij} + \varepsilon_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}$$

for $i = 1, \ldots, a$; $j = 1, \ldots, b$; and $k = 1, \ldots, n$, where $a, b, n \in \{2, 3, \ldots\}$ and ε_{ijk}s are independent errors distributed as either $N(0, \sigma^2)$ or truncated $N(0, \sigma^2)$ with unknown positive standard deviation σ.

(i) Choose several initial values $\hat{\lambda}^{(0)}$s in a non-empty set S, e.g., $S = \{-2, -7/4, -3/2, -5/4, -1, -3/4, -1/2, -1/4, 0, 1/4, 1/2, 3/4, 1, 5/4, 3/2, 7/4, 2\}$.

(ii) For each $\hat{\lambda}^{(0)}$ in S, choose the initial values

$$\begin{aligned}
\hat{\mu}^{(0)} &\equiv \overline{y}_{...}^{(\hat{\lambda}^{(0)})}, \\
\widehat{\tau_i}^{(0)} &\equiv \overline{y}_{i..}^{(\hat{\lambda}^{(0)})} - \overline{y}_{...}^{(\hat{\lambda}^{(0)})}, \\
\widehat{\beta_j}^{(0)} &\equiv \overline{y}_{.j.}^{(\hat{\lambda}^{(0)})} - \overline{y}_{...}^{(\hat{\lambda}^{(0)})}, \\
\widehat{(\tau\beta)}_{ij}^{(0)} &\equiv \overline{y}_{ij.}^{(\hat{\lambda}^{(0)})} - \overline{y}_{i..}^{(\hat{\lambda}^{(0)})} - \overline{y}_{.j.}^{(\hat{\lambda}^{(0)})} + \overline{y}_{...}^{(\hat{\lambda}^{(0)})},
\end{aligned}$$

and

$$\hat{\sigma}^{2(0)} \equiv \frac{1}{abn - ab - 1} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} \left[y_{ijk}^{(\hat{\lambda}^{(0)})} - \overline{y}_{ij.}^{(\hat{\lambda}^{(0)})} \right]^2$$

for $i = 1, \ldots, a$; $j = 1, \ldots, b$; and $k = 1, \ldots, n$, where

$$\begin{aligned}
\overline{y}_{...} &\equiv \frac{1}{abn} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk}, \\
\overline{y}_{i..} &\equiv \frac{1}{bn} \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk}, \\
\overline{y}_{.j.} &\equiv \frac{1}{an} \sum_{i=1}^{a} \sum_{k=1}^{n} y_{ijk},
\end{aligned}$$

and

$$\overline{y}_{ij.} \equiv \frac{1}{n} \sum_{k=1}^{n} y_{ijk}.$$

(iii) Denote these $\hat{\theta}^{(0)}$s as $\hat{\theta}^{(0,1)}, \hat{\theta}^{(0,2)}, \ldots, \hat{\theta}^{(0,|S|)}$, where $|S|$ denotes the number of elements in S. Choose $\hat{\theta}^{(0)}$ as $\hat{\theta}^{(0,\ell^*)}$ such that $\hat{\theta}^{(0,\ell^*)} = \max_{1 \le \ell \le |S|} \ell(\hat{\theta}^{(0,\ell)})$

In Example 3.1, when we choose $\hat{\lambda}^{(0)} = -3/4$, we have the largest log-likelihood function for $\hat{\theta}^{(0)}$, $\ell(\hat{\theta}^{(0)}) = 55.6467$, then we use the initial value to iterate the equation (23) in Section 2.2, finally we get the MLEs for θ.

Appendix D

Suppose that the transformed truncated normal mode two-way ANOVA model is

$$y_{ijk}^{(\lambda)} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}$$

for $i = 1, 2, 3$ and $j, k = 1, 2, 3, 4$, where each y_{ijk} has support $(0, \infty)$ and ε_{ijk}s are independent errors distributed as either $N(0, \sigma^2)$ or truncated $N(0, \sigma^2)$ with unknown positive standard deviation σ and with support $(-\infty, -1/\lambda - \mu - \tau_i - \beta_j - (\tau\beta)_{ij})$ for $\lambda < 0$. Thus, the c.d.f. of ε_{ijk}/σ is

$$P_\theta(\{\varepsilon_{ijk}/\sigma < u\}) = \frac{\Phi(u)}{\Phi([-1/\lambda - \mu - \tau_i - \beta_j - (\tau\beta)_{ij}]/\sigma)}.$$

By the probability integral transformation,

$$\frac{\Phi(\varepsilon_{ijk}/\sigma)}{\Phi([-1/\lambda - \mu - \tau_i - \beta_j - (\tau\beta)_{ij}]/\sigma)} \sim \text{uniform}(0, 1),$$

which implies that

$$\Phi^{-1}\left(\frac{\Phi(\varepsilon_{ijk}/\sigma)}{\Phi([-1/\lambda - \mu - \tau_i - \beta_j - (\tau\beta)_{ij}]/\sigma})\right) \sim N(0, 1).$$

Table 1: Survival times (1 unit = 10 hours) of animals in a 3×4 factorial experiment.

	B (Treatment)			
A (Poison)	1	2	3	4
1	0.31	0.82	0.43	0.45
	0.45	1.10	0.45	0.71
	0.46	0.88	0.63	0.66
	0.43	0.72	0.76	0.62
2	0.36	0.92	0.44	0.56
	0.29	0.61	0.35	1.02
	0.40	0.49	0.31	0.71
	0.23	1.24	0.40	0.38
3	0.22	0.30	0.23	0.30
	0.21	0.37	0.25	0.36
	0.18	0.38	0.24	0.31
	0.23	0.29	0.22	0.33

Table 2: MLEs under the false normality assumption and the truncated normality assumption, respectively, for Example 3.1.

MLE	False Normality	Truncated Normality
$\hat{\lambda}$	−0.8073	−0.8077
$\hat{\mu}$	−1.4175	−1.4179
$\hat{\tau}_1$	0.6797	0.6799
$\hat{\tau}_2$	0.2878	0.2879
$\hat{\beta}_1$	−0.7383	−0.7386
$\hat{\beta}_2$	0.6451	0.6453
$\hat{\beta}_3$	−0.2778	−0.2779
$(\widehat{\tau\beta})_{11}$	0.1359	0.1360
$(\widehat{\tau\beta})_{12}$	−0.0658	−0.0659
$(\widehat{\tau\beta})_{13}$	0.2160	0.2161
$(\widehat{\tau\beta})_{21}$	−0.1043	−0.1043
$(\widehat{\tau\beta})_{22}$	0.1142	0.1142
$(\widehat{\tau\beta})_{23}$	−0.1234	−0.1234
$\hat{\sigma}$	0.3567	0.3569

Table 3: MLEs without interaction under the false normality assumption and the truncated normality assumption, respectively, for Example 3.1.

MLE	False Normality	Truncated Normality
$\hat{\lambda}$	-0.7440	-0.7441
$\hat{\mu}$	-1.3576	-1.3577
$\hat{\tau}_1$	0.6393	0.6394
$\hat{\tau}_2$	0.2696	0.2697
$\hat{\beta}_1$	-0.7383	-0.6938
$\hat{\beta}_2$	0.6123	0.6124
$\hat{\beta}_3$	-0.2778	-0.2644
$\hat{\sigma}$	0.3636	0.3637

Table 4: MLEs with $\lambda = -1$ under the false normality assumption and the truncated normality assumption, respectively, for Example 3.1.

MLE	False Normality	Truncated Normality
$\hat{\mu}$	-1.6232	-1.6228
$\hat{\tau}_1$	0.8213	0.8219
$\hat{\tau}_2$	0.3526	0.3524
$\hat{\beta}_1$	-0.8961	-0.8965
$\hat{\beta}_2$	0.7596	0.7608
$\hat{\beta}_3$	-0.3240	-0.3244
$(\widehat{\tau\beta})_{11}$	0.2111	0.2106
$(\widehat{\tau\beta})_{12}$	-0.1211	-0.1193
$(\widehat{\tau\beta})_{13}$	0.2632	0.2626
$(\widehat{\tau\beta})_{21}$	-0.1017	-0.1015
$(\widehat{\tau\beta})_{22}$	0.1126	0.1120
$(\widehat{\tau\beta})_{23}$	-0.1194	-0.1192
$\hat{\sigma}$	0.4235	0.4241

Table 5: MLEs without interaction and with $\lambda = -1$ under the false normality assumption and the truncated normality assumption, respectively, for Example 3.1.

MLE	False Normality	Truncated Normality
$\hat{\mu}$	-1.6232	-1.6216
$\hat{\tau}_1$	0.8213	0.8242
$\hat{\tau}_2$	0.3527	0.3512
$\hat{\beta}_1$	-0.7383	-0.8978
$\hat{\beta}_2$	0.7596	0.7635
$\hat{\beta}_3$	-0.3240	-0.3256
$\hat{\sigma}$	0.4607	0.4627

Table 6: Cycles to failure of worsted yarn: 3^3 factorial experiment without replication.

Factor levels			
x_1	x_2	x_3	Cycles to failure
−1	−1	−1	674
−1	−1	0	370
−1	−1	1	292
−1	0	−1	338
−1	0	0	266
−1	0	1	210
−1	1	−1	170
−1	1	0	118
−1	1	1	90
0	−1	−1	1414
0	−1	0	1198
0	−1	1	634
0	0	−1	1022
0	0	0	620
0	0	1	438
0	1	−1	442
0	1	0	332
0	1	1	220
1	−1	−1	3636
1	−1	0	3184
1	−1	1	2000
1	0	−1	1568
1	0	0	1070
1	0	1	566
1	1	−1	1140
1	1	0	884
1	1	1	360

Table 7: MLEs under the false normality assumption and the truncated normality assumption, respectively, for Example 3.2.

MLE	False Normality	Truncated Normality
$\hat{\lambda}$	-0.2158	-0.2158
$\hat{\beta}_0$	3.4929	3.4929
$\hat{\beta}_1$	0.2142	0.2142
$\hat{\beta}_2$	-0.1626	-0.1626
$\hat{\beta}_3$	-0.0954	-0.0954
$\hat{\beta}_{12}$	0.0541	-0.0541
$\hat{\beta}_{13}$	0.0232	-0.0232
$\hat{\beta}_{23}$	-0.0124	-0.0124
$\hat{\beta}_{11}$	-0.0219	0.0219
$\hat{\beta}_{22}$	-0.0030	0.0030
$\hat{\beta}_{33}$	-0.0164	-0.0164
$\hat{\sigma}$	0.0435	0.0435

Table 8: MLEs without quadratic terms under the false normality assumption and the truncated normality assumption , respectively, for Example 3.2.

MLE	False Normality	Truncated Normality
$\hat{\lambda}$	-0.0363	-0.0363
$\hat{\beta}_0$	5.6577	5.6577
$\hat{\beta}_1$	0.6611	0.6611
$\hat{\beta}_2$	-0.5010	-0.5010
$\hat{\beta}_3$	-0.2950	-0.2950
$\hat{\sigma}$	0.1541	0.1541

Table 9: MLEs with $\lambda = 0$ under the false normality assumption and the truncated normality assumption , respectively, for Example 3.2.

MLE	False Normality	Truncated Normality
$\hat{\beta}_0$	6.4763	6.4763
$\hat{\beta}_1$	0.8324	0.8324
$\hat{\beta}_2$	−0.6310	−0.6310
$\hat{\beta}_3$	−0.3716	−0.3716
$\hat{\beta}_{12}$	−0.0383	−0.0383
$\hat{\beta}_{13}$	−0.0684	−0.0684
$\hat{\beta}_{23}$	−0.0208	−0.0208
$\hat{\beta}_{11}$	−0.1275	−0.1275
$\hat{\beta}_{22}$	−0.0176	−0.0176
$\hat{\beta}_{33}$	−0.0466	−0.0466
$\hat{\sigma}$	0.1758	0.1758

Table 10: MLEs with $\lambda = 0$ and without quadratic terms under the false normality assumption and the truncated normality assumption, respectively, for Example 3.2.

MLE	False Normality	Truncated Normality
$\hat{\beta}_0$	6.3486	6.3486
$\hat{\beta}_1$	0.8323	0.8323
$\hat{\beta}_2$	-0.6310	-0.6310
$\hat{\beta}_3$	-0.3716	-0.3716
$\hat{\sigma}$	0.1950	0.1950

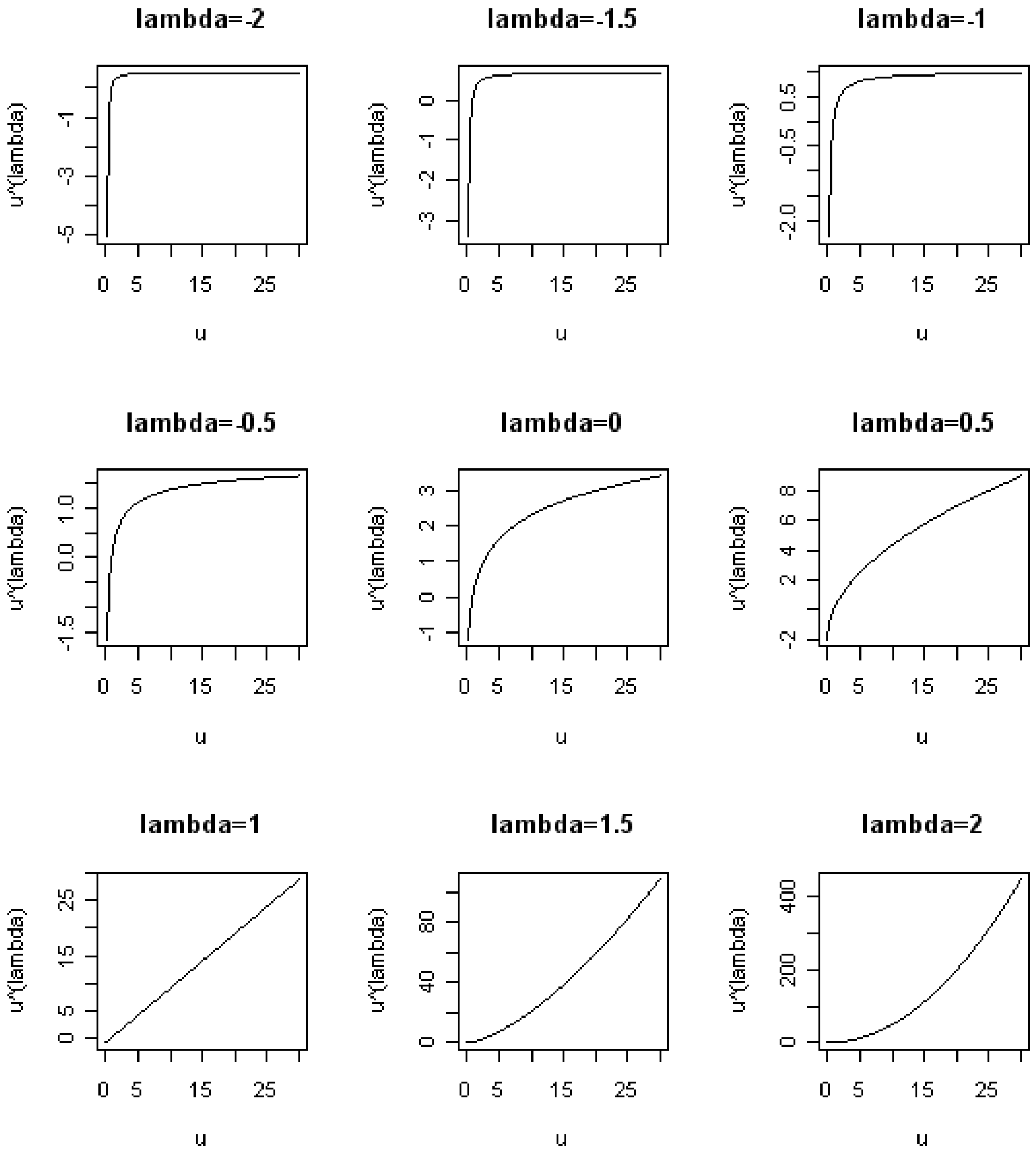
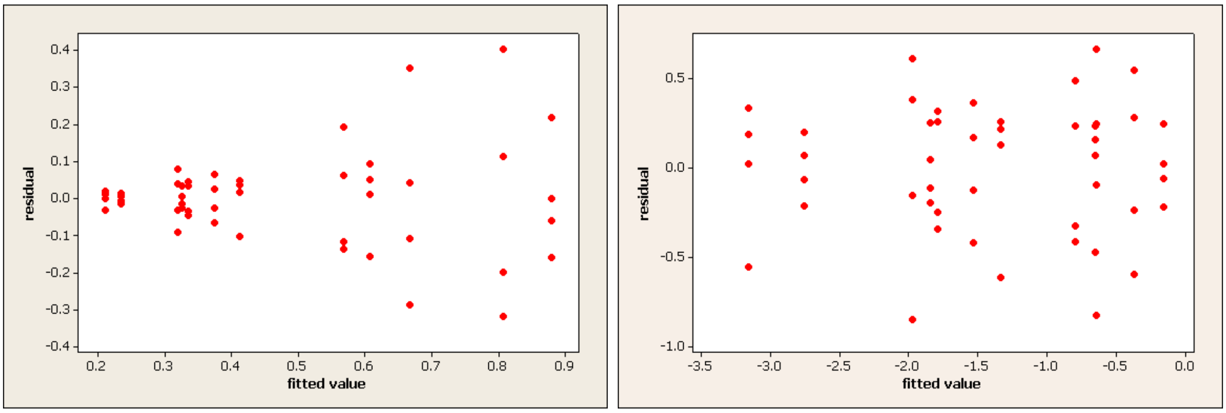


Figure 1: Some different modified power transformations.

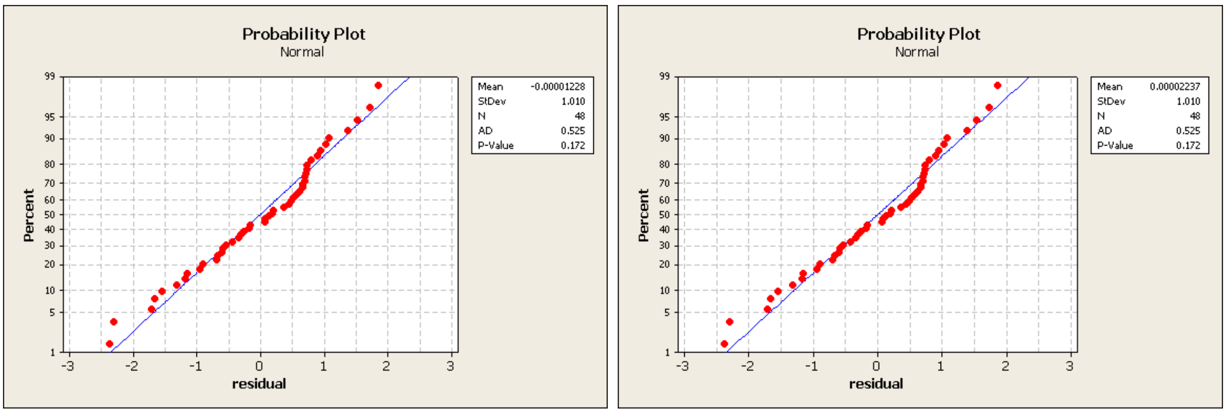


(a) (b)

Figure 2:

(a) Residual plot against fitted values for the original data under the two-way ANOVA effects model for Example 3.1.

(b) Residual plot against fitted values for the transformed data under the Box-Cox transformed mode regression model for Example 3.1.

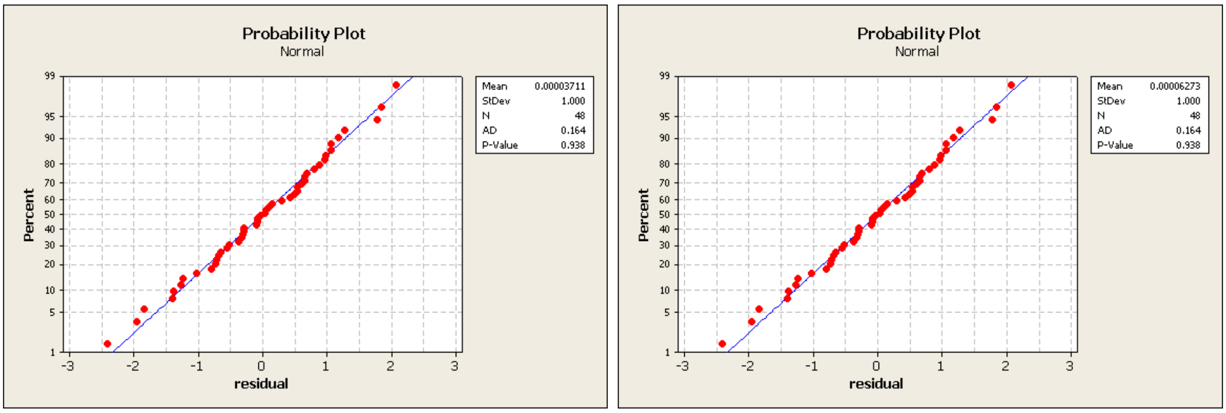


(a) (b)

Figure 3:

(a) Normal probability plot under the false normality assumption for Example 3.1.

(b) Normal probability plot under the truncated normality assumption for Example 3.1.

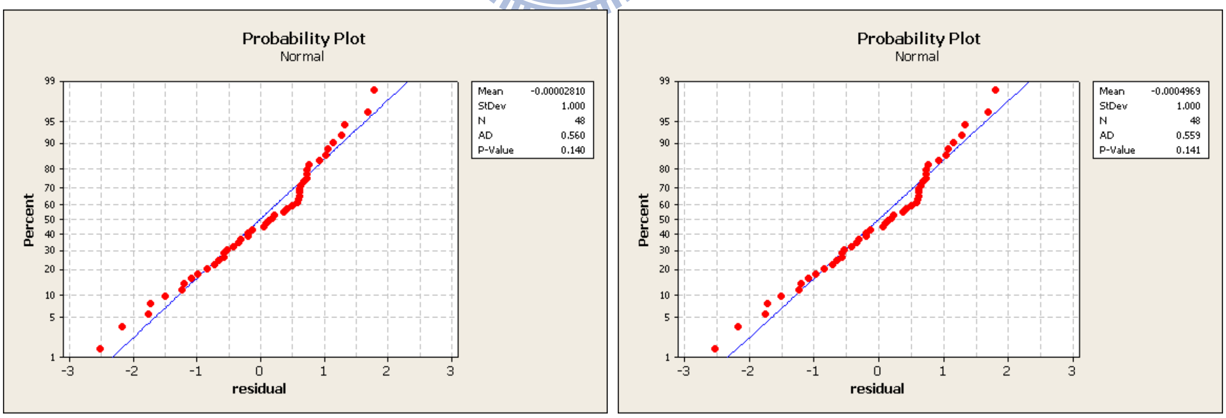


(a) (b)

Figure 4:

(a) Normal probability plot under the false normality assumption without interactions for Example 3.1.

(b) Normal probability plot under the truncated normality assumption without interactions for Example 3.1.

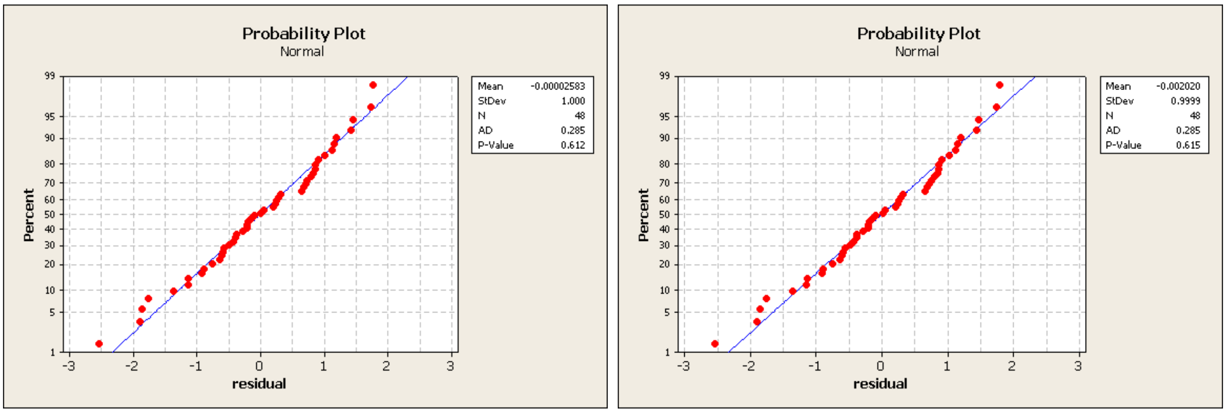


(a) (b)

Figure 5:

(a) Normal probability plot under the false normality assumption with $\lambda = -1$ for Example 3.1.

(b) Normal probability plot under the truncated normality assumption with $\lambda = -1$ for Example 3.1.

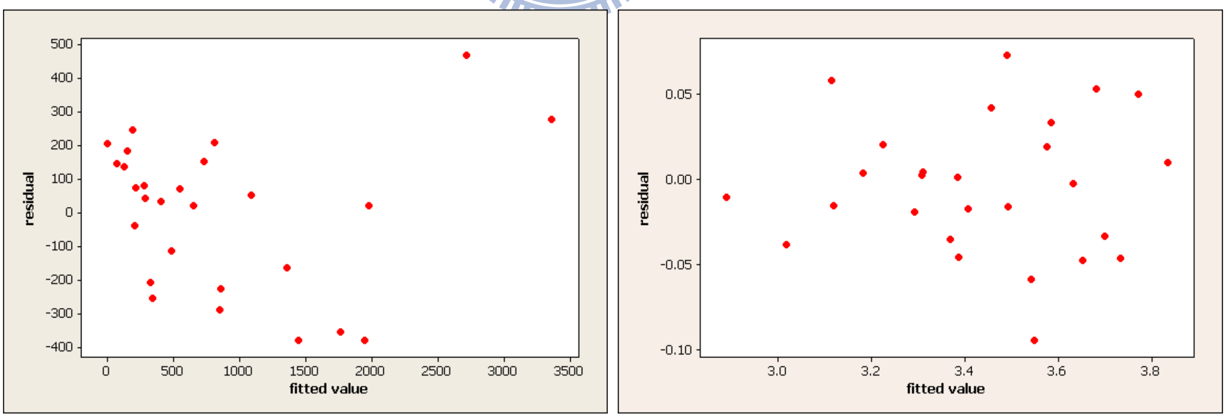


(a) (b)

Figure 6:

(a) Normal probability plot under the false normality assumption without interactions and with $\lambda = -1$ for Example 3.1.

(b) Normal probability plot under the truncated normality assumption without interactions and with $\lambda = -1$ for Example 3.1.

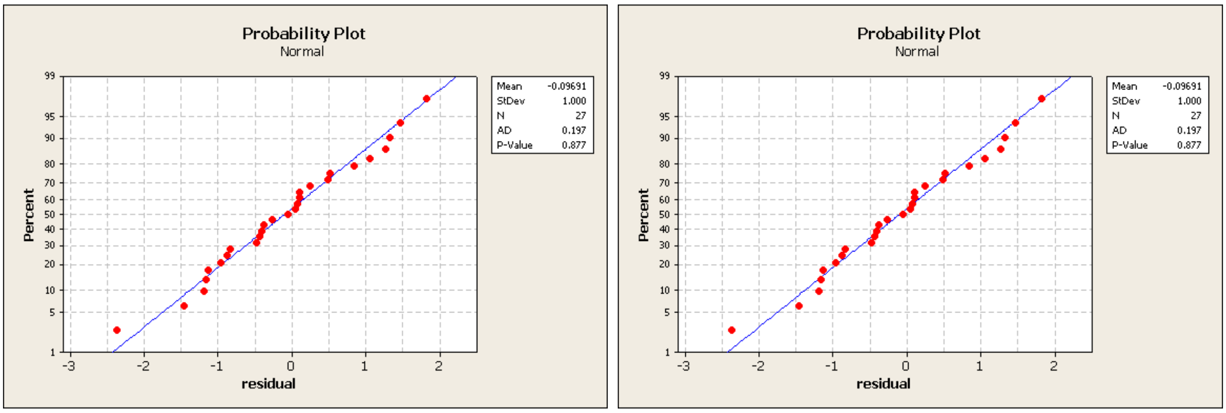


(a) (b)

Figure 7:

(a) Residual plot against fitted values for the original data under the quadratic regression model for Example 3.2.

(b) Residual plot against fitted values for the transformed data under the Box-Cox transformed mode regression model for Example 3.2.

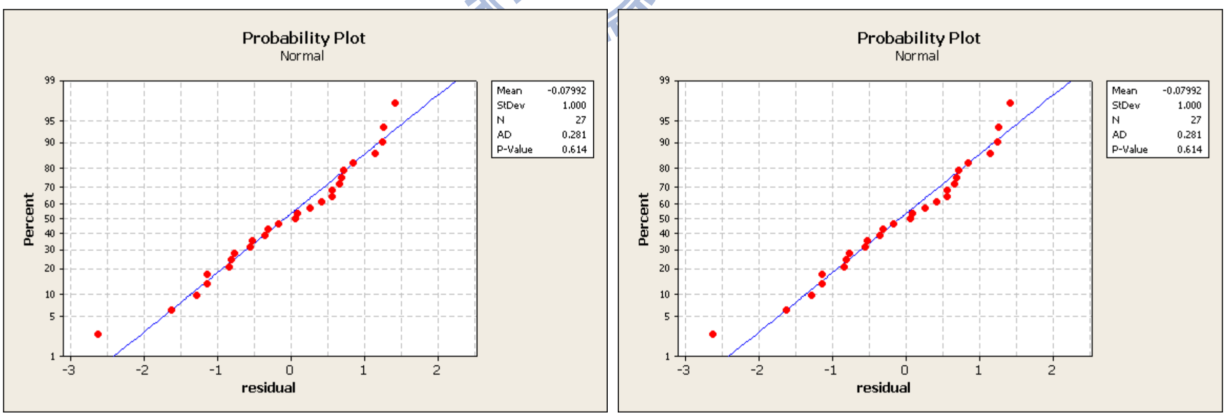


(a) (b)

Figure 8:

(a) Normal probability plot under the false normality assumption for Example 3.2.

(b) Normal probability plot under the truncated normality assumption for Example 3.2.

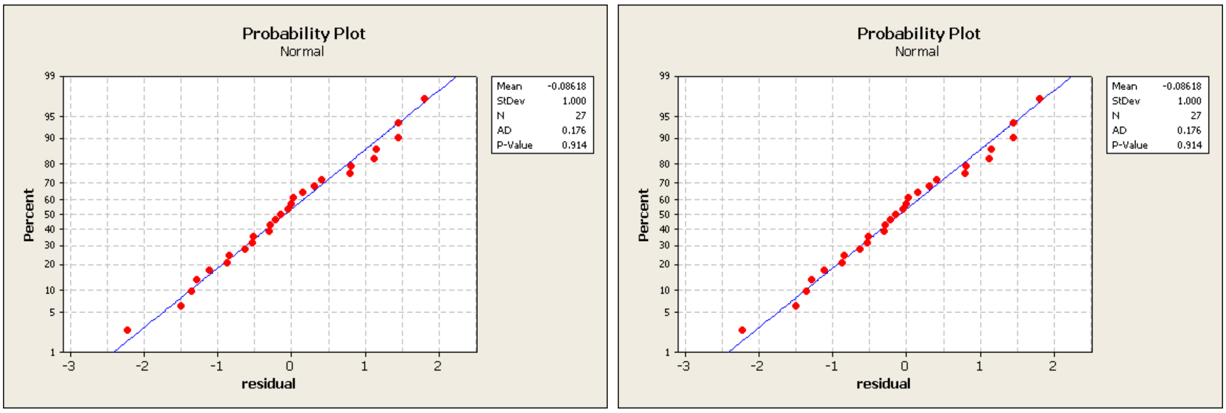


(a) (b)

Figure 9:

(a) Normal probability plot under the false normality assumption without quadratic effects and interactions for Example 3.2.

(b) Normal probability plot under the truncated normality assumption without quadratic effects and interactions for Example 3.2.

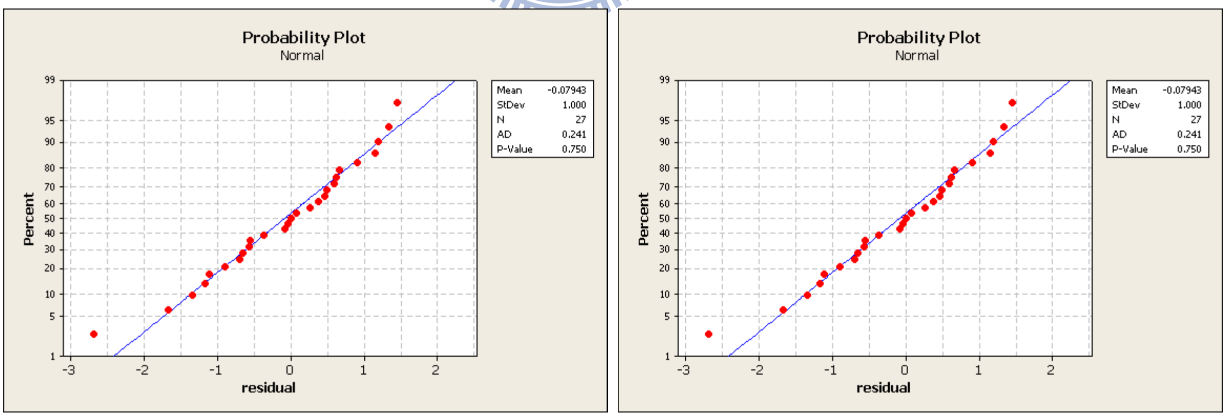


(a) (b)

Figure 10:

(a) Normal probability plot under the false normality assumption with $\lambda = 0$ for Example 3.2.

(b) Normal probability plot under the truncated normality assumption with $\lambda = 0$ for Example 3.2.



(a) (b)

Figure 11:

(a) Normal probability plot under the false normality assumption without quadratic effects and interactions and with $\lambda = 0$ for Example 3.2.

(b) Normal probability plot under the truncated normality assumption without quadratic effects and interactions and with $\lambda = 0$ for Example 3.2.