TAIWANESE JOURNAL OF MATHEMATICS
Vol. 7, No. 2, pp. 261-273, June 2003

LIKELIHOOD INFERENCE UNDER THE GENERAL RESPONSE TRANSFORMATION MODEL WITH HETEROSCEDASTIC ERRORS

Chih-Rung Chen and Lih-Chung Wang

Abstract. In this paper, we propose the likelihood inference under the general response transformation model with heteroscedastic errors when the range of the response transformation is possibly different from the whole real line. Three commonly used families of response transformations are reviewed to illustrate the importance and applicability of the proposed model.

1. Introduction

For modeling independent continuous data, it is a common practice simply to assume the following regression model: For $i = 1, \dots, n$,

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

where y_i is the observation for subject i, x_i is a known covariate vector for subject i, β is a finite-dimensional regression parameter vector, f is a known regression function of both x_i and β, and ε_i's are i.i.d. $N(0, \sigma^2)$ errors with unknown variance $\sigma^2 > 0$. Then $f(x_i; \beta)$ is not only the mean of observation y_i, but also its median for $i = 1, \dots, n$.

When there exist heteroscedastic errors and/or departures from normality in the data, a popular approach is to transform the response. Originally, the response transformation was proposed both as a means of achieving homoscedasticity and approximate normality and for inducing a simpler linear model for the transformed

Received January 9, 2002; revised March 6, 2002.
Communicated by S. B. Hsu.
2000 *Mathematics Subject Classification*: 62A10.
Key words and phrases: Likelihood inference, response transformation, heteroscedasticity, power transformation, folded power transformation, modulus power transformation.

response (Box and Cox, 1964). In such situations, we may assume the following response transformation model rather than model (1) for modeling independent continuous data: For $i = 1, \ldots, n$,

$$h(y_i; \lambda) = f(x_i; \beta) + \varepsilon_i, \tag{2}$$

where λ is a finite-dimensional response transformation parameter vector, $h(\cdot; \lambda)$ is a known strictly monotonic and differentiable response transformation, and ε_i's are i.i.d. $N(0, \sigma^2)$ errors with unknown variance $\sigma^2 > 0$. In the following, without loss of generality, we assume that $h(\cdot; \lambda)$ is a known strictly increasing and differentiable response transformation.

When both heteroscedastic errors and departures from normality cannot be removed simultaneously in the data by any single response transformation, model (2) is further generalized to

$$h(y_i; \lambda) = f(x_i; \beta) + g(f(x_i; \beta), z_i; \gamma)\, \varepsilon_i, \tag{3}$$

where z_i is a known covariate vector for subject i, γ is a finite-dimensional variance parameter vector, g is a known positive weight function of $f(x_i; \beta)$, z_i, and γ, and ε_i's are i.i.d. $N(0, 1)$ standardized errors.

However, if the range of response transformation $h(\cdot; \lambda)$ is different from $\mathcal{R}$ $(\equiv (-\infty, \infty))$, ε_i's in model (3) cannot be normally distributed. They don't even have the same distributions, due to the fact that they may have different supports. In this paper, we propose the general response transformation model with heteroscedastic errors by relaxing the assumption that all ε_i's in model (3) are identically and normally distributed.

In Section 2, three commonly used families of response transformations with ranges possibly different from $\mathcal{R}$ are reviewed. The general response transformation model with heteroscedastic errors is proposed. In Section 3, the likelihood inference under the proposed model is discussed thoroughly. Some concluding remarks are given in Section 4.

2. General Response Transformation Model with Heteroscedastic Errors

In this section, first of all, three commonly used families of response transformations with ranges possibly different from $\mathcal{R}$ are reviewed to illustrate the need to extend model (3) as follows:

Example 1. The family of power transformations (Box and Cox, 1964)

$$h(u; \lambda) = (u - \lambda_2)^{(\lambda_1)} = \begin{cases} \dfrac{(u - \lambda_2)^{\lambda_1} - 1}{\lambda_1}, & \lambda_1 \neq 0, \\ \log(u - \lambda_2), & \lambda_1 = 0, \end{cases} \tag{4}$$

is most frequently used in the literature to transform continuous data with supports contained in (λ_2, ∞), where $\lambda = (\lambda_1, \lambda_2)^T$. Then the range $h((\lambda_2, \infty); \lambda)$ of response transformation $h(\cdot; \lambda)$ is $(-\infty, -1/\lambda_1)$ for $\lambda_1 < 0$, $\mathcal{R}$ for $\lambda_1 = 0$, and $(-1/\lambda_1, \infty)$ for $\lambda_1 > 0$, respectively. Similarly, the family of response transformations

$$h(u; \lambda) = -(\lambda_2 - u)^{(\lambda_1)} = \begin{cases} \dfrac{1 - (\lambda_2 - u)^{\lambda_1}}{\lambda_1}, & \lambda_1 \neq 0, \\ -\log(\lambda_2 - u), & \lambda_1 = 0, \end{cases} \tag{5}$$

can be used to transform continuous data with supports contained in $(-\infty, \lambda_2)$, where $\lambda = (\lambda_1, \lambda_2)^T$. Then the range $h((-\infty, \lambda_2); \lambda)$ of response transformation $h(\cdot; \lambda)$ is $(1/\lambda_1, \infty)$ for $\lambda_1 < 0$, $\mathcal{R}$ for $\lambda_1 = 0$, and $(-\infty, 1/\lambda_1)$ for $\lambda_1 > 0$, respectively.

Example 2. The family of folded power transformations (Mosteller and Tukey, 1977)

$$h(u; \lambda) = \begin{cases} \dfrac{(u - \lambda_2)^{\lambda_1} - (\lambda_3 - u)^{\lambda_1}}{\lambda_1}, & \lambda_1 \neq 0, \\ \log\left(\dfrac{u - \lambda_2}{\lambda_3 - u}\right), & \lambda_1 = 0, \end{cases} \tag{6}$$

is frequently used in the literature to transform continuous data with supports contained in (λ_2, λ_3), where $\lambda = (\lambda_1, \lambda_2, \lambda_3)^T$. Then the range $h((\lambda_2, \lambda_3); \lambda)$ of response transformation $h(\cdot; \lambda)$ is $\mathcal{R}$ for $\lambda_1 \cdot\ 0$, and $(-(\lambda_3 - \lambda_2)^{\lambda_1}/\lambda_1, (\lambda_3 - \lambda_2)^{\lambda_1}/\lambda_1)$ for $\lambda_1 > 0$, respectively.

Example 3. The family of modulus power transformations (John and Draper, 1980)

$$\begin{aligned} h(u; \lambda) = &\operatorname{sgn}(u - \lambda_2)\,(|u - \lambda_2| + 1)^{(\lambda_1)} = \\ &\begin{cases} \operatorname{sgn}(u - \lambda_2) \left[\dfrac{(|u - \lambda_2| + 1)^{\lambda_1} - 1}{\lambda_1}\right], & \lambda_1 \neq 0, \\ \operatorname{sgn}(u - \lambda_2)\, \log(|u - \lambda_2| + 1), & \lambda_1 = 0, \end{cases} \end{aligned} \tag{7}$$

is frequently used in the literature to transform continuous data with supports contained in $\mathcal{R}$, where $\lambda = (\lambda_1, \lambda_2)^T$. Then the range $h(\mathcal{R}; \lambda)$ of response transformation $h(\cdot; \lambda)$ is $(1/\lambda_1, -1/\lambda_1)$ for $\lambda_1 < 0$ and $\mathcal{R}$ for $\lambda_1 \geq 0$, respectively.

Examples 1-3 above are three commonly used families of response transformations with ranges possibly different from $\mathcal{R}$ in the literature. In order to cover such kinds of families, we propose the following general response transformation model with heteroscedastic errors to extend model (3) for modeling independent continuous data: For $i = 1, \dots, n$,

$$h(y_i; \lambda) = f(x_i; \beta) + g(f(x_i; \beta), z_i; \gamma)\, \varepsilon_i, \tag{8}$$

where all assumptions are the same as model (3) except that ε_i's are assumed to be independent standardized errors with median 0 and distributions either $N(0, 1)$ or truncation of some $N(c_i(\beta, \lambda, \gamma), 1)$ with $c_i(\beta, \lambda, \gamma) \in \mathcal{R}$.

Note that, for $i = 1, \dots, n$, $h^{-1}(f(x_i; \beta); \lambda)$ is the median of y_i, but not necessarily the mean of y_i, which may have no closed-form formula to be evaluated directly or even may not exist. Thus, we are mainly concerned with the median rather than mean regression problem for original data y_i's. Moreover, for $i = 1, \dots, n$, ε_i is distributed as $N(0, 1)$ if and only if the support of $h(y_i; \lambda)$ is $\mathcal{R}$. When all supports of $h(y_i; \lambda)$'s are $\mathcal{R}$, the proposed model is exactly the same as model (3).

3. Likelihood Inference Under the General Response Transformation Model with Heteroscedastic Errors

In this section, the likelihood inference under the general response transformation model with heteroscedastic errors is discussed thoroughly as follows.

3.1 Maximum Likelihood Estimation

Let $\theta \equiv (\beta^T, \lambda^T, \gamma^T)^T$ be the d-dimensional parameter vector and let Θ be the corresponding parameter space. Assume that Θ is a non-empty open subset of $\mathcal{R}^d$ and that, for $i = 1, \dots, n$, y_i has a known support $(a_1(w_i), a_2(w_i))$ (e.g., $\mathcal{R}$, $(0, \infty)$, or $(0, 1)$) contained in the domain of response transformation $h(\cdot; \lambda)$, where w_i is a known covariate vector for subject i. Set $a_{1i} \equiv a_1(w_i)$ and $a_{2i} \equiv a_2(w_i)$. Let Φ be the cumulative distribution function (c.d.f.) of $N(0, 1)$, let ϕ be the probability density function (p.d.f.) of $N(0, 1)$, and set $(a_1, a_2) \equiv \bigcup_{i=1}^{n} (a_{1i}, a_{2i})$. For $u \in [a_{1i}, a_{2i}]$ and $i = 1, \dots, n$, set

$$e_i(u; \theta) \equiv \frac{h(u; \lambda) - f_i(\beta)}{g_i(\beta, \gamma)},$$

where $h(a_1; \lambda) \equiv \lim_{v \downarrow a_1} h(v; \lambda)$, $h(a_2; \lambda) \equiv \lim_{v \uparrow a_2} h(v; \lambda)$, $f_i(\beta) \equiv f(x_i; \beta)$ and $g_i(\beta, \gamma) \equiv g(f(x_i; \beta), z_i; \gamma)$. Since ε_i's have median 0 and distributions either $N(0, 1)$ or truncation of some $N(c_i(\theta), 1)$ with $c_i(\theta) \in \mathcal{R}$, $c_i(\theta)$ is the root of equation $G_i(t; \theta)|_{t = c_i(\theta)} = 0$ for $i = 1, \dots, n$, where

$$G_i(t; \theta) = \sum_{j=1}^{2} \Phi(e_i(a_{ji}; \theta) - t) - 2\,\Phi(-t)$$

with $\Phi(-\infty) \equiv 0$ and $\Phi(\infty) \equiv 1$. One way to obtain $c_i(\theta)$'s is to utilize the following Newton-Raphson method: For $i = 1, \ldots, n$, first choose a good initial value $c_i^{(0)}(\theta)$ (e.g., $c_i^{(0)}(\theta) = 0$) and then iterate the following equations

$$c_i^{(k+1)}(\theta) = c_i^{(k)}(\theta) - \frac{G_i(c_i^{(k)}(\theta);\theta)}{G_i'(c_i^{(k)}(\theta);\theta)}, \quad k = 0, 1, 2, \ldots,$$

until $c_i^{(k)}(\theta)$'s converge to $c_i(\theta)$, where

$$G_i'(t;\theta) = -\sum_{j=1}^{2} \phi(e_i(a_{ji};\theta) - t) + 2\,\phi(-t)$$

with $\phi(\pm\infty) \equiv 0$.

The p.d.f. of y_i is

$$p_i(y_i;\theta) = \frac{1_{(a_{1i},a_{2i})}(y_i)\,\phi(r_i(y_i;\theta))\,h'(y_i;\lambda)}{g_i(\beta,\gamma)\,[\Phi(r_i(a_{2i};\theta)) - \Phi(r_i(a_{1i};\theta))]} \tag{9}$$

for $i = 1, \ldots, n$, where $1_{(a_{1i},a_{2i})}(y_i) = 1$ for $y_i \in (a_{1i}, a_{2i})$ and 0 otherwise, $r_i(u;\theta) \equiv e_i(u;\theta) - c_i(\theta)$ for $u \in [a_{1i}, a_{2i}]$, and $h'(v;\lambda) \equiv \partial h(v;\lambda)/\partial v$ for $v \in (a_1, a_2)$. Set $e_i(\theta) \equiv e_i(y_i;\theta)$, $r_i(\theta) \equiv r_i(y_i;\theta)$, and $h_i'(\lambda) \equiv h'(y_i;\lambda)$ for $i = 1, \ldots, n$. Then the log-likelihood function $\ell(\theta)$ for θ is $\sum_{i=1}^n \ell_i(\theta)$, where

$$\begin{aligned}\ell_i(\theta) = {} & \log[\phi(r_i(\theta))] + \log[h_i'(\lambda)] - \log[g_i(\beta,\gamma)] \\ & - \log[\Phi(r_i(a_{2i};\theta)) - \Phi(r_i(a_{1i};\theta))].\end{aligned} \tag{10}$$

Assume that there exists the score function $\partial\ell(\theta)/\partial\theta$ ($\equiv S(\theta)$) for θ. Then $S(\theta) = \sum_{i=1}^n \partial\ell_i(\theta)/\partial\theta$, where

$$\begin{aligned}\frac{\partial\ell_i(\theta)}{\partial\theta} = {} & \frac{\phi'(r_i(\theta))\,\frac{\partial}{\partial\theta}r_i(\theta)}{\phi(r_i(\theta))} + \frac{\frac{\partial}{\partial\theta}h_i'(\lambda)}{h_i'(\lambda)} - \frac{\frac{\partial}{\partial\theta}g_i(\beta,\gamma)}{g_i(\beta,\gamma)} \\ & - \frac{\frac{\partial}{\partial\theta}\Phi(r_i(a_{2i};\theta)) - \frac{\partial}{\partial\theta}\Phi(r_i(a_{1i};\theta))}{\Phi(r_i(a_{2i};\theta)) - \Phi(r_i(a_{1i};\theta))}\end{aligned} \tag{11}$$

with $\phi'(t) \equiv d\phi(t)/dt$,

$$\begin{aligned}\frac{\partial r_i(\theta)}{\partial\theta} &= \frac{\partial e_i(\theta)}{\partial\theta} - \frac{\partial c_i(\theta)}{\partial\theta}, \\ \frac{\partial\Phi(r_i(a_{ji};\theta))}{\partial\theta} &= \phi(r_i(a_{ji};\theta))\left[\frac{\partial e_i(a_{ji};\theta)}{\partial\theta} - \frac{\partial c_i(\theta)}{\partial\theta}\right]\end{aligned}$$

for $t \in \mathcal{R}$ and $j = 1, 2$. For $i = 1, \dots, n$, by differentiating both sides of equation $G_i(c_i(\theta); \theta) = 0$ with respect to θ, we obtain $\partial c_i(\theta)/\partial\theta = c_{i1}(\theta)/c_{i2}(\theta)$, where

$$c_{i1}(\theta) \equiv \sum_{j=1}^{2} \phi(r_i(a_{ji}; \theta)) \frac{\partial e_i(a_{ji}; \theta)}{\partial \theta},$$
$$c_{i2}(\theta) \equiv \sum_{j=1}^{2} \phi(r_i(a_{ji}; \theta)) - 2\,\phi(-c_i(\theta)).$$

Assume that there exists the Hessian matrix $\partial^2 \ell(\theta)/\partial\theta\partial\theta^T$ $(\equiv -J(\theta))$ of $\ell(\theta)$. Then $J(\theta) = -\sum_{i=1}^{n} \partial^2 \ell_i(\theta)/\partial\theta\partial\theta^T$ $(\equiv \sum_{i=1}^{n} J_i(\theta))$, where

(12)

$$\begin{aligned}
J_i(\theta) = & -\frac{\phi''(r_i(\theta))\, \frac{\partial}{\partial\theta} r_i(\theta) \frac{\partial}{\partial\theta^T} r_i(\theta) + \phi'(r_i(\theta)) \frac{\partial^2}{\partial\theta\partial\theta^T} r_i(\theta)}{\phi(r_i(\theta))} \\
& + \frac{[\phi'(r_i(\theta))]^2 \frac{\partial}{\partial\theta} r_i(\theta) \frac{\partial}{\partial\theta^T} r_i(\theta)}{\phi^2(r_i(\theta))} - \frac{\frac{\partial^2}{\partial\theta\partial\theta^T} h_i'(\lambda)}{h_i'(\lambda)} + \frac{\frac{\partial}{\partial\theta} h_i'(\lambda)\, \frac{\partial}{\partial\theta^T} h_i'(\lambda)}{[h_i'(\lambda)]^2} \\
& + \frac{\frac{\partial^2}{\partial\theta\partial\theta^T} g_i(\beta, \gamma)}{g_i(\beta, \gamma)} - \frac{\frac{\partial}{\partial\theta} g_i(\beta, \gamma)\, \frac{\partial}{\partial\theta^T} g_i(\beta, \gamma)}{g_i^2(\beta, \gamma)} \\
& + \frac{\frac{\partial^2}{\partial\theta\partial\theta^T} \Phi(r_i(a_{2i}; \theta)) - \frac{\partial^2}{\partial\theta\partial\theta^T} \Phi(r_i(a_{1i}; \theta))}{\Phi(r_i(a_{2i}; \theta)) - \Phi(r_i(a_{1i}; \theta))} \\
& - \frac{\left[\frac{\partial}{\partial\theta} \Phi(r_i(a_{2i}; \theta)) - \frac{\partial}{\partial\theta} \Phi(r_i(a_{1i}; \theta))\right] \left[\frac{\partial}{\partial\theta^T} \Phi(r_i(a_{2i}; \theta)) - \frac{\partial}{\partial\theta^T} \Phi(r_i(a_{1i}; \theta))\right]}{[\Phi(r_i(a_{2i}; \theta)) - \Phi(r_i(a_{1i}; \theta))]^2}
\end{aligned}$$

with $\phi''(t) \equiv d^2\phi(t)/dt^2$,

$$\frac{\partial^2 r_i(\theta)}{\partial\theta\partial\theta^T} = \frac{\partial^2 e_i(\theta)}{\partial\theta\partial\theta^T} - \frac{\partial^2 c_i(\theta)}{\partial\theta\partial\theta^T},$$

$$\begin{aligned}
\frac{\partial^2 \Phi(r_i(a_{ji}; \theta))}{\partial\theta\partial\theta^T} = & \,\phi'(r_i(a_{ji}; \theta)) \left[\frac{\partial e_i(a_{ji}; \theta)}{\partial\theta} - \frac{\partial c_i(\theta)}{\partial\theta}\right] \left[\frac{\partial e_i(a_{ji}; \theta)}{\partial\theta^T} - \frac{\partial c_i(\theta)}{\partial\theta^T}\right] \\
& + \phi(r_i(a_{ji}; \theta)) \left[\frac{\partial^2 e_i(a_{ji}; \theta)}{\partial\theta\partial\theta^T} - \frac{\partial^2 c_i(\theta)}{\partial\theta\partial\theta^T}\right]
\end{aligned}$$

for $t \in \mathcal{R}$ and $j = 1, 2$. Here $\phi'(\pm\infty) \equiv 0$ and

$$\frac{\partial^2 c_i(\theta)}{\partial\theta\partial\theta^T} = \frac{\frac{\partial}{\partial\theta^T} c_{i1}(\theta) - \frac{\partial}{\partial\theta} c_i(\theta)\, \frac{\partial}{\partial\theta^T} c_{i2}(\theta)}{c_{i2}(\theta)}$$

for $i = 1, \ldots, n$, where

$$\begin{aligned}\frac{\partial c_{i1}(\theta)}{\partial \theta^T} &= \sum_{j=1}^{2} \left\{ \phi'(r_i(a_{ji};\theta)) \frac{\partial e_i(a_{ji};\theta)}{\partial \theta} \left[\frac{\partial e_i(a_{ji};\theta)}{\partial \theta^T} - \frac{\partial c_i(\theta)}{\partial \theta^T} \right] \right. \\ &\quad \left. + \phi(r_i(a_{ji};\theta)) \frac{\partial^2 e_i(a_{ji};\theta)}{\partial \theta \partial \theta^T} \right\}, \\ \frac{\partial c_{i2}(\theta)}{\partial \theta^T} &= \sum_{j=1}^{2} \phi'(r_i(a_{ji};\theta)) \left[\frac{\partial e_i(a_{ji};\theta)}{\partial \theta^T} - \frac{\partial c_i(\theta)}{\partial \theta^T} \right] + 2\phi'(-c_i(\theta)) \frac{\partial c_i(\theta)}{\partial \theta^T}.\end{aligned}$$

Assume that

$$\frac{\partial}{\partial \theta} \int_{a_{1i}}^{a_{2i}} p_i(y_i;\theta)\, dy_i = \int_{a_{1i}}^{a_{2i}} \frac{\partial}{\partial \theta} p_i(y_i;\theta)\, dy_i$$

for $i = 1, \ldots, n$. Then $E_\theta(S(\theta)) = 0$. Set $K(\theta) \equiv \sum_{i=1}^{n} [\partial \ell_i(\theta)/\partial\theta][\partial \ell_i(\theta)/\partial\theta^T]$. Assume that $E_\theta([\partial \ell_i(\theta)/\partial\theta]^T[\partial \ell_i(\theta)/\partial\theta]) < \infty$ for $i = 1, \ldots, n$. Then there exists the expected Fisher information $Cov_\theta(S(\theta))$ $(\equiv I(\theta))$ of θ and $E_\theta(K(\theta)) = I(\theta)$. Assume that

$$\frac{\partial^2}{\partial \theta \partial \theta^T} \int_{a_{1i}}^{a_{2i}} p_i(y_i;\theta)\, dy_i = \int_{a_{1i}}^{a_{2i}} \frac{\partial^2}{\partial \theta \partial \theta^T} p_i(y_i;\theta)\, dy_i$$

for $i = 1, \ldots, n$. Then $E_\theta(J_i(\theta)) = Cov_\theta(\partial \ell_i(\theta)/\partial\theta)$ for $i = 1, \ldots, n$, which implies that $E_\theta(J(\theta)) = I(\theta)$.

Assume that there exists a unique MLE $\hat\theta$ of θ. Then $\hat\theta$ solves the score equation $S(\hat\theta) = 0$ for θ. Assume that both of $I(\theta)$ and $J(\theta)$ are continuous functions of θ. Then all of $I(\hat\theta)$, $J(\hat\theta)$ and $K(\hat\theta)$ are nonnegative definite and generally positive definite matrices. But, $I(\hat\theta)$ generally has no closed-form formula to be evaluated directly. Thus, $\hat\theta$ can be obtained by utilizing the following method: First choose a good initial value $\theta^{(0)}$ and then iterate the following equations

$$\theta^{(k+1)} = \theta^{(k)} + M^{-1}\left(\theta^{(k)}\right) S\left(\theta^{(k)}\right), \quad k = 0, 1, 2, \ldots, \tag{13}$$

until $\theta^{(k)}$'s converge to $\hat\theta$. If $M(\theta^{(k)}) = J(\theta^{(k)})$ for $k = 0, 1, 2, \ldots$, it is called the Newton-Raphson method. If $M(\theta^{(k)}) = I(\theta^{(k)})$ for $k = 0, 1, 2, \ldots$, it is called the Fisher scoring method. Since $I(\theta^{(k)})$'s generally have no closed-form formulae, we suggest only to choose $M(\theta^{(k)})$ as either $J(\theta^{(k)})$ or $K(\theta^{(k)})$ for $k = 0, 1, 2, \ldots$. Note that $K(\theta^{(k)})$'s are nonnegative definite and generally positive definite matrices, but $J(\theta^{(k)})$'s are not necessarily nonnegative definite matrices when the initial value $\theta^{(0)}$ is far from $\hat\theta$. Thus, if $M(\theta^{(k)}) = K(\theta^{(k)})$ for $k = 0, 1, 2, \ldots$, a good initial value $\theta^{(0)}$ is usually easier to find but more iterations are needed for convergence than the Newton-Raphson method. Therefore, a stable and quick method to obtain $\hat{\theta}$ is suggested as follows: First choose $M(\theta^{(k)})$ as $K(\theta^{(k)})$ until near convergence and then $J(\theta^{(k)})$ until convergence.

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) $E_\theta(\max_{1 \le i \le n} |\partial \ell_i(\theta)/\partial \theta_j|)/[Cov_\theta(\partial \ell(\theta)/\partial \theta_j)]^{1/2} \to 0$ as $n \to \infty$ for $j = 1, \ldots, d$, where $\theta \equiv (\theta_1, \ldots, \theta_d)^T$;

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

(iv) $[\text{diag}\{I(\theta)\}]^{-1/2} I(\theta) [\text{diag}\{I(\theta)\}]^{-1/2} \to \Sigma(\theta)$ as $n \to \infty$, where $\Sigma(\theta)$ is a positive definite matrix.

Then, by Theorem 1.80 of Prakasa Rao (1999),

$$M^{-1/2}(\theta)\, S(\theta) \xrightarrow{d} N(0, I_d) \tag{14}$$

as $n \to \infty$, where M can be chosen as any of I, J, and K. Assume that

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

as $n \to \infty$. Then, by equation (14) and condition (iii),

$$M^{1/2}(\theta)\left(\hat{\theta} - \theta\right) = M^{-1/2}(\theta)\, S(\theta) + o_p(1) \xrightarrow{d} N(0, I_d) \tag{15}$$

as $n \to \infty$, where M can be chosen as any of I, J, and K. Thus, by equation (15) and condition (i), the MLE $\hat{\theta}$ of θ is a weakly consistent estimator of θ. Assume that $I^{-1/2}(\theta)I(\hat{\theta})I^{-1/2}(\theta) \xrightarrow{p} I_d$, $J^{-1/2}(\theta)J(\hat{\theta})J^{-1/2}(\theta) \xrightarrow{p} I_d$, and $K^{-1/2}(\theta)K(\hat{\theta})K^{-1/2}(\theta) \xrightarrow{p} I_d$ as $n \to \infty$. Then, by equation (15),

$$M^{1/2}\left(\hat{\theta}\right)\left(\hat{\theta} - \theta\right) = M^{1/2}(\theta)\left(\hat{\theta} - \theta\right) + o_p(1) \xrightarrow{d} N(0, I_d) \tag{16}$$

as $n \to \infty$, where M can be chosen as any of I, J, and K.

3.2 Hypothesis Testing and Confidence Regions

In this subsection, let ω $(\equiv (\psi^T, \chi^T)^T)$ be a one-to-one reparameterization of θ such that $|\partial \theta / \partial \omega^T| \neq 0$ 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 null hypothesis $H_0 : \psi = \psi_0$ versus alternative hypothesis $H_1 : \psi \neq \psi_0$.

Set $S_\psi(\chi) \equiv \partial\ell(\theta)/\partial\chi$, $I_\psi(\chi) \equiv Cov_\theta(S_\psi(\chi))$, $J_\psi(\chi) \equiv -\partial S_\psi(\chi)/\partial\chi^T$, and $K_\psi(\chi) \equiv \sum_{i=1}^n [\partial\ell_i(\theta)/\partial\chi][\partial\ell_i(\theta)/\partial\chi^T]$. Then $S_\psi(\chi) = \partial\theta^T/\partial\chi S(\theta)$, $I_\psi(\chi) = \partial\theta^T/\partial\chi I(\theta)\partial\theta/\partial\chi^T$,

$$J_\psi(\chi) = \frac{\partial\theta^T}{\partial\chi}\, J(\theta)\, \frac{\partial\theta}{\partial\chi^T} - \sum_{j=1}^d \frac{\partial^2\theta_j}{\partial\chi\partial\chi^T}\, S_j(\theta),$$

and $K_\psi(\chi) = \partial\theta^T/\partial\chi K(\theta)\partial\theta/\partial\chi^T$, where $S(\theta) \equiv (S_1(\theta), \dots, S_d(\theta))^T$. Assume that there exists a unique MLE $\hat{\chi}_\psi$ of χ given ψ. Then $\hat{\chi}_\psi$ solves the score equation $S_\psi(\hat{\chi}_\psi) = 0$ for χ given ψ. Similarly, $\hat{\chi}_\psi$ can be obtained by using the same technique as in Section 3.1.

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

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

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

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

as $n \to \infty$. Then, by equations (15) and (16),

(17)
$$W(\psi) = S^T(\theta)\, I^{-1/2}(\theta) \left\{ I_d - I^{1/2}(\theta) \frac{\partial\theta}{\partial\chi^T} \left[\frac{\partial\theta^T}{\partial\chi} I(\theta)\, \frac{\partial\theta}{\partial\chi^T}\right]^{-1} \frac{\partial\theta^T}{\partial\chi}\, I^{1/2}(\theta) \right\} I^{-1/2}(\theta)\, S(\theta) + o_p(1) \xrightarrow{p} \chi^2_{d_0}$$

as $n \to \infty$.

Let $\alpha \in (0, 1)$ be fixed. The likelihood ratio test with asymptotic size α is to reject H_0 if and only if $W(\psi_0) > \chi^2_{d_0,1-\alpha}$, where $\chi^2_{d_0,1-\alpha}$ is the $1-\alpha$ quantile of the χ^2 distribution with d_0 degrees of freedom. Therefore, $\{\psi_0 : W(\psi_0) \le \chi^2_{d_0,1-\alpha}\}$ is an asymptotic size $1 - \alpha$ confidence region for ψ.

3.3 Quantile Estimation of a Future Observation

Suppose that

(18) $$h(y_{n+1}; \lambda) = f(x_{n+1}; \beta) + g(f(x_{n+1}; \beta), z_{n+1}; \gamma)\, \varepsilon_{n+1},$$

where y_{n+1} is the future observation for subject $n+1$, both x_{n+1} and z_{n+1} are known covariate vectors for subject $n+1$, and ε_{n+1} is a standardized error independent of $\varepsilon_1, \dots, \varepsilon_n$ with median 0 and distribution as either $N(0, 1)$ or truncation of some $N(c_{n+1}(\theta),1)$ with $c_{n+1}(\theta)\in\mathcal{R}$. Assume that y_{n+1} has a known support $(a_1(w_{n+1}),a_2(w_{n+1}))$ contained in the domain of response transformation $h(\cdot;\lambda)$, where w_{n+1} is a known covariate vector for subject $n+1$. Set $a_{1,n+1}\equiv a_1(w_{n+1})$ and $a_{2,n+1}\equiv a_2(w_{n+1})$. Similarly, $c_{n+1}(\theta)$ can be obtained by using the same technique as in Section 3.1. For $u\in[a_{1,n+1},a_{2,n+1}]$, set

$$e_{n+1}(u;\theta)\equiv\frac{h(u;\lambda)-f_{n+1}(\beta)}{g_{n+1}(\beta,\gamma)}$$

and $r_{n+1}(u;\theta)\equiv e_{n+1}(u;\theta)-c_{n+1}(\theta)$, where $h(a_{1,n+1};\lambda)\equiv\lim_{v\downarrow a_{1,n+1}}h(v;\lambda)$, $h(a_{2,n+1};\lambda)\equiv\lim_{v\uparrow a_{2,n+1}}h(v;\lambda)$, $f_{n+1}(\beta)\equiv f(x_{n+1};\beta)$, and $g_{n+1}(\beta,\gamma)\equiv g(f(x_{n+1};\beta),z_{n+1};\gamma)$.

Let $\alpha\in(0,1)$ be fixed, let $\Phi_{n+1}(\cdot;\theta)$ be the c.d.f. of ε_{n+1}, and let $q_{n+1,\alpha}(\theta)$ be the α quantile of y_{n+1}. Then

$$q_{n+1,\alpha}(\theta)=h^{-1}\left(f_{n+1}(\beta)+g_{n+1}(\beta,\gamma)\,\Phi_{n+1}^{-1}(\alpha;\theta);\lambda\right),\tag{19}$$

where

$$\Phi_{n+1}^{-1}(t;\theta)=\Phi^{-1}((1-t)\,\Phi(r_{n+1}(a_{1,n+1};\theta))+t\,\Phi(r_{n+1}(a_{2,n+1};\theta)))+c_{n+1}(\theta)$$

for $t\in\mathcal{R}$.

Assume that $h'(q_{n+1,\alpha}(\theta);\lambda)\neq 0$. Then

$$\frac{\partial q_{n+1,\alpha}(\theta)}{\partial\theta}=\frac{\frac{\partial}{\partial\theta}f_{n+1}(\beta)+\Phi_{n+1}^{-1}(\alpha;\theta)\,\frac{\partial}{\partial\theta}g_{n+1}(\beta,\gamma)+g_{n+1}(\beta,\gamma)\,\frac{\partial}{\partial\theta}\Phi_{n+1}^{-1}(\alpha;\theta)}{h'(q_{n+1,\alpha}(\theta);\lambda)}-\frac{\left.\frac{\partial}{\partial\theta}h(u;\lambda)\right|_{u=q_{n+1,\alpha}(\theta)}}{h'(q_{n+1,\alpha}(\theta);\lambda)},$$

where

$$\frac{\partial\Phi_{n+1}^{-1}(t;\theta)}{\partial\theta}=\frac{(1-t)\,\frac{\partial}{\partial\theta}\Phi(r_{n+1}(a_{1,n+1};\theta))+t\,\frac{\partial}{\partial\theta}\Phi(r_{n+1}(a_{2,n+1};\theta))}{\phi(\Phi_{n+1}^{-1}(t;\theta)-c_{n+1}(\theta))}+\frac{\partial c_{n+1}(\theta)}{\partial\theta}$$

for $t\in\mathcal{R}$. Similarly, all of $\partial\Phi(r_{n+1}(a_{1,n+1};\theta))/\partial\theta$, $\partial\Phi(r_{n+1}(a_{2,n+1};\theta))/\partial\theta$, and $\partial c_{n+1}(\theta)/\partial\theta$ can be obtained by using the same techniques as in Section 3.1.

Note that the MLE of $q_{n+1,\alpha}(\theta)$ is $q_{n+1,\alpha}(\hat{\theta})$. Assume that $\partial q_{n+1,\alpha}(\theta)/\partial\theta$ is a continuous function of θ. Then, by equations (15) and (16),

$$\left[\frac{\partial q_{n+1,\alpha}(\theta)}{\partial\theta^T}M^{-1}(\theta)\frac{\partial q_{n+1,\alpha}(\theta)}{\partial\theta}\right]^{-1/2}\left[q_{n+1,\alpha}\left(\hat{\theta}\right)-q_{n+1,\alpha}(\theta)\right]\xrightarrow{d}N(0,1),\tag{20}$$

$$(21)\qquad \left[\left.\frac{\partial q_{n+1,\alpha}(\theta)}{\partial \theta^T}\right|_{\theta=\hat{\theta}} M^{-1}\left(\hat{\theta}\right) \left.\frac{\partial q_{n+1,\alpha}(\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}} \right]^{-1/2} \left[q_{n+1,\alpha}\left(\hat{\theta}\right) - q_{n+1,\alpha}(\theta)\right] \xrightarrow{d} N(0,1)$$

as $n \to \infty$, where M can be chosen as any of I, J and K.

Let $\alpha_1 \in [0, \alpha]$ be fixed (e.g., 0, $\alpha/2$ or α). Then $[q_{n+1,\alpha_1}(\theta), q_{n+1,1-\alpha+\alpha_1}(\theta)]$ is a $1-\alpha$ prediction interval for y_{n+1} with MLE $[q_{n+1,\alpha_1}(\hat{\theta}), q_{n+1,1-\alpha+\alpha_1}(\hat{\theta})]$.

4. Discussion

In this final section, when the range of the response transformation is possibly different from $\mathcal{R}$, the inappropriateness of the likelihood inference under model (3) is shown to demonstrate the importance of our work.

First of all, suppose that model (3) holds. Then it is exactly the same as the proposed model. Thus, the likelihood inference under the proposed model in Section 3 can be used. In particular, we have $(a_{1i}, a_{2i}) = (a_1, a_2)$, $c_i(\theta) = 0$ and $\Phi(r_i(a_{2i};\theta)) - \Phi(r_i(a_{1i};\theta)) = 1$ for $i = 1, \ldots, n$. Then the p.d.f. of y_i is

$$(22)\qquad \frac{1_{(a_1,a_2)}(y_i)\, \phi(e_i(y_i;\theta))\, h'(y_i;\lambda)}{g_i(\beta,\gamma)}$$

for $i = 1, \ldots, n$. Set $e_i(\theta) = e_i(y_i;\theta)$ for $i = 1, \ldots, n$. Then the log-likelihood function for θ is $\ell^*(\theta) = \sum_{i=1}^n \ell_i^*(\theta)$, the score function for θ is $\partial\ell^*(\theta)/\partial\theta = \sum_{i=1}^n \partial\ell_i^*(\theta)/\partial\theta$ ($\equiv S^*(\theta)$), the Hessian matrix of $\ell^*(\theta)$ is $\partial^2\ell^*(\theta)/\partial\theta\partial\theta^T = \sum_{i=1}^n \partial^2\ell_i^*(\theta)/\partial\theta\partial\theta^T$ ($\equiv -\sum_{i=1}^n J_i^*(\theta) \equiv -J^*(\theta)$), and the MLE θ^* of θ solves the score equation $S^*(\theta^*) = 0$ for θ, where

$$(23)\qquad \ell_i^*(\theta) = \log[\phi(e_i(\theta))] + \log[h_i'(\lambda)] - \log[g_i(\beta,\gamma)],$$

$$(24)\qquad \frac{\partial\ell_i^*(\theta)}{\partial\theta} = \frac{\phi'(e_i(\theta))\,\frac{\partial}{\partial\theta}e_i(\theta)}{\phi(e_i(\theta))} + \frac{\frac{\partial}{\partial\theta}h_i'(\lambda)}{h_i'(\lambda)} - \frac{\frac{\partial}{\partial\theta}g_i(\beta,\gamma)}{g_i(\beta,\gamma)},$$

$$(25)\qquad \begin{aligned} J_i^*(\theta) = &-\frac{\phi''(e_i(\theta))\,\frac{\partial}{\partial\theta}e_i(\theta)\,\frac{\partial}{\partial\theta^T}e_i(\theta) + \phi'(e_i(\theta))\,\frac{\partial^2}{\partial\theta\partial\theta^T}e_i(\theta)}{\phi(e_i(\theta))} \\ &+\frac{[\phi'(e_i(\theta))]^2\,\frac{\partial}{\partial\theta}e_i(\theta)\,\frac{\partial}{\partial\theta^T}e_i(\theta)}{\phi^2(e_i(\theta))} - \frac{\frac{\partial^2}{\partial\theta\partial\theta^T}h_i'(\lambda)}{h_i'(\lambda)} \\ &+\frac{\frac{\partial}{\partial\theta}h_i'(\lambda)\frac{\partial}{\partial\theta^T}h_i'(\lambda)}{[h_i'(\lambda)]^2} + \frac{\frac{\partial^2}{\partial\theta\partial\theta^T}g_i(\beta,\gamma)}{g_i(\beta,\gamma)} - \frac{\frac{\partial}{\partial\theta}g_i(\beta,\gamma)\frac{\partial}{\partial\theta^T}g_i(\beta,\gamma)}{g_i^2(\beta,\gamma)}. \end{aligned}$$

Next, suppose that the range of response transformation $h(\cdot;\lambda)$ is possibly different from $\mathcal{R}$. In such situations, the proposed model can hold, but model (3) cannot hold. Now consider the following two different cases:

Case 1. Suppose that the sample size n is fixed. When the proposed model holds, the likelihood inference under the proposed model in Section 3 is correct. However, the likelihood inference under model (3) is incorrect because model (3) cannot hold.

Case 2. Suppose that the sample size n tends to infinity. When the proposed model holds, the MLE $\hat{\theta}$ of θ under the proposed model is consistent, asymptotically normally distributed and generally asymptotically efficient if all proposed conditions in Section 3 hold. Now assume that some particular model holds (e.g., the proposed model). Since model (3) cannot hold, $S^*(\theta)$ is generally of order $O(n)+O_p(n^{1/2})$ but not $o_p(n)$ as $n\to\infty$ and each eigenvalue of $[J^*(\theta)]^{-1}$ is of order $O(n^{-1})+O_p(n^{-3/2})$ but not $o_p(n^{-1})$ as $n\to\infty$. If the MLE θ^* of θ under model (3) is a consistent estimator of θ, then $\theta^*-\theta$ is generally asymptotically equivalent to $[J^*(\theta)]^{-1}S^*(\theta)$ as $n\to\infty$. Since $[J^*(\theta)]^{-1}S^*(\theta)$ is generally of order $O(1)+O_p(n^{-1/2})$ but not $o_p(1)$ as $n\to\infty$, the MLE θ^* of θ under model (3) is generally an inconsistent estimator of θ.

By Cases 1 and 2, when the range of the response transformation is possibly different from $\mathcal{R}$, the likelihood inference under model (3) 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 3 can be used.

Acknowledgements

The authors would like to thank the referee for many valuable comments and suggestions on revising the paper.

References

1. O. E. Barndorff-Nielsen and D. R. Cox, *Inference and Asymptotics*, Chapman & Hall, London, 1994.

2. G. E. P. Box and D. R. Cox, An analysis of transformations, *J. Roy. Statist. Soc. Ser. B* **26** (1964), 211-243.

3. R. D. Cook and S. Weisberg, *Applied Regression Including Computing and Graphics*, John Wiley & Sons, New York, 1999.

4. D. R. Cox and D. V. Hinkley, *Theoretical Statistics*, Chapman & Hall, London, 1974.

5. N. R. Draper and H. Smith, *Applied Regression Analysis*, John Wiley & Sons, New York, 1998.

6. J. A. John and N. R. Draper, An alternative family of transformations, *Applied Statistics* **29** (1980), 190-197.

7. F. Mosteller and J. W. Tukey, *Data Analysis and Linear Regression*, Addison-Wesley, Reading, Mass, 1977.

8. B. L. S. Prakasa Rao, *Semimartingales and their Statistical Inference*, Chapman & Hall/CRC, Boca Raton, 1999.

Chih-Rung Chen
Institute of Statistics
National Chiao Tung University
Hsinchu 30050, Taiwan
E-mail: cchen@stat.nctu.edu.tw

Lih-Chung Wang
Department of Applied Mathematics
National Dong Hwa University
Hualien 974, Taiwan
E-mail: lcwang@mail.ndhu.edu.tw