AISM (2006) 58: 457–473
DOI 10.1007/s10463-005-0027-3

Malay Ghosh · Gauri Sankar Datta
Dalho Kim · Trevor J. Sweeting

Likelihood-based inference for the ratios of regression coefficients in linear models

Received: 22 October 2004 / Revised: 23 June 2005 / Published online: 17 June 2006

Abstract We consider the standard linear multiple regression model in which the parameter of interest is the ratio of two regression coefficients. Our setup includes a broad range of applications. We show that the $1 - \alpha$ confidence interval for the interest parameter based on the profile, conditional profile, modified profile or adjusted profile likelihood can potentially become the entire real line, while appropriately chosen integrated likelihoods do not suffer from this drawback. We further explore the asymptotic length of confidence intervals in order to compare integrated likelihood-based proposals. The analysis is facilitated by an orthogonal parameterization.

Keywords Adjusted profile likelihood · Adjustments to profile likelihood · Conditional profile likelihood · Expected length of confidence interval · Integrated likelihood · Orthogonal transformation · Profile likelihood

1 Introduction

There is a large class of important statistical problems that can broadly be described under the general heading of inference about the ratio of regression coefficients

M. Ghosh
Department of Statistics, University of Florida, Gainesville, FL 32611, USA
E-mail: ghoshm@stat.ufl.edu

G. S. Datta (✉)
Department of Statistics, University of Georgia, Athens, GA 30602, USA
E-mail: gauri@stat.uga.edu

D. Kim
Department of Statistics, Kyungpook National University, Taegu, South Korea
E-mail: dalkim@kyungpook.ac.kr

T. J. Sweeting
Department of Statistical Science, University College London, WC1E 6BT, London, UK
E-mail: trevor@stats.ucl.ac.uk

in the general linear model. This class includes the calibration problem, ratio of two means or Fieller–Creasy problem (cf. Fieller, 1954; Creasy, 1954), slope-ratio assay, parallel-line assay and bioequivalence. The general nature of such problems was recognized several decades ago, as documented for example in the excellent treatise of Finney (1978). There are many articles on these specific problems based on a frequentist perspective. However, solutions to these problems have typically encountered serious difficulties. For example, confidence regions for the ratio of two normal means based on Fieller (1954) pivot may be the entire real line.

We consider here likelihood-based intervals for the ratio of two regression coefficients, which is our interest parameter. We discuss the profile likelihood and its modifications, such as the conditional profile likelihood (Cox and Reid, 1987), modified profile likelihood (Barndorff-Nielsen, 1983) and adjusted profile likelihood (McCullagh and Tibshirani, 1990), and compare these against a certain class of integrated likelihoods (Kalbfleisch and Sprott, 1970; Berger et al. 1999). All these likelihoods may be viewed as special cases of adjustments to the profile likelihood (cf. DiCiccio and Stern, 1994). Wallace (1958) and Scheffé (1970) also discuss confidence intervals for the ratio of two regression coefficients. The key to the derivation of the various likelihoods is a non-trivial orthogonal reparameterization of the original parameter vector given by Ghosh et al. (2003) and reviewed here in Sect. 2.

While the conditional, modified and adjusted profile likelihoods have been very effective frequentist tools in eliminating nuisance parameters, they suffer from a drawback for the problems as mentioned earlier. Since all these likelihoods remain bounded away from zero at the end-points of the parameter space, as shown in Sect. 3, the resulting likelihood-based confidence regions could potentially become the entire real line. Explicit conditions which guarantee this are provided in Sect. 4. However, the integrated likelihoods considered in this paper avoid this problem. We further show in Sect. 5 that a confidence set resulting from the integrated likelihood under the one-at-a-time unconditional reference prior has a certain asymptotic superiority over its competitors. We illustrate our results in Sect. 6 with an example on a parallel-line assay and we make some concluding remarks in Sect. 7. The proofs of some technical results in Sects. 3 and 4 are deferred to Appendix.

2 An orthogonal transformation

Consider the general regression model:

$$y_i = \sum_{j=1}^{r} \beta_j x_{ij} + e_i \ , \quad i = 1, \ldots, n, \tag{1}$$

where the errors e_i are independent $\mathcal{N}(0, \sigma^2)$. Here $\beta_j \in (-\infty, \infty)$ for $j \neq 2$ while $\beta_2 \in (-\infty, \infty) - \{0\}$. The parameter of interest is $\theta_1 = \beta_1/\beta_2$. We write $Y = (y_1, \ldots, y_n)^{\mathrm{T}}$, $x_i = (x_{i1}, \ldots, x_{ir})^{\mathrm{T}}$, $i = 1, \ldots, n$, $X^{\mathrm{T}} = (x_1, \ldots, x_n)$, $\beta = (\beta_1, \ldots, \beta_r)^{\mathrm{T}}$, $e = (e_1, \ldots, e_n)^{\mathrm{T}}$. Let $\operatorname{rank}(X) = r < n$. Thus in matrix notation the model can be rewritten as $Y = X\beta + e$.

Ghosh, Yin and Kim (2003) have given a transformation of the parameter vector $(\beta^{\mathrm{T}}, \sigma)$ that results in the orthogonality of θ_1 with the remaining parameters. Define

$S = (s_{ij}), i, j = 1, \ldots, r$ by $S = n^{-1}X^{\mathrm{T}}X$ and $S_{11} = (s_{ij}), i, j = 1, 2,\ S_{12} = S_{21}^{\mathrm{T}} = (s_{ij}), i = 1, 2, j = 3, \ldots, r, S_{22} = (s_{ij}), i, j = 3, \ldots, r$. Since $\mathrm{rank}(X) = r$, S, S_{22} and $C = S_{11} - S_{12}A$ are positive definite, where $A = S_{22}^{-1}S_{21}$. Write $A = (a_{ij}), i = 3, \ldots, r, j = 1, 2$ and $C = (c_{ij}), i, j = 1, 2$. Since C is positive definite, the quantity $Q(\theta_1) = c_{11}\theta_1^2 + 2c_{12}\theta_1 + c_{22}$ is positive for all θ_1. Consider now the transformation:

$$\begin{aligned} &\beta_1 = \theta_1\beta_2,\ \beta_2 = \theta_2 Q^{-1/2}(\theta_1), \\ &\beta_j = \theta_j - \beta_2(a_{j1}\theta_1 + a_{j2}),\ j = 3, \ldots, r,\ \sigma = \theta_{r+1}. \end{aligned} \tag{2}$$

Let $\psi = (\theta_2, \ldots, \theta_{r+1})$. It is shown in Ghosh, Yin and Kim (2003) that θ_1 is orthogonal to ψ and the reparameterized Fisher information matrix of $\theta = (\theta_1, \psi)$ is given by

$$I(\theta) = \frac{n}{\theta_{r+1}^2} \text{ Block diag } \left(\frac{\theta_2^2|C|}{Q^2(\theta_1)}, 1, S_{22}, 2 \right).$$

We will repeatedly use this information matrix for the development of various likelihoods for the parameter of interest θ_1. We treat the rest of the parameters ψ as nuisance parameters.

3 Development of likelihoods

We first derive the profile likelihood of θ_1. Writing $\widehat{\beta} = (\widehat{\beta}_1, \ldots, \widehat{\beta}_r)^{\mathrm{T}}$ as the maximum likelihood estimator (MLE) of β and $\mathrm{SSE} = (Y - X\widehat{\beta})^{\mathrm{T}}(Y - X\widehat{\beta})$, the likelihood is

$$L(\beta_1, \beta_2, \ldots, \beta_r, \sigma) \propto \sigma^{-n} \exp\left[-\frac{1}{2\sigma^2}\left\{ \mathrm{SSE} + (\beta - \widehat{\beta})^{\mathrm{T}}X^{\mathrm{T}}X(\beta - \widehat{\beta}) \right\}\right].$$

Next, writing $\phi^{\mathrm{T}} = (\theta_1, 1)$, $\hat{\gamma}^{\mathrm{T}} = (\hat{\beta}_1, \hat{\beta}_2)$, by an identity for partitioned matrices we get

$$n^{-1}(\beta - \widehat{\beta})^{\mathrm{T}}X^{\mathrm{T}}X(\beta - \widehat{\beta}) = (\beta_2\phi - \hat{\gamma})^{\mathrm{T}}C(\beta_2\phi - \hat{\gamma}) + u^{\mathrm{T}}S_{22}u,$$

where $u^{\mathrm{T}} = (\beta_3 - \widehat{\beta}_3, \ldots, \beta_r - \widehat{\beta}_r) + (\beta_2\theta_1 - \hat{\beta}_1, \beta_2 - \hat{\beta}_2)S_{12}S_{22}^{-1}$. Now, by maximizing with respect to $(\beta_3, \ldots, \beta_r)$, β_2 and σ in succession, one obtains the profile likelihood of θ_1 as

$$L_{\mathrm{PL}}(\theta_1) \propto \left\{ \mathrm{SSE} + n|C|(\widehat{\beta}_2\theta_1 - \widehat{\beta}_1)^2/Q(\theta_1) \right\}^{-n/2}. \tag{3}$$

Remark 1 Note that as $|\theta_1| \to \infty$, $(\widehat{\beta}_2\theta_1 - \widehat{\beta}_1)^2/Q(\theta_1) \to \widehat{\beta}_2^2/c_{11}$ so from Eq. (3) $L_{\mathrm{PL}}(\theta_1)$ is bounded away from 0 as $|\theta_1| \to \infty$. This immediately suggests that any likelihood-based confidence interval for θ_1 can potentially be the entire real line; see Sect. 4.

Remark 2 In many important problems of statistical inference, such as the Neyman–Scott problem, the deficiency of the profile likelihood has been rectified by various adjustments. One such adjustment is the conditional profile likelihood, as proposed by Cox and Reid (1987). Let $L(\theta)$ denote the likelihood function and $l(\theta)$ the log-likelihood function of θ. Let $\tilde{\psi}(\theta_1)$ denote the MLE of ψ for fixed θ_1 and $(\theta_1, \tilde{\psi}(\theta_1)) = \tilde{\theta}(\theta_1)$. It is shown in Appendix that

$$L_{\mathrm{CPL}}(\theta_1) \propto \left\{\mathrm{SSE} + n|C|(\widehat{\beta}_2\theta_1 - \widehat{\beta}_1)^2/Q(\theta_1)\right\}^{-n-r/2}. \tag{4}$$

Clearly $L_{\mathrm{CPL}}(\theta_1)$, like $L_{\mathrm{PL}}(\theta_1)$, remains bounded away from zero as $|\theta_1| \to \infty$. Moreover, due to their similarity in form, both produce identical confidence sets.

Remark 3 A second adjustment to the profile likelihood due to Barndorff-Nielsen (1983) is the modified profile likelihood given by $L_{\mathrm{MPL}}(\theta_1) = L_{\mathrm{PL}}(\theta_1)M(\theta_1)$, where $M(\theta_1)$ is an adjustment factor. This adjustment factor, derived in Appendix, leads to

$$L_{\mathrm{MPL}}(\theta_1) \propto L_{\mathrm{PL}}(\theta_1)Q^{1/2}(\theta_1)(c_{11}\theta_1\hat{\theta}_1 + c_{12}\theta_1 + c_{12}\hat{\theta}_1 + c_{22})^{-1}\tilde{\theta}_{r+1}^{(r+1)}(\theta_1). \tag{5}$$

Thus as $|\theta_1| \to \infty$, $L_{\mathrm{MPL}}(\theta_1)/L_{\mathrm{PL}}(\theta_1)$ tends to a constant depending only on y and X and hence $L_{\mathrm{MPL}}(\theta_1)$ also remains bounded away from 0 as $|\theta_1| \to \infty$.

Remark 4 McCullagh and Tibshirani (1990) proposed an adjustment to the profile likelihood based on unbiased estimating functions. Let $\tilde{\eta}(\theta_1)$ be the maximum likelihood estimator of η for fixed θ_1, where $\eta = (\theta_1, \beta_2, \ldots, \beta_r, \sigma)$. Let $\widetilde{U}(\theta_1) = \{U(\theta_1) - m(\theta_1)\}w(\theta_1)$, where $U(\theta_1) = (\partial/\partial\theta_1)\log L_{\mathrm{PL}}(\theta_1)$ is the score function based on the profile log-likelihood, and $m(\theta_1)$ and $w(\theta_1)$ are chosen such that $E_{\tilde{\eta}(\theta_1)}\widetilde{U}(\theta_1) = 0$ and $V_{\tilde{\eta}(\theta_1)}\widetilde{U}(\theta_1) = -E_{\tilde{\eta}(\theta_1)}(\partial/\partial\theta_1)\widetilde{U}(\theta_1)$. Then

$$\begin{aligned} m(\theta_1) &= E_{\tilde{\eta}(\theta_1)}U(\theta_1), \\ w(\theta_1) &= \{V_{\tilde{\eta}(\theta_1)}U(\theta_1)\}^{-1}\left\{-E_{\tilde{\eta}(\theta_1)}\frac{\partial^2}{\partial\theta_1^2}\log L_{\mathrm{PL}}(\theta_1) + \frac{\partial}{\partial\theta_1}m(\theta_1)\right\}. \end{aligned} \tag{6}$$

Then the adjusted profile likelihood of McCullagh and Tibshirani is given by

$$L_{\mathrm{APL}}(\theta_1) = \exp\left\{\int_{\theta_{1*}}^{\theta_1}\widetilde{U}(t_1)\mathrm{d}t_1\right\}, \tag{7}$$

where θ_{1*} is an arbitrary point in the parameter space of θ_1.
In our case it can be shown that

$$\begin{aligned} \tilde{\beta}_2(\theta_1) &= \frac{\theta_1(c_{11}\hat{\beta}_1 + c_{12}\hat{\beta}_2) + c_{12}\hat{\beta}_1 + c_{22}\hat{\beta}_2}{Q(\theta_1)}, \\ \tilde{\sigma}(\theta_1) &= \left\{\frac{\mathrm{SSE}}{n} + \frac{|C|(\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2}{Q(\theta_1)}\right\}^{1/2}. \end{aligned} \tag{8}$$

The theorem below, proved in Appendix, provides expressions for $m(\theta_1)$ and $w(\theta_1)$.

Theorem 1

$$m(\theta_1) = 0 \textit{ and } w(\theta_1) = \frac{n-r-1}{n} \frac{Q(\theta_1)\tilde{\beta}_2^2(\theta_1)}{Q(\theta_1)\tilde{\beta}_2^2(\theta_1) + \tilde{\sigma}^2(\theta_1)/n}.$$

Let $l_{\text{APL}}(\theta_1) = \log L_{\text{APL}}(\theta_1)$. Then, for $\theta_{1*} < \theta_1$, Theorem 1 and Eq. (7) imply that $|l_{\text{APL}}(\theta_1)| \le \int_{\theta_{1*}}^{\theta_1} |U(t)| \mathrm{d}t$. Since $U(t) = O_p(t^{-2})$ for large $|t|$, it follows that l_{APL} does not diverge to $-\infty$ as $|\theta_1| \to \infty$. In otherwords, L_{APL} also remains positive as $|\theta_1| \to \infty$.

Yet another approach to this problem is the one based on the integrated likelihood as introduced by Kalbfleisch and Sprott (1970) and more recently discussed in Berger, Liseo and Wolpert (1999). We begin with the likelihood for $\eta = (\theta_1, \beta_2, \ldots, \beta_r, \sigma)$:

$$L(\eta) \propto \sigma^{-n} \exp\left[-\frac{n}{2\sigma^2}\left\{ n^{-1}\text{SSE} + (\beta_2\phi - \hat{\gamma})^{\mathrm{T}} C(\beta_2\phi - \hat{\gamma}) + u^{\mathrm{T}} S_{22} u \right\}\right]. \tag{9}$$

The submatrix of the Fisher information corresponding to $(\beta_2, \ldots, \beta_r, \sigma)$ is given by

$$\frac{n}{\sigma^2}\begin{bmatrix} \phi^{\mathrm{T}} S_{11}\phi & \phi^{\mathrm{T}} S_{12} & 0 \\ S_{21}\phi & S_{22} & 0 \\ 0 & 0^{\mathrm{T}} & 2 \end{bmatrix}. \tag{10}$$

Following Berger, Liseo and Wolpert (1999) we calculate the conditional reference integrated likelihood. To this end, first we start with the reference prior $\pi^*(\beta_2, \ldots, \beta_r, \sigma|\theta_1) \propto \sigma^{-r}(\phi^{\mathrm{T}} C\phi)^{1/2} = \sigma^{-r} Q^{1/2}(\theta_1)$ based on the square root of the determinant of Eq. (10). Next, taking the sequence of compact intervals $[-i, i]^{r-1} \times [i^{-1}, i],\ i = 1, 2, 3, \ldots$ for $\beta_2, \ldots, \beta_r, \sigma$, following Berger, Liseo and Wolpert (1999),

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

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

$$\pi^{\text{CR}}(\beta_2, \ldots, \beta_r, \sigma|\theta_1) \propto \sigma^{-r} Q^{1/2}(\theta_1) Q^{-1/2}(\theta_1) = \sigma^{-r}.$$

The corresponding integrated likelihood, after some simplification, is then

$$\begin{aligned} L_{\text{I}}^{\text{CR}}(\theta_1) &\propto \int_0^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} L(\theta_1, \beta_2, \ldots, \beta_r, \sigma)\pi^{\text{CR}} \\ &\quad \times(\beta_2, \ldots, \beta_r, \sigma|\theta_1)\mathrm{d}\beta_2\, \mathrm{d}\beta_3 \cdots \mathrm{d}\beta_r\, \mathrm{d}\sigma \\ &\propto Q^{-1/2}(\theta_1)\left\{\text{SSE} + n|C|(\widehat{\beta}_2\theta_1 - \widehat{\beta}_1)^2/Q(\theta_1)\right\}^{-n/2} \\ &\propto Q^{-1/2}(\theta_1) L_{\text{PL}}(\theta_1). \end{aligned}$$

The advantage of the integrated likelihood over the profile likelihood is that the former tends to 0 as $|\theta_1| \to \infty$ due to the multiplying factor $Q^{-1/2}(\theta_1)$. We note that in the θ-parameterization, the conditional reference prior reduces to $\pi^{\mathrm{CR}}(\theta_2, \ldots, \theta_{r+1}|\theta_1) \propto \theta_{r+1}^{-r} Q^{-1/2}(\theta_1)$.

Remark 5 Interestingly, if one uses instead Jeffreys' prior $\pi^{\mathrm{J}}(\beta_2, \ldots, \beta_r, \sigma|\theta_1) \propto \sigma^{-r} Q^{1/2}(\theta_1)$ as the conditional prior, the resulting integrated likelihood is the same as the profile likelihood and thus has the same disadvantages pointed out earlier. However, for the one-at-a-time conditional reference prior $\pi^{\mathrm{OCR}}(\beta_2, \ldots, \beta_r, \sigma|\theta_1) \propto \sigma^{-1}$, the integrated likelihood is $L_{\mathrm{I}}^{\mathrm{OCR}}(\theta_1) \propto Q^{-1/2}(\theta_1)\Big\{\mathrm{SSE} + n|C|(\widehat{\beta}_2\theta_1 - \widehat{\beta}_1)^2/Q(\theta_1)\Big\}^{-n-r+1/2}$, which, due to the factor $Q^{-1/2}(\theta_1)$, also tends to 0 as $|\theta_1| \to \infty$. Clearly there are other choices. For example, for the one-at-a-time Berger–Bernardo unconditional reference prior $\pi^{\mathrm{R}}(\theta_2, \ldots, \theta_{r+1}|\theta_1) \propto \theta_{r+1}^{-1} Q^{-1}(\theta_1)$, as found in Ghosh, Yin and Kim (2003), the integrated likelihood is $L_{\mathrm{I}}^{\mathrm{R}}(\theta_1) \propto Q^{-1}(\theta_1)\Big\{\mathrm{SSE} + n|C|(\widehat{\beta}_2\theta_1 - \widehat{\beta}_1)^2/Q(\theta_1)\Big\}^{-n-r+1/2}$.

4 Likelihood-based confidence regions

We begin with the derivation of a confidence region for θ_1 using the profile likelihood. Such a region is obtained from the corresponding likelihood ratio test. In particular, if $\hat{\theta}_1$ denotes the maximum likelihood estimator of θ_1, then writing $\lambda_{\mathrm{PL}}(\theta_1) = L_{\mathrm{PL}}(\theta_1)/L_{\mathrm{PL}}(\hat{\theta}_1)$ and $l_{\mathrm{PL}}(\theta_1) = \log L_{\mathrm{PL}}(\theta_1)$, we get

$$\begin{aligned} -2 \log \lambda_{\mathrm{PL}}(\theta_1) &= 2\{l_{\mathrm{PL}}(\hat{\theta}_1) - l_{\mathrm{PL}}(\theta_1)\} \\ &= n \log\left\{1 + \frac{n|C|(\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2/Q(\theta_1)}{\mathrm{SSE}}\right\} \end{aligned}$$

which is monotonically increasing in $F = n|C|(\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2 Q^{-1}(\theta_1)/\mathrm{MSE}$, where $\mathrm{MSE} = \mathrm{SSE}/(n-r)$ and the latter has the F-distribution with $1, n-r$ degrees of freedom. Thus the acceptance region for $H_0 : \theta_1 = \theta_{10}$ against the alternative $H_1 : \theta_1 \neq \theta_{10}$ is given by

$$A(\theta_{10}) = \left\{y : \frac{n|C|(\hat{\beta}_2\theta_{10} - \hat{\beta}_1)^2/Q(\theta_{10})}{\mathrm{MSE}} \leq F_{1,n-r;\alpha}\right\},$$

where $F_{1,n-r;\alpha}$ denotes the upper $100(1-\alpha)\%$ point of the F-distribution with $1, n-r$ degrees of freedom. The corresponding confidence region is then given by

$$C(Y) = \{\theta_1 : F \leq F_{1,n-r;\alpha}\}. \tag{11}$$

We now find conditions under which $C(Y)$ becomes the entire real line. Recall $\phi = (\theta_1, 1)^{\mathrm{T}}$, and $\hat{\gamma} = (\hat{\beta}_1, \hat{\beta}_2)^{\mathrm{T}}$. The following lemma is proved in Appendix.

Lemma 1 $sup_{\theta_1}(\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2/Q(\theta_1) = \hat{\gamma}^{\mathrm{T}} C \hat{\gamma}/|C|$.

Table 1 $P(F_{2,n-r}(\zeta) \leq F_{1,n-r;\alpha}/2)$ for different choices of $n-r$, ζ and α

$n-r$	α	ζ				
		0.001	0.01	0.1	1	10
5	0.01	0.973	0.973	0.968	0.914	0.311
	0.05	0.878	0.876	0.861	0.709	0.048
	0.10	0.774	0.771	0.749	0.549	0.012
10	0.01	0.969	0.968	0.961	0.875	0.109
	0.05	0.867	0.865	0.845	0.653	0.015
	0.10	0.758	0.755	0.729	0.499	0.004
15	0.01	0.967	0.967	0.958	0.858	0.067
	0.05	0.862	0.860	0.839	0.632	0.009
	0.10	0.753	0.750	0.722	0.482	0.003

Remark 6 From Eq. (11) and Lemma 1, it follows that $C(Y)$ becomes the real line if and only if $n\hat{\gamma}^{\mathrm{T}}C\hat{\gamma}/\mathrm{MSE} \leq F_{1,n-r;\alpha}$. Since $n\hat{\gamma}^{\mathrm{T}}C\hat{\gamma}/(2\mathrm{MSE})$ is distributed as the non-central F with degrees of freedom 2 and $n-r$ and non-centrality parameter $\zeta = n\gamma^{\mathrm{T}}C\gamma/(2\sigma^2)$, where $\gamma = (\beta_1, \beta_2)^{\mathrm{T}}$, the probability of $C(Y)$ being the real line is given by $P(F_{2,n-r}(\zeta) \leq F_{1,n-r;\alpha}/2)$. Table 1 gives these probabilities for different choices of $n-r$, ζ and α and shows very clearly that for small values of ζ there is a high probability that $C(Y)$ is the entire real line.

Next, the maximum likelihood estimator of θ_1 based on the conditional profile likelihood given in Eq. (4) continues to be $\hat{\beta}_1/\hat{\beta}_2 = \hat{\theta}_1$. Hence, writing $l_{\mathrm{CPL}}(\theta_1) = \log L_{\mathrm{CPL}}(\theta_1)$ and $\lambda_{\mathrm{CPL}}(\theta_1) = L_{\mathrm{CPL}}(\theta_1)/L_{\mathrm{CPL}}(\hat{\theta}_1)$, one gets

$$\begin{aligned} -2\log\lambda_{\mathrm{CPL}}(\theta_1) &= 2\{l_{\mathrm{CPL}}(\hat{\theta}_1) - l_{\mathrm{CPL}}(\theta_1)\} \\ &= (n-r)\log\left\{1 + \frac{n|C|(\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2 Q^{-1}(\theta_1)}{\mathrm{SSE}}\right\}. \end{aligned}$$

Thus the conditional profile likelihood-based confidence region is the same as that based on profile likelihood and, as pointed out in Remark 6, can become the entire real line.

A similar phenomenon for the adjusted profile likelihood is given in Lemma 2, proved in Appendix. Let $\hat{\theta}_{1,\mathrm{APL}} = \arg\sup_{\theta_1} L_{\mathrm{APL}}(\theta_1)$ and let $\lambda_{\mathrm{APL}}(\theta_1) = L_{\mathrm{APL}}(\theta_1)/L_{\mathrm{APL}}(\hat{\theta}_{1,\mathrm{APL}})$.

Lemma 2

$$sup_{\theta_1}\{-2\,log_{\mathrm{e}}\lambda_{\mathrm{APL}}(\theta_1)\} \leq K\hat{\gamma}^{\mathrm{T}}C\hat{\gamma}/MSE,$$

where K is a positive constant that depends only on n and C.

Remark 7 Lemma 2 states that an adjusted profile likelihood-based confidence set becomes the entire real line for a given α if $K\hat{\gamma}^{\mathrm{T}}C\hat{\gamma}/\mathrm{MSE}$ does not exceed the upper $100\alpha\%$ point of the distribution of $-2\log_{\mathrm{e}}\lambda_{\mathrm{APL}}$. For the asymptotic distribution of $-2\log\lambda_{\mathrm{APL}}(\theta_1)$ we write

$$\begin{aligned} -2\log\lambda_{\mathrm{APL}}(\theta_1) &= 2\{l_{\mathrm{PL}}(\hat{\theta}_{1,\mathrm{APL}}) - l_{\mathrm{PL}}(\theta_1)\} \\ &\quad -2\int_{\theta_1}^{\hat{\theta}_1}(1-w(t))U(t)dt. \end{aligned} \tag{12}$$

But $1 - w(t) = \tilde{\sigma}^2(t)/\{n\tilde{\beta}_2^2(t)Q(t) + \tilde{\sigma}^2(t)\}$. As $|t| \to \infty$, it follows from Eq. (8) that both $\tilde{\beta}_2^2(t)Q(t)$ and $\tilde{\sigma}^2(t)$ are bounded in probability. Since $U(t) = O_p(|t|^{-2})$, by Eq. (12),

$$-2 \log \lambda_{\text{APL}}(\theta_1) = 2\{l_{\text{PL}}(\hat{\theta}_{1,\text{APL}}) - l_{\text{PL}}(\theta_1)\} + O_p(n^{-1}). \tag{13}$$

Since $2\{l_{\text{PL}}(\hat{\theta}_{1,\text{APL}}) - l_{\text{PL}}(\theta_1)\}$ is asymptotically distributed as χ_1^2, it follows from Eq. (13) that $-2 \log \lambda_{\text{APL}}(\theta_1)$ has the same asymptotic distribution. Thus asymptotic confidence regions for θ_1 can be constructed using percentiles of the χ_1^2 distribution.

The confidence regions for θ_1 based on certain integrated likelihoods are bounded. This is because $Q(\theta_1) \to \infty$ as $|\theta_1| \to \infty$. Hence, writing $\hat{\theta}_{1,\text{I}}^{\text{CR}} = \arg\sup_{\theta_1} L_{\text{I}}^{\text{CR}}(\theta_1)$, $-2 \log \lambda_{\text{I}}^{\text{CR}}(\theta_1) = 2\{\log L_{\text{I}}(\hat{\theta}_{1,\text{I}}^{\text{CR}}) - \log L_{\text{I}}^{\text{CR}}(\theta_1)\} \to \infty$ as $|\theta_1| \to \infty$. The same is true for the integrated likelihoods $L_{\text{I}}^{\text{OCR}}(\theta_1)$ and $L_{\text{I}}^{\text{R}}(\theta_1)$ introduced in the previous section. Finally, to $O(n^{-1})$ the integrated likelihood ratio statistics obtained from $L_{\text{I}}^{\text{CR}}(\theta_1)$, $L_{\text{I}}^{\text{OCR}}(\theta_1)$ and $L_{\text{I}}^{\text{R}}(\theta_1)$ are all asymptotically χ_1^2 and these may be used to obtain approximate confidence intervals for θ_1.

Suppose $s = \sqrt{\text{MSE}}$. Since the F statistic occurring in $-2 \log \lambda_{\text{PL}}(\theta_1)$ is equal to t^2, where

$$t = \frac{\sqrt{n|C|}(\widehat{\beta}_1 - \widehat{\beta}_2\theta_1)}{s\sqrt{Q(\theta_1)}} \tag{14}$$

has the t-distribution with $n-r$ degrees of freedom, we see that the profile likelihood ratio statistic delivers equitailed confidence intervals whenever these are not the entire real line. In contrast, all three integrated likelihood ratios will deliver intervals with unequal tail probabilities. However, we can apply a modification which, to $O(n^{-1})$, will produce equitailed intervals if desired. Furthermore, these intervals will always be finite. The necessary modification is an $O(n^{-1/2})$ bias correction to the signed root log-likelihood ratio statistic; cf. Barndorff-Nielsen (1994). In our case it is relatively straightforward to derive the adjustment by direct calculation.

Consider the conditional reference integrated likelihood ratio and let $R_{\text{I}}^{\text{CR}} = \text{sign}(\hat{\theta}_{1,\text{I}}^{\text{CR}} - \theta_1)\{-2 \log \lambda_{\text{I}}^{\text{CR}}(\theta_1)\}^{1/2}$ be the signed log-likelihood ratio statistic. It is shown in Appendix that $R_{\text{I}}^{\text{CR}} - a^{\text{CR}}$ is standard normal to $O(n^{-1})$, where

$$a^{\text{CR}} = -\frac{sQ'(\hat{\theta}_1)}{2\widehat{\beta}_2\{Q(\hat{\theta}_1)n|C|\}^{1/2}}. \tag{15}$$

Hence an approximate $(1-\alpha)$ confidence interval for θ_1 is $\{\theta_1 : |R_{\text{I}}^{\text{CR}} - a^{\text{CR}}| \le z_{\alpha/2}\}$, where z_α is the upper α point of the standard normal distribution. Furthermore this interval is equitailed to $O(n^{-1})$ and always finite, since $R_{\text{I}}^{\text{CR}} \to \mp\infty$ as $\theta_1 \to \pm\infty$. Note that $\widehat{\beta}_2^2 Q(\hat{\theta}_1)$ in the denominator of Eq. (15) should be computed as $c_{11}\widehat{\beta}_1^2 + 2c_{12}\widehat{\beta}_1\widehat{\beta}_2 + c_{22}\widehat{\beta}_2^2$ to avoid numerical instability when $\widehat{\beta}_2$ is near zero. The corresponding adjustments for the Jeffreys' and unconditional reference integrated likelihood ratio statistics are easily seen to be $a^{\text{OCR}} = a^{\text{CR}} = (1/2)a^{\text{R}}$.

The parameter space Ω in the ratios problem excludes parameter vectors with $\beta_2 = 0$. Since the asymptotic confidence intervals obtained in this paper are non-uniform in neighborhoods of $\beta_2 = 0$, approximate $100(1-\alpha)\%$ coverage is not guaranteed for every value of θ_1 for a given sample size n. However, a quasi-Bayesian definition of an asymptotic $100(1-\alpha)\%$ confidence interval $S_\alpha(Y)$ for θ_1 would be $P_\pi(\theta_1 \in S_\alpha(Y)) \to 1-\alpha$ for every proper prior distribution π on Ω. The requirement of some proper prior distribution on Ω is a mild assumption, since the value $\beta_2 = 0$ would not normally be regarded as being of any special significance.

5 Asymptotic expected volumes of confidence sets

Test statistics and statistical procedures are often compared based on expected lengths of associated confidence intervals (cf. Mukerjee and Reid, 2001). In this section we compare the various adjusted likelihoods via asymptotic expected lengths of the confidence intervals for θ_1. The necessary quantities are based on asymptotic distributions of the likelihood ratio statistics and are not equivalent to limits of the expected lengths. Indeed, for the present class of problems, the latter are always infinity (Gleser and Hwang, 1987).

Mukerjee and Reid (1999) and Datta and DiCiccio (2001) considered confidence sets by inverting approximate $1-\alpha+o(n^{-1})$ acceptance regions of likelihood ratio tests obtained via maximization of the adjusted likelihoods. Let $C_u(Y)$ ($C(Y)$, respectively) be a $1-\alpha+o(n^{-1})$ likelihood ratio confidence set of θ_1 based on the profile likelihood function (adjustments to the profile likelihood, respectively) of θ_1. To compare these adjustments, the aforementioned authors have developed an expression for the change T in the asymptotic expected volume of $C(Y)$ relative to the asymptotic expected volume of $C_u(Y)$. Mukerjee and Reid (1999) showed that T remains the same for Cox and Reid, Barndorff-Nielsen or McCullagh and Tibshirani adjustment. Therefore, in our comparison we consider only the Cox and Reid (1987) adjustment.

We now write down an expression for T. Writing $D_u \equiv \partial/\partial\theta_u$, we define $\lambda_{uvw} = E_\theta\{D_u D_v D_w l(\theta)\}$. Due to orthogonality of θ_1 with θ_i, $i = 2, \ldots, r+1$, from the expression for $2nT$ given by Eq. (20) of Datta and DiCiccio (2001) we get

$$T = \frac{\delta^2}{2I_{11}} + \sum_{u=1}^{r+1} I^{1u} D_u(\delta) - \frac{1}{2}\sum_{u=1}^{r+1}\sum_{s=1}^{r+1}\sum_{t=1}^{r+1} I^{1u} I^{st}\delta\{2D_t(I_{us}) \\ +\lambda_{ust}\} + \frac{\lambda_{111}\delta}{2I_{11}^2} + o(n^{-1}).$$

Using $I^{1i} = 0$ for $i = 2, \ldots, r+1$, $I^{2j} = 0$ for $j = 3, \ldots, r+1$, $\lambda_{122} = 0$, $\lambda_{1ij} = 0$ for $i, j = 3, \ldots, r+1$, in our case, T further simplifies to

$$T = \frac{\delta^2}{2I_{11}} + D_1(\delta I_{11}^{-1}) + o(n^{-1}). \tag{16}$$

Here δ depends on the adjustment term. We denote the δ corresponding to the Cox and Reid adjustment by δ^0 and the δ corresponding to an integrated likelihood for a conditional prior $\pi(\psi|\theta_1)$ by δ^π. From Eqs. (23) and (26) of Datta and DiCiccio (2001) it follows that

$$\delta^0 = \frac{1}{2}\sum_{i=2}^{r+1}\sum_{j=2}^{r+1} I^{ij}\lambda_{1ij}, \quad \delta^\pi = \delta^0 + D_1(\log \pi).$$

Using $\lambda_{1ij} = 0$ for $i, j = 3, \ldots, r+1$, $\lambda_{122} = 0$, $I^{2i} = 0$ for $i = 3, \ldots, r+1$, it follows that

$$\delta^0 = 0, \quad \delta^\pi = D_1(\log \pi). \tag{17}$$

Thus for a conditional prior of the form $\pi_w(\psi|\theta_1) \propto Q^w(\theta_1)g(\psi)$ for some constant w and for a general function $g(\psi)$ we get

$$\delta^{\pi_w} = w\frac{Q'(\theta_1)}{Q(\theta_1)}. \tag{18}$$

Note that δ^{π_w} depends on the prior π_w only through w and θ_1.

Using $I_{11}^{-1} = \theta_{r+1}^2 Q^2(\theta_1)/(n\theta_2^2|C|)$ we get from Eqs. (16) and (18) that the expression for T corresponding to an integrated likelihood for the prior π_w, denoted by T_w, is given by

$$\begin{aligned} T_w &= \frac{\theta_{r+1}^2 w^2\{Q'(\theta_1)\}^2}{2n\theta_2^2|C|} + \frac{w}{n}D_1\left\{\frac{Q'(\theta_1)Q(\theta_1)\theta_{r+1}^2}{\theta_2^2|C|}\right\} + o(n^{-1}) \\ &= \frac{\theta_{r+1}^2}{2n\theta_2^2|C|^2}[(w^2+2w)\{Q'(\theta_1)\}^2 + 2wQ''(\theta_1)Q(\theta_1)] + o(n^{-1}). \end{aligned} \tag{19}$$

The class of priors given by $\pi_w(\theta)$ is quite general and by suitably choosing $g(\cdot)$ and w we get Jeffreys' prior, the one-at-a-time reference prior, the conditional reference prior, the first and second order quantile matching priors and a highest posterior density matching prior for θ_1. In particular, a highest posterior density matching prior is given by $|\theta_2|Q(\theta_1)$ with $w = 1$. The details are given in Ghosh, Yin and Kim (2003). For the other priors mentioned above, w is -1, $-1/2$ or 0. In particular the one-at-a-time unconditional reference prior π^{R} has $w = -1$ and the conditional reference prior π^{CR} has $w = -1/2$. Note that $T_{-1} < T_{-1/2} < 0 < T_1$, where 0 is the value of T for Cox–Reid adjustment. Since the integrated likelihoods corresponding to π^{R} and π^{CR} produce confidence sets with asymptotic expected length shorter than the integrated likelihood based on the above highest posterior density matching prior, in our numerical studies we do not consider the last prior. Note, however, that unlike other objective priors, this prior results in a confidence set which has also posterior coverage equal to $1 - \alpha + o(n^{-1})$.

6 Numerical results

We begin with an example from Finney (1978) involving a parallel-line assay. Consider an experiment in which p doses $(x_{11}, \ldots, x_{1p})$ of a standard drug S is assayed m times and q doses $(x_{21}, \ldots, x_{2q})$ of a test drug T is assayed u times so that a set $\{Z_{1ik}, i = 1, \ldots, p; k = 1, \ldots, m; Z_{2jk}, j = 1, \ldots, q; k = 1, \ldots, u\}$ of $n = pm + qu$ observations are obtained. The assumed model for a parallel-line assay is

$$Z_{1ik} = \alpha + \beta x_{1i} + \epsilon_{1ik}, \quad k = 1, \ldots, m; i = 1, \ldots, p,$$
$$Z_{2jk} = \alpha + \beta(x_{2j} + \rho) + \epsilon_{2jk}, \quad k = 1, \ldots, u; j = 1, \ldots, q,$$

where ϵ_{1ik} and ϵ_{2jk} are independent $\mathcal{N}(0, \sigma^2)$. This model is a special case of Eq. (1), where $\theta_1 = \rho$ is the parameter of interest. In the example considered in (Finney, 1978, p. 105), n=36.

Table 2 gives the exact confidence interval using the F statistic and the likelihood-based confidence intervals for ρ with α=0.01, 0.05 and 0.10 based on the profile likelihood, identical with the conditional profile likelihood, and the three integrated likelihoods based on the conditional reference prior, the one-at-a-time conditional and unconditional reference priors, along with the corresponding lengths. Notice that we have two kinds of confidence intervals based on the integrated likelihoods, the upper one based on the Chi-square approximation and the lower one based on the equitailed approximation. We have also included the HPD credible sets based on the one-at-a-time reference prior π^{R} for the sake of comparison with the likelihood-based confidence intervals. The HPD credible intervals are slightly longer than the other confidence intervals but they have posterior coverage

Table 2 Likelihood-based confidence intervals for ρ with α=0.01, 0.05 and 0.10 using Finney's data

α	0.01		0.05		0.10	
	Interval (length)	Coverage	Interval (length)	Coverage	Interval (length)	Coverage
Exact	(-0.307, 0.676) (0.987)	0.989	(-0.181, 0.536) (0.717)	0.948	(-0.121, 0.471) (0.592)	0. 898
PL	(-0.280, 0.646) (0.926)	0.985	(-0.162, 0.516) (0.678)	0.936	(-0.106, 0.455) (0.561)	0.880
IL with	(-0.279, 0.640) (0.919)	0.984	(-0.162, 0.512) (0.674)	0.934	(-0.106, 0.451) (0.557)	0.877
CR prior	(-0.275, 0.645) (0.920)	0.984	(-0.159, 0.516) (0.675)	0.935	(-0.103, 0.455) (0.558)	0.878
IL with	(0.294, 0.656) (0.950)	0.987	(-0.173, 0.523) (0.696)	0.942	(-0.115, 0.460) (0.575)	0 .888
OCR prior	(-0.289, 0.661) (0.950)	0.987	(-0.169, 0.527) (0.696)	0.942	(-0.111, 0.464) (0.575)	0.888
IL with	(-0.293, 0.650) (0.943)	0.987	(-0.173, 0.518) (0.691)	0.940	(-0.115, 0.456) (0.571)	0.886
R prior	(-0.284, 0.660) (0.944)	0.987	(-0.165, 0.527) (0.692)	0.940	(-0.107, 0.464) (0.571)	0.886
HPD with R prior	(-0.316, 0.675) (0.991)	0.990	(-0.188, 0.534) (0.722)	0.950	(-0.127, 0.469) (0.596)	0.900

Table 3 Confidence sets for ρ with coverage probability $1-\alpha$ based on three integrated likelihoods using 1,000 simulated data

	IL under CR prior		IL under OCR prior		IL under R prior	
1-α	Coverage probability	Expected length	Coverage probability	Expected length	Coverage probability	Expected length
0.99	0.996 (0.0020)	79.3824 (2.3069)	0.998 (0.0014)	79.6600 (2.2224)	0.998 (0.0014)	15.5772 (0.2491)
0.95	0.959 (0.0063)	19.4912 (0.5551)	0.965 (0.0058)	19.5617 (0.5354)	0.966 (0.0057)	7.2921 (0.1128)
0.90	0.918 (0.0087)	10.8141 (0.3007)	0.925 (0.0083)	10.8453 (0.2901)	0.922 (0.0085)	5.1105 (0.0768)

probabilities right on the target and higher than the posterior coverage probabilities of all the other intervals in our consideration.

It is evident from Table 2 that in all the situations considered, the integrated likelihood under the conditional reference prior provides intervals with shortest lengths, although the integrated likelihoods under the one-at-a-time conditional and unconditional reference priors lead to intervals which are only slightly longer. In this case $n\hat{\gamma}^{\mathrm{T}}C\hat{\gamma}/(2\mathrm{MSE}) = 45.92$ is quite large. It is therefore no surprise that a profile likelihood-based confidence interval for ρ becomes the entire real line only for very small values of α which has very limited practical interest.

In contrast, next we undertake a simulation study, generating data from the same parallel-line assay with β_1=0.25, β_2=0.25, β_3=1, σ^2=3 and n, p, q, m, u as before. In most of the runs the confidence set for ρ is an interval. However, sometimes it is the union of two intervals, in which case the length of the confidence set is the sum of the lengths of these two intervals. Based on these simulations, Table 3 provides the coverage probabilities and the asymptotic expected lengths of the confidence sets for ρ using the three integrated likelihoods. The simulation standard errors are given within parentheses. As seen from Table 3, confidence sets based on the unconditional reference prior have lengths much shorter than the ones based on the conditional reference prior and the one-at-a-time conditional reference prior.

Based on the theoretical findings of Sects. 4 and 5, the two examples presented here and others that we have considered but have not reported, it is our recommendation to use an integrated likelihood based on the one-at-a-time unconditional reference prior for the construction of confidence regions for θ_1.

7 Concluding remarks

In this paper we have considered likelihood-based inference for ratios of regression coefficients in linear models, which includes a large class of problems, as indicated in Sect. 1. We have demonstrated both analytically and numerically situations where it becomes imperative to use an integrated likelihood if one is to avoid confidence sets being the entire real line. Although the repeated sampling confidence interpretation is not invalidated by the occasional occurrence of such sets, their final precision is very different from their initial precision, since a confidence level of, say, 95% is attached to an interval in which θ_1 is known to lie with probability one. For the ratios problem our recommended approach is the integrated likelihood based on the one-at-a-time unconditional reference prior of Berger and Bernardo (1992).

We note that it is not lack of knowledge of σ that gives rise to the results in this paper. Exactly the same behavior ensues when σ is known. In particular, for the Fieller–Creasy problem with means $\theta\mu$ and μ and $\sigma = 1$, both the profile and conditional profile likelihoods are proportional to

$$\exp\left[-\frac{1}{2}\left\{\frac{n(\bar{x}-\theta\bar{y})^2}{1+\theta^2}\right\}\right].$$

In contrast, the integrated likelihood multiplies the above by a factor $(\theta^2+1)^{-1/2}$ or $(\theta^2+1)^{-1}$, depending on whether one is using π^{CR} or π^{R}. More generally, the phenomena described in this paper are not confined to the normal regression model (Eq. (1)). For example, it can be shown that in the Fieller–Creasy problem above when the error distributions are Student t, the profile likelihood does not tend to zero as $|\theta| \to \infty$. On the other hand, the information matrix is readily available and the conditional priors associated with the various integrated likelihoods discussed in Sect. 3 are unchanged. Analysis of the exact integrated likelihood is difficult, but an analysis of the Laplace approximation, as given in Sweeting (1987), for example, to the integrated likelihood reveals similar behavior to that obtained in the normal case. A full analysis of location-scale and other regression models involves additional technical difficulties and is a topic for future research.

Acknowledgements Ghosh and Kim's research was partially supported by NSF Grants SBR-9810968 and SBR-9911485. Datta's research was partially supported by NSF Grants DMS-0071642, SES-0241651 and NSA Grant MDA904-03-1-0016. Sweeting's research was partially supported by EPSRC Grant GR/R24210/01.

Appendix

Derivation of Eq. (4)

Clearly, $L_{\mathrm{PL}}(\theta_1)$ given by Eq. (3) is equal to $L(\tilde{\theta}(\theta_1))$. It can be checked that

$$\left\{\left(-\frac{\partial^2 l(\theta)}{\partial\theta_k\partial\theta_m}\bigg|_{\tilde{\theta}(\theta_1)}\right)\right\}_{k,m=2,\ldots,r+1} = n\tilde{\theta}_{r+1}^{-2}(\theta_1)\ \mathrm{Diag}\ (1, S_{22}, 2). \tag{20}$$

Due to the orthogonality of θ_1 and ψ Eqs. (3), (20) and Eq. (10) of Cox and Reid (1987) lead to the conditional profile likelihood of θ_1 as

$$L_{\mathrm{CPL}}(\theta_1) \propto L_{\mathrm{PL}}(\theta_1)\{\tilde{\theta}_{r+1}(\theta_1)\}^r \propto \left\{\mathrm{SSE} + n|C|(\widehat{\beta}_2\theta_1 - \widehat{\beta}_1)^2/Q(\theta_1)\right\}^{-n-r/2},$$

which is Eq. (4).

Derivation of Eq. (5)

To derive $M(\theta_1)$ we first write the logarithm of the original likelihood as $l \equiv l(\theta_1, \psi, \hat{\theta}_1, \hat{\psi})$, where $\hat{\theta}_1$ and $\hat{\psi}$ are the maximum likelihood estimators of θ_1 and ψ. Then

$$M(\theta_1) = \left|-\frac{\partial^2 l}{\partial\psi\partial\hat{\psi}^T}\Big|_{\tilde{\theta}(\theta_1)}\right|^{-1} \left|-\frac{\partial^2 l}{\partial\psi\partial\psi^T}\Big|_{\tilde{\theta}(\theta_1)}\right|^{1/2} \tag{21}$$

by Eq. (8.7) of Barndorff-Nielsen (1994). It follows after considerable algebra that

$$\begin{aligned}\left|-\frac{\partial^2 l}{\partial\psi\partial\hat{\psi}^{\mathrm{T}}}\Big|_{\tilde{\theta}(\theta_1)}\right| &= g(\hat{\theta})Q^{-1/2}(\theta_1)\tilde{\theta}_{r+1}^{-(2r+1)}(\theta_1)\\ &\quad\times(c_{11}\theta_1\hat{\theta}_1 + c_{12}\theta_1 + c_{12}\hat{\theta}_1 + c_{22}),\end{aligned} \tag{22}$$

where $g(\hat{\theta})$ only involves $\hat{\theta}$. This derivation is carried out by expressing l as $l = -n\log\theta_{r+1} - (2\theta_{r+1}^2)^{-1}[n\hat{\theta}_{r+1}^2 + \{\beta(\theta_1,\psi) - \beta(\hat{\theta}_1,\hat{\psi})\}^{\mathrm{T}}(X^{\mathrm{T}}X)\{\beta(\theta_1,\psi) - \beta(\hat{\theta}_1,\hat{\psi})\}] + h(y, X)$, where h does not involve any parameters, $\beta(\theta_1,\psi)$ is β expressed as a function of θ_1 and ψ as given in Eq. (2). Then Eqs. (20), (21) and (22) give Eq. (5).

Proof of Theorem 1

Let $\tau_1(\theta_1) = \hat{\beta}_2\theta_1 - \hat{\beta}_1 = (-1,\theta_1)\hat{\gamma}$ and $\tau_2(\theta_1) = Q(\theta_1)\tilde{\beta}_2(\theta_1) = (c_{11}\hat{\beta}_1 + c_{12}\hat{\beta}_2)\theta_1 + c_{21}\hat{\beta}_1 + c_{22}\hat{\beta}_2 = (\theta_1, 1)C\hat{\gamma}$. It follows after some algebraic simplifications that

$$U(\theta_1) = -\frac{n^2|C|\tau_1(\theta_1)\tau_2(\theta_1)}{Q^2(\theta_1)\{\mathrm{SSE} + n|C|\tau_1^2(\theta_1)/Q(\theta_1)\}}. \tag{23}$$

From the distribution of $\hat{\gamma}$, it follows that $(\tau_1(\theta_1),\ \tau_2(\theta_1))^{\mathrm{T}}$ is bivariate normal with means 0 and $Q(\theta_1)\beta_2$, variances $(\sigma^2/n)|C|^{-1}Q(\theta_1)$ and $(\sigma^2/n)Q(\theta_1)$, and covariance $n^{-1}(-1,\theta_1)C^{-1}C(\theta_1,1)^T = 0$. Thus $\tau_1(\theta_1)$ and $\tau_2(\theta_1)$ are independently distributed. Further, SSE $\sim \sigma^2\chi^2_{n-r}$ and is distributed independently of $\hat{\gamma}$, and hence of $(\tau_1(\theta_1), \tau_2(\theta_1))$. Thus,

$$\begin{aligned} m(\theta_1) &= E_{\tilde{\eta}(\theta_1)}\{U(\theta_1)\}\\ &= -\frac{n^2|C|}{Q^2(\theta_1)}E_{\tilde{\eta}(\theta_1)}\left\{\frac{\tau_1(\theta_1)}{\mathrm{SSE} + n|C|\tau_1^2(\theta_1)/Q(\theta_1)}\right\}E_{\tilde{\eta}(\theta_1)}\{\tau_2(\theta_1)\} = 0, \end{aligned} \tag{24}$$

utilizing the symmetry of $\tau_1(\theta_1)$ around 0 and the independence of $\tau_1(\theta_1)$ and SSE. Next

$$\begin{aligned} V_{\tilde{\eta}(\theta_1)}\{U(\theta_1)\} &= \frac{n^3|C|}{Q^3(\theta_1)}E_{\tilde{\eta}(\theta_1)}\left\{\frac{n|C|\tau_1^2(\theta_1)/Q(\theta_1)}{(\mathrm{SSE} + n|C|\tau_1^2(\theta_1)/Q(\theta_1))^2}\right\}\\ &\quad\times E_{\tilde{\eta}(\theta_1)}\{\tau_2^2(\theta_1)\}. \end{aligned} \tag{25}$$

Since $n|C|\tau_1^2(\theta_1)Q^{-1}(\theta_1)/\sigma^2$ and SSE/σ^2 are independent Chi-squares with respective degrees of freedom 1 and $n-r$, after some simplifications we get from Eq. (25),

$$V_{\tilde{\eta}(\theta_1)}\{U(\theta_1)\} = \frac{n^2|C|}{(n-r+1)(n-r-1)Q^2(\theta_1)} \frac{Q(\theta_1)\tilde{\beta}_2^2(\theta_1)+\tilde{\sigma}^2(\theta_1)/n}{\tilde{\sigma}^2(\theta_1)/n}. \tag{26}$$

Finally, we calculate

$$U'(\theta_1) = \frac{2}{n}U^2(\theta_1) - \frac{n^2|C|\{\tau_1'(\theta_1)\tau_2(\theta_1)+\tau_2'(\theta_1)\tau_1(\theta_1)\}}{Q^2(\theta_1)\{\mathrm{SSE}+n|C|\tau_1^2(\theta_1)/Q(\theta_1)\}} + \frac{2n^2|C|Q'(\theta_1)\tau_1(\theta_1)\tau_2(\theta_1)}{Q^3(\theta_1)\{\mathrm{SSE}+n|C|\tau_1^2(\theta_1)/Q(\theta_1)\}}.$$

Proceeding as in Eq. (24),

$$E_{\tilde{\eta}(\theta_1)}\{-U'(\theta_1)\} = -\frac{2}{n}V_{\tilde{\eta}(\theta_1)}\{U(\theta_1)\} + \frac{n^2|C|}{Q^2(\theta_1)}E_{\tilde{\eta}(\theta_1)}\left\{\frac{\tau_1'(\theta_1)\tau_2(\theta_1)+\tau_2'(\theta_1)\tau_1(\theta_1)}{\mathrm{SSE}+n|C|\tau_1^2(\theta_1)/Q(\theta_1)}\right\}.$$

But $\tau_1'(\theta_1) = \hat{\beta}_2$ and $\tau_2'(\theta_1) = c_{11}\hat{\beta}_1 + c_{12}\hat{\beta}_2$. Solving $\tau_1(\theta_1) = \hat{\beta}_2\theta_1 - \hat{\beta}_1$ and $\tau_2(\theta_1) = (c_{11}\theta_1 + c_{12})\hat{\beta}_1 + (c_{12}\theta_1 + c_{22})\hat{\beta}_2$, we get

$$\hat{\beta}_1 = \frac{\theta_1\tau_2(\theta_1) - (c_{12}\theta_1 + c_{22})\tau_1(\theta_1)}{Q(\theta_1)}, \quad \hat{\beta}_2 = \frac{(c_{11}\theta_1 + c_{12})\tau_1(\theta_1) + \tau_2(\theta_1)}{Q(\theta_1)}.$$

Now, after much algebraic simplification,

$$\tau_1'(\theta_1)\tau_2(\theta_1) + \tau_2'(\theta_1)\tau_1(\theta_1) = Q^{-1}(\theta_1)\{\tau_2^2(\theta_1) - |C|\tau_1^2(\theta_1) + 2\tau_1(\theta_1)\tau_2(\theta_1)(c_{11}\theta_1 + c_{12})\}.$$

Hence, as in Eq. (24),

$$E_{\tilde{\eta}(\theta_1)}\{-U'(\theta_1)\} = -\frac{2}{n}V_{\tilde{\eta}(\theta_1)}\{U(\theta_1)\} + \frac{n^2|C|}{Q^3(\theta_1)}E_{\tilde{\eta}(\theta_1)}\left\{\frac{\tau_2^2(\theta_1) - |C|\tau_1^2(\theta_1)}{\mathrm{SSE}+n|C|\tau_1^2(\theta_1)/Q(\theta_1)}\right\}.$$

Next, by Eq. (26) and the remarks after Eq. (25),

$$E_{\tilde{\eta}(\theta_1)}\{-U'(\theta_1)\} = \frac{n^2|C|\tilde{\beta}_2^2(\theta_1)}{(n-r+1)Q(\theta_1)\tilde{\sigma}^2(\theta_1)}. \tag{27}$$

The expression for $w(\theta_1)$ in Theorem 1 now follows from Eqs. (24), (26), (27) and (6).

Proof of Lemma 1

Write $d^{\mathrm{T}} = (\hat{\beta}_2, -\hat{\beta}_1)$. Then

$$\sup_{\theta_1} (\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2/Q(\theta_1) = \sup_{\phi\neq 0} \frac{(\phi^{\mathrm{T}}d)^2}{\phi^{\mathrm{T}}C\phi} = d^{\mathrm{T}}C^{-1}d = \hat{\gamma}^{\mathrm{T}}C\hat{\gamma}/|C|\,,$$

as required.

Proof of Lemma 2

Since $0 \le w(t) \le 1$ we have

$$0 \le -2\,\log\,\lambda_{\mathrm{APL}}(\theta_1) \le 2\int_{-\infty}^{\infty} |U(t)|\mathrm{d}t. \tag{28}$$

Recalling the definitions of τ_1 and τ_2 immediately before Eq. (23), write $\tau_1(t) = (-1, t)C^{-1/2}C^{1/2}\hat{\gamma}$ and $\tau_2(t) = (t, 1)C^{1/2}C^{1/2}\hat{\gamma}$. Then, by the Schwarz inequality,

$$|\tau_1(t)| \le \{(-1, t)C^{-1}(-1, t)^{\mathrm{T}}\}^{1/2}(\hat{\gamma}^{\mathrm{T}}C\hat{\gamma})^{1/2} = \{Q(t)/|C|\}^{1/2}(\hat{\gamma}^{\mathrm{T}}C\hat{\gamma})^{1/2},$$

$$|\tau_2(t)| \le \{(t, 1)C(t, 1)^{\mathrm{T}}\}^{1/2}(\hat{\gamma}^{\mathrm{T}}C\hat{\gamma})^{1/2} = Q^{1/2}(t)(\hat{\gamma}^{\mathrm{T}}C\hat{\gamma})^{1/2}.$$

Hence, from Eq. (23),

$$|U(t)| \le n^2|C|^{1/2}(\hat{\gamma}^{\mathrm{T}}C\hat{\gamma})Q^{-1}(t)/\mathrm{SSE}. \tag{29}$$

Since $Q^{-1}(t)$ is integrable over $(-\infty, \infty)$, Lemma 2 follows from Eqs. (28) and (29).

Derivation of the asymptotic distribution of $R_{\mathrm{I}}^{\mathrm{CR}}$

Write $g(\theta_1) = -\frac{1}{2}\,\log\,Q(\theta_1)$. Then

$$g(\hat{\theta}_{1,\mathrm{I}}^{\mathrm{CR}}) - g(\theta_1) = (\hat{\theta}_{1,\,\mathrm{I}}^{\mathrm{CR}} - \theta_1)g'(\theta_1) + O(n^{-1})\,.$$

Since $\hat{\theta}_{1,\mathrm{I}}^{\mathrm{CR}} - \hat{\theta}_1 = O(n^{-1})$, it follows from Eq. (14) that $(\hat{\theta}_{1,\mathrm{I}}^{\mathrm{CR}} - \theta_1)g'(\theta_1) = a^{\mathrm{CR}}t + O(n^{-1})$, where a^{CR} is given in Eq. (15). Also

$$-2\,\log\,\lambda_{\mathrm{PL}}(\theta_1) = n\,\log\{1 + t^2/(n-r)\} = t^2 + O(n^{-1})\,,$$

which gives $-2\,\log\,\lambda_{\mathrm{PL}}(\hat{\theta}_{1,\mathrm{I}}^{\mathrm{CR}}) = O(n^{-1})$, again using $\hat{\theta}_{1,\mathrm{I}}^{\mathrm{CR}} - \hat{\theta}_1 = O(n^{-1})$. Therefore

$$\begin{aligned}-2\,\log\,\lambda_{\mathrm{I}}^{\mathrm{CR}}(\theta_1) &= 2\{g(\hat{\theta}_{1,\,\mathrm{I}}^{\mathrm{CR}}) - g(\theta_1)\} - 2\,\log\,\lambda_{\mathrm{PL}}(\theta_1) + 2\,\log\,\lambda_{\mathrm{PL}}(\hat{\theta}_{1,\mathrm{I}}^{\mathrm{CR}}) \\ &= 2a^{\mathrm{CR}}t + t^2 + O(n^{-1}) = (t + a^{\mathrm{CR}})^2 + O(n^{-1})\,. \end{aligned} \tag{30}$$

It now follows straightforwardly from Eq. (30) that, to $O(n^{-1})$, $R_{\mathrm{I}}^{\mathrm{CR}} = t + a^{\mathrm{CR}}$ and hence $R_{\mathrm{I}}^{\mathrm{CR}} - a^{\mathrm{CR}}$ is standard normal to $O(n^{-1})$.

References

Barndorff-Nielsen, O.E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika*, *70*, 345–365.

Barndorff-Nielsen, O.E. (1994). Adjusted versions of profile likelihood and directed likelihood, and extended likelihood. *Journal of the Royal Statistical Society Series B*, *56*, 125–140.

Berger, J. O., Bernardo, J. M. (1992). On the development of reference priors (with discussion). In J. M. Bernardo, J. O. Berger, A. P. Dawid, A. F. M. Smith (Eds.), Bayesian Statistics (Vol. 4 pp. 35–60). Oxford: Oxford University Press.

Berger, J.O., Liseo, B., Wolpert, R. (1999). Integrated likelihood methods for eliminating nuisance parameters (with discussion). *Statistical Science*, *14*, 1–22.

Cox, D.R., Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society Series B*, *49*, 1–39.

Creasy, M.A. (1954). Limits for the ratio of the means. *Journal of the Royal Statistical Society Series B*, *16*, 186–194.

Datta, G.S., and DiCiccio, T.J. (2001). On expected volumes of multidimensional confidence sets associated with the usual and adjusted likelihoods. *Journal of the Royal Statistical Society Series B*, *63*, 691–703.

DiCiccio, T.J., and Stern, S.E. (1994). Frequentist and Bayesian Bartlett correction of test statistics based on adjusted profile likelihoods. *Journal of the Royal Statistical Society Series B*, *56*, 397–408.

Fieller, E.C. (1954). Some problems in interval estimation. *Journal of the Royal Statistical Society Series B*, *16*, 175–185.

Finney, D.J. (1978). *Statistical method in biological assay*. Charles Griffin and Company: London.

Ghosh, M., Yin, M., and Kim, Y-H. (2003). Objective Bayesian inference for ratios of regression coefficients in linear models. *Statistica Sinica*, *13*, 409–422.

Gleser, L.J., and Hwang, J.T. (1987). The non-existence of 100(1-α)% confidence sets of finite expected diameters in error-in-variable and related models. *The Annals of Statistics*, *15*, 1351–1362.

Kalbfleisch, J.D., and Sprott,D.A. (1970). Application of likelihood methods to models involving large number of parameters (with discussion). *Journal of the Royal Statistical Society Series B*, *32*, 175–208.

McCullagh, P., Tibshirani, R. (1990). A simple method for the adjustment of profile likelihoods. *Journal of the Royal Statistical Society Series B*, *52*, 325–344.

Mukerjee, R., Reid, N. (1999). On confidence intervals associated with the usual and adjusted likelihoods. *Journal of the Royal Statistical Society Series B*, *61*, 945–954.

Mukerjee, R., Reid, N. (2001). Comparison of test statistics via expected lengths of associated confidence intervals. *Journal of Statistical Planning and Inference*, *97*, 141–151.

Scheffé, H. (1970). Multiple testing versus multiple estimation. Improper confidence sets. Estimation of directions and ratios. *The Annals of Mathematical Statistics*, *41*, 1–29.

Sweeting, T. J. (1987). Discussion of paper by D. R. Cox and N. Reid. *Journal of the Royal Statistical Society Series B 49*, 20–21.

Wallace, D. L. (1958). Intersection region confidence procedures with an application to the location of the maximum in quadratic regression. *The Annals of Mathematical Statistics*, *29*, 455–475.