Journal of Korean
Data & Information Science Society
2004, Vol. 15, No. 1 pp. 251 ~261

Some Remarks on the Likelihood Inference for the Ratios of Regression Coefficients in Linear Model

Yeong-Hwa Kim[1)] · Wan-Yeon Yang[2)] · M. J. Kim[3)] · C. G. Park[4)]

Abstract

The paper focuses primarily on the standard linear multiple regression model where the parameter of interest is a ratio of two regression coefficients. The general model includes the calibration model, the Fieller-Creasy problem, slope-ratio assays, parallel-line assays, and bioequivalence. We provide an orthogonal transformation (*cf.* Cox and Reid (1987)) of the original parameter vector. Also, we give some remarks on the difficulties associated with likelihood based confidence interval.

Keywords : Conditional profile likelihood, Integrated likelihood, Orthogonal transformation, Profile likelihood, Ratio of regression coefficients

1. Introduction

There is a large class of important statistical problems which can be broadly described under the general heading of inference about the ratios of regression coefficients in a general linear model. The calibration model, ratio of two means or the Fieller-Creasy problem, slope-ratio assay, and parallel-line assay are included in this class. The general nature of this problem was recognized several decades ago, as documented for example in the excellent treatise of Finney (1978).

The frequentist approaches for these problems have typically confronted with serious difficulties. As an example, confidence intervals for ratios of two normal means based on Fieller's pivot (1954) may fail to exist in many circumstances. On other occasions, a confidence set for this ratio may be the union of two disjoint

1) First Author : Assistant Professor, Department of Statistics, Chung-Ang University, Seoul, Korea.
E-mail : gogators@cau.ac.kr
2) Associate Professor, Department of Applied Statistics, Kyungwon University, Sungnam, Korea.
3) Ph.D. Candidate, Department of Statistics, University of Florida, Gainesville, Florida, U.S.A.
4) Researcher, Department of Statistics, Seoul National University, Seoul, Korea.

unbounded intervals, and in extreme situations, may even be the entire real line. The same phenomenon occurs as in the other problems described above.

A second general approach is one based on likelihoods. We discuss the profile likelihood, and its variants such as conditional profile likelihood (Cox and Reid, 1987; Barndorff-Nielsen, 1983) or adjusted profile likelihood (McCullagh and Tibshirani, 1990).

In Section 2, we find the orthogonal transformation of the original parameter vector in the sense of Cox and Reid(1987). In Section 3, we first find the profile likelihood for the parameter of interest, then find various adjustments to this same based on this orthogonal transformation. In Section 4, we find also an integrated likelihood following the prescription of Berger, Liseo and Wolpert (1999). The adjusted likelihood as well as integrated likelihood are quite similar, and we point out same of the difficulties associated with likelihood based confidence interval. The proof of a theorem is deferred to the Appendix.

2. The Orthogonal Transformation

Consider the general regression model

$$y_i = \sum_{j=1}^{r} \beta_j x_{ij} + \epsilon_i \ , \ (i = 1, \ldots, n) \qquad .1)$$

where the errors ϵ_i's are iid $N(0, \sigma^2)$. Here $\beta_j \in (-\infty, \infty)$ for $j \neq 2$, while $\beta_2 \in (-\infty, \infty) - \{0\}$ and the parameter of interest is $\theta_1 = \beta_1/\beta_2$. We write $\boldsymbol{y} = (y_1, \ldots, y_n)^T$, $\boldsymbol{x_i} = (x_{i1}, \ldots, x_{ip})^T$, $(i = 1, \ldots, n)$, $\boldsymbol{X^T} = (\boldsymbol{x_1}, \ldots, \boldsymbol{x_n})$, $\underset{\sim}{\beta} = (\beta_1, \ldots, \beta_n)^T$, and, $\boldsymbol{e} = (\epsilon_1, \ldots, \epsilon_n)^T$. Thus the model can be rewritten in matrix notation as $\boldsymbol{Y} = \boldsymbol{X}\underset{\sim}{\beta} + \boldsymbol{e}$, where we assume that the rank$(\boldsymbol{X}) = r < n$.

First we introduce a transformation of the parameter vector $(\underset{\sim}{\beta}, \sigma)$ which results in the orthogonality of θ_1 with the remaining nuisance parameters. We begin with the Fisher information matrix

$$\boldsymbol{I}(\underset{\sim}{\beta}, \sigma) = \begin{bmatrix} ((s_{jl})) & \boldsymbol{0^T} \\ \boldsymbol{0} & 2n \end{bmatrix} \qquad (2.2)$$

where $s_{jl} = \sum_{i=1}^{n} x_{ij} x_{il}$, $(j, l = 1, \ldots, r)$. Consider the transformation

$$\begin{aligned} &\beta_1 = \theta_1 \theta_2 h(\theta_1) \ , \ \beta_2 = \theta_2 h(\theta_1) \ , \ \beta_j = \theta_j - \theta_2 g_j(\theta_1) \ , \ j = 3, , .., r \\ &\sigma = \theta_{r+1} \ . \end{aligned} \qquad (2.3)$$

Then the Jacobian matrix is given by

$$\boldsymbol{J}=\begin{bmatrix} \theta_2\{\theta_1 h'(\theta_1)+h(\theta_1)\} & \theta_2 h'(\theta_1) & -\theta_2 g_3{}'(\theta_1) & \cdots & -\theta_2 g_r{}'(\theta_1) & 0 \\ \theta_1 h(\theta_1) & h(\theta_1) & -g_3(\theta_1) & \cdots & -g_r(\theta_1) & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \tag{2.4}$$

Theorem 1. Let $g_j(\theta_1) = h(\theta_1)(a_{j1}\theta_1 + a_{j2})$, $j = 3, \ldots, r$, where

$$\begin{bmatrix} a_{31} & a_{32} \\ \vdots & \vdots \\ a_{r1} & a_{r2} \end{bmatrix} = \begin{bmatrix} s_{33} & \cdots & s_{3r} \\ \vdots & \ddots & \vdots \\ s_{3r} & \cdots & s_{rr} \end{bmatrix}^{-1} \begin{bmatrix} s_{13} & s_{23} \\ \vdots & \vdots \\ s_{1r} & s_{2r} \end{bmatrix} . \tag{2.5}$$

Then parametric orthogonality holds by choosing

$$h(\theta_1) = Q^{-1/2}(\theta_1),$$

where $Q^{-1/2}(\theta_1) = c_{11}\theta_1^2 + 2c_{12}\theta_1 + c_{22}$, $c_{11} = s_{11} - \sum_{j=3}^{r} s_{1j}a_{j1}$, $c_{12} = s_{12} - \sum_{j=3}^{r} s_{1j}a_{j2}$, and $c_{22} = s_{22} - \sum_{j=3}^{r} s_{2j}a_{j2}$. The proof of this theorem is deferred to the Appendix. Based on this transformation and writing $\boldsymbol{C}=\begin{pmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{pmatrix}$, it follows from (2.1)-(2.4) that the reparametrized Fisher information matrix is given by

$$\boldsymbol{I}(\theta) = \frac{1}{\theta_{r+1}^2}\begin{bmatrix} \dfrac{\theta_2^2 |\boldsymbol{C}|^2}{Q^2(\theta_1)} & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & s_{33} & \cdots & s_{3r} & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & s_{r3} & \cdots & s_{rr} & 0 \\ 0 & 0 & 0 & \cdots & 0 & 2n \end{bmatrix} \tag{2.6}$$

Remark 1. If we write $\boldsymbol{X}^T\boldsymbol{X}=\begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 \\ \boldsymbol{A}_2^T & \boldsymbol{A}_3 \end{bmatrix}$, where $\boldsymbol{A}_1 = \begin{pmatrix} s_{11} & s_{12} \\ c_{12} & c_{22} \end{pmatrix}$, $\boldsymbol{A}_2 = \begin{pmatrix} s_{13} & \cdots & s_{1r} \\ s_{23} & \cdots & s_{3r} \end{pmatrix}$, and $\boldsymbol{A}_3 = \begin{pmatrix} s_{33} & \cdots & s_{3r} \\ \vdots & \ddots & \vdots \\ s_{r3} & \cdots & s_{rr} \end{pmatrix}$, then from (2.4), it is easy to see that $\boldsymbol{C}= \boldsymbol{A}_1 - \boldsymbol{A}_2\boldsymbol{A}_3^{-1}\boldsymbol{A}_2^T$. Also, since $\mathrm{rank}(\boldsymbol{X}^T\boldsymbol{X}) = \mathrm{rank}(\boldsymbol{X}) = r$, it is positive definite. Then $\boldsymbol{C}$ is also positive definite.

3. Likelihood Analysis

We begin with the derivation of the profile likelihood of θ_1 . This is done in several steps as following. First writing $\hat{\underline{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_r)^T$ as the unrestricted MLE of $\underline{\beta}$ and $SSE = (\boldsymbol{y} - \boldsymbol{X}\hat{\underline{\beta}})^T(\boldsymbol{y} - \boldsymbol{X}\hat{\underline{\beta}})$, one can write the likelihood

$$L(\beta_1, \beta_2, \ldots, \beta_r, \sigma) \propto \sigma^{-n} \exp\left[-\frac{1}{2\sigma^2}\{SSE + (\underline{\beta} - \hat{\underline{\beta}})^T \boldsymbol{X}^T \boldsymbol{X}(\underline{\beta} - \hat{\underline{\beta}})\}\right] \tag{3.1}$$

Next, using a basic identity of quadratic form

$$\begin{aligned}
&(\underline{\beta} - \hat{\underline{\beta}})^T \boldsymbol{X}^T \boldsymbol{X}(\underline{\beta} - \hat{\underline{\beta}}) \\
&= c_{11}(\beta_1 - \hat{\beta}_1)^2 + 2c_{12}(\beta_1 - \hat{\beta}_1)(\beta_2 - \hat{\beta}_2) + c_{22}(\beta_2 - \hat{\beta}_2)^2 + \boldsymbol{u}^T \boldsymbol{A}_3 \boldsymbol{u} \\
&= c_{11}(\beta_2\theta_1 - \hat{\beta}_1)^2 + 2c_{12}(\beta_2\theta_1 - \hat{\beta}_1)(\beta_2 - \hat{\beta}_2) + c_{22}(\beta_2 - \hat{\beta}_2)^2 + \boldsymbol{u}^T \boldsymbol{A}_3 \boldsymbol{u}
\end{aligned}$$

where $\boldsymbol{u} = (\beta_3 - \hat{\beta}_3, \ldots, \beta_r - \hat{\beta}_r)^T$. Hence for fixed θ_1, β_2 and σ , the MLE of β_k is $\hat{\beta}_k$ for every $k = 3, \ldots, r$, and these do not depend θ_1, β_2 and σ. Hence, the profile likelihood of θ_1, β_2 and σ is given by

$$\begin{aligned}
&L_{PL}(\theta_1, \beta_2, \sigma) \\
&\propto \sigma^{-n} \exp\left[-\frac{1}{2\sigma^2}\{SSE + c_{11}(\beta_2\theta_1 - \hat{\beta}_1)^2 + 2c_{12}(\beta_2\theta_1 - \hat{\beta}_1)(\beta_2 - \hat{\beta}_2) + c_{22}(\beta_2 - \hat{\beta}_2)^2\}\right]
\end{aligned} \tag{3.2}$$

It follows from (3.2) that for fixed θ_1 and σ, the MLE of β_2 is

$$\hat{\beta}_2(\theta_1, \sigma) = \hat{\beta}_2(\theta_1) = Q^{-1/2}(\theta_1)[\theta_1(c_{11}\hat{\beta}_1 + c_{12}\hat{\beta}_2) + c_{12}\hat{\beta}_1 + c_{22}\hat{\beta}_2] \tag{3.3}$$

Also, after some algebraic simplifications,

$$\begin{aligned}
&c_{11}(\hat{\beta}_2(\theta_1)\theta_1 - \hat{\beta}_1)^2 + 2c_{12}(\hat{\beta}_2(\theta_1)\theta_1 - \hat{\beta}_1)(\hat{\beta}_2(\theta_1) - \hat{\beta}_2) + c_{22}(\hat{\beta}_2(\theta_1) - \hat{\beta}_2)^2 \\
&= |\boldsymbol{C}|^2(\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2 / Q(\theta_1)
\end{aligned} \tag{3.4}$$

This leads to

$$L_{PL}(\theta_1, \sigma) \propto \sigma^{-n} \exp\left[-\frac{1}{2\sigma^2}\{SSE + |\boldsymbol{C}|^2(\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2 / Q(\theta_1)\}\right] . \tag{3.5}$$

Now, for fixed θ_1 , the MLE of σ is

$$\hat{\sigma}(\theta_1) = n^{-1/2}\left[SSE + |\boldsymbol{C}|^2(\hat{\beta_2}\theta_1 - \hat{\beta_1})^2/Q(\theta_1)\right]^{-1/2} \quad . \tag{3.6}$$

Thus, the profile likelihood of θ_1 is given by

$$L_{PL}(\theta_1) \propto \left[SSE + |\boldsymbol{C}|^2(\hat{\beta_2}\theta_1 - \hat{\beta_1})^2/Q(\theta_1)\right]^{-n/2} \quad . \tag{3.7}$$

Remark 2. It may be noted from (3.7) that as $|\theta_1| \to \infty$, $(\hat{\beta_2}\theta_1 - \hat{\beta_1})^2/Q(\theta_1) \to \hat{\beta_2}/c_{11}$. Thus $L_{PL}(\theta_1)$ is bounded away from 0 as $|\theta_1| \to \infty$. This immediately leads to the fact that any likelihood-based confidence interval for θ_1 can potentially the entire real line. For the ratio of normal means problem, this fact was first observed by Liseo (1993) for known σ , and subsequently by Yin and Ghosh (2000) for unknown σ .

Remark 3. Due to orthogonality of θ_1 with $(\theta_2, \ldots, \theta_r, \theta_{r+1})$, from (3.7) and (2.6), the conditional profile likelihood (CPL) proposed by Cox and Reid (1987) of θ_1 is given by

$$\begin{aligned} L_{CPL}(\theta_1) &\propto L_{PL}(\theta_1)\,|\hat{\theta}_{r+1}|^{-r} \\ &\propto \left[SSE + |\boldsymbol{C}|^2(\hat{\beta_2}\theta_1 - \hat{\beta_1})^2/Q(\theta_1)\right]^{-\frac{n+r}{2}} \quad . \end{aligned} \tag{3.8}$$

Clearly, $L_{CPL}(\theta_1)$ suffers from the same drawback as $L_{PL}(\theta_1)$ in this case as $|\theta_1| \to \infty$.

Remark 4. Another adjustment of the profile likelihood is proposed in McCullagh and Tibshirani (1990) based on the idea of unbiased estimating functions. Suppose θ is the parameter of interest and ψ is the nuisance parameter. Denote the score function derived from the profile log-likelihood by $U(\theta) = \dfrac{\partial}{\partial\theta}\log L_{PL}(\theta_1)$ where $L_{PL}(\theta_1) = L(\theta, \hat{\psi}(\theta))$, $\hat{\psi}(\theta)$ being the MLE of ψ for fixed θ . A property of the regular maximum likelihood score function is that it has zero mean, and its variance is the negative of the expected derivative matrix, expectation being computed at the true parameter value. McCullagh and Tibshirani (1990) propose adjusting $U(\theta)$ so that these properties hold when expectations and derivatives are computed at $(\theta, \hat{\psi}(\theta))$ rather than at the true parameter point.

To this end, let $\tilde{U}(\theta) = [V(\theta) - m(\theta)]w(\theta)$, where $m(\theta)$ and $w(\theta)$ are so chosen that $E_{\theta,\hat{\psi}(\theta)}\tilde{U}(\theta) = 0$ and $V_{\theta,\hat{\psi}(\theta)}(\tilde{U}(\theta)) =- E_{\theta,\hat{\psi}(\theta)}[\frac{\partial}{\partial\theta}\tilde{U}(\theta)]$. this leads to the solutions

$$\begin{aligned} m(\theta) &= E_{\theta,\hat{\psi}(\theta)}\, U(\theta) \\ w(\theta) &= [- E_{\theta,\hat{\psi}(\theta)} \frac{\partial^2}{\partial\theta^2} \log \mathrm{L}_{\mathrm{PL}}(\theta) + \frac{\partial}{\partial\theta} \mathrm{m}(\theta)] \,/\, \mathrm{V}_{\theta,\hat{\psi}}[\mathrm{U}(\theta)] \end{aligned} \tag{3.9}$$

Then the adjusted profile log-likelihood is given by

$$L_{APL}(\theta) = \exp\left[\int_{\theta} \tilde{U}(t)\,dt\right] \quad . \tag{3.10}$$

In the present example with $\theta = \theta_1$ and $\psi = (\theta_2, \ldots, \theta_r, \theta_{r+1})$, it follows after some simplifications that

$$U(\theta_1) = \frac{\partial \log L_{PL}(\theta_1)}{\partial\theta_1} =- \frac{n}{2} \frac{2(\hat{\beta}_2\theta_1 - \hat{\beta}_1)[\hat{\beta}_2(c_{12}\theta_1 + c_{22}) + \hat{\beta}_1(c_{11}\theta_1 + c_{12})]}{[SSE + |\boldsymbol{C}|^2(\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2/Q(\theta_1)]Q(\theta_1)} \ . \tag{3.11}$$

Next observing that $V\begin{pmatrix}\hat{\beta}_1 \\ \hat{\beta}_2\end{pmatrix} = \sigma^2\begin{bmatrix} c_{11} & c_{12} \\ c_{12} & c_{22}\end{bmatrix}^{-1} = \frac{\sigma^2}{|\boldsymbol{C}|}\begin{bmatrix} c_{22} & -c_{12} \\ -c_{12} & c_{11}\end{bmatrix}$, it follow that

$$\begin{aligned} &Cov[\hat{\beta}_2\theta_1 - \hat{\beta}_1\,,\ \hat{\beta}_2(c_{12}\theta_1 + c_{22}) + \hat{\beta}_1(c_{11}\theta_1 + c_{12})] \\ &= \frac{\sigma^2}{|\boldsymbol{C}|}[\theta_1(c_{12}\theta_1 + c_{22})c_{11} - (c_{11}\theta_1 + c_{12})c_{22} - \theta_1(c_{11}\theta_1 + c_{12})c_{12} - (c_{12}\theta_1 + c_{22})c_{12}] = 0 \end{aligned} \tag{3.12}$$

This proves the independence of $\hat{\beta}_2\theta_1 - \hat{\beta}_1$ and $\hat{\beta}_2(c_{12}\theta_1 + c_{22}) + \hat{\beta}_1(c_{11}\theta_1 + c_{12})$. Also, since $\hat{\beta}_2\theta_1 - \hat{\beta}_1$ is $N\left(0, \frac{\sigma^2}{|\boldsymbol{C}|}Q(\theta_1)\right)$, it follows that

$$E\left[\frac{\hat{\beta}_2\theta_1 - \hat{\beta}_1}{SSE + |\boldsymbol{C}|^2(\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2/Q(\theta_1)}\right] = 0.$$

Hence, from (3.11) and (3.12), $m(\theta_1) = E[U(\theta_1)] = 0$. Also, it can be shown after much algebra that $w(\theta_1) = k + O(n^{-1})$, where $k(>0)$ is some positive constant not depending on θ_1. Thus, $\tilde{U}(\theta_1) = k^{-1}U(\theta_1) + O(n^{-1})$. This shows that $\log L_{APL} = k^{-1}\log L_{PL}(\theta_1) + O(n^{-1})$ proving thereby that $L_{APL}(\theta_1)$ suffers from the same drawback as $L_{PL}(\theta_1)$ when $|\theta_1| \to \infty$. However, such an integrated likelihood cannot be identified with any proper posterior of θ_1 since $\int_{-\infty}^{\infty} L_I(\theta_1)\,d\theta_1 = +\infty$.

4. Concluding Remark

In many important problems of statistical inference (e.g. the Neyman-Scott problem), the deficiency of the profile likelihood has been modified by various adjustments. We consider such as the conditional profile likelihood and the adjusted profile likelihood in Section 3. However, as we saw in Section 3, such likelihoods remain bounded away from zero at the end points of the parameter space, and accordingly, any likelihood-based interval for the ratio of means could potentially become the entire real line. Earlier, this problem has been noticed in the Liseo (1993) and Yin and Ghosh (2000) for the special Fieller-Creasy problem. Interestingly, we find here that an integrated likelihood derived under the algorithm of Berger, Liseo and Wolpert (1996) avoids this drawback.

To find the integrated likelihood approach proposed in Berger, Liseo and Wolpert (1999), we begin with the likelihood

$$L(\theta_1, \beta_2, \ldots, \beta_r, \sigma) \propto \sigma^{-n} exp[-\frac{1}{2\sigma^2}\{SSE + c_{11}(\beta_2\theta_1 - \hat{\beta_1})^2 + 2c_{12}(\beta_2\theta_1 - \hat{\beta_1})(\beta_2 - \hat{\beta_2}) + c_{22}(\beta_2 - \hat{\beta_2})^2 + \boldsymbol{u}^T\boldsymbol{A}_3\boldsymbol{u}\}] \ .$$

The corresponding Fisher information matrix is given by

$$\boldsymbol{I}(\theta_1, \beta_2, \ldots, \beta_r, \sigma) = \frac{1}{\sigma^2}\begin{bmatrix} c_{11}\beta_2^2 & \beta_2(c_{11}+c_{12}) & \mathbf{0}^T & 0 \\ \beta_2(c_{11}+c_{12}) & Q(\theta_1) & \mathbf{0}^T & 0 \\ \mathbf{0} & \mathbf{0} & \boldsymbol{A}_3 & \mathbf{0} \\ 0 & 0 & \mathbf{0}^T & 2n \end{bmatrix} .$$

Following Berger, Liseo and Wolpert (1999), we begin with the development of conditional reference integrated likelihood. To this end, first we start with the reference priors $\pi^*(\beta_2, \ldots, \beta_r, \sigma \mid \theta_1) \propto \sigma^{-r} Q^{1/2}(\theta_1)$ based on the positive square root of the determinant of the appropriate submatrix of the Fisher information matrix. Next, taking the sequence of compact intervals $[-i, i] \times [-i, i] \times \cdots \times [-i, i] \times [i^{-1}, i]$, $i = 1, 2, 3, \ldots$ for $(\beta_2, \ldots, \beta_r, \sigma)$, following Berger, Liseo and Wolpert (1999)

$$k_i^{-1}(\theta_1) = \int_{i^{-1}}^{i} \int_{-i}^{i} \cdots \int_{-i}^{i} \sigma^{-r} Q^{1/2}(\theta_1) d\beta_2 d\beta_3 \ldots d\beta_r d\sigma \propto Q^{1/2}(\theta_1) \quad .$$

Thus, $\lim_{i \to \infty} k_i(\theta_1)/k_i(\theta_{10}) \propto Q^{-1/2}(\theta_1)$ where θ_{10} is a fixed value of θ_1 . Now from Berger, Liseo and Wolpert (1999), the conditional reference prior is given by

$$\pi^R(\beta_2, \ldots, \beta_r, \sigma \mid \theta_1) \propto \sigma^{-r} Q^{1/2}(\theta_1) Q^{-1/2}(\theta_1) = \sigma^{-r} \quad .$$

The corresponding integrated likelihood after some simplification is then

$$\begin{aligned} L_I(\theta_1) &\propto \int_0^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \pi^R(\beta_2, \ldots, \beta_r, \sigma \mid \theta_1)\, d\beta_2 d\beta_3 \ldots d\beta_r d\sigma \\ &\propto Q^{-1/2}(\theta_1)[SSE + |\boldsymbol{C}|^2 (\hat{\beta}_2 \theta_1 - \hat{\beta}_1)^2 / Q(\theta_1)]^{-n/2} \end{aligned}$$

which multiplies the profile likelihood by the factor $Q^{-1/2}(\theta_1)$. The advantage of the integrated likelihood over the profile likelihood is that it tends to 0 as $|\theta_1| \to \infty$ due to the multiplying factor $Q^{-1/2}(\theta_1)$.

5. Appendix

Proof of Theorem 1

Let $\boldsymbol{I}(\theta) = \boldsymbol{J}\boldsymbol{I}(\underline{\beta}, \sigma)\boldsymbol{J}^T$. Then

$$(\boldsymbol{JI})_{11} = \frac{\theta_2}{\theta_{r+1}^2}[s_{11}\{\theta_1 h'(\theta_1) + h(\theta_1)\} + s_{12}h'(\theta_1) - \sum_{j=3}^{r} s_{1j}g'_j(\theta_1)]$$

$$(\boldsymbol{JI})_{12} = \frac{\theta_2}{\theta_{r+1}^2}[s_{12}\{\theta_1 h'(\theta_1) + h(\theta_1)\} + s_{22}h'(\theta_1) - \sum_{j=3}^{r} s_{2j}g'_j(\theta_1)]$$

$$(\boldsymbol{JI})_{1i} = \frac{\theta_2}{\theta_{r+1}^2}[s_{1i}\{\theta_1 h'(\theta_1) + h(\theta_1)\} + s_{2i}h'(\theta_1) - \sum_{j=3}^{r} s_{i1j}g'_j(\theta_1)] \ , \quad i = 3, \ldots, r$$

$$(\boldsymbol{JI})_{1,r+1} = 0 \quad .$$

Hence, if

$$\begin{bmatrix} g'_3(\theta_1) \\ \vdots \\ g'_r(\theta_1) \end{bmatrix} = \begin{bmatrix} s_{33} & \cdots & s_{3r} \\ \vdots & \ddots & \vdots \\ s_{3r} & \cdots & s_{rr} \end{bmatrix}^{-1} \begin{bmatrix} s_{13} & s_{23} \\ \vdots & \vdots \\ s_{1r} & s_{2r} \end{bmatrix} \begin{bmatrix} \theta_1 h'(\theta_1) + h(\theta_1) \\ h'(\theta_1) \end{bmatrix} \quad ,$$

then one has $(\boldsymbol{JI})_{1i} = 0, \ i = 3, \ldots, r$. So pick

$$\begin{aligned} \begin{bmatrix} g_3(\theta_1) \\ \vdots \\ g_r(\theta_1) \end{bmatrix} &= \begin{bmatrix} s_{33} & \cdots & s_{3r} \\ \vdots & \ddots & \vdots \\ s_{3r} & \cdots & s_{rr} \end{bmatrix}^{-1} \begin{bmatrix} s_{13} & s_{23} \\ \vdots & \vdots \\ s_{1r} & s_{2r} \end{bmatrix} \begin{bmatrix} \theta_1 h(\theta_1) \\ h(\theta_1) \end{bmatrix} \\ &= \begin{bmatrix} a_{31} & a_{32} \\ \vdots & \vdots \\ a_{r1} & a_{r2} \end{bmatrix} \begin{bmatrix} \theta_1 h(\theta_1) \\ h(\theta_1) \end{bmatrix} \\ &= h(\theta_1) \begin{bmatrix} a_{31}\theta_1 + a_{32} \\ \vdots \\ a_{r1}\theta_1 + a_{r2} \end{bmatrix} \quad . \end{aligned}$$

Now, since $\boldsymbol{I}(\theta) = \boldsymbol{JI}(\underset{\sim}{\beta}, \sigma)\boldsymbol{J}^T$, one has

$$(\boldsymbol{I}(\theta))_{1i} = (\boldsymbol{JI})_{1i} = 0 \ , \ i = 3, .., r, r+1 \ .$$

Similarly,

$$(\boldsymbol{I}(\theta))_{2i} = (\boldsymbol{JI})_{2i} = 0 \ , \ i = 3, .., r, r+1 \ .$$

Note that

$$
\begin{aligned}
(\boldsymbol{I}(\theta))_{12} &= (\boldsymbol{JIJ^T})_{12} \\
&= \frac{\theta_2 h(\theta_1)}{\theta_{r+1}^2}[s_{11}\theta_1\{\theta_1 h'(\theta_1)+h(\theta_1)\}+2s_{12}\theta_1 h'(\theta_1)-\theta_1\sum_{j=3}^{r} s_{1j}g'_j(\theta_1) \\
&\quad + s_{12}h(\theta_1)+s_{22}h'(\theta_1)-\sum_{j=3}^{r} s_{2j}g'_j(\theta_1)] \\
&= \frac{\theta_2 h(\theta_1)}{\theta_{r+1}^2}[s_{11}\theta_1\{\theta_1 h'(\theta_1)+h(\theta_1)\}+2s_{12}\theta_1 h'(\theta_1) \\
&\quad -\theta_1\sum_{j=3}^{r} s_{1j}\{a_{j1}[\theta_1 h'(\theta_1)+h(\theta_1)]+a_{j2}h'(\theta_1)\} \\
&\quad + s_{12}h(\theta_1)s_{22}h'(\theta_1)-\sum_{j=3}^{r} s_{2j}\{a_{j1}[\theta_1 h'(\theta_1)+h(\theta_1)]+a_{j2}h'(\theta_1)\}] \quad .
\end{aligned}
$$

We need $\boldsymbol{I}_{12}=0$ to satisfy the condition of orthogonality. Thus, we require

$$
\begin{aligned}
0 &= h'(\theta_1)[\theta_1^2(s_{11}-\sum_{j=3}^{r} s_{1j}a_{j1})+\theta_1(2s_{12}-\sum_{j=3}^{r} s_{1j}a_{j2}-\sum_{j=3}^{r} s_{2j}a_{j1}) \\
&\quad + s_{22}-\sum_{j=3}^{r} s_{2j}a_{j2}]+h(\theta_1)[\theta_1(s_{11}-\sum_{j=3}^{r} s_{1j}a_{j1})+s_{12}-\sum_{j=3}^{r} s_{2j}a_{j1}] \ .
\end{aligned}
$$

Also $\sum_{j=3}^{r} s_{1j}a_{j2}=\sum_{j=3}^{r} s_{2j}a_{j1}$, since

$$
\begin{aligned}
\sum_{j=3}^{r} s_{1j}a_{j2} &= (s_{13}\cdots s_{1r})\begin{bmatrix} s_{33} & \cdots & s_{3r} \\ \vdots & \ddots & \vdots \\ s_{3r} & \cdots & s_{rr}\end{bmatrix}^{-1}\begin{bmatrix} s_{23} \\ \vdots \\ s_{2r}\end{bmatrix} \\
\sum_{j=3}^{r} s_{2j}a_{j1} &= (s_{23}\cdots s_{2r})\begin{bmatrix} s_{33} & \cdots & s_{3r} \\ \vdots & \ddots & \vdots \\ s_{3r} & \cdots & s_{rr}\end{bmatrix}^{-1}\begin{bmatrix} s_{13} \\ \vdots \\ s_{1r}\end{bmatrix} \ .
\end{aligned}
$$

Hence, one needs to find $h(\theta_1)$ from

$$
\begin{aligned}
0 &= h'(\theta_1)[\theta_1^2(s_{11}-\sum_{j=3}^{r} s_{1j}a_{j1})+2\theta_1(s_{12}-\sum_{j=3}^{r} s_{1j}a_{j2})+s_{22}-\sum_{j=3}^{r} s_{2j}a_{j2}] \\
&\quad + h(\theta_1)[\theta_1(s_{11}-\sum_{j=3}^{r} s_{1j}a_{j1})+s_{12}-\sum_{j=3}^{r} s_{1j}a_{j2}] \quad .
\end{aligned}
$$

Thus,

$$\frac{h'(\theta_1)}{h(\theta_1)} = -\frac{\theta_1(s_{11} - \sum_{j=3}^{r} s_{1j}a_{j1}) + s_{12} - \sum_{j=3}^{r} s_{1j}a_{j2}}{\theta_1^2(s_{11} - \sum_{j=3}^{r} s_{1j}a_{j1}) + 2\theta_1(s_{12} - \sum_{j=3}^{r} s_{1j}a_{j2}) + s_{22} - \sum_{j=3}^{r} s_{2j}a_{j2}}$$

or equivalently

$$\log h(\theta_1) = -\frac{1}{2}\left[\theta_1^2(s_{11} - \sum_{j=3}^{r} s_{1j}a_{j1}) + 2\theta_1(s_{12} - \sum_{j=3}^{r} s_{1j}a_{j2}) + s_{22} - \sum_{j=3}^{r} s_{2j}a_{j2}\right] .$$

This leads to

$$\begin{aligned} h(\theta_1) &= \left[\theta_1^2(s_{11} - \sum_{j=3}^{r} s_{1j}a_{j1}) + 2\theta_1(s_{12} - \sum_{j=3}^{r} s_{1j}a_{j2}) + s_{22} - \sum_{j=3}^{r} s_{2j}a_{j2}\right]^{-\frac{1}{2}} \\ &= [c_{11}\theta_1^2 + 2c_{12}\theta_1 s_{12} + c_{22}]^{-1/2} \quad . \end{aligned}$$

References

1. Barndorff-Nielsen, O. E.(1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* , 70, 345-365.
2. Berger, J. O., Liseo, B., and Wolpert, R.(1999). Integrated likelihood methods for eliminating nuisance parameters (with discussion). *Statistical Science*, 14, 1-28.
3. Cox, D. R. and Reid, N.(1987). Parameter orthogonality and approximate conditional inference. *JRSS* B, 49, 1-39.
4. Filler, E. C.(1954). Some problems in interval estimation. *JRSS* B, 16, 175-185.
5. Finney, D. J.(1978). *Statistical Methods in Biological Assay,* Charles Griffin & Co.
6. Liseo, B.(1993). Elimination of nuisance parameters with reference priors. *Biometrika* , 80, 295-304.
7. McCullagh, P. and Tibshirani, R.(1990). A simple method for the adjustment of profile likelihoods. *JRSS* B, 52, 325-344.
8. Yin, M. and Ghosh, M.(2000). Bayesian and Likelihood Inference for the Generalized Fieller-Creasy problem. In *Empirical Bayes and Likelihood Inference*. (Edited by A. E. Ahmed and N. Reid), 121-139. Springer Verlag: New York.

[received date : Nov. 2003, Accepted date : Feb. 2004]