Statistica Sinica **14**(2004), 547-570

KURTOSIS AND CURVATURE MEASURES FOR NONLINEAR REGRESSION MODELS

L. M. Haines, T. E. O'Brien and G. P. Y. Clarke

University of KwaZulu-Natal, Loyola University Chicago and Agriculture Western Australia

Abstract: An expression for the second-order approximation to the kurtosis associated with the least squares estimate of an individual parameter in a nonlinear regression model is derived, and connections between this and various other measures of curvature are made. Furthermore a means of predicting the reliability of the commonly-used Wald confidence intervals for individual model parameters, based on measures of skewness and kurtosis, is developed. Numerous examples illustrating the theoretical results are provided.

Key words and phrases: Bias, confidence intervals, parameter-effects curvature, relative overlap, skewness.

1. Introduction

There has been considerable interest over the past twenty years in developing measures of curvature for nonlinear regression models which in some way quantify the deviation of the model from linearity. Specifically, in a landmark paper in 1980, Bates and Watts built on the seminal work of Beale (1960) and introduced relative intrinsic and parameter-effects curvatures which provide global measures of the nonlinearity of the model. However these measures are not always helpful when the individual parameters in the model are of interest (see e.g., Cook and Witmer (1985)) and as a consequence a number of researchers have developed measures of curvature which are specifically associated with the individual parameters. In particular Ratkowsky (1983) suggested examining the skewness and kurtosis of the least squares estimates of the parameters by means of simulation and Hougaard (1985) reinforced this idea by deriving a formula for the second-order approximation to skewness. Further Cook and Goldberg (1986) and Hamilton (1986) extended the ideas of Bates and Watts (1980) to accommodate individual parameters, while Clarke (1987) introduced a marginal curvature measure derived from the second-order approximation to the profile likelihood. The list seems rather daunting and the curvature measures diverse and unconnected. In fact Clarke (1987), and more recently Kang and Rawlings (1998), have identified certain relationships between the measures, but these connections are limited.

The usefulness of measures of curvature for individual model parameters can best be gauged by how well the measures predict the coincidence between the Wald and the likelihood-based confidence intervals. Ratkowsky (1983, 1990) provided rules of thumb relating to the closeness-to-normality of the individual parameter estimates, and thus to the accuracy of the Wald intervals, which are based on the skewness and kurtosis of the parameter estimates obtained by simulation and which are supported by a wealth of practical experience. Additionally Cook and Goldberg (1986) suggested a connection between their individual parameter-effects curvature measure and the reliability of the Wald confidence intervals but this claim has to a greater extent been refuted by the findings of van Ewijk and Hoekstra (1994). Clarke (1987) specifically designed his marginal curvature measure to reflect the closeness or otherwise of the Wald and the profile likelihood-based intervals, but the measure is approximate and does not perform well in all cases. In summary, none of the existing curvature measures associated with individual model parameters are entirely satisfactory in terms of assessing the reliability of the Wald confidence intervals. Interestingly Cook and Tsai (1990) approached the problem of assessing the closeness or otherwise of Wald and profile likelihood intervals more directly, by invoking a method introduced by Hodges (1987) which is based on determining the confidence levels associated with the smallest Wald interval containing, and the largest Wald interval contained in, the likelihood-based interval.

The aim of the present study is two-fold, first to develop a formula for the second-order approximation to kurtosis for the least squares estimates of the individual model parameters and to relate this to other measures of curvature, and second to develop a measure based on skewness and kurtosis which predicts well the closeness or otherwise of the Wald and the likelihood-based confidence intervals. The paper is organized as follows. Relevant notation and background are given in the next section and the new results on kurtosis are derived in Section 3. Various measures of curvature are related to the second-order approximations to skewness and kurtosis in Section 4, and a composite measure for assessing the accuracy of the Wald confidence intervals is developed in Section 5. Several illustrative examples are provided in Section 6 and the Fieller-Creasy problem is treated in detail in Section 7. Some brief conclusions are drawn in Section 8.

2. Preliminaries

Consider the nonlinear regression model

$$y_i = \eta(x_i, \theta) + \epsilon_i, \quad i = 1, \ldots, n,$$

where y_i is an observation taken at a value x_i of the explanatory variable x, θ is a $p \times 1$ vector of unknown parameters, $\eta(x_i, \theta)$ is a function nonlinear in and three times differentiable with respect to θ, and the ϵ_i are error terms independently distributed as $N(0, \sigma^2)$. Let D be the $(n \times p)$ matrix with (i, a)th entry $\partial\eta_i/\partial\theta_a$ where $\eta_i = \eta(x_i, \theta)$, let $G = D^T D$, and introduce a matrix K, not necessarily unique, such that $K^T D^T DK = I$. Furthermore, define B to be the 3-dimensional array of order $(p \times p \times p)$ with (l, rs)th element given by

$$B_{l,rs} = \sum_{a,b,c} \sum_i K_{la} K_{rb} K_{sc} \frac{\partial\eta_i}{\partial\theta_a} \frac{\partial^2\eta_i}{\partial\theta_b \partial\theta_c},$$

and let B_l denote the lth face of that array for $l = 1, \ldots, p$. Also define C to be the 4-dimensional array of order $(p \times p \times p \times p)$ with (l, rst)th element given by

$$C_{l,rst} = \sum_{a,b,c,d} \sum_i K_{la} K_{rb} K_{sc} K_{td} \frac{\partial\eta_i}{\partial\theta_a} \frac{\partial^3\eta_i}{\partial\theta_b \partial\theta_c \partial\theta_d},$$

and let C_l denote the lth 3-dimensional symmetric array within C for $l = 1, \ldots, p$.

Next, consider an individual parameter θ_j with maximum likelihood estimate $\hat{\theta}_j$ and define $t_j = \hat{\theta}_j - \theta_j$ for $j = 1, \ldots, p$. Let $z = K^T D^T e$ and $w = He$, with H an orthonormal matrix such that $K^T D^T H = 0$. Clearly $z \sim N(0, \sigma^2 I)$ independently of w. Then, following Clarke (1980), t_j can be expanded in a power series in terms of z and w. Specifically, suppose that terms in w are neglected, since these are associated with the intrinsic nonlinearity of the model, and suppose also that terms in z of degree four and higher are ignored. Then the series can be written as

$$t_j = s_{10} + s_{20} + s_{30}^b + s_{30}^c + \cdots, \tag{2.1}$$

where the subscripts u and v in the terms s_{uv} denote the degree in z and in w, respectively, and the superscripts b and c for s_{30} identify terms involving the matrices B and C separately. Furthermore, let k_j^T denote the jth row of K, b the vector with lth element $z^T B_l z = \sum_{r,s} B_{l,rs} z_r z_s$, and c the vector with lth element $\sum_{r,s,t} C_{l,rst} z_r z_s z_t$ for $l = 1, \ldots, p$. Then it immediately follows from Clarke (1980) that $s_{10} = k_j^T z$ and that $s_{20} = -(1/2) k_j^T b = -(1/2) z^T M z$, where $M = \sum_l K_{jl} B_l$. Also

$$s_{30}^b = \frac{1}{2} b^T M z = \frac{1}{2} \sum_l (z^T B_l z)(m_l^T z), \tag{2.2}$$

where m_l^T denotes the lth row of M, and

$$\begin{aligned} s_{30}^c = -\frac{1}{6} k_j^T c &= -\frac{1}{6} \sum_{r,s,t} \sum_l K_{jl} C_{l,rst} z_r z_s z_t \\ &= -\frac{1}{6} \sum_r (z^T N_r z) z_r = -\frac{1}{6} \sum_r (z^T N_r z)(e_r^T z), \end{aligned} \tag{2.3}$$

where N_r is the rth face of the symmetric 3-dimensional array $N = \sum_l K_{jl}C_l$, and e_r is a vector with rth element equal to 1 and all other elements 0.

The second-order approximation to the bias in $\hat{\theta}_j$ is thus given by

$$E(s_{20}) = -\frac{1}{2}E(z^T Mz) = -\frac{1}{2}\sigma^2 tr(M) = -\frac{1}{2}\sigma^2 \sum_r \sum_l K_{jl}B_{l,rr}$$

in accord with the findings of Box (1971), Bates and Watts (1980) and Clarke (1980). Further, following Clarke (1980), the second-order approximation to the variance of $\hat{\theta}_j$ is given by

$$E[s_{10}^2] + E[(s_{20} - E[s_{20}])^2] + 2E[s_{10}s_{30}^b] + 2E[s_{10}s_{30}^c],$$

where $E[s_{10}^2] = \sigma^2 g^{jj}$ with g^{jj} the (jj)th element of G^{-1}, and where the three trailing terms can be summarized as $\sigma^4 v_{j,add}$ with $v_{j,add}$ equal to

$$\sum_{r,s}\sum_{l,m}\{\frac{1}{2}K_{jl}K_{jm}B_{l,rs}B_{m,rs} + K_{jl}K_{jr}B_{l,mr}B_{m,ss} + 2K_{jl}K_{js}B_{l,mr}B_{m,rs}\}$$
$$-\sum_{r,s}\sum_{l} K_{jl}K_{jr}C_{l,rss}.$$

The required approximation can thus be written succinctly as $\sigma^2 g^{jj} + \sigma^4 v_{j,add}$. The third central moment of $\hat{\theta}_j$ can be expressed as $E[(t_j - E[t_j])^3]$ and the term contributing to the second order approximation to this moment is given by

$$\begin{aligned} 3E[s_{10}^2(s_{20} - E[s_{20}])] &= 3 \text{ Cov}(s_{10}^2, s_{20}) \\ &= -3\sigma^4 k_j^T M k_j \\ &= -3\sigma^4 \sum_{r,s}\sum_{l} K_{jr}K_{js}K_{jl}B_{l,rs}. \end{aligned}$$

It then follows that the second-order approximation to the coefficient of skewness for $\hat{\theta}_j$ can be expressed as $\gamma_1 = -3\sigma\Gamma$ where, in the notation of Clarke (1987), $\Gamma = (g^{jj})^{-3/2}k_j^T M k_j$. This result is in agreement with the findings of Hougaard (1982, 1985). Many of the above results are also presented clearly and concisely by Seber and Wild (1989).

3. New Results on Kurtosis

The second-order approximation to the fourth central moment of $\hat{\theta}_j$, and thus equivalently to $E[(t_j - E[t_j])^4]$, follows immediately from (2.1) and is given by

$$E[s_{10}^4] + 6E[s_{10}^2(s_{20} - E[s_{20}])^2] + 4E[s_{10}^3 s_{30}^b] + 4E[s_{10}^3 s_{30}^c]. \qquad (3.1)$$

Now $E[s_{10}^4]=E[(k_j^T z)^4]=3\sigma^4(g^{jj})^2$. Also, since $s_{10}=k_j^T z$ and $s_{20}=(1/2)z^T Mz$, it follows from (A.1) in the Appendix that

$$E[s_{10}^2(s_{20}-E[s_{20}])^2]=E[s_{10}^2]E[(s_{20}-E[s_{20}])^2]+2\sigma^6 k_j^T M^2 k_j,$$

where the term $k_j^T M^2 k_j$ is given explicitly by $\sum_{r,s,t}\sum_{l,m} K_{jr}K_{jt}K_{jl}K_{jm}B_{l,rs}$ $B_{m,st}$ and, in the notation of Clarke (1987), by $(g^{jj})^3\Gamma^l\Gamma^l$. Furthermore it follows from the expressions for s_{30}^b and s_{30}^c given in (2.2) and (2.3), respectively, and by invoking result (A.2) in the Appendix, that

$$E[s_{10}^3 s_{30}^b]=3E[s_{10}^2]E[s_{10}s_{30}^b]+3\sigma^6\sum_l(k_j^T m_l)(k_j^T B_l k_j),$$

$$E[s_{10}^3 s_{30}^c]=3E[s_{10}^2]E[s_{10}s_{30}^c]-\sigma^6\sum_r(k_j^T e_r)(k_j^T N_r k_j),$$

where the summations $\sum_l(k_j^T m_l)(k_j^T B_l k_j)$ and $\sum_r(k_j^T e_r)(k_j^T N_r k_j)$ are given explicitly in terms of elements of the matrices K, B and C by $\sum_{r,s,t}\sum_{l,m}K_{jr}K_{js}K_{jt}$ $K_{jm}B_{m,lr}B_{l,st}$ and $\sum_{r,s,t}\sum_l K_{jr}K_{js}K_{jt}K_{jl}C_{l,rst}$ and, in the notation of Clarke (1987), by $(g^{jj})^3\Gamma^l\Gamma_l$ and $(g^{jj})^3\kappa$ respectively. It thus follows, by gathering the above expressions for the terms in (3.1) together and by recalling that $\sigma^4 v_{j,add}$ is given by $E[(s_{20}-E[s_{20}])^2]+2E[s_{10}s_{30}^b]+2E[s_{10}s_{30}^c]$, that the required approximation to the fourth central moment of $\hat{\theta}_j$ can be expressed succinctly as

$$3\sigma^4(g^{jj})^2+6\sigma^6 g^{jj}v_{j,add}+12\sigma^6(g^{jj})^2\beta_a, \tag{3.2}$$

where $\beta_a=g^{jj}(\Gamma^l\Gamma^l+\Gamma^l\Gamma_l-(1/3)\kappa)$. Furthermore, a second-order approximation to the coefficient of kurtosis can be obtained from (3.2) and from the square of the variance of $\hat{\theta}_j$, $(\sigma^2 g^{jj}+\sigma^4 v_{j,add}^2+\cdots)^2$, expanded as $\sigma^4(g^{jj})^2+2\sigma^6 g^{jj}v_{j,add}$. Specifically the approximation is given by

$$3+\frac{12\sigma^6(g^{jj})^2\beta_a}{\sigma^4(g^{jj})^2+2\sigma^6 g^{jj}v_{j,add}}$$

and further, since it is reasonable to assume that $\left|2\sigma^2 v_{j,add}/g^{jj}\right|<1$, by $3+12\sigma^2\beta_a$. Thus the second-order approximation to the excess kurtosis is simply $\gamma_2=12\sigma^2\beta_a$.

4. Relating Curvature Measures

Cook and Goldberg (1986) developed subset intrinsic and subset parameter-effects curvature measures for the parameters of a nonlinear regression model. These hold for an individual parameter taken without loss of generality to be θ_p, the last element of the vector θ, and are given by $\Gamma_s^\tau(\theta_p)=\sigma|B_{p,pp}|$ and $\Gamma_s^\eta(\theta_p)=2\sigma\{\sum_{r=1}^{p-1}B_{p,pr}^2\}^{\frac{1}{2}}$, respectively. The measures can be related to the terms involved in the second-order approximations to skewness and kurtosis by observing that for K an upper triangular matrix, $k_p^T = [0 \ \cdots \ 0 \ K_{pp}]$ and $g^{pp} = K_{pp}^2$. It then follows immediately that $\Gamma_s^\tau(\theta_p) = \sigma|\Gamma|$ in accord with the result derived by Kang and Rawlings (1998) and implied by Clarke (1987). In addition it follows that $\Gamma_s^\eta(\theta_p) = 2\sigma\{g^{pp}\Gamma^l\Gamma^l - \Gamma^2\}^{\frac{1}{2}}$ and thus that the relation between intrinsic subset curvature and kurtosis would seem to be rather convoluted. It is however tempting to surmise that the latter expression is incomplete and should, more correctly, be given by $\Gamma_s^\eta(\theta_p) = 2\sigma\{\beta_a - \Gamma^2\}^{\frac{1}{2}}$. Cook and Goldberg (1986) also introduced a measure combining the subset intrinsic and the parameter-effects subset curvatures, termed the total subset curvature, and given by $\Gamma_s = [\Gamma_s^\tau(\theta_p)^2 + \Gamma_s^\eta(\theta_p)^2]^{1/2}$.

Clarke (1987) developed approximate confidence limits for an individual parameter θ_j which are based on an expansion of the profile likelihood and which can be expressed in the form

$$\theta_j - \hat{\theta}_j = (g^{jj}\sigma^2)^{\frac{1}{2}}c\{1 - \frac{1}{2}\Gamma\sigma c + \frac{1}{2}\beta\sigma^2 c^2 + \cdots\}, \tag{4.1}$$

where $\beta = \beta_a - \Gamma^2$ and c is an appropriately chosen critical value. In addition, $(1/2)\sigma|\Gamma|$ was introduced as a measure of marginal curvature (see also Kang and Rawlings (1998)). The relation of the terms in this expression to those in the second-order approximations to the coefficients of skewness and kurtosis for $\hat{\theta}_j$ is immediate. Furthermore it is interesting to note that the "skewness" of the confidence interval for θ_j defined by (4.1) is captured by $\Gamma\sigma$ and the "flatness" by $\beta_a\sigma^2$, and thus that these terms are in turn directly related to the skewness and kurtosis for $\hat{\theta}_j$, respectively. In fact (4.1) can be rewritten as

$$\theta_j - \hat{\theta}_j = (g^{jj}\sigma^2)^{\frac{1}{2}}c\{1 + \frac{1}{6}\gamma_1 c + \frac{1}{72}(3\gamma_2 - 4\gamma_1^2)c^2 + \cdots\}. \tag{4.2}$$

The approximation to the coefficient of skewness derived by Hougaard (1985), the subset parameter-effects curvature measure of Cook and Goldberg (1986) and the marginal curvature of Clarke (1987) are the same up to multiplying constants. However, it should also be emphasized that the curvature measures discussed here are all derived assuming that the intrinsic nonlinearity of the model is negligible. This is not an unreasonable assumption in general and can in any case be appraised by examining the value of the maximum or the root-mean-square intrinsic curvature for the particular model of interest (Bates and Watts (1980)).

Ratkowsky (1983), following earlier work of Gillis and Ratkowsky (1978), suggested assessing the nonlinearity associated with an individual parameter θ_j by appraising the normality or otherwise of the distribution of the maximum likelihood estimate $\hat{\theta}_j$. Specifically, the approach involves estimating the coefficients of skewness and kurtosis for $\hat{\theta}_j$ by simulation and invoking tests of significance commonly associated those coefficients. The method is not entirely satisfactory however. In particular, simulations are time-consuming and the nonlinear optimization routine used in the estimation of the parameters may not converge for all simulated data sets. Also the standard errors associated with the estimates of the coefficients, and specifically that for the excess kurtosis, are generally large. In addition, the tests of significance associated with the coefficients are not necessarily diagnostic (Horswell and Looney (1993), Rayner, Best and Mathews (1995)). Some of these concerns are illustrated by means of the following example.

Example 4.1. *The Michaelis-Menten reaction.* Bliss and James (1966) reported a data set from enzyme kinetics which is well modelled by the Michaelis-Menten equation. Specifically $\eta(x,\theta) = \theta_1 x/(\theta_2 + x)$, where $\eta(x,\theta)$ represents the velocity of the reaction at the substrate concentration x and θ_1 and θ_2 are unknown parameters. There were six observations and the maximum likelihood estimates of the parameters are given by $\hat{\theta} = \{0.6904, 0.5965\}$ and $\hat{\sigma}^2 = 0.000184$. The maximum intrinsic curvature for the model, Γ^N, was found to be 0.050 which, following the rule of thumb given by Bates and Watts (1980), is deemed to be negligible. 1,000 data sets were generated from this model, with the x-values set to those reported in Bliss and James (1966) and with the true parameter values for θ and σ^2 taken to be the maximum likelihood estimates of the original data, and estimates of the mean, the variance and the coefficients of skewness and excess kurtosis for $\hat{\theta}$ obtained. The process was repeated 10,000 times and the results are summarized in Table 1(a). It is clear from the standard errors recorded in that table that the simulated estimate of excess kurtosis is highly variable, an observation which underscores the fragility of the tests suggested by Ratkowsky (1983). Indeed conclusions drawn from a single simulation, or indeed from hundreds of simulations, are expected to be unconvincing. The second-order approximations to skewness and kurtosis given by $-3\Gamma\hat{\sigma}$ and $12\beta_a\hat{\sigma}^2$, respectively, are included in Table 1(b) and are remarkably close to the simulated values.

Ratkowsky (1990) suggested that the second-order approximation to the coefficient of skewness derived by Hougaard (1985) should replace the estimate obtained by simulation since the former is computed more readily and is in general more reliable. Indeed Hougaard's approximation is incorporated into version 8.0 of the software package SAS®. It now follows immediately from the findings of the present study that the second-order approximation to the excess kurtosis derived in Section 3 is also to be preferred to the simulated estimate. It should be emphasized however that the approximations for skewness and kurtosis ignore high-order terms in σ^2 and terms involving intrinsic nonlinearity, and are therefore not necessarily always close to the true values.

Table 1. Mean, variance, coefficient of skewness and excess kurtosis for the maximum likelihood estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ for Example 4.1.

parameter	mean	variance	skewness	kurtosis
$\hat{\theta}_1$	0.6924 ($\pm$0.00117)	0.001387 ($\pm$0.000065)	0.2967 ($\pm$0.08633)	0.1799 ($\pm$0.2363)
$\hat{\theta}_2$	0.6008 ($\pm$0.00218)	0.004803 ($\pm$0.00023)	0.3888 ($\pm$0.09073)	0.2901 ($\pm$0.28649)

(a) Results from simulating 1,000 data sets 10,000 times, with standard errors given in parenthesis.

parameter	mean	variance	skewness	kurtosis
$\hat{\theta}_1$	0.6924	0.001356	0.2898	0.1817
$\hat{\theta}_2$	0.6008	0.004658	0.3792	0.2837

(b) Second-order approximations.

5. Relating Skewness and Kurtosis to Nonlinearity

An attractive approach to assessing the nonlinearity associated with a parameter of interest is that based on a comparison of the closeness or otherwise of the Wald and the profile likelihood intervals for that parameter. This idea is presented in the paper of Cook and Tsai (1990), following the earlier and more general results of Jennings (1986) and Hodges (1987), and complements methods based on measures of curvature. It is interesting therefore to relate the measures of skewness and kurtosis developed in Sections 2 and 3 to the discrepancy between the Wald and the profile likelihood intervals of an individual parameter and hence, albeit tentatively, to the nonlinearity associated with that parameter. This can be achieved by using the second-order approximation to the profile likelihood developed by Clarke (1987), and given in (4.1), and by invoking specific measures of closeness of Wald and profile likelihood-based intervals. The idea is similar in principle to that introduced in a more general setting by Cook and Tsai (1990).

The aim of linking skewness and kurtosis for an individual parameter with the agreement or otherwise of the Wald and profile likelihood-based intervals is two-fold. First the values of skewness and kurtosis can be used jointly to provide an indication of how well a particular Wald interval approximates that based on the profile likelihood. Cook and Weisberg (1990) emphasize however that calculating intervals based on profile likelihoods, particularly in the context of confidence curves, is not a difficult task. Indeed, as pointed out by one of the referees, if the calculation of the profile likelihood for a given parameter proves difficult, this could well be an indication of an inappropriate model or of high intrinsic nonlinearity associated with that parameter. Nevertheless it is possible that a practitioner is interested in setting a single confidence interval to a given parameter and has available the appropriate Wald confidence interval and values of the measures of skewness and kurtosis. In such cases an indication of how close or otherwise the Wald interval is to that based on the profile likelihood could well be useful.

Second, it is appealing to use the relationship between skewness and kurtosis and the discrepancy between Wald and profile likelihood-based intervals for an individual parameter to develop a rule of thumb for assessing how large or small those measures are and, more generally but possibly somewhat tentatively, to formulate a measure of nonlinearity associated with the parameter of interest. This idea should be treated with some caution however. In particular Cook and Tsai (1990) and Cook and Weisberg (1990) emphasize that, in examining the closeness or otherwise of Wald and profile likelihood intervals for an individual parameter, it is highly desirable to examine such intervals over a wide range of confidence levels and not at just one level. Clearly cognizance must be taken of this.

Two methods for quantifying the closeness or otherwise of the Wald interval to the interval based on the profile likelihood, or rather on Clarke's approximation to that likelihood, are explored here in terms of skewness and kurtosis. The one approach is based on the notion of approximate relative overlap introduced by van Ewijk and Hoekstra (1994), and the other on the contours method of Hodges (1987).

5.1. Approximate relative overlap

In order to appraise the reasonableness or otherwise of the linear approximation to a nonlinear model in terms of an individual parameter, van Ewijk and Hoekstra (1994) introduced the notion of relative overlap which is defined as the ratio of the intersection of the Wald and profile likelihood-based confidence intervals to the union of those intervals. It is thus possible to derive an approximation to the exact relative overlap using the profile likelihood-based confidence interval developed by Clarke (1987), and to formulate this in terms of the measures of skewness and kurtosis γ_1 and γ_2, respectively. Specifically, the Wald interval for the parameter θ_j is given by $\hat{\theta}_j \pm c\sigma(g^{jj})^{1/2}$. Further, following Clarke (1987) and invoking (4.1), the lower and upper limits of the profile likelihood-based interval are approximated by $\hat{\theta}_j - c\sigma(g^{jj})^{1/2}\{1+\Gamma_t+B_t\}$ and $\hat{\theta}_j + c\sigma(g^{jj})^{1/2}\{1-\Gamma_t+B_t\}$, respectively, where $\Gamma_t = (1/2)\Gamma\sigma c$, $B_t = (1/2)\beta\sigma^2c^2$ and $c > 0$ is an appropriate critical value. Equivalently these limits follow immediately from (4.2) and hence are given by

$$\hat{\theta}_j - c\sigma(g^{jj})^{\frac{1}{2}}\{1 - \frac{\gamma_{1c}}{6} + \frac{1}{72}(3\gamma_{2c} - 4\gamma_{1c}^2)\}$$

and
$$\hat{\theta}_j + c\sigma(g^{jj})^{\frac{1}{2}}\{1 + \frac{\gamma_{1c}}{6} + \frac{1}{72}(3\gamma_{2c} - 4\gamma_{1c}^2)\},$$

respectively, where $\gamma_{1c} = \gamma_1 c$ and $\gamma_{2c} = \gamma_2 c^2$. Expressions for the approximate relative overlap calculated from these intervals depend on the ordering of the associated limits and are summarized, together with the attendant conditions for the expressions to hold, in Table 2.

Table 2. Expressions for the approximate relative overlap (ARO) based on the Wald (W) and the approximate profile likelihood-based (P) confidence limits in terms of γ_{1c} and γ_{2c}.

order of limits	ARO	conditions	relation of γ_{2c} and γ_{1c} to p
PWWP	$\frac{72}{72+3\gamma_{2c}-4\gamma_{1c}^2}$	$\gamma_{2c} > \frac{4}{3}\gamma_{1c}(\gamma_{1c}-3)$ and $\gamma_{2c} > \frac{4}{3}\gamma_{1c}(\gamma_{1c}+3)$	$\gamma_{2c} = \frac{4}{3}\gamma_{1c}^2 + 24\frac{(1-p)}{p}$
PWPW	$\frac{144+3\gamma_{2c}-4\gamma_{1c}^2+12\gamma_{1c}}{144+3\gamma_{2c}-4\gamma_{1c}^2-12\gamma_{1c}}$	$\gamma_{1c} < 0, \gamma_{2c} < \frac{4}{3}\gamma_{1c}(\gamma_{1c}-3)$ and $\gamma_{2c} > \frac{4}{3}\gamma_{1c}(\gamma_{1c}+3)$	$\gamma_{2c} = \frac{4}{3}\gamma_{1c}^2 - \frac{4(1+p)}{1-p}\gamma_{1c} - 48$
WPWP	$\frac{144+3\gamma_{2c}-4\gamma_{1c}^2-12\gamma_{1c}}{144+3\gamma_{2c}-4\gamma_{1c}^2+12\gamma_{1c}}$	$\gamma_{1c} > 0, \gamma_{2c} > \frac{4}{3}\gamma_{1c}(\gamma_{1c}-3)$ and $\gamma_{2c} < \frac{4}{3}\gamma_{1c}(\gamma_{1c}+3)$	$\gamma_{2c} = \frac{4}{3}\gamma_{1c}^2 + \frac{4(1+p)}{1-p}\gamma_{1c} - 48$
WPPW	$\frac{72+3\gamma_{2c}-4\gamma_{1c}^2}{72}$	$\gamma_{2c} < \frac{4}{3}\gamma_{1c}(\gamma_{1c}-3)$ and $\gamma_{2c} < \frac{4}{3}\gamma_{1c}(\gamma_{1c}+3)$	$\gamma_{2c} = \frac{4}{3}\gamma_{1c}^2 - 24(1-p)$

It is also interesting to consider the region in the space of $(\gamma_{1c}, \gamma_{2c})$-pairs within which the relative overlap is greater than or equal to some fixed value p with $0 < p < 1$. This area, denoted B_p, is bounded by the parabolas specified in the last column of Table 2 and is illustrated for $p = 0.90$ in Figure 1. However values of γ_{1c} and γ_2 which fall within the area B_p are awkward to specify explicitly. Consider therefore the largest rectangle completely enclosed by the boundaries of the region B_p and specified by $-6(1-p) \leq \gamma_{1c} \leq 6(1-p)$ and $-24(1-p)(2p-1) \leq \gamma_{2c} \leq 24(1-p)/p$. Then the ranges of γ_{1c} and γ_{2c} so defined provide a conservative approximation to the values of γ_{1c} and γ_{2c} for which the approximate relative overlap is greater than or equal to p. These ranges are listed in Table 3 for selected values of p and the associated rectangle for $p = 0.90$ is graphed in Figure 1. Finally, note that the probabilities associated with the largest rectangle enclosed by B_p with boundaries γ_{1c} and γ_{2c} are given by $p_1 = 1 - (1/6)\,|\gamma_{1c}|$ and

$$p_2 = \begin{cases} \dfrac{24}{24+\gamma_{2c}} & \text{for } \gamma_{2c} \geq 0, \\ \dfrac{3}{4} + \dfrac{1}{4}\sqrt{1 + \dfrac{1}{3}\gamma_{2c}} & \text{for } -3 \leq \gamma_{2c} < 0, \end{cases}$$

respectively, and hence that the probability $P_{\min} = \min\{p_1, p_2\}$ can be introduced in place of approximate relative overlap.

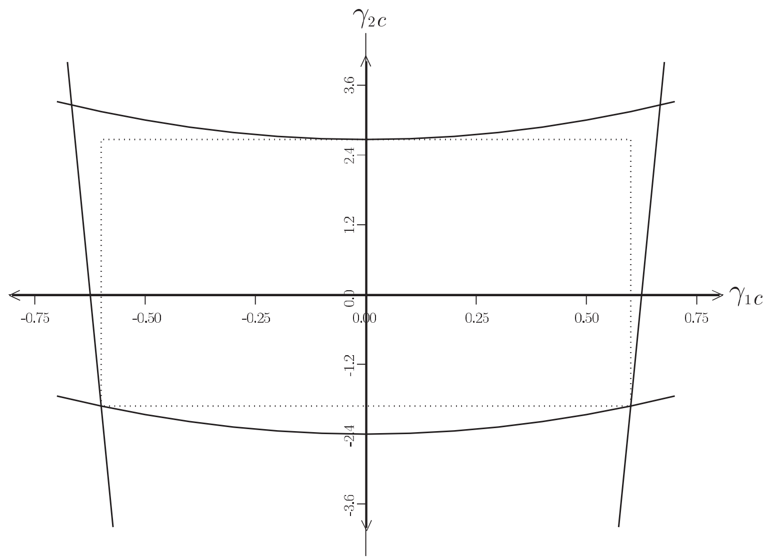

Figure 1. The region in the space of $(\gamma_{1c}, \gamma_{2c})$-pairs within which the relative overlap of the Wald and likelihood-based confidence intervals is greater than or equal to 90% (solid line), together with the largest rectangle enclosed by the boundaries of that region (dashed line).

Table 3. Ranges of γ_{1c} and γ_{2c} which define the largest rectangle enclosed by B_p.

p	range for γ_{1c}	range for γ_{2c}
90%	$-0.60 \leq \gamma_{1c} \leq 0.60$	$-1.92 \leq \gamma_{2c} \leq 2.67$
95%	$-0.30 \leq \gamma_{1c} \leq 0.30$	$-1.08 \leq \gamma_{2c} \leq 1.26$
98.3%	$-0.10 \leq \gamma_{1c} \leq 0.10$	$-0.39 \leq \gamma_{2c} \leq 0.41$
99%	$-0.06 \leq \gamma_{1c} \leq 0.06$	$-0.24 \leq \gamma_{2c} \leq 0.24$

The results summarized in Tables 2 and 3 can be used immediately to assess the closeness or otherwise of Wald and profile likelihood-based intervals for a given parameter. Specifically, consider setting a single confidence interval to that parameter and suppose that the approximate measures of skewness and kurtosis are available. Then Clarke's approximation to the limits for the profile likelihood can be calculated and the approximate relative overlap found using the appropriate formula from Table 2. Alternatively, and more simply, the discrepancy in the intervals can be assessed directly from γ_{1c} and γ_{2c} by using Table 3 to identify the approximate relative overlap associated with those values. Note that relative overlap between the Wald and the profile likelihood intervals has a ready interpretation and thus the practitioner can decide whether or not the approximate value so calculated is reasonable.

It should be noted here that Clarke (1987) recommended using marginal curvature, which is equivalent to scaled skewness, to indicate whether the Wald, the adjusted or the exact profile likelihood-based confidence intervals for an individual parameter should be calculated but the measure does not perform well in all cases. In the present study this idea is, in essence, extended by introducing a joint measure incorporating both scaled skewness and kurtosis. At the same time, in view of the findings of Cook and Tsai (1990) and Cook and Weisberg (1990), the calculation of adjusted Wald confidence intervals as advocated by Clarke (1987) is not considered.

It is tempting to suggest that the approximate relative overlap of the Wald and profile likelihood-based intervals for a given parameter, and thus the measures of skewness and kurtosis jointly, can be used to assess the nonlinearity associated with that parameter. However, as noted already, Cook and Weisberg (1990) caution against basing such measures on intervals at a single level of confidence. Hence, as a rough rule of thumb, it is tentatively recommended that the nonlinearity associated with a parameter be deemed negligible if the approximate relative overlap for the 99% confidence interval exceeds 95%, that it be considered moderate if the overlap for the 95% confidence interval exceeds 95% but that for the 99% interval does not, and that otherwise the nonlinearity be taken to be severe.

These results can also be translated into a rough rule of thumb for assessing the measures of skewness and kurtosis. Specifically suppose that a relative overlap of the Wald and profile likelihood intervals of 95% or higher is considered satisfactory. Then, if the critical value specified by c is equal to 3 corresponding to a deviation from the parameter estimate of three standard errors, it follows from Table 3 that the measures of skewness and kurtosis fall in the intervals

$$-0.1 \leq \gamma_1 \leq 0.1 \quad \text{and} \quad -0.12 \leq \gamma_2 \leq 0.14, \tag{5.1}$$

and if c is equal to 2 these measures fall in the ranges

$$-0.15 \leq \gamma_1 \leq 0.15 \quad \text{and} \quad -0.27 \leq \gamma_2 \leq 0.315, \tag{5.2}$$

respectively. Thus a rough rule of thumb would be to take the measures of skewness and kurtosis, either individually or jointly, to be negligible if they fall within the limits specified in (5.1), to be moderate if they fall outside those limits but within the limits given in (5.2), and to be severe otherwise. It is interesting to note that the rule of thumb for γ_1 alone derived here is in accord with, but slightly more conservative than, that advocated by Ratkowsky (1983).

5.2 The contours method

An approach to appraising the closeness or otherwise of likelihood-based and approximate confidence regions for the parameters of a model, termed the contours method, was introduced by Hodges (1987) and extended to include subsets of parameters by Cook and Tsai (1990). In the case of a single parameter, the method reduces to finding the confidence levels associated with the smallest Wald interval containing the interval based on the profile likelihood, say $W_{\max}$, and the largest Wald interval contained in the likelihood-based interval, say $W_{\min}$, and comparing these levels with the nominal confidence level. In the present context, suppose that interest centers on setting a specified confidence interval to a parameter of interest and suppose that the associated critical value is c. Then, if the Clarke approximation to the profile likelihood is invoked, it is straightforward to show that the critical value associated with $W_{\max}$ is given, at least approximately, by

$$c^\star_{\max} = c\{1 + \frac{|\gamma_{1c}|}{6} + \frac{1}{72}(3\gamma_{2c} - 4\gamma_{1c}^2)\},$$

and that with $W_{\min}$ by

$$c^\star_{\min} = c\{1 - \frac{|\gamma_{1c}|}{6} + \frac{1}{72}(3\gamma_{2c} - 4\gamma_{1c}^2)\}.$$

Confidence levels associated with $c^\star_{\max}$ and $c^\star_{\min}$, denoted $1-\alpha_{\max}$ and $1-\alpha_{\min}$ respectively, can immediately be calculated. As indicated by Hodges (1987) and by Cook and Tsai (1990), these levels have a natural interpretation and the practitioner can therefore decide, depending on the particular model setting, whether or not they are satisfactorily close to the nominal level.

6. Examples

The results and recommendations of the previous sections are now illustrated by means of selected examples of model-data settings taken from the literature. The Fieller-Creasy problem, which is amenable to a more extensive algebraic treatment, is considered in the following section.

6.1 The Michaelis-Menten model

Consider fitting the Michaelis-Menten model to the enzyme kinetic data of Bliss and James (1966), as described in Example 4.1, and specifically consider setting 95% confidence intervals to the individual parameters θ_1 and θ_2. The appropriate critical value c is $t_{4;0.025} = 2.7765$, the 2.5% critical t value with 4 degrees of freedom, and the least squares estimates of the parameters, the associated standard errors, the scaled skewness and kurtosis, the approximate and exact relative overlaps of the Wald and likelihood-based confidence intervals and the confidence levels associated with the Wald intervals $W_{\min}$ and $W_{\max}$ are summarized in Table 4. The approximate relative overlaps of the confidence intervals, or equivalently the values of γ_{1c} and γ_{2c}, indicate very clearly that the Wald limits are not satisfactory and that the profile likelihood-based intervals should be calculated for both parameters. This is affirmed by the values of the Wald and likelihood-based confidence limits of $(0.5882, 0.7926)$ and $(0.5989, 0.8105)$, respectively, for θ_1 and of $(0.4070, 0.7861)$ and $(0.4336, 0.8295)$, respectively, for θ_2 (Clarke (1987)).

Table 4. Curvature measures and relative overlap of the 95% confidence intervals for the individual parameters of the Michaelis-Menten model.

θ_p	$\hat{\theta_p}$	$SE(\hat{\theta_p})$	γ_{1c}	γ_{2c}	ARO	ERO	$1-\alpha_{\min}$	$1-\alpha_{\max}$
θ_1	0.6904	0.0368	0.8038	1.4005	0.8757	0.8701	0.9308	0.9674
θ_2	0.5965	0.0683	1.0528	2.1866	0.8408	0.8341	0.9233	0.9713

It is reassuring to observe that the approximate and the exact relative overlaps of the Wald and likelihood-based confidence intervals are in close agreement, indicating that Clarke's approximation to the profile likelihood works well for this example. Note also that the confidence levels corresponding to the Wald intervals $W_{\min}$ and $W_{\max}$ are not particularly close to the nominal level of 0.95, thereby supporting the recommendation that intervals based on the profile likelihood should be calculated for both parameters.

6.2. The Mitscherlich model

Consider the biomedical oxygen demand data set 1 from Draper and Smith (1981, p. 522) and specifically, following van Ewijk and Hoekstra (1994), consider fitting the three-parameter Mitscherlich model

$$\eta(x, \theta) = \theta_1 + \theta_2 e^{\theta_3 x}, \tag{6.1}$$

where $\eta(x, \theta)$ represents biomedical oxygen demand and x represents time, to that data. Suppose that interest centers on setting 95% confidence intervals to the three parameters of the model. The maximum intrinsic curvature for the model, Γ^N, was computed to be 0.121 and is negligible. The critical value c is taken to be $t_{4;0.025} = 2.7765$ and parameter estimates, the associated standard errors, the terms Γ_t and B_t, and the approximate and exact relative overlaps of the Wald and likelihood-based intervals for the individual parameters are presented in Table 5(a). The results for the parameter θ_2 are disturbing. Specifically, there is a sharp disparity between the approximate and exact relative overlaps of the Wald and the profile likelihood-based confidence intervals. One possible explanation for this apparent anomaly is that the value of the term B_t for the parameter θ_2 is almost twice that of Γ_t, and thus the neglect of high-order terms in the expansion of the profile likelihood as a power series may not be justified (see also Clarke (1987, Example 3.2)). Alternatively, and more persuasively, the

Mitscherlich model could well be inappropriate. In fact Draper and Smith (1981, p.522) indicate that the exponential decay model with $\eta(x,\theta) = \theta_1(1 - e^{\theta_3 x})$ is physically meaningful for biomedical oxygen demand data and, indeed, Bates and Watts (1988, p.41) fitted this model to such data. In the present case the fit of the exponential decay model to the data is good and, in addition, the approximate and exact relative overlaps of the Wald and likelihood-based 95% confidence intervals are in close agreement for both parameters.

Table 5. Curvature measures and relative overlap of the 95% confidence intervals for the individual parameters of the Mitscherlich model.

θ_p	$\hat{\theta}_p$	$SE(\hat{\theta}_p)$	Γ_t	B_t	ARO	$P_{\min}$	ERO
θ_1	222.627	14.927	-0.5325	0.3263	0.6275	0.4675	0.4973
θ_2	-191.303	10.397	0.0721	0.1392	0.8778	0.8699	0.5722
θ_3	-0.2996	0.0675	0.0643	0.0257	0.9385	0.9357	0.9378

(a) Biomedical oxygen demand data.

θ_p	$\hat{\theta}_p$	$SE(\hat{\theta}_p)$	Γ_t	B_t	ARO	$P_{\min}$	ERO
θ_1	539.084	9.4323	-0.2405	0.0797	0.7927	0.7595	0.7791
θ_2	-307.548	9.7182	0.1305	0.0560	0.8806	0.8695	0.8647
θ_3	-0.0155	0.00126	0.0776	0.0183	0.9259	0.9224	0.9265

(b) Potato data.

To further allay concerns with regard to the setting of confidence limits to the parameters of the Mitscherlich model, a second data set, taken from Pimentel-Gomes (1953) and recorded as data set 6 in Ratkowsky (1983, p.102), was examined. The data comprise yields of potatoes for varying amounts of fertilizer, and interest again centers on setting 95% confidence intervals to the parameters. The intrinsic curvature for the model-data setting is negligible, the value of c is taken to be $t_{2;0.025} = 4.3027$, and details of the parameter estimates, the terms Γ_t and B_t and the relative overlap are given in Table 5(b). The parameters θ_1 and θ_2 are clearly significantly different and the three-parameter model fits the data well. In addition for the parameter θ_2 the value of the term B_t is less than half that of Γ_t, indicating that high-order terms in σ^2 in the expansion of the profile likelihood can be neglected. Furthermore the agreement between approximate and relative overlap for parameter θ_2, and indeed for all the parameters, is excellent. Note in addition that the value of $P_{\min}$ is close to, and thus a reasonable approximation for, the relative overlap.

Overall the results of this example underscore the need to treat approximate profile likelihood-based confidence limits associated with individual parameters for which $B_t > \Gamma_t$ with considerable caution and, in addition and arguably more importantly, to ensure that an appropriate model is fitted to the data.

6.3. The logistic model

Consider the data on bean root cells introduced by Ratkowsky (1983, p.88) and in particular consider fitting the three-parameter logistic model

$$\eta(x,\theta) = \frac{\theta_1}{1+e^{\theta_2(x-\theta_3)}},$$

where $\eta(x,\theta)$ represents water content of the cell and x the distance of the cell from the root tip, to this data. Note that the maximum intrinsic curvature associated with the overall model, $\Gamma^N = 0.107$, is negligible. Details of the parameter estimates and their standard errors, the approximate relative overlap of the 95% and the 99% Wald and profile likelihood confidence intervals, labelled ARO_{95} and ARO_{99} respectively, the skewness and kurtosis and the total subset curvature are summarized in Table 6. Note that for this example good agreement between approximate and exact relative overlap of the confidence intervals was observed and the exact values are therefore not recorded.

Table 6. Parameter estimates, approximate relative overlap, skewness and kurtosis and the total subset curvature relating to the individual parameters of the logistic model.

θ	$\hat{\theta}$	SE	ARO_{95}	ARO_{99}	γ_1	γ_2	Γ_s
θ_1	21.509	0.4154	0.9519	0.9335	0.1362	0.0634	0.1020
θ_2	-0.6222	0.0446	0.9098	0.8766	-0.2618	0.1575	0.0914
θ_3	6.3604	0.1388	0.9736	0.9633	0.0740	0.0421	0.1260

Cook and Goldberg (1986) suggested that total subset curvature, Γ_s, be used to appraise the closeness or otherwise of the Wald and profile likelihood-based confidence intervals for an individual parameter, with values "substantially less" than the inverse of the appropriate critical value indicating close-to-linear behaviour. In the present example the values of Γ_s recorded in Table 6 are indeed substantially smaller than the values of $1/c$ of 0.4592 and 0.3274 for the 95% and 99% confidence intervals respectively, indicating that the Wald intervals are satisfactory in all cases. In contrast the values of the approximate relative overlap indicate that intervals based on the profile likelihoods should be calculated at both the 95% and 99% confidence levels for parameter θ_2, at the 99% level for θ_1, and that the Wald intervals are acceptable otherwise. The discrepancy in the conclusions relating to the closeness or otherwise of the Wald and the likelihood-based intervals drawn from the approximate relative overlaps and from the total subset curvatures are disturbing. Furthermore it is also clear from Table 6 that total subset curvature and approximate relative overlap do not appear to be related in an obvious way. Similar observations were made by van Ewijk and Hoekstra (1994), prompting them to refute the use of total subset curvature in assessing the appropriateness or otherwise of Wald confidence intervals.

In order to assess the nonlinearity associated with an individual parameter more broadly, the rules of thumb based on approximate relative overlap and developed in Section 5.1 can be invoked. Specifically these indicate that the nonlinearity for the parameter θ_3 is negligible, that for θ_1 is moderate and that for θ_2 tends to be severe. Furthermore, as explained in Section 5.1, these rules of thumb can be translated into rules of thumb for the skewness and kurtosis of the distribution of the corresponding estimates. In the present case it is clear that the nonlinearity exhibited by the parameters θ_1 and θ_2 can be identified with skewness rather than kurtosis.

Finally, as an aside, it is interesting to note that van Ewijk and Hoekstra (1994) fitted the Mitscherlich model of Example 6.2 to the current data set. This is a curious choice of model in that the data clearly exhibit a sigmoidal growth pattern whereas the model function (6.1) has no point of inflection. In fact the fit is very poor. As a consequence, measures of curvature associated with fitting the Mitscherlich model to the bean root cell data set and the conclusions drawn from them, in particular by van Ewijk and Hoekstra (1994), may well be unreliable and possibly misleading. This observation again highlights the importance of fitting an appropriate model to a particular data set.

6.4. The linear logistic model

van Ewijk and Hoekstra (1994) examined a large number of ecotoxicity data sets and advocated fitting the linear logistic model

$$\eta(x,\theta) = \frac{\theta_1 \left\{1 + \frac{1}{2}(e^{\theta_2\theta_3} - 1)e^{x-\theta_4}\right\}}{1 + e^{\theta_2(x-\theta_4+\theta_3)}},$$

where $\eta(x,\theta)$ represents plant growth and x the natural log of the concentration of chemical compound to such data. Note that the parameter θ_4 corresponds to the natural log of the ED50 and is of particular interest in these examples. The intrinsic curvature for many of the model-data settings is high and the linear logistic model was therefore fitted to one of the data sets provided by van Ewijk and Hoekstra (1994) for which the maximum intrinsic curvature $\Gamma^N = 0.450$ is relatively small, namely data set 73. The critical value of c is $t_{10;0.025} = 2.2281$ and parameter estimates, scaled total subset curvature and measures of relative overlap of the Wald and likelihood-based 95% confidence intervals for the model parameters are presented in Table 7. It is immediately clear that the approximate relative overlap and the probability $P_{\min}$ are good indicators of exact relative overlap, whereas total subset curvature is not. Furthermore for the parameter of interest, θ_4, the value of the approximate relative overlap indicates that the 95% confidence limits be taken as the those based on the profile likelihood. This recommendation is supported by the fact that the Wald and the exact likelihood-based intervals are given by $(1.0122, 1.7409)$ and $(1.0386, 1.7947)$, respectively.

Table 7. Parameter estimates, scaled total subset curvature and relative overlap for the parameters of the linear logistic model.

θ_p	$\hat{\theta_p}$	$SE(\hat{\theta_p})$	$c\Gamma_s$	ARO	$P_{\min}$	ERO
θ_1	0.0469	0.0023	0.0936	0.9906	0.9905	0.9932
θ_2	1.4476	0.0504	0.4039	0.8672	0.8561	0.8565
θ_3	2.1689	0.4089	0.4942	0.9567	0.9567	0.9768
θ_4	1.3769	0.1636	0.4672	0.9029	0.8967	0.8964

7. The Fieller-Creasy Problem

Cook and Witmer (1985) formulated the Fieller-Creasy problem relating to the estimation of the ratio of two population means as a nonlinear regression model. Specifically suppose that random samples of size n are drawn independently from two normal populations with means θ_1 and $\theta_1\theta_2$, respectively, and with a common variance σ^2. Then the observations can be modelled as

$$y_i = \theta_1 x_i + \theta_1\theta_2(1 - x_i) + \epsilon_i \qquad i = 1, \ldots, 2n,$$

where x_i is an indicator variable equal to 1 for the population with mean θ_1 and 0 for the population with mean $\theta_1\theta_2$, and with error terms ϵ_i independently distributed as $N(0, \sigma^2)$. Interest centers on the parameter θ_2 which represents the ratio of the two population means.

The exact relative overlap between the Wald and the profile likelihood-based confidence intervals for the parameter θ_2 can be obtained directly from the explicit expressions for these intervals derived by Cook and Witmer (1985). Specifically this overlap is given by

$$\begin{cases} \dfrac{\sqrt{r}\hat{\theta}_2+(1-r)\sqrt{1+\hat{\theta}_2^2}+\sqrt{1+\hat{\theta}_2^2-r}}{-\sqrt{r}\hat{\theta}_2+(1-r)\sqrt{1+\hat{\theta}_2^2}+\sqrt{1+\hat{\theta}_2^2-r}} & \text{for } \hat{\theta}_2 < -\sqrt{\frac{r}{4-r}}, \\ \dfrac{(1-r)\sqrt{1+\hat{\theta}_2^2}}{\sqrt{1+\hat{\theta}_2^2-r}} & \text{for } -\sqrt{\frac{r}{4-r}} < \hat{\theta}_2 < \sqrt{\frac{r}{4-r}}, \\ \dfrac{-\sqrt{r}\hat{\theta}_2+(1-r)\sqrt{1+\hat{\theta}_2^2}+\sqrt{1+\hat{\theta}_2^2-r}}{\sqrt{r}\hat{\theta}_2+(1-r)\sqrt{1+\hat{\theta}_2^2}+\sqrt{1+\hat{\theta}_2^2-r}} & \text{for } \hat{\theta}_2 > \sqrt{\frac{r}{4-r}}, \end{cases}$$

where $r = (c^2\sigma^2)/(n\hat{\theta}_1^2)$, with c an appropriate critical value and $\hat{\theta}_1$ and $\hat{\theta}_2$ the maximum likelihood estimators of θ_1 and θ_2, respectively. Note that the confidence limits associated with the profile likelihood only define an interval provided $r < 1$. From these results it is straightforward to calculate values of $\hat{\theta}_1$ and $\hat{\theta}_2$ for which the exact relative overlap is equal to a given value p. Note in particular that $\hat{\theta}_1 = \pm[\sigma c/(n(1-p^2))^{1/2}]$ when $\hat{\theta}_2 = 0$ and that $\hat{\theta}_2$ approaches plus or minus infinity as $\hat{\theta}_1$ approaches $[(1+p)+(9-14p+9p^2)^{1/2}]/(4(1-p)) \times (\sigma c)/n^{1/2}$ from below, and minus that value from above. For the case with $\sigma = 0.1$, $n = 10$ and $c = 2$, plots of the values of $\hat{\theta}_1$ and $\hat{\theta}_2$ for which the exact relative overlap is $80\%, 85\%, 90\%$ and 95% are shown as solid lines in Figure 2. It is interesting to note that, in accord with the findings of Clarke (1987), the Wald confidence interval does not provide a good approximation to the profile likelihood-based interval for small values of the parameter θ_1.

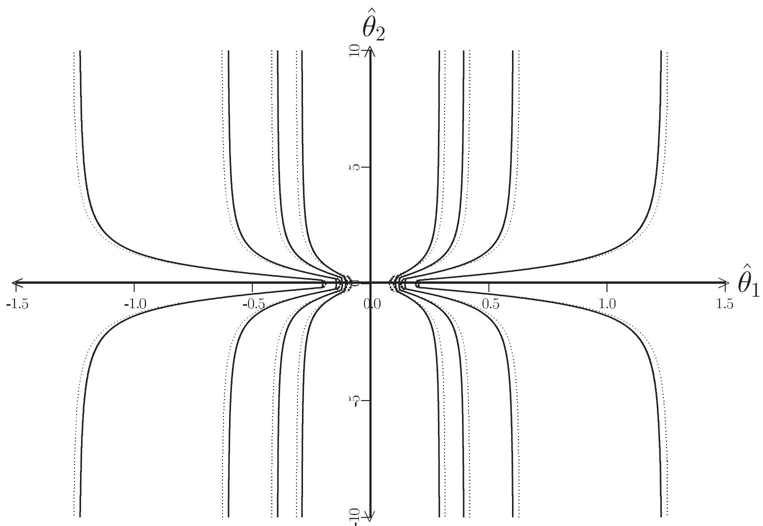

Figure 2. Values of the $(\hat{\theta}_1, \hat{\theta}_2)$-pairs for the Fieller-Creasy problem with $c = 2$, $\sigma = 0.1$ and $n = 10$ for which the exact relative overlap (solid lines) and the approximate relative overlap based on the rules of thumb for skewness and kurtosis (dotted lines) are 80%, 85%, 90% and 95%.

The second-order approximations to the scaled coefficient of skewness and the scaled excess kurtosis for the parameter θ_2, written in terms of the maximum likelihood estimates $\hat{\theta}_1$ and $\hat{\theta}_2$, follow immediately from the expressions

$$\Gamma = -\frac{2\theta_2}{\theta_1\sqrt{n(1+\theta_2^2)}} \qquad \text{and} \qquad \beta = \frac{1+2\theta_2^2}{n\theta_1^2(1+\theta_2^2)}$$

derived by Clarke (1987) and evaluated at $\hat{\theta}_1$ and $\hat{\theta}_2$, and are given by

$$\gamma_{1c} = \frac{6\hat{\theta}_2\sqrt{r}}{\sqrt{1+\hat{\theta}_2^2}} \qquad \text{and} \qquad \gamma_{2c} = \frac{12r(1+6\hat{\theta}_2^2)}{(1+\hat{\theta}_2^2)},$$

respectively. The approximation to the relative overlap described in Section 5.1, and based on Clarke's approximation to the likelihood-based confidence limits, can then be found by substituting these expressions for γ_{1c} and γ_{2c} into the appropriate formulae in Table 2. Note that in this example the order of the limits WPPW cannot occur and, in accord with this, $\beta > 0$. The exact and the approximate relative overlaps of the Wald and profile likelihood-based confidence intervals were examined for a large range of possible values of r and $\hat{\theta}_2$, and thus of problem settings, and clearly indicated that the approximation to the relative overlap consistently overestimates the true overlap, and also that this approximation was excellent for overlap values greater than 95%. The latter observation reinforces the appropriateness of the rules of thumb based on relative overlap and developed in Section 5.1.

The excess kurtosis for this example is non-negative and the rules of thumb $|\gamma_{1c}| \leq 6(1-p)$ and $\gamma_{2c} \leq (24(1-p))/p$ developed in Section 5.1 can therefore be reformulated in terms of $\hat{\theta}_1$ and $\hat{\theta}_2$ as

$$\hat{\theta}_2^2 \leq \frac{\hat{\theta}_1^2}{(a_s^2-\hat{\theta}_1^2)} \quad \text{where} \quad a_s = \frac{c\sigma}{\sqrt{n}(1-p)}, \tag{7.1}$$

$$\hat{\theta}_2^2 \leq \frac{\hat{\theta}_1^2 - a_k^2}{(6a_k^2-\hat{\theta}_1^2)} \quad \text{where} \quad a_k = \frac{c\sigma\sqrt{p}}{\sqrt{2n(1-p)}}, \tag{7.2}$$

respectively. Note that for equality in (7.1), $\hat{\theta}_2 = 0$ when $\hat{\theta}_1 = 0$, and $\hat{\theta}_2 \to \pm\infty$ as $\hat{\theta}_1 \to \pm a_s$, and note also that for equality in (7.2), $\hat{\theta}_1 = \pm a_k$ when $\hat{\theta}_2 = 0$ and $\hat{\theta}_2 \to \pm\infty$ as $\hat{\theta}_1 \to \pm 6^{1/2}a_k$. For the case where $\sigma = 0.1$, $n = 10$ and $c = 2$, plots for which these rules of thumb for p values of $80\%, 85\%, 90\%$ and 95% hold simultaneously are presented as dotted lines in Figure 2. The similarity in the patterns for exact and approximate relative overlaps exhibited in that figure is striking, indicating that the rule of thumb based simultaneously on skewness and kurtosis performs well in this particular case. It is also interesting to examine the impact of skewness and kurtosis separately. To this end, plots of the values of $\hat{\theta}_1$ and $\hat{\theta}_2$ for which the rules of thumb based on skewness and kurtosis with $p = 80\%$ hold separately are shown in Figure 3. From these it is clear that the rules of thumb for scaled skewness and kurtosis do not coincide and indeed that neither measure, singly, provides an adequate indicator of relative overlap.

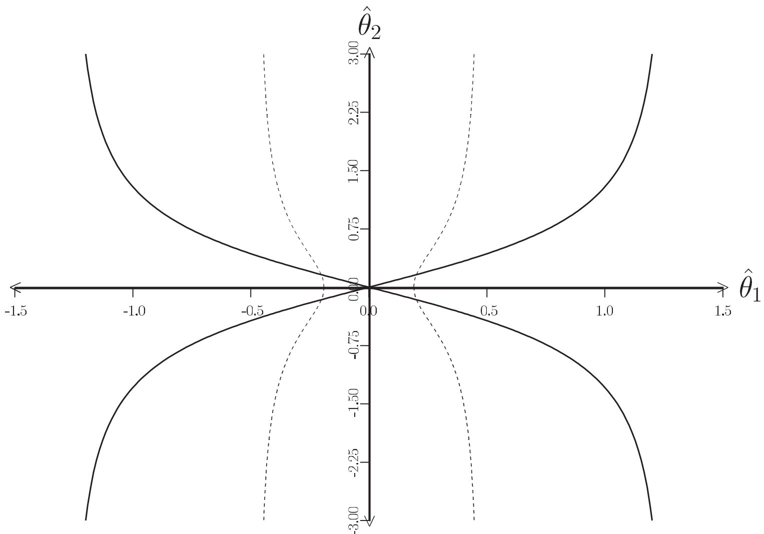

Figure 3. Values of the $(\hat{\theta}_1, \hat{\theta}_2)$-pairs for the Fieller-Creasy problem with $c = 2$, $\sigma = 0.1$ and $n = 10$ for which the equalities in (7.2) and (7.3) relating to scaled skewness (solid lines) and kurtosis (dashed lines) for a relative overlap of $\rho = 95\%$ are satisfied.

Finally it is straightforward to show that the scaled total subset curvature for the parameter θ_2 is given by $(2\sigma c)/(|\theta_1|\, n^{1/2}) = 2r^{1/2}$. Thus, for a fixed value of θ_1, this measure remains constant regardless of the value of θ_2. However, since exact relative overlap varies as θ_2 changes, it is clear that total subset curvature is unreliable as a predictor of that overlap. Thus, for example, for $r = 0.1$ the exact relative overlap is equal to 94.87% when $\theta_2 = 0$ but approaches 71.46% for very large values of $|\theta_2|$, whereas the total subset curvature under each of these settings is the same.

8. Conclusions

One of the main features of this paper is the derivation of an algebraic expression for the second-order approximation to kurtosis for the least squares estimate of an individual parameter in a nonlinear regression model. This result complements those already established for the bias and the skewness associated with such estimates (Box (1971) and Hougaard (1985)). Furthermore the expression for approximate kurtosis is immediately useful in that it is readily and rapidly computed, in contrast to estimates obtained by simulation (Ratkowsky (1983)). As an aside, it is interesting to note that the second-order approximations to skewness and kurtosis are shown to be closely related to a broad range of measures of curvature for individual parameters through certain building block terms taken from Clarke (1987).

A second feature of the present study is the formulation of rules of thumb for assessing the closeness or otherwise of Wald and profile likelihood-based confidence intervals and, more broadly, the nonlinearity associated with an individual parameter of a nonlinear regression model. The rules are derived in terms of approximate skewness and kurtosis and are translated into simple rules relating to those measures. The rules of thumb for appraising the disparity between Wald and likelihood-based confidence intervals for an individual parameter are summarized in Tables 2 and 3 and are directed towards the practitioner who has available a Wald interval and is uncertain as to whether or not to calculate the interval based on the profile likelihood. The rules of thumb for appraising the nonlinearity associated with an individual parameter, and hence the corresponding measures of skewness and kurtosis, are presented in Section 5.1 but it should be noted that these are somewhat tentative.

A number of examples are given in this paper and many are of interest in their own right. From these it is clear that the rules of thumb developed in the present study perform well but that there are two important caveats. First the intrinsic nonlinearity associated with the nonlinear regression model of interest should be negligible, and second, the chosen model should be appropriate for and provide a good fit to the data. In other words, spurious results relating to curvature may well be obtained if either of these two conditions fails to hold (van Ewijk and Hoekstra (1994)).

A particular drawback to the calculation of measures of curvature for nonlinear regression models, and of the second-order approximations to skewness and excess kurtosis in particular, is the tedious algebra needed to obtain the requisite first, second and third-order derivatives of the expected response with respect to the parameters. This problem is to some extent alleviated today by the ready availability of symbolic algebra packages which can in turn be linked to a range of statistical and programming languages, and also by routines for calculating the required derivatives numerically. However for the practitioner the computations still remain time-consuming and tedious. Thus work is currently in progress to develop quick and easy-to-use software for calculating a comprehensive suite of curvature and related measures for any nonlinear regression model and for the individual parameters within such a model.

Acknowledgement

Linda Haines was generously supported by funding from the University of KwaZulu-Natal and the National Research Foundation, South Africa. Tim O'Brien gratefully acknowledges the financial support of the Katholieke Universiteit Leuven (Belgium), which funded his sabbatical leave for 2001-2002. The authors would like to thank P. H. van Ewijk and J. A. Hoekstra for making available to them the data sets used in their paper, and an associate editor and two referees for their insightful and helpful comments.

Appendix. Expectations for Quadratic Forms

Moments of quadratic forms in the variable $z \sim N(0, \sigma^2 I)$ can be derived routinely from the cumulant generating function and, in turn, used to obtain expectations of products of those forms. Specifically, consider the two quadratic forms $q_1 = z^T Q_1 z$ and $q_2 = z^T Q_2 z$. Then $E[(q_1 - E[q_1])(q_2 - E[q_2])^2] = 8\ \sigma^6\ tr(Q_1 Q_2^2)$ and thus

$$E[q_1(q_2 - E[q_2])^2] = E[q_1]E[(q_2 - E[q_2])^2] + 8\sigma^6 tr(Q_1 Q_2^2). \tag{A.1}$$

Consider also the three quadratic forms q_1, q_2 and $q_3 = z^T Q_3 z$. Then

$$E[(q_1 - E[q_1])(q_2 - E[q_2])(q_3 - E[q_3])] = 4\ \sigma^6 \{tr(Q_1 Q_2 Q_3) + tr(Q_1 Q_3 Q_2)\}$$

and thus

$$\begin{aligned} E(q_1 q_2 q_3) = 4\ \sigma^6 \{tr(Q_1 Q_2 Q_3) + tr(Q_1 Q_3 Q_2)\} + E[q_1]\mathrm{Cov}(q_2, q_3) + \\ E[q_2]\mathrm{Cov}(q_1, q_3) + E[q_3]\mathrm{Cov}(q_1, q_2) + E[q_1]E[q_2]E[q_3]. \end{aligned}$$

It now follows that

$$\begin{aligned} E[(a^T z)^3 (b^T z)(z^T Q z)] &= E[(z^T a a^T z)(z^T a b^T z)(z^T Q z)] \\ &= 3\sigma^6 \{2(a^T a)(a^T Q b) + (a^T a)(a^T b) tr Q + 2(a^T b)(a^T Q a)\} \end{aligned}$$

and hence, since $E[(a^T z)^2] = \sigma^2 (a^T a)$ and $E[(a^T z)(b^T z)(z^T Q z)] = \sigma^4 \{2a^T Q b + (a^T b) tr Q\}$, that

$$E[(a^T z)^3 (b^T z)(z^T Q z)] = 3E[(a^T z)^2] E[(a^T z)(b^T z)(z^T Q z)] + 6\sigma^6 (a^T b)(a^T Q a). \tag{A.2}$$

References

Bates, D. M. and Watts, D. G. (1980). Relative curvature measures of nonlinearity (with discussion). *J. Roy. Statist. Soc. Ser. B* **42**, 1-25.

Bates, D. M. and Watts, D. G. (1988). *Nonlinear Regression Analysis and its Applications*. Wiley, New York.

Beale, E. M. L. (1960). Confidence regions in non-linear estimation (with discussion). *J. Roy. Statist. Soc. Ser. B* **22**, 41-88.

Bliss, C. I. and James, A. T. (1966). Fitting the rectangular hyperbola. *Biometrics* **22**, 573-602.

Box, M. J. (1971). Bias in nonlinear estimation (with discussion). *J. Roy. Statist. Soc. Ser. B* **33**, 171-201.

Clarke, G. P. Y. (1980). Moments of the least squares estimators in a non-linear regression model. *J. Roy. Statist. Soc. Ser. B* **42**, 227-237.

Clarke, G. P. Y. (1987). Marginal curvatures and their usefulness in the analysis of nonlinear regression models. *J. Amer. Statist. Assoc.* **82**, 844-850.

Cook, R. D. and Goldberg, M. L. (1986). Curvatures for parameter subsets in nonlinear regression. *Ann. Statist.* **14**, 1399-1418.

Cook, R. D. and Tsai, C-L. (1990). Diagnostics for assessing the accuracy of normal approximations in exponential family nonlinear models. *J. Amer. Statist. Assoc.* **85**, 770-777.

Cook, R. D. and Weisberg, S. (1990). Confidence curves in nonlinear regression. *J. Amer. Statist. Assoc.* **85**, 544-551.

Cook, R. D. and Witmer, J. A. (1985). A note on parameter-effects curvature. *J. Amer. Statist. Assoc.* **80**, 872-878.

Draper, N. R. and Smith, H. (1981). *Applied Regression Analysis*. Second Edition. Wiley, New York.

Gillis, P. R. and Ratkowsky, D. A. (1978). The behaviour of estimators of the parameters of various yield-density relationships. *Biometrics* **34**, 191-198.

Hamilton, D. (1986). Confidence regions for parameter subsets in nonlinear regression. *Biometrika* **73**, 57-64.

Hodges, J. S. (1987). Assessing the accuracy of normal approximations. *J. Amer. Statist. Assoc.* **82**, 149-154.

Horswell, R. L. and Looney, S. W. (1993). Diagnostic limitations of skewness coefficients in assessing departures from univariate and multivariate normality. *Comm. Statist. Simulation Comput.* **22**, 437-459.

Hougaard, P. (1982). Parametrizations of non-linear models. *J. Roy. Statist. Soc. Ser. B* **44**, 244-252.

Hougaard, P. (1985). The appropriateness of the asymptotic distribution in a nonlinear regression model in relation to curvature. *J. Roy. Statist. Soc. Ser. B* **47**, 103-114.

Jennings, D. E. (1986). Judging inference adequacy in logistic regression. *J. Amer. Statist. Assoc.* **81**, 471-476.

Kang, G. and Rawlings, J. O. (1998). Marginal curvatures for functions of parameters in nonlinear regression. *Statist. Sinica* **8**, 467-476.

Pimentel-Gomes, F. (1953). The use of Mitscherlich's regression law in the analysis of experiments with fertilizers. *Biometrics* **9**, 498-516.

Ratkowsky, D. A. (1983). *Nonlinear Regression Modeling*. Marcel Dekker, New York.

Ratkowsky, D. A. (1990). *Handbook of Nonlinear Regression Models*. Marcel Dekker, New York.

Rayner, J. C. W., Best, D. J. and Mathews, K. L. (1995). Interpreting the skewness coefficient. *Comm. Statist. Theory Method* **24**, 593-600.

Seber, G. A. F. and Wild, C. J. (1989). *Nonlinear Regression*. Wiley, New York.

van Ewijk, P. H. and Hoekstra, J. A. (1994). Curvature measures and confidence intervals for the linear logistic model. *Appl. Statist.* **43**, 477-487.

School of Mathematics, Statistics and Information Technology, University of KwaZulu-Natal, Private Bag X01, Scottsville 3209, Pietermaritzburg, South Africa.

E-mail: haines@nu.ac.za

Loyola University Chicago, Department of Mathematics and Statistics, 6525 N. Sheridan Road, Chicago, Illinois 60626, U.S.A.

E-mail: tobrien@math.luc.edu

Agriculture Western Australia, 3 Baron Court Road, South Perth WA 6151, Australia.

E-mail: pclarke@agric.wa.gov.au

(Received April 2002; accepted September 2003)