Ann. Inst. Statist. Math.
Vol. 54, No. 2, 235–244 (2002)

QUASI-PROFILE LOGLIKELIHOODS FOR UNBIASED ESTIMATING FUNCTIONS

G. ADIMARI AND L. VENTURA

Dipartimento di Scienze Statistiche, Università di Padova, via C. Battisti 241, 35121 Padova, Italy

(Received April 3, 2000; revised October 20, 2000)

Abstract. This paper presents a new quasi-profile loglikelihood with the standard kind of distributional limit behaviour, for inference about an arbitrary one-dimensional parameter of interest, based on unbiased estimating functions. The new function is obtained by requiring the corresponding quasi-profile score function to have bias and information bias of order $O(1)$. We illustrate the use of the proposed pseudo-likelihood with an application to robust inference in linear models.

Key words and phrases: Estimating equation, M-estimator, profile likelihood, quasi-likelihood, second Bartlett identity.

1. Introduction

Consider a sample $\underline{y} = (y_1, \dots, y_n)$ of n independent observations with distribution function $F(\underline{y}; \theta)$ depending on an unknown parameter $\theta \in \Theta \subseteq R^d$, $d \geq 1$. Let $\Psi(\underline{y}; \theta) = \sum_{i=1}^n \psi(y_i; \theta)$ be an unbiased estimating function for θ based on $\underline{y}$. Occasionally, we shall write Ψ_θ and ψ_θ for $\Psi(\underline{y}; \theta)$ and $\psi(y; \theta)$, respectively. The estimator of θ corresponding to Ψ_θ is defined as a root $\hat{\theta}$ of the estimating equation $\Psi(\underline{y}; \theta) = 0$. Under broad conditions which we will assume throughout this paper (see e.g. Barndorff-Nielsen and Cox (1994), Section 9.2) it can be shown that $\hat{\theta}$ is consistent and asymptotically normal, with mean θ and variance $B(\theta)^{-1}\Omega(\theta)(B(\theta)^{-1})^{\mathrm{T}}$, where $B(\theta) = -E\{\Psi_{\theta/\theta}\}$, $\Omega(\theta) = \mathrm{var}\{\Psi_\theta\} = E\{\Psi_\theta \Psi_\theta^{\mathrm{T}}\}$ and the symbol / as a subscript indicates differentiation.

Let $l_Q(\theta) = l_Q(\theta; \underline{y})$ be a scalar function whose gradient with respect to θ equals Ψ_θ, i.e. $l_Q(\theta) = \int_c^\theta \Psi(\underline{y}; t)dt$, where c is an arbitrary constant. When $l_Q(\theta)$ exists, it may be thought of as a quasi-loglikelihood for θ and it may be used, in analogy with ordinary loglikelihood, for setting quasi-likelihood tests and confidence regions. Actually, the relation

$$\mathrm{var}\{\Psi_\theta\} = -E\{\Psi_{\theta/\theta}\}, \tag{1.1}$$

that is known as the second Bartlett identity when Ψ_θ is the usual score function (see Bartlett (1953a, 1953b)) does not hold in general. It is however possible to make relation (1.1) hold, by considering the linear transformation

$$\Psi_{\theta 1} = \Psi_1(\underline{y}; \theta) = A(\theta)\Psi_\theta, \tag{1.2}$$

where the matrix $A(\theta)$ is such that

$$A(\theta)^{\mathrm{T}} = -\mathrm{var}\{\Psi_\theta\}^{-1}E\{\Psi_{\theta/\theta}\} = \Omega(\theta)^{-1}B(\theta) \tag{1.3}$$

(see McCullagh (1991), Section 11.7). Since $A(\theta)$ is nonsingular for all θ, the estimating functions $\Psi_\theta = 0$ and $\Psi_{\theta 1} = 0$ have the same solution. If a quasi-loglikelihood function satisfies (1.1) many asymptotic considerations are simplified. In particular, the quasi-observed information has the usual relation with the asymptotic covariance matrix of the estimator $\hat{\theta}$ and the quasi-likelihood ratio statistic has a standard χ^2 distribution. Quasi-likelihood has been introduced in the context of generalized linear models (see McCullagh and Nelder (1989)). In this case relation (1.1) is verified if the variance function is correctly specified and, following Godambe (1976), the quasi-score is an optimal unbiased estimating function. For a survey of quasi-likelihood and estimating functions see Heyde (1997) and Desmond (1997).

When $d = 1$, a quasi-loglikelihood for θ, corresponding to the modified estimating function (1.2), given by

$$\bar{l}_Q(\theta) = \sum_{i=1}^{n} \int_{c}^{\theta} A(t)\psi(y_i;t)dt,$$

is usually easy to derive. In view of this, for setting quasi- likelihood confidence regions or for testing hypotheses, the quasi- likelihood ratio statistic

$$W_Q(\theta) = 2\{\bar{l}_Q(\hat{\theta}) - \bar{l}_Q(\theta)\} = 2\sum_{i=1}^{n} \int_{\theta}^{\hat{\theta}} A(t)\psi(y_i;t)dt \tag{1.4}$$

may be used. For instance, confidence regions with nominal coverage $1 - \alpha$ for θ can be constructed as $\{\theta : W_Q(\theta) \leq \chi^2_{1;1-\alpha}\}$, where $\chi^2_{1;1-\alpha}$ is the $(1 - \alpha)$-quantile of the χ^2_1 distribution. Alternatively, the directed quasi-likelihood $r_Q(\theta) = \text{sgn}(\hat{\theta} - \theta)\{W_Q(\theta)\}^{1/2}$, which is approximately standard normal, may be used.

When $d > 1$, a quasi-loglikelihood for θ does not exist in general. A necessary and sufficient condition for the existence is that the matrix $\Psi_{\theta 1/\theta}$ be symmetric. Nevertheless, the problem of nonexistence may be overcome when the interest parameter is a scalar component of θ. For this case Barndroff-Nielsen (1995) proposes a quasi-profile loglikelihood with the standard kind of distributional limit behaviour. However, as will be discussed in Section 2, the modification of the estimating function needed to achieve the usual asymptotic behaviour and, in particular, the asymptotic χ^2_1 distribution for the quasi-profile likelihood ratio statistic, may lead to some interpretation problems as well as computational difficulties.

To avoid such drawbacks, in this paper we propose an alternative quasi- profile loglikelihood for an arbitrary one-dimensional parameter of interest. Such a function, called *adjusted quasi-profile loglikelihood*, is obtained by a scale adjustment of the estimating function for the scalar parameter of interest only, aimed at obtaining a quasi-profile score function with properties similar to those of the ordinary profile score, i.e. with bias and information bias of order $O(1)$. An application example, discussed in Section 3, illustrates the use of the proposed pseudo-likelihood function for robust inference in linear models.

2. Quasi-profile loglikelihood functions

Suppose that θ is partitioned as $\theta = (\tau, \lambda)$ into a scalar parameter of interest τ and a $(d - 1)$-dimensional nuisance parameter λ. The estimating function Ψ_θ is similarly partitioned as $(\Psi_\tau, \Psi_\lambda)$, where $\Psi_\tau = \Psi_\tau(\underline{y};\theta)$ and $\Psi_\lambda = \Psi_\lambda(\underline{y};\theta)$ are the estimating functions corresponding to τ and λ, respectively. This means that, for instance, if λ is known, Ψ_τ may be used as an estimating function for τ.

To define a quasi-profile loglikelihood for τ, Barndorff-Nielsen (1995) assumes that the estimating function Ψ_θ is multiplied by the matrix $A(\theta)$ so that relation (1.1) is satisfied. Consequently, the resulting $\Psi_{\theta 1}$ is partitioned as

$$\Psi_{\theta 1} = \begin{pmatrix} \Psi_{\tau 1} \\ \Psi_{\lambda 1} \end{pmatrix} = \begin{pmatrix} A_{\tau,\tau}\Psi_\tau + A_{\tau,\lambda}\Psi_\lambda \\ A_{\lambda,\tau}\Psi_\tau + A_{\lambda,\lambda}\Psi_\lambda \end{pmatrix}, \tag{2.1}$$

where $A_{\tau,\tau}$, $A_{\tau,\lambda}$, $A_{\lambda,\tau}$ and $A_{\lambda,\lambda}$ are, respectively, the (τ,τ), (τ,λ), (λ,τ) and (λ,λ) blocks of the matrix $A(\theta)$. Let $\bar{\lambda}_\tau$ be the estimate for λ derived from $\Psi_{\lambda 1}$ when τ is considered as known, i.e. $\Psi_{\lambda 1}(\underline{y};\tau,\bar{\lambda}_\tau) = 0$. For an arbitrary estimating function $\Psi_{\theta 1}$ so specified, Barndorff-Nielsen defines the quasi-profile score for τ by $\Psi_{\tau 1}(\underline{y};\tau,\bar{\lambda}_\tau)$ and the corresponding quasi-profile loglikelihood function for τ by

$$\bar{l}_{QP}(\tau) = \int_c^\tau \Psi_{\tau 1}(\underline{y};t,\bar{\lambda}_t)dt. \tag{2.2}$$

This pseudo-likelihood has properties similar to the ordinary profile likelihood, since the quasi-profile likelihood ratio statistic and the quasi-profile directed likelihood, under regularity conditions of the standard type, have the usual asymptotic distributions (see Barndorff-Nielsen (1995)). Then, (2.2) may be used for setting quasi-likelihood intervals for τ, for testing hypotheses, etc.

However, due to transformation (1.2), some conceptual and practical difficulties may arise in using the quasi-profile loglikelihood (2.2). In fact, transformation (1.2) mixes, in general, the components of the original estimating function Ψ_θ. As a consequence, the interpretation of the components of the new estimating function $\Psi_{\theta 1}$ cannot be clear, since, in general, in (2.1) the original partition of the estimating function Ψ_θ into the estimating equation for the interest parameter and the one for the nuisance parameter is no longer respected. Moreover, the partial estimator $\bar{\lambda}_\tau$ does not coincide, in general, with the estimator of λ that actually would be used if τ was known, i.e. with the solution in λ of $\Psi_\lambda(\underline{y};\tau,\lambda)$. Finally, the use of $\Psi_{\theta 1}$ instead of Ψ_θ can increase the computational task needed to make inference about τ.

Observe that all these difficulties vanish when the matrix $A(\theta)$ is such that $A_{\lambda,\tau} = 0$. This condition on the matrix $A(\theta)$ is equivalent to the condition

$$E\{\Psi_\lambda^{\mathrm{T}}\Psi_\lambda\}E\{\Psi_{\tau/\lambda}\} = E\{\Psi_\tau\Psi_\lambda\}E\{\Psi_{\lambda/\lambda}\} \tag{2.3}$$

on the estimating function Ψ_θ. Relation (2.3) is obtained by looking for a transformation of the form (χ,ω), with $\chi = \chi(\tau)$ and $\omega = \omega(\tau,\lambda)$, such that $A_{\omega,\chi} = 0$, motivated as in Cox and Reid (1987) for orthogonal reparameterizations. Clearly, condition (2.3) is not verified in general.

In this paper we adopt a more natural criterion for the construction of a quasi-profile loglikelihood for τ, which is based essentially on a suitable adjustment to the estimating function of the interest parameter only. Let $\hat{\lambda}_\tau$ be the partial estimator of λ corresponding to Ψ_λ, i.e. $\Psi_\lambda(\underline{y};\tau,\hat{\lambda}_\tau) = 0$. When Ψ_θ is the usual score of the loglikelihood function, $\tilde{\Psi}_\tau$ is the ordinary profile score function. Here and in the following, the symbol $\sim$ indicates that a function of θ is evaluated at $(\tau,\hat{\lambda}_\tau)$ and, by convention, the operation $\sim$ is taken to be always the first carried out. Without this convention a symbol such as $\tilde{\Psi}_{\tau/\tau}$ would be ambiguous. It is well-known that, unlike the full score function, the mean of the profile score function is not in general exactly 0 and its variance does not satisfy the second Bartlett identity. However, its bias and information bias are both typically of order $O(1)$ (see McCullagh and Tibshirani (1990)). In view of this, for an arbitrary estimating function Ψ_τ, we propose to substitute the unknown parameter λ with its partial estimate $\hat{\lambda}_\tau$, obtaining the equivalent of an ordinary profile score function $\tilde{\Psi}_\tau = \Psi_\tau(\underline{y}; \tau, \hat{\lambda}_\tau)$. Then, we adjust $\tilde{\Psi}_\tau$ so that its bias and information bias are of order $O(1)$, as for the ordinary profile score function.

The pseudo-profile score function $\tilde{\Psi}_\tau$ has bias $E\{\tilde{\Psi}_\tau\}$ and information bias

$$\operatorname{var}\{\tilde{\Psi}_\tau\} + E\{\tilde{\Psi}_{\tau/\tau}\}. \tag{2.4}$$

In the Appendix we show that, under standard conditions, $E\{\tilde{\Psi}_\tau\}$ is of order $O(1)$, while (2.4) is of order $O(n)$. Essentially, we generalize the calculations of McCullagh and Tibshirani (1990) to an arbitrary profile estimating equation and we propose a scale adjustment to the pseudo-profile score function that reduces its information bias to order $O(1)$. The scale adjustment yields an estimating function of the form $\tilde{\Psi}_{\tau 2} = \Psi_{\tau 2}(\underline{y}; \tau, \hat{\lambda}_\tau) = w(\tau, \hat{\lambda}_\tau)\tilde{\Psi}_\tau$, where $w(\cdot,\cdot)$ is a suitable function, given in (2.7), resulting from the leading term of

$$\left\{\frac{\partial}{\partial\tau}E\{\tilde{\Psi}_\tau\} - E\{\tilde{\Psi}_{\tau/\tau}\}\right\} / \operatorname{var}\{\tilde{\Psi}_\tau\} \tag{2.5}$$

(see McCullagh and Tibshirani (1990), Section 3). Finally, let

$$l_{QP}(\tau) = \int_c^\tau w(t, \hat{\lambda}_t)\Psi_\tau(\underline{y}; t, \hat{\lambda}_t)dt \tag{2.6}$$

be the *adjusted quasi-profile loglikelihood function* for τ. This function, which represents an alternative to the quasi-profile loglikelihood (2.2), has some properties of the ordinary profile loglikelihood. In particular, we show that the adjusted quasi-profile likelihood ratio statistic $W_{QP}(\tau) = 2\{l_{QP}(\hat{\tau}) - l_{QP}(\tau)\}$ has approximately a standard χ^2_1 distribution.

To give $w(\tau, \lambda)$ explicitly, in the following it is convenient to use index notation. The components of λ are denoted by λ^a, the corresponding components of Ψ_λ are Ψ_a and the derivatives of Ψ_τ and Ψ_a with respect to the components of λ are denoted by

$$\Psi_{\tau/a} = \frac{\partial}{\partial\lambda^a}\Psi_\tau, \quad \Psi_{\tau/ab} = \frac{\partial^2}{\partial\lambda^a\partial\lambda^b}\Psi_\tau, \quad \Psi_{a/b} = \frac{\partial}{\partial\lambda^b}\Psi_a \text{ and } \Psi_{a/bc} = \frac{\partial^2}{\partial\lambda^b\partial\lambda^c}\Psi_a,$$

where the indices a, b, $c, \ldots$ range over $1, \ldots, d-1$. For the expected values of these derivatives, we use the notation $\nu_{\tau/a} = E\{\Psi_{\tau/a}\}$, $\nu_{\tau/ab} = E\{\Psi_{\tau/ab}\}$, $\nu_{a/b} = E\{\Psi_{a/b}\}$ and $\nu_{a/bc} = E\{\Psi_{a/bc}\}$ and we assume that these quantities are of order $O(n)$. Further, the zero-mean variables Ψ_τ, Ψ_a, $\Psi_{\tau/a} - \nu_{\tau/a}$, etc., are assumed to be of order $O_p(n^{1/2})$. These assumptions are typically satisfied in practice, when Ψ_θ behaves asymptotically like the sum of n independent random variables. In addition, $\kappa^{a/b}$ denotes the inverse matrix of $-\nu_{a/b}$ and $\nu_{\tau/\tau} = E\{\Psi_{\tau/\tau}\}$.

By using the expansions in the Appendix, we find $E\{\tilde{\Psi}_\tau\} = m(\tau, \lambda) + O(n^{-1})$, where $m(\cdot,\cdot)$ is of order $O(1)$ and has the expression given in (A.4). The expansion for (2.5) is more complicated. By using the results in the Appendix, we find

$$(2.7)\qquad w(\tau,\lambda)=\frac{-\nu_{\tau/\tau}-\kappa^{b/a}\nu_{\tau/a}\nu_{b/\tau}}{E\{\Psi_\tau^2\}+2\nu_{\tau/a}\kappa^{b/a}E\{\Psi_\tau\Psi_b\}+\nu_{\tau/a}\nu_{\tau/b}\kappa^{c/a}\kappa^{d/b}E\{\Psi_c\Psi_d\}}.$$

Observe that, when relation (2.3) holds, quasi-profile loglikelihoods (2.2) and (2.6) coincide. In fact, in view of (2.3) we have that $A_{\lambda,\tau}=0$, $\bar{\lambda}_\tau=\hat{\lambda}_\tau$ and the $A_{\tau,\tau}$ block of the matrix $A(\theta)$, which is in general given by

$$A_{\tau,\tau}=\frac{-\nu_{\tau/\tau}+E\{\Psi_\tau\Psi_\lambda\}E\{\Psi_\lambda\Psi_\lambda^T\}^{-1}\nu_{\lambda/\tau}}{E\{\Psi_\tau^2\}-E\{\Psi_\tau\Psi_\lambda\}E\{\Psi_\lambda\Psi_\lambda^T\}^{-1}E\{\Psi_\lambda\Psi_\tau\}},$$

reduces to

$$A_{\tau,\tau}=\frac{-\nu_{\tau/\tau}-\kappa^{b/a}\nu_{\tau/a}\nu_{b/\tau}}{E\{\Psi_\tau^2\}+\nu_{\tau/a}\kappa^{b/a}E\{\Psi_\tau\Psi_b\}}$$

which is the same expression that one obtains for $w(\tau,\lambda)$.

To show that, under regularity conditions of standard type, $W_{QP}(\tau)$ is approximately χ_1^2-distributed, we consider its Taylor expansion about $\hat{\tau}$. In view of some obvious simplifications, we find $W_{QP}(\tau)=-(\hat{\tau}-\tau)^2w(\tau,\hat{\lambda}_\tau)\tilde{\Psi}_{\tau/\tau}+o_p(1)$. Further, since the scale adjustment to the pseudo-profile score function $\tilde{\Psi}_\tau$ is such that $w(\tau,\hat{\lambda}_\tau)=-E\{\tilde{\Psi}_{\tau/\tau}\}/\text{var}\{\tilde{\Psi}_\tau\}+o(1)$ (see equations (A.5)–(A.7) in the Appendix), we have

$$-(\hat{\tau}-\tau)^2w(\tau,\hat{\lambda}_\tau)\tilde{\Psi}_{\tau/\tau}=(\hat{\tau}-\tau)^2\frac{E\{\tilde{\Psi}_{\tau/\tau}\}}{\text{var}\{\tilde{\Psi}_\tau\}}\tilde{\Psi}_{\tau/\tau}+o_p(1)=(\hat{\tau}-\tau)^2V_{\tau,\tau}^{-1}+o_p(1),$$

where $V_{\tau,\tau}$ denotes the (τ,τ) block of the matrix $B(\theta)^{-1}\Omega(\theta)(B(\theta)^{-1})^{\mathrm{T}}$. Since $\hat{\tau}$ is asymptotically normal with mean τ and variance $V_{\tau,\tau}$, the conclusion concerning the asymptotic distribution of $W_{QP}(\tau)$ follows.

3. Example: robust inference in linear models

Let $y_i=(x_i,z_i)$, $i=1,\dots,n$, be independent and identically distributed observations from a random vector $Y=(X,Z)$ such that $Z=X^{\mathrm{T}}\beta+e$, where β is an unknown vector belonging to R^{d-1}, $d\geq 2$, and e is independent of X and has distribution $F(\cdot;\sigma)=F_0(\cdot/\sigma)$ symmetric around 0, depending on a scale parameter σ. Let $\theta=(\beta,\sigma)$ and let $K(x)$ be the distribution of X on R^{d-1}.

A wide class of robust M-estimators for regression and scale parameters is defined by estimating functions of the form

$$(3.1)\qquad \Psi(\underline{y};\beta,\sigma)=\sum_{i=1}^n\psi(y_i;\beta,\sigma)=\begin{pmatrix}\sum_i s(x_i)\psi_\beta\{r_iv(x_i)\}x_i\\ \sum_i\psi_\sigma(r_i)\end{pmatrix},$$

where $r_i=(z_i-x_i^{\mathrm{T}}\beta)/\sigma$ and $s(\cdot)$, $v(\cdot)$, $\psi_\beta(\cdot)$, $\psi_\sigma(\cdot)$ are appropriate functions (see Hampel *et al.* (1986), Chapter 6). In particular, when $s(x)=v(x)=1$ and $\psi_\beta(\cdot)=\psi_{HF}(\cdot;k_1)$ we obtain the Huber (1973) estimator for regression, where $\psi_{HF}(u;k_1)=u\min\{1,k_1/|u|\}$, for some positive constant k_1. Alternatively, the choice $s(x)=1/v(x)$, $v(x)=||x||$ and $\psi_\beta(\cdot)=\psi_{HF}(\cdot;k_1)$ defines the so-called Hampel-Krasker estimator (see Maronna *et al.* (1979)). Unlike the Huber estimator, the Hampel-Krasker estimator is not very sensitive to points with high leverage. A popular choice for the function ψ_σ is $\psi_\sigma(\cdot) = \psi^2_{HF}(\cdot; k_2) - \gamma(k_2)$, for appropriate constants k_2 and $\gamma(k_2)$, which correspond to Huber's Proposal 2 (Huber (1964)).

Let $\dot\psi(u) = \partial\psi(u)/\partial u$. For a general M-estimator defined by (3.1) with ψ_β and ψ_σ odd and even functions, respectively, we have $\Omega(\beta,\sigma) = \Omega = \mathrm{diag}(\Omega_{\beta,\beta}, \Omega_{\sigma,\sigma})$, where $\Omega_{\sigma,\sigma} = n\int \psi^2_\sigma(r)dF_0(r)$ and $\Omega_{\beta,\beta} = n\int s^2(x)g_1(x)xx^{\mathrm{T}}dK(x)$, with $g_1(x) = \int \psi^2_\beta\{rv(x)\}dF_0(r)$. Moreover $B(\beta,\sigma) = (1/\sigma)B$, where $B = \mathrm{diag}(B_{\beta,\beta}, B_{\sigma,\sigma})$, with $B_{\sigma,\sigma} = n\int r\dot\psi_\sigma(r)dF_0(r)$, $B_{\beta,\beta} = n\int s(x)v(x)g_2(x)xx^{\mathrm{T}}dK(x)$ and $g_2(x) = \int \dot\psi_\beta\{rv(x)\}$ $dF_0(r)$. Therefore, (1.3) can be written as $A^{\mathrm{T}}(\beta,\sigma) = \Omega^{-1}(\beta,\sigma)B(\beta,\sigma) = (1/\sigma)\Omega^{-1}B = (1/\sigma)A = (1/\sigma)\,\mathrm{diag}(A^{\mathrm{T}}_{\beta,\beta}, A_{\sigma,\sigma})$: in this special case, the matrix $A(\beta,\sigma)$ depends on σ only.

Suppose we are interested in making inference only about a scalar component β_j $(1 \le j \le p)$ of β. If we consider the Huber estimator we find that $g_1(x) = g_1 = \int \psi^2_{HF}(r;k_1)dF_0(r)$, $\Omega_{\beta,\beta} = ng_1\int xx^{\mathrm{T}}dK(x)$, $g_2(x) = g_2 = \int \dot\psi_{HF}(r;k_1)dF_0(r)$, $B_{\beta,\beta} = ng_2\int xx^{\mathrm{T}}dK(x)$ so that the matrix A is diagonal and $A_{\beta_j,\beta_j} = g_2/g_1$. Therefore, in this case, the adjusted quasi-profile loglikelihood for β_j coincides with Barndorff-Nielsen's one and has expression

$$(3.2)\ l_{QP}(\beta_j) = \frac{g_2}{g_1}\sum_{i=1}^n x_{ij}\int_c^{\beta_j}\frac{1}{\hat\sigma_b}\psi_{HF}\left(\frac{y_i - \hat\beta_{1b}x_{i1} - \cdots - bx_{ij} - \cdots - \hat\beta_{pb}x_{ip}}{\hat\sigma_b}; k_1\right)db,$$

where x_{ij} is the j-th element of the vector x_i and $\hat\beta_{qb}$, $q \neq j$, $\hat\sigma_b$ are the estimates for β_q, $q \neq j$, and σ when β_j is considered as known and set equal to b. For a Gaussian model the factor A_{β_j,β_j} is

$$\frac{\Phi(k_1) - \Phi(-k_1)}{2[k_1^2\Phi(-k_1) - k_1\phi(k_1) + \{\Phi(k_1) - 1/2\}]},$$

where $\Phi(\cdot)$ denotes the standard normal distribution and $\phi(\cdot)$ its density. Observe that, in general, the adjusted quasi-profile loglikelihood for a regression parameter coincides with Barndorff-Nielsen's one for any M-estimator for which $s(x) = v(x) = 1$ and ψ_β odd. The general expression for the factor g_2/g_1 is $\int \dot\psi_\beta(r)dF_0(r)/\int \psi^2_\beta(r)dF_0(r)$.

If we consider the Hampel-Krasker estimator we have $g_1(x) = \int \psi^2_{HF}(r||x||;k_1)$ $dF_0(r)$, $\Omega_{\beta,\beta} = n\int\{g_1(x)/||x||^2\}xx^{\mathrm{T}}dK(x)$, $g_2(x) = \int \dot\psi_{HF}(r||x||;k_1)dF_0(r)$, and $B_{\beta,\beta} = n\int g_2(x)xx^{\mathrm{T}}dK(x)$. Thus, in general, the $A_{\lambda,\tau}$ block of the matrix A is not null. The adjusted quasi-profile loglikelihood is given by

$$l_{QP}(\beta_j) = w\sum_{i=1}^n x_{ij}\int_c^{\beta_j}\frac{1}{||x_i||\hat\sigma_b}\psi_{HF}\left(||x_i||\frac{y_i - \hat\beta_{1b}x_{i1} - \cdots - bx_{ij} - \cdots - \hat\beta_{pb}x_{ip}}{\hat\sigma_b}; k_1\right)db.$$

Using (2.7), the constant w can be written as

$$(3.3)\qquad w = \frac{B_{\beta_j,\beta_j} - \xi^{\mathrm{T}}_{\beta_j}B^{-1}_{(-j)}\xi_{\beta_j}}{\Omega_{\beta_j,\beta_j} - 2\xi^{\mathrm{T}}_{\beta_j}B^{-1}_{(-j)}\eta_{\beta_j} + \xi^{\mathrm{T}}_{\beta_j}B^{-1}_{(-j)}\Omega_{(-j)}B^{-\mathrm{T}}_{(-j)}\xi_{\beta_j}},$$

where $B_{\beta j,\beta j}$ is the j-th diagonal element of B, ξ_{β_j} is the j-th column of the matrix B without its j-th element, $B_{(-j)}$ denotes the matrix B without the j-th column and the

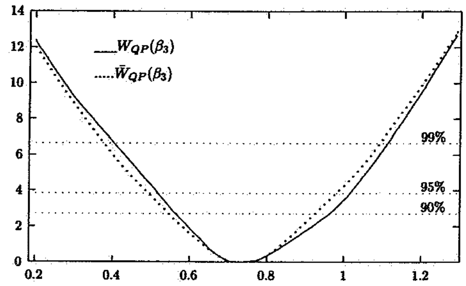

Fig. 1. Adjusted quasi-profile loglikelihood ratio function W_{QP} and Barndorff-Nielsen's quasi-profile loglikelihood ratio function $\bar{W}_{QP}$ for the parameter β_3 of the model from the Scottish Hill Races data.

j-th row and η_{β_j} is the j-th column of Ω without its j-th element. In this case, matrix B is symmetric.

In the usual formalization, one considers a linear model with fixed (not random) carriers $x_1, \dots, x_n$. In such a situation, for a general M-estimator defined by (3.1) with ψ_β and ψ_σ odd and even functions, respectively, we have that $\text{var}\{\Psi(\underline{y}; \beta, \sigma)\} = \Omega^* = \text{diag}(\Omega^*_{\beta,\beta}, \Omega^*_{\sigma,\sigma})$ and $-E\{\partial\Psi(\underline{y}; \beta, \sigma)/\partial(\beta, \sigma)^{\mathrm{T}}\} = (1/\sigma)B^*$, with $B^* = \text{diag}(B^*_{\beta,\beta}, B^*_{\sigma,\sigma})$, where $\Omega^*_{\beta,\beta} = \sum_i s^2(x_i) g_1(x_i) x_i x_i^{\mathrm{T}}$, $\Omega^*_{\sigma,\sigma} = n \int \psi^2_\sigma(r) dF_0(r)$, $B^*_{\beta,\beta} = \sum_i s(x_i) v(x_i) g_2(x_i)$ $x_i x_i^{\mathrm{T}}$ and $B^*_{\sigma,\sigma} = n \int \dot{\psi}_\sigma(r) r dF_0(r)$. Consequently, in the case of fixed carriers, l_{QP} for β_j, computed from the Huber estimator, has the same expression, given by (3.2), as in the case of random carriers. In contrast, to obtain $l_{QP}(\beta_j)$ from the Hampel-Krasker estimator when carriers are fixed we have to calculate the factor w by replacing matrix Ω and B in (3.3) with Ω^* and B^*, respectively.

To illustrate an application to some real data, Fig. 1 gives the plot of the adjusted quasi-profile loglikelihood ratio function

$$\text{(3.4)}\quad W_{QP}(\beta_3) = 2\{l_{QP}(\hat{\beta}_3) - l_{QP}(\beta_3)\}$$
$$= 2w \sum_{i=1}^{n} x_{i3} \int_{\beta_3}^{\hat{\beta}_3} \frac{1}{||x_i||\hat{\sigma}_b} \psi_{HF}\left(||x_i|| \frac{z_i - \hat{\beta}_{1b}x_{i1} - \hat{\beta}_{2b}x_{i2} - bx_{i3}}{\hat{\sigma}_b}; k_1 \right) db$$

for the parameter β_3 of the model $z_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i$, computed from the Scottish Hill Races data (discussed in Staudte and Sheather (1990), p. 265). The variables considered are the record time in minutes (z_i), the distance in miles (x_{i2}) and the climb in feet/100 (x_{i3}). The sample size is $n = 35$. Carriers are considered as fixed and a Gaussian model is assumed as the central one. The Hampel-Krasker estimator is used with $k_1 = 1.1$, $\psi_\sigma(\cdot) = \psi^2_{HF}(\cdot; k_2) - \gamma(k_2)$ and $k_2 = 0.6$. Figure 1 also gives the plot of Barndorff-Nielsen's quasi-profile loglikelihood ratio function $\bar{W}_{QP}(\beta_3) = 2\{\bar{l}_{QP}(\hat{\beta}_3) - \bar{l}_{QP}(\beta_3)\}$, computed from the same data. Moreover, Table 1 gives the results of a Monte Carlo experiment (based on 5000 trials) performed to assess the coverage error of the nominal $1 - \alpha$ confidence intervals for β_3, based on the adjusted quasi-profile loglikelihood ratio (3.4). For this experiment, the parameters β_1, β_2, β_3 are set equal to -4, 6 and 0.7, respectively. Errors e_i are generated from three different distributions: the standard normal $N(0,1)$, the standard normal contaminated by a $N(4,1)$ and the standard normal contaminated by a $N(0,25)$. We consider a contamination model of the form $F_\varepsilon = (1-\varepsilon)F + \varepsilon G$, where $G(\cdot)$ denotes the contaminating distribution. The contamination percentage ε is set at 5%. A simulation experiment has

Table 1. Empirical coverage probabilities of the confidence intervals for β_3 based on the adjusted quasi-profile loglikelihood.

	$1-\alpha$		
distribution	0.990	0.950	0.900
$N(0,1)$	0.991	0.954	0.903
$N(0,1)$ cont. by $N(4,1)$	0.991	0.952	0.898
$N(0,1)$ cont. by $N(0,25)$	0.992	0.953	0.894

also been made to evaluate coverage probabilities of confidence intervals from Barndorff-Nielsen's quasi-profile loglikelihood. Only the $N(0,1)$ distribution has been considered for the e_i's. For nominal 0.90, 0.95 and 0.99 coverage probabilities we obtained empirical coverages probabilities 0.903, 0.953 and 0.990, respectively.

4. Final remarks

The adjusted quasi-profile loglikelihood l_{QP} discussed in this paper represents an alternative to the quasi-profile loglikelihood $\bar{l}_{QP}$ suggested by Barndorff-Nielsen (1995), for inference about an arbitrary one-dimensional parameter of interest, based on unbiased estimating functions. In some particular situations, functions l_{QP} and $\bar{l}_{QP}$ coincide. Generally, our experience, based on the application example discussed in Section 3 and other simulation experiments not reported here, suggests that l_{QP} and $\bar{l}_{QP}$ perform very closely and allow inference with similar level of accuracy. However, as pointed out in Section 2, the use of l_{QP} is preferable because it avoids some conceptual and practical difficulties that arise in using $\bar{l}_{QP}$. In particular, from a practical point of view, our experience indicates that the computation of the loglikelihood ratio statistic from $\bar{l}_{QP}$ can be difficult even in relatively simple cases as the one considered in the application example of Section 3. Essentially, this occurs because when we are using Barndorff-Nielsen's approach, the estimating equation which gives the partial estimate $\bar{\lambda}_\tau$ is more complicated to solve numerically.

Another theoretical concern is with possible location adjustements designed to improve the asymptotic properties. The adjusted quasi-profile loglikelihood is based only on a scale adjustment of the estimating function for the scalar parameter of interest, aimed at obtaining a quasi-profile score function with properties similar to those of the ordinary profile score. A location adjustment is not necessary since the bias of the quasi-profile score $\tilde{\Psi}_\tau$ is already of the proper order $O(1)$. For the ordinary profile score function several additive adjustments that reduce its bias have been proposed, including Bartlett (1955), Cox and Reid (1987) and McCullagh and Tibshirani (1990), and the bias reducing properties of these adjustments are discussed further by Levin and Kong (1990) and DiCiccio *et al.* (1996). Following these approaches, it would be interesting to consider an additive adjustment to the adjusted quasi-profile score function, based on the first-order bias expansion (A.4). This would yield to a modified quasi-profile loglikelihood which, in view of (A.4), appears relatively easy to compute and which presumably could be of some importance in small samples.

Acknowledgements

The helpful comments of Alessandra Salvan and the referees are gratefully acknowledged. Partial support for this work was provided by the grants from Ministero dell'Università e della Ricerca Scientifica e Tecnologica, Italy.

Appendix

A Taylor expansion for the quasi-profile score function $\tilde{\Psi}_\tau$ about the true parameter value gives

$$\text{(A.1)} \qquad \tilde{\Psi}_\tau = \Psi_\tau + (\hat{\lambda}_\tau - \lambda)^a \Psi_{\tau/a} + \frac{1}{2}(\hat{\lambda}_\tau - \lambda)^{ab}\Psi_{\tau/ab} + O_p(n^{-1/2}),$$

where $(\hat{\lambda}_\tau - \lambda)^{ab} = (\hat{\lambda}_\tau - \lambda)^a(\hat{\lambda}_\tau - \lambda)^b$. Under the usual regularity conditions, which assure that the global estimator $\hat{\theta}$ is consistent and asymptotically normal, the summands on the right-hand side of (A.1) are $O_p(n^{1/2})$, $O_p(n^{1/2})$ and $O_p(1)$, respectively.

An expansion for $(\hat{\lambda}_\tau - \lambda)^a$ is obtained by expanding the estimating equation $\Psi_\lambda(\tau, \hat{\lambda}_\tau) = 0$ around the true parameter value and next by inverting the resulting expression into an asymptotic expansion for $(\hat{\lambda}_\tau - \lambda)^a$. We find

$$\text{(A.2)} \quad (\hat{\lambda}_\tau - \lambda)^a = \kappa^{b/a}\Psi_b + \frac{1}{2}\kappa^{d/a}\kappa^{e/b}\kappa^{f/c}\nu_{d/bc}\Psi_e\Psi_f + \kappa^{c/a}\kappa^{d/b}H_{c/b}\Psi_d + O_p(n^{-1}),$$

where $H_{c/b} = \Psi_{c/b} - \nu_{c/b}$. The sample size does not appear explicitly here but is incorporated into the random variables and their expected values. Thus, $\kappa^{b/a} = O(n^{-1})$, $\nu_{d/bc} = O(n)$ and $H_{c/b} = O_p(n^{1/2})$.

Now, substituting (A.2) into equation (A.1) and collecting terms of the same asymptotic order, we obtain

$$\text{(A.3)} \quad \begin{aligned} \tilde{\Psi}_\tau &= \Psi_\tau + \kappa^{b/a}\nu_{\tau/a}\Psi_b + \kappa^{b/a}H_{\tau/a}\Psi_b + \kappa^{d/a}\kappa^{c/b}H_{d/b}\nu_{\tau/a}\Psi_c \\ &\quad + \frac{1}{2}\kappa^{f/a}\kappa^{d/b}\kappa^{e/c}\nu_{\tau/a}\nu_{f/bc}\Psi_d\Psi_e + \frac{1}{2}\kappa^{c/a}\kappa^{d/b}\nu_{\tau/ab}\Psi_c\Psi_d + O_p(n^{-1/2}), \end{aligned}$$

where $H_{\tau/a} = \Psi_{\tau/a} - \nu_{\tau/a} = O_p(n^{1/2})$. An expansion for the mean of $\tilde{\Psi}_\tau$ is readily obtained by taking termwise expectations in (A.3). Then we find $E\{\tilde{\Psi}_\tau\} = m(\tau, \lambda) + O(n^{-1})$, where $m(\tau, \lambda)$ is of order $O(1)$ and is given by

$$\text{(A.4)} \quad \begin{aligned} m(\tau, \lambda) &= \kappa^{b/a}E\{\Psi_b\Psi_{\tau/a}\} + \nu_{\tau/a}\kappa^{c/a}\kappa^{d/b}E\{\Psi_d\Psi_{c/b}\} \\ &\quad + \frac{1}{2}\nu_{\tau/a}\nu_{d/bc}\kappa^{d/a}\kappa^{e/b}\kappa^{f/c}E\{\Psi_e\Psi_f\} + \frac{1}{2}\nu_{\tau/ab}\kappa^{d/a}\kappa^{c/b}E\{\Psi_d\Psi_c\}. \end{aligned}$$

The first-order bias expansion (A.4) is simple since it involves only the first two derivatives with respect to λ of the estimating functions. There is a formal similarity between equation (A.4) and the expression for the bias of the ordinary profile score function given in McCullagh and Tibshirani (1990).

The expansion for the scale adjustment (2.5) is more complicated. For the variance of the quasi-profile score function $\tilde{\Psi}_\tau$ we find

$$\text{(A.5)} \quad \text{var}\{\tilde{\Psi}_\tau\} = E\{\Psi_\tau^2\} + 2\kappa^{b/a}\nu_{\tau/a}E\{\Psi_\tau\Psi_b\} + \kappa^{c/a}\kappa^{d/b}\nu_{\tau/a}\nu_{\tau/b}E\{\Psi_c\Psi_d\} + O(1),$$

where the three summands on the right-hand side of (A.5) are of order $O(n)$. Its derivation is similar to that for the mean expansion (A.3) and is not given here. For the numerator of the scale adjustment (2.5) we find that

$$\frac{\partial}{\partial\tau}E\{\tilde{\Psi}_\tau\} = O(1) \tag{A.6}$$

and

$$-E\{\tilde{\Psi}_{\tau/\tau}\} = -\nu_{\tau/\tau} - \kappa^{b/a}\nu_{\tau/a}\nu_{b/\tau} + O(1), \tag{A.7}$$

where the two summands on the right hand side of (A.7) are of order $O(n)$. Putting equations (A.5), (A.6) and (A.7) together, we find that the adjusted quasi-profile score function has the form $w(\tau, \hat{\lambda}_\tau)\tilde{\Psi}_\tau$, where $w(\tau, \lambda)$ is given by (2.7).

References

Barndorff-Nielsen, O. E. (1995). Quasi profile and directed likelihoods from estimating functions, *Ann. Inst. Statist. Math.*, **47**, 461–464.

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

Bartlett, M. S. (1953*a*). Approximate confidence intervals, *Biometrika*, **40**, 12–19.

Bartlett, M. S. (1953*b*). Approximate confidence intervals. II. More than one unknown parameter, *Biometrika*, **40**, 306–317.

Bartlett, M. S. (1955). Approximate confidence intervals. III. A bias correction, *Biometrika*, **42**, 201–204.

Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion), *J. Roy. Statist. Soc. Ser. B*, **49**, 1–39.

Desmond, A. F. (1997). Optimal estimating functions, quasi-likelihood and statistical modelling, *J. Statist. Plann. Inference*, **60**, 77–121.

DiCiccio, T. J., Martin, M. A., Stern, S. E. and Young, G. A. (1996). Information bias and adjusted profile likelihoods, *J. Roy. Statist. Soc. Ser. B*, **58**, 189–203.

Godambe, V. P. (1976). Conditional likelihood and unconditional optimum estimating equations, *Biometrika*, **63**, 277–284.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics. The Approach Based on Influence Functions*, Wiley, New York.

Heyde, C. (1997). *Quasi-likelihood and Its Application. A General Approach to Optimal Parameter Estimation*, Springer, New York.

Huber, P. J. (1964). Robust estimation of a location parameter, *Ann. Math. Statist.*, **35**, 73–101.

Huber, P. J. (1973). Robust regression: Asymptotics, conjectures, and Monte Carlo, *Ann. Statist.*, **1**, 799–821.

Levin, B. and Kong, F. (1990). Bartlett's bias correction to the profile score function is a saddlepoint correction, *Biometrika*, **77**, 219–221.

Maronna, R. A., Bustos, O. and Yohai, V. J. (1979). Bias- and efficiency-robustness of general M-estimators for regression with random carriers, *Smoothing Techniques for Curve Estimation* (eds. T. Gasser and M. Rosenblatt), Lecture Notes in Math., **757**, 91–116, Springer, Berlin-Heidelberg-New York.

McCullagh, P. (1991). Quasi-likelihood and estimating functions, *Statistical Theory and Modelling* (eds. D. V. Hinkley, N. Reid and E. J. Snell), 265–286, Chapman and Hall, London.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed., Chapman and Hall, London.

McCullagh, P. and Tibshirani, R. (1990). A simple method for the adjustment of profile likelihoods, *J. Roy. Statist. Soc. Ser. B*, **52**, 325–344.

Staudte, R. G. and Sheather, S. J. (1990). *Robust Estimation and Testing*, Wiley, New York.