Default prior distributions from quasi- and quasi-profile likelihoods

Laura Ventura
Department of Statistical Sciences
University of Padua
Italy

Stefano Cabras, Walter Racugno
Department of Mathematics and Informatics
University of Cagliari
Italy

Abstract: In some problems of practical interest, a standard Bayesian analysis can be difficult to perform. This is true, for example, when the class of sampling parametric models is unknown or if robustness with respect to data or to model misspecifications is required. These situations can be usefully handled by using a posterior distribution for the parameter of interest which is based on a pseudo-likelihood function derived from estimating equations, i.e. on a quasi-likelihood, and on a suitable prior distribution.

The aim of this paper is to propose and discuss the construction of a default prior distribution for a scalar parameter of interest to be used together with a quasi-likelihood function. We show that the proposed default prior can be interpreted as a Jeffreys-type prior, since it is proportional to the square-root of the expected information derived from the quasi-likelihood. The frequentist coverage of the credible regions, based on the proposed procedure, is studied through Monte Carlo simulations in the context of robustness theory and of generalized linear models with overdispersion.

Keywords: Estimating equation, Jeffrey's prior, Matching prior, Nuisance parameter, Pseudo-likelihood, Robustness

Contents

Department of Statistical Sciences
Via Cesare Battisti, 241
35121 Padova
Italy

tel: +39 049 8274168

fax: +39 049 8274170

http://www.stat.unipd.it

Corresponding author:
Laura Ventura
tel: +39 049 827 4177
ventura@stat.unipd.it
http://homes.stat.unipd.it/
ventura/

Default prior distributions from quasi- and quasi-profile likelihoods

Laura Ventura
Department of Statistical Sciences
University of Padua
Italy

Stefano Cabras, Walter Racugno
Department of Mathematics and Informatics
University of Cagliari
Italy

Abstract: In some problems of practical interest, a standard Bayesian analysis can be difficult to perform. This is true, for example, when the class of sampling parametric models is unknown or if robustness with respect to data or to model misspecifications is required. These situations can be usefully handled by using a posterior distribution for the parameter of interest which is based on a pseudo-likelihood function derived from estimating equations, i.e. on a quasi-likelihood, and on a suitable prior distribution.

The aim of this paper is to propose and discuss the construction of a default prior distribution for a scalar parameter of interest to be used together with a quasi-likelihood function. We show that the proposed default prior can be interpreted as a Jeffreys-type prior, since it is proportional to the square-root of the expected information derived from the quasi-likelihood. The frequentist coverage of the credible regions, based on the proposed procedure, is studied through Monte Carlo simulations in the context of robustness theory and of generalized linear models with overdispersion.

Keywords: Estimating equation, Jeffrey's prior, Matching prior, Nuisance parameter, Pseudo-likelihood, Robustness

1   Introduction

The theory and the use of estimating equations and of the related quasi- and quasi-profile likelihood functions has received much attention in recent years; see, among others, Liang and Zeger (1995), Barndorff-Nielsen (1995), Desmond (1997), Heyde (1997), Adimari and Ventura (2002), Severini (2002), Wang and Hanfelt (2003), Jorgensen and Knudsen (2004), and Bellio *et al.* (2008). In addition, Lin (2006) and Greco *et al.* (2008) discussed the use of quasi-likelihood functions in the Bayesian setting. In this paper we derive a default prior distribution for a parameter of interest to be used together with a quasi- or a quasi-profile likelihood function derived from estimating functions. The corresponding posterior distribution allows to deal with

those problems in which standard Bayesian analyses are difficult to perform or not available.

Let $y = (y_1, \ldots, y_n)$ be a random sample of size n drawn from $F_\theta = F(y; \theta)$, $\theta \in \Theta \subseteq \mathbb{R}^p, p \geq 1$. Bayesian inference on θ, or a scalar component of θ, is usually based on the likelihood function $L(\theta) = L(\theta; y)$ and on a suitable prior $\pi(\theta)$ on θ, which is often derived throught default procedures, in order to avoid the difficulties related to the elicitation. However, complementary to likelihood-based procedures, in many situations of practical interest it is preferable to base inference on estimating equations and on the related quasi- and quasi-profile likelihood functions. This is true, for example, in the context of robustness theory when stability with respect to small deviations from the assumed model is required (see Hampel *et al.*, 1986), or in the context of generalized linear models with overdispersion or random effects (see McCullagh and Nelder, 1989). For inference about the parameter of interest, a quasi-likelihood function $L_Q(\theta)$ can be defined, with the standard limiting behaviour (McCullagh, 1991, Adimari and Ventura, 2002). In a Bayesian approach, if one treats $L_Q(\theta)$ as a true likelihood then a posterior distribution for θ can be considered, given by

$$\pi_Q(\theta|y) \propto \pi(\theta) L_Q(\theta) . \tag{1}$$

Some properties of $\pi_Q(\theta|y)$ follow from the used $L_Q(\theta)$. For instance, if $L_Q(\theta)$ is derived from a robust estimating equation, then the posterior distribution (1) provides inferential procedures which are reliable when the underlying distribution lies in a neighborhood of F_θ (Greco *et al.*, 2008).

Although the use of a pseudo-likelihood function, in the Bayesian inference cannot be considered as orthodox, it is actually widely shared. Papers which discuss the use of pseudo-likelihoods in the Bayesian perspective are Efron (1993), Bertolino and Racugno (1992, 1994), Raftery *et al.* (1996), Fraser *et al.* (2003), Cabras *et al.* (2006), Lin (2006), Chang and Mukerjee (2006), Greco *et al.* (2008) and Pauli *et al.* (2008). Papers which are more specifically related to the validation of a posterior distribution based on an alternative likelihood are Monahan and Boos (1992), Severini (1999), Lazar (2003), Schennach (2005), Racugno *et al.* (2008) and Ventura *et al.* (2008).

In this paper, we show that the proposed default priors to be used in (1) assume an expression analogue to the Jeffreys priors, since $\pi(\theta)$ turns to be proportional to the square root of the expected quasi-information. This information is related to the asymptotic variance of the quasi-maximum likelihood estimator. Such result agrees with the Welch and Peers (1963) solution in a model with a scalar parameter and with the Ventura *et al.* (2008) solution for a scalar parameter of interest in the presence of nuisance parameters. The proposed procedure is illustrated in the context of robust scale-location models and of log-linear models for count data with overdispersion. Monte Carlo studies are also provided in order to investigate the frequentist coverage of the credible regions based on $\pi_Q(\theta|y)$.

The remainder of the paper is organized as follows. Section 2 briefly reviews the background theory on estimating equations and on the related quasi- and quasi-profile likelihood functions. In Section 3 the proposed default priors are discussed.

Section 4 illustrates two examples and provides Monte Carlo studies. Section 5 is devoted to some final remarks.

2   Estimating equations and quasi-likelihoods

Let $\Psi_\theta = \Psi(y;\theta) = \sum_{i=1}^{n} \psi(y_i;\theta)$ be an unbiased estimating function for θ based on y, i.e. such that $E_\theta(\Psi_\theta) = 0$, where $\psi(\cdot)$ is given function and $E_\theta(\cdot)$ denotes expectation under F_θ. A general M-estimator of θ is defined as the solution $\tilde\theta$ of the estimating equation $\Psi_\theta = 0$ (see, e.g., Hampel *et al.*, 1986). The class of M-estimators includes well-known estimators: for example, if Ψ_θ is the likelihood score function, then $\tilde\theta$ is the maximum likelihood estimator. Under regularity conditions assumed throughout this paper (see e.g. Huber, 1981), M-estimators are consistent and asymptotically normally distributed with mean θ and covariance matrix

$$V(\theta) = M^{-1}\Omega(M^{-1})^{\mathsf{T}} , \qquad (2)$$

where $M = M(\theta) = -E_\theta(\partial\Psi_\theta/\partial\theta^{\mathsf{T}})$ and $\Omega = \Omega(\theta) = E_\theta(\Psi_\theta\Psi_\theta^{\mathsf{T}})$. Moreover, if $\psi(y;\theta)$ is bounded, the M-estimator $\tilde\theta$ is B-robust (see, e.g., Hampel *et al.*, 1986), i.e. it has bounded influence function.

A quasi-likelihood for θ based on Ψ_θ (see e.g. McCullagh, 1991) is given by

$$L_Q(\theta) = \exp\left\{ \sum_{i=1}^{n} \int_{c_0}^{\theta} A(t)\psi(y_i;t)\,dt \right\} , \qquad (3)$$

where $A(\theta)^{\mathsf{T}} = \Omega^{-1}M$ and c_0 is an arbitrary constant. When $p = 1$, a quasi likelihood for θ is usually easy to derive. However, when $p > 1$, $L_Q(\theta)$ exists if and only if the matrix M is symmetric. Matrix $A(\theta)$ in (3) allows to obtain a quasi-observed information with the usual relation with $V(\theta)$ and a quasi-likelihood ratio statistic $W_Q(\theta) = 2\{\ell_Q(\tilde\theta) - \ell_Q(\theta)\}$ with a standard χ_p^2 distribution, where $\ell_Q(\theta) = \log L_Q(\theta)$. Moreover, $A(\theta)$ does not modify the robustness properties of $\tilde\theta$ because it does not change its influence function. According to the results obtained in Heritier and Ronchetti (1994), such robustness properties will carry over to quasi-likelihood based inferential procedures.

Assume now θ partitioned as $\theta = (\tau, \lambda)$, where τ is a scalar parameter of interest and λ a $(p-1)-$dimensional nuisance parameter. Similarly, $\Psi_\theta = (\Psi_\tau, \Psi_\lambda)$, where $\Psi_\tau = \sum_{i=1}^{n} \psi_\tau(y_i;\tau,\lambda)$ and $\Psi_\lambda = \sum_{i=1}^{n} \psi_\lambda(y_i;\tau,\lambda)$ are the estimating functions corresponding to τ and λ, respectively. Let $\tilde\lambda_\tau$ be the M-estimate for λ derived from Ψ_λ for fixed τ, i.e. from $\Psi_\lambda = 0$. A quasi-profile likelihood for τ (Adimari and Ventura, 2002) is given by

$$L_{QP}(\tau) = \exp\left\{ \sum_{i=1}^{n} \int_{c_0}^{\tau} w(t,\tilde\lambda_t)\psi_\tau(y_i;t,\tilde\lambda_t)\,dt \right\} , \qquad (4)$$

where

$$w(\tau,\lambda) = \frac{-\nu_{\tau\tau} - \kappa^{ba}\nu_{\tau a}\nu_{b\tau}}{E_\theta(\Psi_\tau^2) + 2\nu_{\tau a}\kappa^{ba}E_\theta(\Psi_\tau\Psi_b) + \nu_{\tau a}\nu_{\tau b}\kappa^{ca}\kappa^{db}E_\theta(\Psi_c\Psi_d)} . \qquad (5)$$

In (5) index notation has been used: components of λ and Ψ_λ are denoted by λ^a and Ψ_a, respectively, and derivatives of Ψ_τ and Ψ_a with respect to the components of λ are $\Psi_{\tau a} = (\partial/\partial\lambda^a)\Psi_\tau$ and $\Psi_{ab} = (\partial/\partial\lambda^b)\Psi_a$, where $a, b = 1, \ldots, p-1$. Moreover, $\nu_{\tau\tau} = E_\theta((\partial/\partial\tau)\Psi_\tau)$, $\nu_{\tau a} = E_\theta(\Psi_{\tau a})$, $\nu_{ab} = E_\theta(\Psi_{ab})$, and κ^{ab} is the inverse matrix of $-\nu_{ab}$. The quasi-profile likelihood (4) has all the desired standard first-order properties. In particular, the quasi-maximum likelihood estimator $\tilde{\tau}$ is consistent and asymptotically normal, with mean τ and variance

$$V_{\tau\tau}(\tau) = (M^{\tau\tau})^2\Omega_{\tau\tau} + 2M^{\tau\lambda}\Omega_{\lambda\tau}(M^{\tau\tau})^\mathsf{T} + M^{\tau\lambda}\Omega_{\lambda\lambda}(M^{\tau\lambda})^\mathsf{T}\Big|_{\lambda=\tilde{\lambda}_\tau} \; , \qquad (6)$$

where $M^{\tau\tau}$ and $M^{\tau\lambda}$ denote the (τ,τ)-element and the (τ,λ)-block of the inverse of $M(\theta)$, respectively, and $\Omega_{\tau\tau}, \Omega_{\lambda\tau}$ and $\Omega_{\lambda\lambda}$ are the blocks of Ω. When the dimension of λ is large relative to n (see e.g. Di Ciccio $et~al.$, 1996, Liang and Zeger, 1995, Severini, 2002), it may be preferable to consider a modification of $L_{QP}(\tau)$ (see Bellio $et~al.$, 2008), which is given by

$$L_{QPM}(\tau) = L_{QP}(\tau) \exp\left\{ -\int_{c_0}^{\tau} w(t, \tilde{\lambda}_t)m(t, \tilde{\lambda}_t)\, dt \right\} \; , \qquad (7)$$

where $m(\tau, \lambda)$ is of order $O(1)$ and is given by

$$
\begin{aligned}
m(\tau, \lambda) &= \kappa^{ba}E_\theta(\Psi_b\Psi_{\tau a}) + \nu_{\tau a}\kappa^{ca}\kappa^{db}E_\theta(\Psi_d\Psi_{cb}) + \frac{1}{2}\nu_{\tau a}\nu_{dbc}\kappa^{da}\kappa^{eb}\kappa^{fc}E_\theta(\Psi_e\Psi_f) \\
&\quad + \frac{1}{2}\nu_{\tau ab}\kappa^{da}\kappa^{cb}E_\theta(\Psi_d\Psi_c) \; .
\end{aligned}
\qquad (8)
$$

Note that the first-order bias correction (8) involves only the first two derivatives with respect to λ of $(\Psi_\tau, \Psi_\lambda)$.

3   A default prior from quasi-likelihoods

Let us assume that the parameter of interest τ is a scalar function of θ, $\tau = \tau(\theta)$ say. If θ is scalar, we have $\tau = \theta$; otherwise, $\theta = (\tau, \lambda)$. Let us denote with $\ell^*(\tau) = \log L^*(\tau)$ a pseudo-loglikelihood function for τ, i.e. a function of the parameter of interest and of the data with properties similar to a genuine likelihood. In particular, in the context of this paper, $\ell^*(\tau)$ may be given by (3) when $\theta = \tau$, or by (4) or (7), when $\theta = (\tau, \lambda)$.

Let $\ell^*(\tau)$ be differentiable with derivatives $\ell^{*1}(\tau)$, $\ell^{*2}(\tau)$, $\ell^{*3}(\tau)$. For the expected values of these derivatives, we use the notation $\nu_2 = E_\theta(\ell^{*2}(\tau))$ and $\nu_{1,1} = E_\theta(\ell^{*1}(\tau)^2)$, which we assume of order $O(n)$. These assumptions are typically satisfied in practice, when the pseudo-score function $\ell^{*1}(\tau)$ behaves asymptotically like the sum of n independent random variables. Under these regularity conditions, $\ell^*(\tau)$ can be expanded in a Taylor series about its maxima, i.e. $\tilde{\tau}$, giving

$$\ell^*(\tau) = \ell^*(\tilde{\tau}) - \frac{1}{2}(\tau - \tilde{\tau})^2 j^*(\tilde{\tau}) + R_n \; ,$$

where $j^*(\tau) = -\ell^{*2}(\tau)$ is the pseudo-observed information and R_n is a remainder term of order $O_p(n^{-1/2})$.

In general, unlike as it happens with a genuine likelihood, the identity $\nu_2 + \nu_{1,1} = 0$, analogous to the information identity, does not hold. However, for the pseudo-likelihood functions considered in this paper, i.e. (3), (4) and (7), we have $\nu_2 + \nu_{1,1} = o(n)$ and

$$j^*(\tau) = i^*(\tau) + o_p(n) \ , \tag{9}$$

where $i^*(\tau) = \nu_2^2/\nu_{1,1}$ is the expected pseudo-information. The asymptotic variance of $\tilde{\tau}$ can be expressed as

$$i^*(\tau)^{-1} = \frac{\nu_{1,1}}{\nu_2^2} \ . \tag{10}$$

In the context of the quasi-likelihood and the quasi-profile likelihood, (10) reduces to (2) and (6), respectively.

A default prior distribution for τ can be obtained defining an uniform distribution that takes into account the geometry of the parameter space (see e.g. Ghosh *et al.*, 2007, Chapter 5). Let $\phi = \phi(\tau)$ be a smooth one-to-one transformation. Then

$$i_\phi^*(\phi_0) = i^*(\tau_0) \left(\frac{d\tau}{d\phi}\right)^2 \ ,$$

where τ_0 is a fixed value, $\phi_0 = \phi(\tau_0)$ and $i_\phi^*(\cdot)$ denotes the expected pseudo-information in the new parameterization. If $\phi(\tau)$ is chosen so that $(d\tau/d\phi)^2 \propto (i^*(\tau_0))^{-1/2}$, then $i_\phi^*(\phi_0)$ will be a constant. Thus, the metric for which a locally uniform prior is approximately noninformative can be obtained from the relationship $d\phi/d\tau \propto i^*(\tau)^{1/2}$. This, in turn, implies that the corresponding default prior invariant under one-to-one transformation of τ is

$$\pi(\tau) \propto i^*(\tau)^{1/2} \ . \tag{11}$$

This means that a parametrization invariant prior distribution for τ, derived from a pseudo-likelihood function, is proportional to the square root of the pseudo-expected information. The expression of this prior is analogous to the Jeffreys prior.

In the context of quasi-likelihood functions, when $\theta = \tau$ the default prior (11) leads to the posterior distribution

$$\pi(\theta|y) \propto V(\theta)^{1/2} L_Q(\theta) \ . \tag{12}$$

If $\theta = (\tau, \lambda)$, the posterior distribution based on a quasi-profile likelihood and (11) is given by

$$\pi(\tau|y) \propto V_{\tau\tau}(\tau)^{1/2} L_{QP}(\tau) \ , \tag{13}$$

or by

$$\pi(\tau|y) \propto V_{\tau\tau}(\tau)^{1/2} L_{QPM}(\tau) \ . \tag{14}$$

These results agree with the Welch and Peers (1963) solution in a model with a scalar parameter and with the Ventura *et al.* (2008) solution for a scalar parameter of interest in the presence of nuisance parameters based on pseudo-likelihoods.

4   Examples and simulation results

In this section we discuss two examples to illustrate the use of (13) and (14) and to evaluate the frequentist coverage of the corresponding credible sets. Interest is focused on quasi-profile likelihood functions derived from estimating equations in the context of a robust scale–location model and of a log-linear model for count data with overdispersion. In the first example, a proper posterior distribution for the parameter of interest is available, but it is not robust with respect to model misspecifications (see Greco *et al.*, 2008). In the second example a genuine likelihood function is not available and the comparison with a marginal posterior distribution is not possible. In both the simulation studies, the number of Monte Carlo replications is set to 5000.

4.1   Robust inference for scale-location model

Let $\theta = (\tau, \lambda)$, where $\tau \in \mathrm{I\!R}$ is a location parameter and $\lambda > 0$ a scale parameter. In this case we have $F(y; \tau, \lambda) = F_0((y - \tau)/\lambda)$, $\psi(y; \tau, \lambda) = \psi((y - \tau)/\lambda)$, $\Omega(\tau, \lambda) = n \int \psi(x)\psi(x)^\mathsf{T} dF_0(x) = \Omega$, $M(\tau, \lambda) = (n/\lambda) \int \dot\psi(x)\rho(x)^\mathsf{T} dF_0(x) = M/\lambda$, where $\rho(x)$ is the column vector $(1, x)$, and $A(\tau, \lambda)^\mathsf{T} = (1/\lambda)\Omega^{-1}M = A^\mathsf{T}/\lambda$.

Consider robust inference about τ when λ is the nuisance parameter. For a symmetric model, a well-known location and scale M-estimator is the Huber estimator (Hampel *et al.*, 1986, Section 4.2). The Huber's estimator for (τ, λ) is the solution of the unbiased estimating equations $\Psi_\tau = \sum_{i=1}^n \psi_k((y_i - \tau)/\lambda)$, where $\psi_k(x) = x \min\{1, k/|x|\}$ for some $k > 0$, and $\Psi_\lambda = \sum_{i=1}^n \psi_{k_1}((y_i - \tau)/\lambda)^2 - k_2$, where k_1 and k_2 are appropriate constants.

In general, when Ψ_τ and Ψ_λ are odd and even functions, respectively, the quasi-profile likelihood (4) for τ is given by (Adimari and Ventura, 2001, Bellio *et al.*, 2008)

$$L_{QP}(\tau) = \exp\left\{ A_{\tau\tau} \int_{c_0}^\tau \sum_{i=1}^n \frac{1}{\tilde\lambda_t} \psi_\tau\left(\frac{y_i - t}{\tilde\lambda_t}\right) dt \right\} , \tag{15}$$

where $\tilde\lambda_\tau$ is the estimate for λ derived from Ψ_λ when τ is considered as known. The factor $A_{\tau\tau} = M_{\tau\tau}/\Omega_{\tau\tau}$ is the diagonal element of the matrix A corresponding to τ. When the central model is the normal one and the Huber estimator is used, then the factor $A_{\tau\tau}$ has expression

$$A_{\tau\tau} = \frac{\Phi(k) - \Phi(-k)}{2(k^2\Phi(-k) - k\phi(k) + \Phi(k) - 1/2)} ,$$

where $\phi(x)$ and $\Phi(x)$ are the density and the cumulative density function of the standard normal model $N(0, 1)$, respectively.

It is straightforward to show that, in this case, the default prior (11) reduces to $\pi(\tau) \propto \tilde\lambda_\tau^{-1}$, so that the posterior distribution (13) is given by

$$\pi(\tau|y) \propto L_{QP}(\tau)\,\tilde\lambda_\tau^{-1} . \tag{16}$$

The normalizing constant of (16) is computationally demanding, because of the integral in (15), and it has been obtained using numerical integration after the

approximation of the posterior kernel $L_{QP}(\tau)\tilde{\lambda}_\tau^{-1}$ in a grid of 100 points with a fitted smooth spline on them.

Table 1 gives the frequentist coverages for 95% posterior credible intervals for the pseudo-posterior (16) compared with the classical confidence intervals for the mean of a normal distribution, which are also the posterior credible intervals with the usual Jeffreys' prior $\pi_J(\tau,\lambda) \propto \lambda^{-2}$. The numerical study is carried out under two different distributions: the $N(0,1)$, and the $N(0,1)$ contaminated by a $N(0,100)$. The contamination percentage is set at 10%. The Huber estimators are used with $k = k_1 = 1.5$ and $k_2 = 0$. From Table 1, it can be noted that inference based on $\pi(\tau|y)$ is quite satisfactory when the model is correctly specified and is preferable to $\pi_J(\tau|y)$ under departures from the standard normal model.

	\multicolumn{2}{c}{$N(0,1)$}	\multicolumn{2}{c}{$N(0,1)$ cont. by $N(0,100)$}
n	$\pi(\tau|y)$	$\pi_J(\tau|y)$	$\pi(\tau|y)$	$\pi_J(\tau|y)$
5	0.993	0.966	0.957	0.975
15	0.984	0.951	0.954	0.976

Table 1: Frequentist coverages of 95% posterior credible intervals for the location-scale model.

4.2 Overdispersion in count data

Let us consider a log-linear regression with overdispersion for count data (see Mc-Cullagh and Nelder, 1989). The responses y_i are realizations of independent random variables with mean $\mu_i = \exp(x_i^\mathsf{T}\beta)$, $\beta \in \mathbb{R}^p$, $p \geq 1$, $i = 1,\ldots,n$. We focus on the situation where the variance is assumed to be a quadratic function of the mean, i.e. $V_i = \mu_i(1+\alpha)$, with $\alpha > 0$, $i = 1,\ldots,n$.

An unbiased estimating function for β is the score function from the Poisson likelihood, given by

$$\Psi_\beta = \sum_{i=1}^{n}(y_i - \mu_i)\,x_i^\mathsf{T} \ .$$

An estimating function for α can be obtained from the method of moments (Lawless, 1987), that is

$$\Psi_\alpha = \sum_{i=1}^{n}\frac{(y_i - \mu_i)^2}{V_i} - (n-p) \ .$$

Assume that the parameter of interest τ is a scalar regression coefficient, i.e. $\tau = \beta_j\,(1 \leq j \leq p)$. In the presence of overdispersion, a genuine likelihood for τ is not available, but it is possibile to resort to a quasi-profile likelihood (see Bellio *et al.*, 2008). Let Ψ_τ and Ψ_λ be unbiased estimating functions for τ and λ, respectively, with $\tau = \beta_j$ and $\lambda = (\beta_1,\ldots,\beta_{j-1},\beta_{j+1},\ldots,\beta_d,\alpha)$. In this setting, for inference about τ we use (7), which accounts for the consequences of the overdispersion for inference on τ.

Since the assumptions on the variance are typical of a negative binomial model for the response, this distribution can be used for computing the adjustments in (7) and the prior (11).

We apply this procedure to the Ames Salmonella data (see Lawless, 1987, and Bellio *et al.*, 2008) and we use this design matrix to assess the coverage of posterior credible intervals in a Monte Carlo study. In this example y_i, $i = 1, \ldots, n$, represents the number of revertant colonies on a plate, and covariates are based on the dose level of quinoline on the plate. We assume

$$\log(\mu_i) = \beta_0 + \beta_1 x_i + \tau \log(x_i + 10) \ ,$$

for $i = 1, \ldots, 18$.

The prior distribution (11) does not have in this case an analytical form and it can be approximated by averaging the observed pseudo-information for τ on 100 parametric bootstrap samples from the negative binomial model, with τ chosen in a grid of points and λ fixed. This average is supposed to approximate $\hat{i}_{PS}(\tau)$ and so $\pi_{PS}(\tau)$ at point τ.

In order to evaluate the frequentist coverages of 95% posterior credible intervals based on (14), we ran a simulation study generating data under a negative binomial model. In particular, the data were simulated according to the same design and parameter values similar to those obtained from the Ames Salmonella data (see Bellio *et al.*, 2008). The frequentist coverage of credible intervals of $\pi(\tau|y)$ is 0.968, quite near to the nominal 95%. Other scenarios are showed in second and third rows of Table 2. We note that in all cases considered, despite the small dimension of the sample size, credible intervals of $\pi(\tau|y)$ have accurate frequentist coverages.

τ	$\lambda = (\beta_0, \beta_1, \alpha)$	$\pi(\tau|y)$
0.3	(1,0,2)	0.968
0	(1,0,0.01)	0.956
0	(1,0,0.05)	0.945

Table 2: Frequentist coverages of 95% posterior credible intervals in the log-linear model for count data.

5   Final remarks

In this paper a Jeffreys-type prior for a parameter of interest is derived and discussed for Bayesian inference based quasi- and quasi-profile likelihood functions. Other examples of pseudo-likelihoods for which (9) is satisfied, and for which the default prior can be derived, are the marginal, the conditional, the profile and its modified versions, the approximate conditional, the empirical likelihood functions; see, e.g., Barndorff-Nielsen (1995), DiCiccio *et al.* (1996), Severini (1998), and Adimari and Ventura (2002). For the marginal, the conditional and the modified profile likelihoods in Ventura *et al.* (2008) it is shown that these Jefrreys-type priors have also matching properties.

Acknowledgments

References

Adimari, G., Ventura, L. (2002). Quasi-profile loglikelihoods for unbiased estimating functions, *Ann. Inst. Statist. Math.*, **54**, 235–244.

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

Bellio, R., Greco, L., Ventura, L. (2008). Adjusted quasi-profile likelihoods from estimating functions, *J. Statist. Plan. Inf.*, **138**, 3059–3068.

Bertolino, F., Racugno, W. (1992). Analysis of the linear correlation coefficient using pseudo-likelihoods, *J. It. Statist. Soc.*, **1**, 33–50.

Bertolino, F., Racugno, W. (1994). Robust Bayesian analysis of variance and the χ^2-test by using marginal likelihoods, *The Statistician*, **43**, 191–201.

Cabras, S., Racugno, W., Ventura, L. (2006). Bayesian inference on the scalar skew-normal distribution, *Proceedings of COMPSTAT 2006, 17th Symposium of IASC-ERS*, 1373–1380.

Chang, H., Mukerjee, R. (2006). Probability matching propoerty of adjusted likelihoods, *Statist. Prob. Letters*, **76**, 838–842.

Desmond, A.F. (1997). Optimal estimating functions, quasi-likelihood and statistical modelling, *J. Statist. Plann. Inf.*, **60**, 77-104.

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

Efron, B. (1993). Bayes and likelihood calculations from confidence intervals, *Biometrika*, **80**, 3–26.

Greco, L., Racugno, W., Ventura L. (2008). Robust likelihood functions in Bayesian inference, *J. Statist. Plan. Inf.*, **138**, 1258–1270.

Fraser, D.A.S., Reid, N., Wong, A., Yun Yi, G.(2003). Direct Bayes for interest parameters, *Valencia 7*, 529–533.

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

Heritier, S., Ronchetti, E. (1994). Robust bounded-influence tests in general parametric models, *J. Amer. Statist. Assoc.*, **89**, 897–904.

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

Huber, P.J. (1981). *Robust Statistics*, Wiley, NewYork.

Jorgensen, B., Knudsen, S.J. (2004). Parameter orthogonality and bias adjustment for estimating functions, *Scand. J. Statist.*, **31**, 93-114.

Lazar, N.A. (2003). Bayesian empirical likelihood, *Biometrika*, **90**, 319–326.

Liang, K.Y., Zeger, S.L. (1995). Inference based on estimating functions in the presence of nuisance parameters. *Statist. Sci.*, **10**, 158-173.

Lin, L. (2006). Quasi Bayesian likelihood, *Statistical Methodology*, **3**, 444–455.

McCullagh, P. (1991). *Quasi-likelihood and estimating functions*, In: Hinkley DV, Reid N, Snell EJ (eds) Statistical theory and modelling, 265–286. Chapman and Hall, London.

McCullagh, P., Nelder, J.A. (1989). *Generalized linear models*, Chapman and Hall, London.

Monahan, J.F., Boos, D.D. (1992). Proper likelihoods for Bayesian analysis, *Biometrika*, **79**, 271–278.

Pauli, F., Racugno, W., Ventura, L. (2008). Composite likelihoods in the Bayesian inference: an application, *Atti della XLIV Riunione Scientifica della Societa Italiana di Statistica*, June 25-27 2008, CD.

Racugno, W., Salvan, A., Ventura, L. (2008). Pseudo-Bayesian analysis in regression models, *Atti della XLIV Riunione Scientifica della Societa Italiana di Statistica*, June 25-27 2008, CD.

Raftery, A.E., Madigan, D., Volinsky, C.T. (1996). Accounting for model uncertainty in survival analysis improves predictive performance, *Bayesian Statistics*, 323–349.

Severini, T.A. (1998). Likelihood functions for inference in the presence of a nuisance parameter, *Biometrika*, **85**, 507–522.

Severini, T.A. (1999). On the relationship between Bayesian and non-Bayesian elimination of nuisance parameters, *Statistica Sinica*, **9**, 713–724.

Severini, T.A. (2002). Modified estimating functions, *Biometrika*, **89**, 333- 343.

Schennach, S.M. (2005). Bayesian exponentially tilted empirical likelihood, *Biometrika*, **92**, 31–46.

Ventura, L., Cabras, S., Racugno, W. (2008), Prior distributions from pseudo-likelihoods in the presence of nuisance parameters, *J. Amer. Statist. Assoc.*, to appear.

Wang, M., Hanfelt, J.J. (2003). Adjusted profile estimating function, *Biometrika*, **90**, 845-858.

Welch, B., Peers, H.W. (1963). On formulae for confidence points based on integrals of weighted likelihoods, *J. Roy. Statist. Soc.* B, **25**, 318–329.

Working Paper Series
Department of Statistical Sciences, University of Padua

You may order paper copies of the working papers by emailing wp@stat.unipd.it

Most of the working papers can also be found at the following url: http://wp.stat.unipd.it

Department of Statistical Sciences
University of Padua
Italy