The Annals of Statistics
1994, Vol. 22, No. 1, 340–356

POTENTIAL FUNCTIONS AND CONSERVATIVE ESTIMATING FUNCTIONS

BY BING LI[1] AND PETER MCCULLAGH[2]

Pennsylvania State University and University of Chicago

A quasiscore function, as defined by Wedderburn and by McCullagh, frequently fails to have a symmetric derivative matrix. Such a score function cannot be the gradient of any potential function on the parameter space; that is, there is no "quasilikelihood." Without a likelihood function it is difficult to distinguish good roots from bad roots or to set satisfactory confidence limits. From a different perspective, a potential function seems to be essential in order to give the theory an approximate Bayesian interpretation. The purpose of this paper is to satisfy these needs by developing a method of projecting the true score function onto a class of conservative estimating functions. By construction, a potential function for the projected score exists having many properties of a log-likelihood function.

1. Introduction. An unbiased estimating function $g(\theta, y)$ is defined to be a function of the data y and parameter θ having zero mean for all θ. In other words, $E_\theta\{g(\theta, Y)\} = 0$ for all θ. One purpose of an estimating function is to produce an estimate $\hat{\theta}$ of the parameter from data y, the estimate being obtained as a root of the equation $g(\theta, y) = 0$. Consequently, if the parameter θ is p-dimensional, it is necessary at a minimum that the range of g be p-dimensional with nonvanishing derivative matrix.

A quasiscore function is a linear unbiased estimating function based only on the first two moments of the observations. Suppose that the $n \times 1$ random variable Y has mean vector μ and covariance matrix V. Both are known functions of the p-dimensional parameter θ, and $V(\cdot)$ is a positive definite matrix. The quasiscore function is

$$q(\theta, y) = \{\dot{\mu}(\theta)\}^T \{V(\theta)\}^{-1} \{y - \mu(\theta)\}, \tag{1}$$

where $\dot{\mu}$ is an $n \times p$ matrix with components $\partial\mu_i / \partial\theta_r$. It is easily verified that

$$i = \operatorname{cov}_\theta(q) = -E_\theta(\partial q / \partial\theta) = \dot{\mu}^T V^{-1} \dot{\mu} \tag{2}$$

This matrix plays the role of Fisher information exactly as in fully parametric inference. Under the usual kinds of limiting conditions the asymptotic covariance matrix of $\hat{\theta}$ is i^{-1}.

Received June 1992; revised August 1993.
[1]Research supported by NSF Grant DMS-93-06738.
[2]Research supported by NSF Grant NSF-91-01333.
AMS 1991 *subject classifications.* Primary 62J12; secondary 62A10, 62A15.
Key words and phrases. Conservative vector field, quasilikelihood, linear estimating function, potential function.

The quasiscore function (1) can be derived as a projection of the true score function, $s(\theta,y) = \partial \log p_\theta(y)/\partial\theta$, onto the linear space spanned by $\{y_1 - \mu_1(\theta), \ldots, y_n - \mu_n(\theta)\}$ [see McLeish (1984) and McLeish and Small (1992)]. The projection is accomplished using the inner product $\langle X, Y\rangle = E_\theta(XY)$. Since $\langle s_r, Y_i - \mu_i\rangle = \partial\mu_i/\partial\theta_r$ and $\langle Y_i - \mu_i, Y_j - \mu_i\rangle = V_{ij}$, a direct application of the projection formula gives (1). The derivation of (1) by projection of the true unknown score function onto the space of linear estimating functions is sometimes put forward as a kind of finite-sample optimality property inherited from closeness to the true unknown score function [see Godambe and Heyde (1987)].

Although the projection (1) has many properties of a likelihood score function [Wedderburn (1974) and McCullagh (1983)], there is an important property of the likelihood score function that the quasiscore function does not inherit. Unlike the likelihood score function, which is by definition the gradient of the log-likelihood, the projection (1) often is not a gradient of any potential function. Note that, although $E(\partial q/\partial\theta)$ is by (2) symmetrical, the derivative matrix $\partial q/\partial\theta$ may not be symmetrical even at $\hat{\theta}$. In general, therefore, there can be no "quasi-log-likelihood" $Q(\theta,y)$ such that $\partial Q/\partial\theta = q(\theta,y)$ [see McCullagh and Nelder (1989) and McCullagh (1991)]. See also Firth and Harris (1991) for a discussion about the impossibility of constructing deviances from the nonconservative quasiscore of a multiplicative error model. The nonexistence of such a potential function makes the comparison of parameter values awkward, particularly when the estimating function has multiple roots. A potential function, if it existed, could be used to distinguish local maxima from minima. From a different perspective, a potential function seems to be essential in order to give the theory an approximate Bayesian interpretation.

The purpose of this paper is to satisfy these needs by developing a theory of conservative estimating functions that are linear in the observations. By construction, a potential function or quasilikelihood exists having the estimating function as gradient vector.

2. Definition of conservative quasiscore. Let Θ be the p-dimensional parameter space and let $\mathcal{Y}$ be the n-dimensional sample space. Suppose that on $\mathcal{Y}$ is defined a class of probability distributions $\{P_\theta\colon \theta \in \Theta\}$ with densities p_θ, and on Θ is defined a prior distribution Π with density $\pi(\theta)$. Let $L^2(P_\theta\Pi)$ be the class of all estimating functions $g\colon \Theta \times \mathcal{Y} \to R^p$ such that $\int_\Theta \int_\mathcal{Y} g^T g \, dP_\theta \, d\Pi < \infty$. We shall use g^r, g^s and so on to denote components of an estimating equation g, and g_i, g_j and so on to denote different estimating equations.

Evidently, $L^2(P_\theta\Pi)$ is a vector space over the real constants. In other words, if λ_1 and λ_2 are scalars (not dependent on θ and y), then for g_1 and g_2 in $L^2(P_\theta\Pi)$, $\lambda_1 g_1 + \lambda_2 g_2$ is in $L^2(P_\theta\Pi)$. For each pair of elements g_1 and g_2 in $L^2(P_\theta\Pi)$, we define their inner product as $\langle g_1, g_2\rangle_\pi = E_\pi E_\theta(g_1^T g_2)$. It follows that $L^2(P_\theta\Pi)$ is complete in terms of the metric associated with the inner product; thus $L^2(P_\theta\Pi)$ is a Hilbert space. Now let $\mathcal{G}$ be the subclass of $L^2(P_\theta\Pi)$, which consists of functions that are linear in y, unbiased and conservative. By conservative we mean that there is a potential function Q_g such that $\partial Q_g/\partial\theta_s = g^s$. This is equivalent to the statement that the line integral of g over Θ is path independent. If g is continuously differentiable, an equivalent version is

$$\partial g^r/\partial\theta_s = \partial g^s/\partial\theta_r \quad \text{for } r,s = 1,\dots,p. \tag{3}$$

It is easy to verify that $\mathcal{G}$ is a linear manifold in $L^2(P_\theta\Pi)$. We shall assume that $\mathcal{G}$ is closed in terms of the metric associated with $\langle\cdot,\cdot\rangle_\pi$.

DEFINITION 1. Suppose that the true score function $s(\theta,y) = \partial l/\partial\theta$ is in $L^2(P_\theta\Pi)$. We call the projection of s onto $\mathcal{G}$ the conservative quasiscore, and its potential function the quasilikelihood function.

To demonstrate that there is no conflict of terminology here, it is necessary to show that if the quasiscore (1) happens to be conservative, then it is the conservative quasiscore. Otherwise the definition would give rise to a different quasilikelihood function.

PROPOSITION 1. *If q is in $\mathcal{G}$, then, independently of the choice of Π, q is the conservative quasiscore.*

PROOF. We write the projection of s onto $\mathcal{G}$ as q^*. It suffices to show that, if q is in $\mathcal{G}$, then $q = q^*$ a.e. $P_\theta\Pi$ for an arbitrary Π. Letting g be an arbitrary element in $\mathcal{G}$, then

$$\langle g,q\rangle_\pi = E_\pi E_\theta\left(q^Tg\right) = \sum_{r=1}^{p} E_\pi E_\theta\left(q^r g^r\right). \tag{4}$$

Since $g \in \mathcal{G}$, g^r is necessarily of the form $\{a_r(\theta)\}^T y + b_r(\theta)$, with $a_r(\theta)$ in R^n and $b_r(\theta)$ in R^1. So $E_\theta(g^r q^r) = (\partial\mu/\partial\theta_r)^T a_r(\theta)$. If the differentiation with respect to θ and the integration over $\mathcal{Y}$ commute, then this reduces to $E_\theta(q^r g^r) = E_\theta(s^r g^r)$. Hence the right-hand side of (4) is $\langle s,g\rangle_\pi$. Thus (4) implies that $\langle q - s,g\rangle_\pi = 0$, for all g in $\mathcal{G}$; that is, $q = q^*$ a.e. $P_\theta\Pi$. Note that the argument follows independently of the choice of Π. □

In applications, $\mathcal{G}$ is approximated by a finite-dimensional space $\mathcal{F}$, say, which is always closed. Let $\{g_1, \dots, g_m\}$ be a basis in $\mathcal{F}$. Then the projection of s onto $\mathcal{F}$ can be written explicitly as

$$q^*_{\mathcal{F}} = \left\{\langle s,g_i\rangle_\pi\right\}^T\left\{\langle g_i,g_j\rangle_\pi\right\}^{-1}\{g_i\}, \tag{5}$$

where $\{\langle s,g_i\rangle_\pi\}$ is an $m \times 1$ vector, $\{\langle g_i,g_j\rangle_\pi\}$ is an $m \times m$ matrix and $\{g_i\}$ is an $m \times 1$ vector with entries being p-dimensional estimating equations. Since a linear unbiased g must take the form

$$G(\theta)^T\{y - \mu(\theta)\} \tag{6}$$

for some $(n \times p)$-dimensional matrix $G(\theta)$, the inner products needed for the projection (5) are

$$\langle s,g\rangle_\pi = \operatorname{tr}\left\{E_\pi\left(G^T\dot{\mu}\right)\right\} \quad \text{and} \quad \langle g_1,g_2\rangle_\pi = \operatorname{tr}\left\{E_\pi\left(G_1^T V G_2\right)\right\}.$$

These inner products, similar to those for the quasiscore function, depend on P_θ only through the first two moments of Y.

3. Some properties of conservative quasiscore function. By construction, q^* is an unbiased estimating equation. Since $s - q^*$ is orthogonal to q^*, $\langle q^*, q^*\rangle_\pi = \langle s, q^*\rangle_\pi$. It follows then that

(7) $$\operatorname{tr}\left\{E\left(q^* q^{*T}\right)\right\} = -\operatorname{tr}\{E(\partial q^*/\partial\theta)\},$$

with expectation taken over both Y and θ. Note that the matrices $E(q^*q^{*T})$ and $-E(\partial q^*/\partial\theta)$ are not in general equal. The projection argument guarantees only symmetry of the derivative matrix and equality of the traces.

An optimum property of q^* is obtained in terms of the efficiency criterion $e(g)$ in the spirit of Godambe (1960) and Bhapkar (1972). Since θ is estimated as a root of g, it is desirable that the root be close to the true θ. From efficiency and power considerations it is equally desirable that $|g|$ should be as large as possible at all false parameter values. Bearing in mind that $g(\cdot)$ is random, these requirements are captured in the one-dimensional case by the efficiency criterion

$$e(g) = \frac{\{\operatorname{tr} E_\pi E_\theta(\partial g/\partial\theta)\}^2}{\operatorname{tr}\{E_\pi E_\theta(gg^T)\}}.$$

This is an efficiency criterion in the sense of Pitman (1949) and Noether (1955): it depends only on the local behavior of $g(\cdot)$, ignoring entirely the global nature of the estimating function. An optimal conservative estimating function is naturally defined as the one in $\mathcal{G}$ that maximizes this quantity.

We now demonstrate that q^* is efficient according to this criterion. Since $s - q^*$ is perpendicular to $\mathcal{G}$,

(8) $$\langle g, q^*\rangle_\pi = \langle g, s\rangle_\pi = -\operatorname{tr} E_\pi E_\theta\left(\frac{\partial g}{\partial\theta}\right).$$

By the Cauchy–Schwarz inequality,

(9) $$\frac{\langle g, q^*\rangle_\pi^2}{\langle g, g\rangle_\pi} \le \langle q^*, q^*\rangle_\pi.$$

Substitution of the last expression of (8) into the left-hand side of (9) gives $e(g)$, and by (7) the right-hand side of (9) becomes $e(q^*)$. Hence $e(g) \le e(q^*)$ for all g in $\mathcal{G}$.

4. Construction of the conservative quasiscore. We now present a method of constructing a finite-dimensional space, $\mathcal{G}_K$, say, of linear, unbiased and conservative estimating equations. We shall demonstrate, under certain conditions, that the projection of s onto $\mathcal{G}_K$ approximates the conservative quasiscore q^* if K is large. First, we need a necessary and sufficient condition for an estimating equation to be linear, unbiased and conservative. In what follows, $A(\theta) = (A_1(\theta), \dots, A_n(\theta))^T$ represents an n-dimensional column vector, $\dot{A} \equiv \{\partial A_i / \partial \theta_s\}$ represents the $\{n \times p\}$-dimensional matrix of the partial derivatives and so on.

LEMMA 1. *Let g be an estimating equation in $L^2(P_\theta \Pi)$ having continuous partial derivatives. Then g is linear, unbiased and conservative if and only if*

$$g = \dot{A}^T \{y - \mu(\theta)\} \quad \text{for some } A \text{ satisfying } \dot{A}^T \dot{\mu} = \dot{\mu}^T \dot{A}. \tag{10}$$

In this case, the potential function of g can be written as

$$\begin{aligned} Q_g(\theta, y) &= A^T(\theta)\{y - \mu(\theta)\} \\ &\quad + \int^\theta \left\{ \sum_{i=1}^n A_i(t) \frac{\partial \mu_i}{\partial t_1} \right\} dt_1 + \cdots + \left\{ \sum_{i=1}^n A_i(t) \frac{\partial \mu_i}{\partial t_p} \right\} dt_p, \end{aligned} \tag{11}$$

where the second term is the (path independent) indefinite line integral in the parameter space.

PROOF. Linearity and unbiasedness of g implies (6). By conservativeness and continuous differentiablity, g must also satisfy (3). Substituting (6) into (3), we obtain

$$\sum_{i=1}^n \frac{\partial G_{ri}}{\partial \theta_s}(y_i - \mu_i) - \sum_{i=1}^n G_{ri} \frac{\partial \mu_i}{\partial \theta_s} = \sum_{i=1}^n \frac{\partial G_{si}}{\partial \theta_r}(y_i - \mu_i) - \sum_{i=1}^n G_{si} \frac{\partial \mu_i}{\partial \theta_r}, \tag{12}$$

where G_{rj} is the (r,j)th element of G^T. Taking expectation (E_θ) of both sides of (12) gives $G^T \dot{\mu} = \dot{\mu}^T G$. Substituting this into (12), we have

$$\sum_{i=1}^n \frac{\partial G_{ri}}{\partial \theta_s}(y_i - \mu_i) = \sum_{i=1}^n \frac{\partial G_{si}}{\partial \theta_r}(y_i - \mu_i) \quad \text{or} \quad \sum_{i=1}^n \left(\frac{\partial G_{ri}}{\partial \theta_s} - \frac{\partial G_{si}}{\partial \theta_r} \right)(y_i - \mu_i) = 0. \tag{13}$$

Since (13) holds for all y in the n-dimensional sample space, $\partial G_{ri}/\partial \theta_s = \partial G_{si}/\partial \theta_r$ for all $r, s = 1, \dots, p$ and $i = 1, \dots, n$, and for all $\theta \in \Theta$. Hence, for each $i, (G_{i1}, \dots, G_{ip})^T$ is the gradient vector of some function $A_i(\theta)$. This proves the necessity. Sufficiency and expression (11) can be easily verified. □

For computational purposes it is convenient to consider the special case in which

$$\mu_i(\theta) = Q_i(\theta) e^{R_i(\theta)}, \tag{14}$$

for some polynomials $Q_i(\theta)$ of degree K_Q and $R_i(\theta)$ of degree K_R. Models of this type arise in many applications. In the examples of Section 6, $Q_i(\theta)$ is linear and $R_i(\theta)$ is constant. For log-linear models $Q_i(\theta) = 1$ and $R_i(\theta)$ is linear. We seek a solution to (10) in which $A_i(\theta)$ has the form

$$A_i(\theta) = P_i(\theta)e^{-R_i(\theta)}, \tag{15}$$

where $P_i(\theta)$ is a polynomial of degree K with coefficients $\mathbf{C}_i$ to be determined. Substitution of (15) and (14) into the second expression (10) gives

$$\begin{aligned}\sum_{i=1}^{n} &\left(\frac{\partial P_i}{\partial \theta_r} - P_i\frac{\partial R_i}{\partial \theta_r}\right)\left(\frac{\partial Q_i}{\partial \theta_s} + Q_i\frac{\partial R_i}{\partial \theta_s}\right)\\ &- \left(\frac{\partial P_i}{\partial \theta_s} - P_i\frac{\partial R_i}{\partial \theta_s}\right)\left(\frac{\partial Q_i}{\partial \theta_r} + Q_i\frac{\partial R_i}{\partial \theta_r}\right) = 0,\end{aligned} \tag{16}$$

for $r < s, r, s = 1, \ldots, p$. On the left-hand side of (16) are polynomials in θ with coefficients linear in $\mathbf{C} \equiv \{\mathbf{C}_1, \ldots, \mathbf{C}_n\}$, that is

$$\sum_{i_1,\ldots,i_p} L^{(r,s)}_{i_1,\ldots,i_p}(\mathbf{C})\theta_1^{i_1}\cdots\theta_p^{i_p} = 0, \qquad r < s,\ r,s = 1,\ldots,p,$$

where $i_1, \ldots, i_p$ run through the index set

$$I = \{(i_1, \ldots, i_p):\ i_1 \geq 0, \ldots, i_p \geq 0; i_1 + \cdots + i_p \leq K + K_Q + 2K_R - 2\},$$

and $L^{(r,s)}_{i_1,\ldots,i_p}(\mathbf{C})$ are linear functions of $\mathbf{C}$. Hence condition (16) is equivalent to a system of linear equations,

$$L^{(r,s)}_{i_1,\ldots,i_p}(\mathbf{C}) = 0, \qquad r < s,\ r,s = 1, \ldots, p, (i_1, \ldots, i_p) \in I. \tag{17}$$

The solution space of (17) corresponds to all possible polynomial n-tuples of degree not higher than K that satisfy (16). Thus, if we let $\mathcal{G}_K$ be

$$\{g:\ g = \dot{A}^T(y - \mu), A_i = P_i e^{-R_i}, (P_1, \ldots, P_n) \text{ satisfy } (16),\ \text{degree of } P_i\text{'s} \leq K\},$$

then a basis in $\mathcal{G}_K$, say, $\{g_1, \ldots, g_m\}$, can be obtained by solving (17). The projection of s onto $\mathcal{G}_K$ can then be calculated using (5). We write the projection as q_K^*.

We now justify the projection onto the finite-dimensional space $\mathcal{G}_K$ by showing that the projection q_K^* approximates q^* as K tends to infinity. For this purpose, we make some assumptions (possibly stronger than necessary) about Π and the functions in $\mathcal{G}$. We assume that there is a subset Θ_0 of Θ such that (i) $\Pi(\Theta_0) = 1$; (ii) if $\mathcal{G}^0 = \{g:\ g = G^T(y - \mu)\}$ is a subset of $\mathcal{G}$ and if, for each g in $\mathcal{G}^0$, the entries of G are analytic functions on Θ_0, then $\mathcal{G}^0$ is dense in $\mathcal{G}$ in terms of the metric associated with $\langle\cdot,\cdot\rangle_\pi$. These conditions are satisfied if, for instance, Θ_0 is a closed rectangle and, for each $g = G^T(y - \mu)$ in $\mathcal{G}$, the entries of G are squared-integrable with respect to Π.

THEOREM 1. *If μ has the form* (14), *and if the assumptions made in the last paragraph hold, then* $\|q_K^* - q^*\|_\pi \to 0$ *as* $K \to \infty$.

PROOF. Since q^* is in $\mathcal{G}$, there is a g in $\mathcal{G}^0$ such that $\|g - q^*\|_\pi < \varepsilon$ for an arbitrary $\varepsilon > 0$. Since g is in $\mathcal{G}^0$, it has the form (10) for some $A(\theta)$. Let $P_i^K(\theta)$ be the Kth-order Taylor approximation of $A_i e^{R_i}$ (whose Taylor expansion also has a convergence region that contains Θ_0). We write $P_i^K e^{-R_i}$ as A_i^K, $(A_1^K, \dots, A_n^K)^T$ as A^K and the estimating equation $(\partial A^K/\partial\theta)^T(y-\mu)$ as g_K. Evidently g_K is in $\mathcal{G}_K$. Since $\{P_i^K\}$ and $\{\partial P_i^K/\partial\theta_s\}$ converge uniformly on Θ_0 to $A_i e^{R_i}$ and $\partial(A_i e^{R_i})/\partial\theta_s$, as $K \to \infty$, and $\sup_{\Theta_0}\{e^{-R_i}\}$ and $\sup_{\Theta_0}\{|\partial e^{-R_i}/\partial\theta_s|\}$ are finite,

$$\begin{aligned}\sup_{\Theta_0}\left\{\left|\frac{\partial A_i}{\partial\theta_s} - \frac{\partial A_i^K}{\partial\theta_s}\right|\right\} &= \sup_{\Theta_0}\left\{\left|e^{-R_i}\left\{\frac{\partial(A_i e^{R_i})}{\partial\theta_s} - \frac{\partial P_i^K}{\partial\theta_s}\right\} + \frac{\partial e^{-R_i}}{\partial\theta_s}\left(A_i e^{R_i} - P_i^K\right)\right|\right\}\\ &\le \sup_{\Theta_0}\left\{\left|e^{-R_i}\right|\right\}\sup_{\Theta_0}\left\{\left|\frac{\partial(A_i e^{R_i})}{\partial\theta_s} - \frac{\partial P_i^K}{\partial\theta_s}\right|\right\}\\ &\quad + \sup_{\Theta_0}\left\{\left|\frac{\partial e^{-R_i}}{\partial\theta_s}\right|\right\}\sup_{\Theta_0}\left\{\left|A_i e^{R_i} - P_i^K\right|\right\} \to 0 \quad \text{as } K \to \infty.\end{aligned}$$

Choosing K so large that $\sup_{\Theta_0}\{|\partial A_i/\partial\theta_s - \partial A_i^K/\partial\theta_s|\} < \varepsilon_1$, we have

$$\|g_K - g\|_\pi = \left\|\left(\frac{\partial A}{\partial\theta} - \frac{\partial A^K}{\partial\theta}\right)^T(Y-\mu)\right\|_\pi \le \varepsilon_1\left\|J^T(Y-\mu)\right\|_\pi \equiv \varepsilon_1 C,$$

where J is an $n \times p$ matrix with entries equal to 1, and $C < \infty$. Since $g_K \in \mathcal{G}_K$, $\|q_K^* - q^*\|_\pi \le \|g_K - q^*\|_\pi \le \|g_K - g\|_\pi + \|g - q^*\|_\pi \le \varepsilon_1 C + \varepsilon \equiv \varepsilon_2$, where ε_2 can be arbitrarily small. □

The examples in Section 6 indicate that q^* can be approximated reasonably well by small K, as small as $K = 3$. If $\mu_i(\theta)$ is not of the form (14), we can approximate it by a polynomial $\nu_i(\theta)$ and then use the above projection method. This approximation introduces a bias into the estimating equation. To take the bias into account, the inner product should be modified as $\langle g_1, g_2\rangle_\pi = \operatorname{tr}\{E_\pi(G_1^T(V+B)G_2)\}$, where $B = (\mu-\nu)(\mu-\nu)^T$.

5. Asymptotic distribution of the quasilikelihood ratio test. Once the projection $q^*(\theta, y)$ is obtained, expression (11) gives the quasilikelihood function $Q^*(\theta, y)$ for some vector $A(\theta)$. We now outline an argument that leads to the asymptotic distribution (under θ) of the quasilikelihood ratio tests $2\{Q^*(\hat\theta, y) - Q^*(\theta, y)\}$ and $2\{Q^*(\hat\theta, y) - Q^*(\tilde\theta, y)\}$, where $\hat\theta$ is obtained by maximizing $Q^*(\theta, y)$ globally on Θ and $\tilde\theta$ by maximizing on a subset $\tilde\Theta$ of Θ, and θ is the true value.

Assume that, as n approaches infinity, $\dot A^T\dot\mu/n$ and $\dot A^T V\dot A/n$ converge to positive definite matrices and that $\hat\theta$ is consistent. Then, under regularity conditions [e.g., Cox and Hinkley (1974), page 294] it can be shown that

$$\sqrt{n}(\widehat{\theta}-\theta) \to_{\mathcal{L}} N\Big(0, n\big(\dot{A}^T\dot{\mu}\big)^{-1}\big(\dot{A}^TV\dot{A}\big)\big(\dot{A}^T\dot{\mu}\big)^{-1}\Big) \equiv N(0,\Sigma). \tag{18}$$

Taking the Taylor expansion of $2\{Q^*(\widehat{\theta},y) - Q^*(\theta,y)\}$ about $\widehat{\theta}$, we get

$$2\{Q^*(\widehat{\theta},y) - Q^*(\theta,y)\} = n(\widehat{\theta}-\theta)^T\Lambda(\widehat{\theta}-\theta) + o_p(1), \tag{19}$$

where Λ is the matrix $\dot{A}^T\dot{\mu}/n$. By (18), the quadratic form $\{\sqrt{n}(\widehat{\theta}-\theta)\}^T\Sigma^{-1} \times\{\sqrt{n}(\widehat{\theta}-\theta)\}$ is asymptotically distributed as χ_p^2 under θ. Therefore,

$$\{\sqrt{n}(\widehat{\theta}-\theta)\}^T\Lambda\{\sqrt{n}(\widehat{\theta}-\theta)\} \to_{\mathcal{L}} \sum_{i=1}^{p}\lambda_i Z_i^2, \tag{20}$$

where the Z_i's are independent $N(0,1)$ random variables and the λ_i's are the eigenvalues of the positive definite matrix $\Sigma\Lambda$ [Johnson and Kotz (1970), page 151]. By (19) and (20) we conclude that

$$2\{Q^*(\widehat{\theta},y) - Q^*(\theta,y)\} \to_{\mathcal{L}} \sum_{i=1}^{p}\lambda_i Z_i^2. \tag{21}$$

Notice that when the quasiscore function $q(\theta,y)$ happens to be conservative, $\Sigma\Lambda$ is the identity matrix, so that the limiting distribution reduces to χ_p^2.

Suppose now that θ of dimension p is partitioned into two components, $\theta = (\psi,\lambda)$, where ψ of dimension p_1 is the parameter of interest and λ of dimension $p_2 = p - p_1$ is a nuisance parameter. Now consider the composite null hypothesis H_0: $\psi = \psi_0$ versus the alternative H_A: $\psi \neq \psi_0$, in the presence of nuisance parameter λ. Let the matrices Σ and Λ be partitioned into

$$\begin{pmatrix}\Sigma_{\psi\psi} & \Sigma_{\psi\lambda}\\ \Sigma_{\lambda\psi} & \Sigma_{\lambda\lambda}\end{pmatrix} \quad\text{and}\quad \begin{pmatrix}\Lambda_{\psi\psi} & \Lambda_{\psi\lambda}\\ \Lambda_{\lambda\psi} & \Lambda_{\lambda\lambda}\end{pmatrix},$$

respectively, with $\Sigma_{\psi\psi}$ of dimension $p_1\times p_1$, and so on; let $(\widehat{\psi},\widehat{\lambda})$ be the global maximizer of $Q^*(\theta,y)$ and let $(\psi_0,\widetilde{\lambda})$ be the conditional maximizer under H_0. It then follows that

$$2\Big\{Q^*\big(\widehat{\psi},\widehat{\lambda},y\big) - Q^*\big(\psi_0,\widetilde{\lambda},y\big)\Big\} \to_{\mathcal{L}} \sum_{i=1}^{p_1}\tau_i Z_i^2, \tag{22}$$

where $\tau_1,\ldots,\tau_{p_1}$ are the eigenvalues of the matrix $(\Lambda_{\psi\psi} - \Lambda_{\psi\lambda}\Lambda_{\lambda\lambda}^{-1}\Lambda_{\lambda\psi})\Sigma_{\psi\psi}$. If $q(\theta,y)$ is conservative, then the limiting distribution in (22) reduces to $\chi_{p_1}^2$.

The percentage points of the limiting distributions in (21) and (22) can be calculated using the algorithms given by Griffiths and Hill (1985). The preceding argument also indicates that if the matrix Λ and Σ^{-1} are approximately equal, then the asymptotic distributions in (21) and (22) can be approximated by χ^2 distributions.

6. Examples. Two examples are presented to illustrate the application of the conservative quasiscore and its potential function. Both examples are assumed to have an underlying probabilistic model so that the present approach can be compared with maximum likelihood. In practice, the quasilikelihood approach is more useful when there is a plausible model for the first two moments, but the underlying probabilistic mechanism is unclear or not computationally tractable. The results are also compared with those obtained using the usual quasilikelihood approach, which does not yield likelihood functions in these examples.

EXAMPLE 1 (Voter transition probabilities). This example concerns the estimation of voter transition probabilities based only on the vote totals of each of two parties, C and L, say, in two successive elections. Let θ_1 be the probability that a voter who votes for party C in the first election also votes for C in the second election, and let X_1 be the number of such voters. Similarly, θ_2 is the probability that a voter who previously voted for L subsequently switches to C, and X_2 is the number of such voters. We only observe the vote totals m_1 for C and m_2 for L in the first election, and vote totals Y for C and m_1+m_2-Y for L in the second. With transition probabilities θ_1 and θ_2 being the parameters of interest, we condition on the vote totals of election 1 and assume that $X_1 \sim B(m_1, \theta_1)$ and $X_2 \sim B(m_2, \theta_2)$. This example is given in McCullagh and Nelder [(1989), page 337] to illustrate the nonconservativeness of the quasiscore function.

The quasilikelihood score function has components

$$(23)\quad q^1(\theta, Y) = \sum_i \frac{m_{i1}\{Y_i - \mu_{i1}(\theta)\}}{V_i(\theta)} \quad \text{and} \quad q^2(\theta, Y) = \sum_i \frac{m_{i2}\{Y_i - \mu_{i2}(\theta)\}}{V_i(\theta)},$$

where $\mu_i(\theta) = m_{i1}\theta_1 + m_{i2}\theta_2$ and $V_i(\theta) = m_{i1}\theta_1(1-\theta_1) + m_{i2}\theta_2(1-\theta_2)$. The functions (q^1, q^2) do not form a conservative vector field because $\partial q^1/\partial\theta_2 \neq \partial q^2/\partial\theta_1$.

We now construct the conservative quasiscore function as described in Section 4. We choose $\pi(\theta)$ to be uniform over the parameter space $(0,1)\times(0,1)$. Observe that $\mu_i(\theta)$ is of the form (14) with $Q_i(\theta) = m_{i1}\theta_1 + m_{i2}\theta_2$ and $R_i(\theta) = 0$. So according to (15) we let $A_i(\theta) = \sum C^i_{\mu\nu}\theta_1^\mu\theta_2^\nu$, where the summation is over the set $\{(\mu,\nu)\colon \mu \geq 0, \nu \geq 0, 0 \leq \mu+\nu \leq 3\}$. Thus the conservative condition (16) becomes $\sum_{i=1}^3\{m_{i2}(\partial P_i/\partial\theta_1) - m_{i1}(\partial P_i/\partial\theta_2)\} = 0$. This is a system of equations involving only the polynomials $\partial P_i/\partial\theta_1$ and $\partial P_i/\partial\theta_2$ of order not higher than 2. Equating the coefficients of these polynomials to zero, as described in Section 4, the linear system of equations (17) becomes

$$(24)\qquad \mathbf{AC} = \mathbf{0},$$

where $\mathbf{C}^T = (\mathbf{C}_1, \mathbf{C}_2, \mathbf{C}_3)$, with $\mathbf{C}_i = (C^i_{10}, C^i_{01}, C^i_{20}, C^i_{11}, C^i_{02}, C^i_{30}, C^i_{21}, C^i_{12}, C^i_{03})$, and

$\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3)$, with

$$\mathbf{A}_i = \begin{pmatrix} m_{i2} & -m_{i1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2m_{i2} & -m_{i1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & m_{i2} & -2m_{i1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3m_{i2} & -m_{i1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2m_{i2} & -2m_{i1} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & m_{i2} & -3m_{i1} \end{pmatrix}.$$

Notice that in this case the C^i_{00}'s do not appear in $\partial A_i(\theta)/\partial\theta_1$ and $\partial A_i(\theta)/\partial\theta_2$, so they are irrelevant for constructing $q^*(\theta,y)$, and hence do not appear in (24).

The solution space of (24) corresponds to all polynomials $P(\theta) = \{P_1(\theta), P_2(\theta), P_3(\theta)\}$ of order not higher than 3 that satisfy the conservativeness condition (16), which result in a basis $\{g_1, \ldots, g_{21}\}$ of the family $\mathcal{G}_3$. Projection of s onto $\mathcal{G}_3$ using (5) gives $q_3^*(\theta,y)$. Finally, we use the Newton–Raphson method to obtain the estimate $\widetilde{\theta}$. The covariance matrix of $\widetilde{\theta}$ is estimated by substituting $\widetilde{\theta}$ into Σ in expression (18). The results of the calculations using the usual quasilikelihood method (QL) and the likelihood method based on the conservative quasiscore (CQL) are summarized in Table 1.

To compare the efficiency of CQL and QL, it is fairer to evaluate the two covariance matrices Σ/n in (18) and i^{-1} in (2) at the same value of θ. Evaluating both at $\widetilde{\theta}$, we get

$$\Sigma(\widetilde{\theta})/n = \begin{pmatrix} 0.24217 & -0.22665 \\ -0.22665 & 0.23600 \end{pmatrix}, \qquad i^{-1}(\widetilde{\theta}) = \begin{pmatrix} 0.24212 & -0.22661 \\ -0.22661 & 0.23595 \end{pmatrix}.$$

The difference is negligible, indicating that there is little loss of efficiency by using CQL rather than QL. Furthermore, the information identity holds approximately for q_3^* at the estimate $\widetilde{\theta}$:

$$\left\{E_\theta\left(-\frac{\partial q_3^*}{\partial\theta}\right)\right\}_{\widetilde{\theta}} = \begin{pmatrix} 39.7 & 38.6 \\ 38.6 & 42.1 \end{pmatrix} \approx \begin{pmatrix} 38.6 & 38.1 \\ 38.1 & 42.3 \end{pmatrix} = \left\{E_\theta(q_3^* q_3^{*T})\right\}_{\widetilde{\theta}}.$$

Therefore the asymptotic distribution of the quasilikelihood ratio based on Q^* can be approximated fairly well by χ^2.

In Figure 1, we compare the true likelihood and the potential function of the conservative quasiscore using contour plots and perspective plots. The maximum likelihood estimate $(\widehat{\theta}_1, \widehat{\theta}_2) = (0.2, 1.0)$ is on the boundary of Θ, but the maximum of Q^* is in the interior of Θ. Nevertheless, the shapes of the two likelihood functions—in particular, the direction of the longer axis of the contours—are very alike.

EXAMPLE 2 (Nonhomogeneous Poisson process). Suppose that on the time interval $[0, T]$ is defined a Poisson process with intensity $\lambda(t) = \exp(\theta_1 + \theta_2 t)$,

TABLE 1
Comparison of QL and CQL for Example 1

Methods	Point estimates	Covariance matrices
QL	$\bar{\theta}_1 = 0.363$ $\bar{\theta}_2 = 0.837$	$\begin{pmatrix} 0.239 & -0.223 \\ -0.223 & 0.233 \end{pmatrix}$
CQL	$\tilde{\theta}_1 = 0.368$ $\tilde{\theta}_2 = 0.832$	$\begin{pmatrix} 0.242 & -0.227 \\ -0.227 & 0.236 \end{pmatrix}$

of which we observe only the numbers of events $Y_1, \dots, Y_n$ that fall in certain overlapping intervals $I_1, \dots, I_n$ contained in $[0, T]$. Based on the data $\{I_1, \dots, I_n; Y_1, \dots, Y_n\}$, we want to estimate θ_1 and θ_2. In the simulation, θ_1 and θ_2 are taken to be 0.7 and 0.5, T is 6.5 and $I_1, \dots, I_{10}$ are intervals with length $\Delta = 0.8$ and centered at $x_i = Ti/(n+1) \equiv \delta i = 0.5909i$, $i = 1, \dots, 10$. Two hundred random variables are simulated from the nonhomogeneous Poisson process $N(t)$. The numbers of events of the process $N(t)$ that occur in intervals $I_1, \dots, I_{10}$ are recorded to be $1, 5, 4, 5, 7, 8, 12, 11, 15, 27$.

The likelihood function is obtained by partitioning the 10 overlapping intervals $\{I_1, \dots, I_{10}\}$ into 19 consecutive and nonoverlapping intervals

$$\{J_1, \dots, J_{19}\} = \{I_1 \cap I_2^c, I_1 \cap I_2, I_2 \cap I_1^c \cap I_3^c, \dots, I_9 \cap I_8^c \cap I_{10}^c, I_9 \cap I_{10}, I_{10} \cap I_9^c\}.$$

Letting $Y_1^*, \dots, Y_{19}^*$ be the unobserved random variables representing the numbers of events that fall in $J_1, \dots, J_{19}$, we have

$$P_\theta\{Y_1 = y_1, \dots, Y_{10} = y_{10}\}$$
$$= \sum \prod_{i=1}^{19} \left\{ (y_i^*!)^{-1} \left(\int_{J_i} \exp(\theta_1 + \theta_2 t)\, dt \right)^{y_i^*} \exp\left(- \int_{J_i} \exp(\theta_1 + \theta_2 t)\, dt \right) \right\},$$

where the summation is over the set

$$\left\{ (y_1^*, \dots, y_{19}^*) \colon y_1^* + y_2^* = y_1, \dots, y_{18}^* + y_{19}^* = y_{10}, y_j^* \geq 0, j = 1, \dots, 19 \right\}. \tag{25}$$

For simplicity, we approximate the integrals $\int_{J_i} \exp(\theta_1 + \theta_2 t)\, dt$ by $c_i \times$ (the length of J_i), where c_i is the intensity at the center of J_i. The number of elements in the set (25) is about 45 million, so the computation involved is heavy, not to mention that the counts simulated are moderate. One can also obtain the likelihood function by considering this as a missing-data problem and using the EM algorithm [Dempster, Laird and Rubin (1977)], in which the complete data is $\{y_1^*, \dots, y_{19}^*\}$ and the observed data is $\{y_1, \dots, y_{10}\}$.

We now apply the QL approach. The first two moments of $Y_1, \dots, Y_{10}$ are given by

$$\mu_i(\theta) = \int_{I_i} \exp(\theta_1 + t\theta_2)\, dt \quad \text{and} \quad V_{ij}(\theta) = \int_{I_i \cap I_j} \exp(\theta_1 + t\theta_2)\, dt,$$

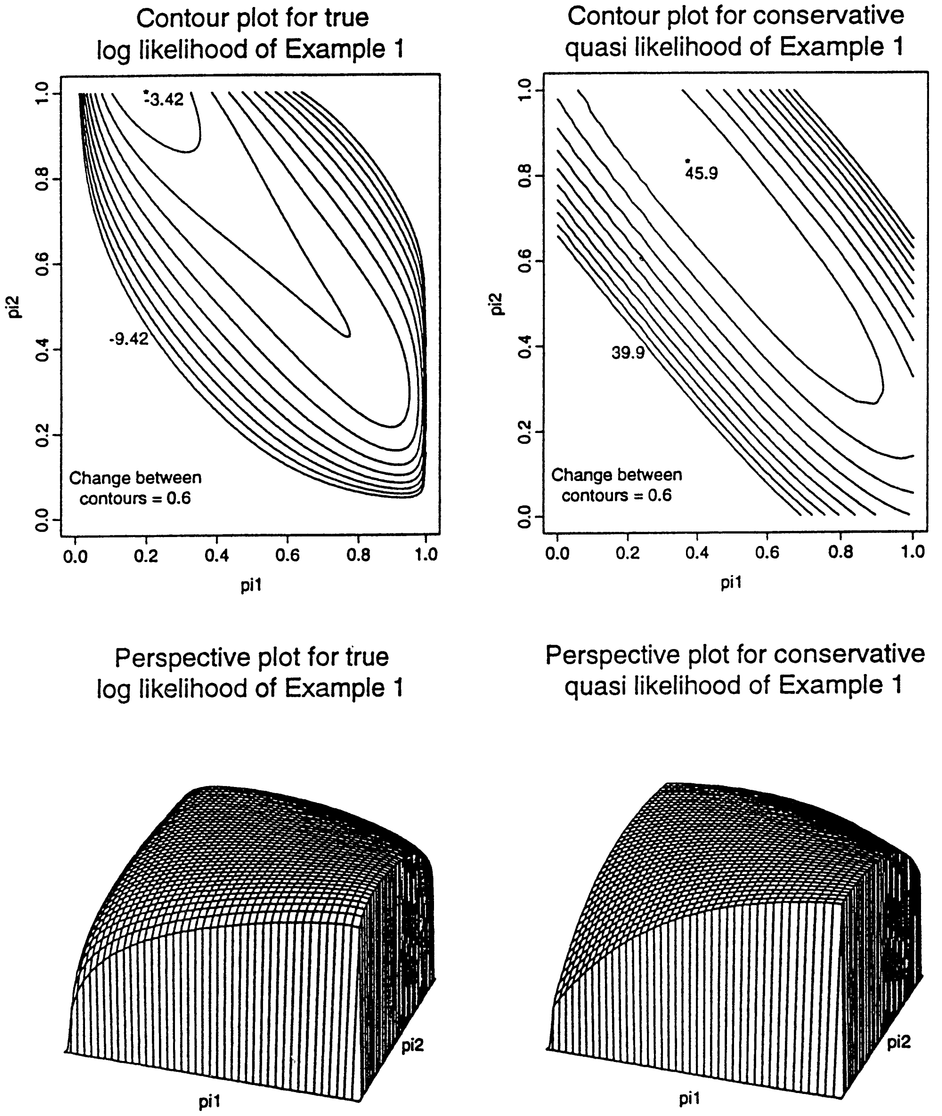

FIG. 1. *Comparison of true likelihood and conservative quasilikelihood for Example 1.*

which, as before, are approximated by the product of the intensity at the centers of the intervals and the length of those intervals; that is,

$$\begin{aligned}\mu_i(\theta) &= \Delta\exp(\theta_1 + i\delta\theta_2),\\ V_{ij}(\theta) &= (\Delta - |i-j|\delta)^+ \exp\Big[\theta_1 + \{i\delta + \chi(|i-j|=1)\delta/2\}\theta_2\Big],\end{aligned} \tag{26}$$

where χ is an indicator function. The quasilikelihood method is then applied to get the estimates $\overline{\theta}_1$ and $\overline{\theta}_2$ and their covariance matrix. As in the previous example, since the quasilikelihood score functions do not form a conservative system, no corresponding likelihood exists.

Finally, we apply the CQL approximation using $K = 3$. By calculations similar to Example 1, we obtain the linear system of equations (17) as $\mathbf{AC} = \mathbf{0}$, with

$$\mathbf{A} = \begin{pmatrix}\mathbf{B}_1 \cdots \mathbf{B}_{10}\\ \mathbf{A}_1 \cdots \mathbf{A}_{10}\end{pmatrix}, \qquad \mathbf{C}^T = (\mathbf{C}_1, \ldots, \mathbf{C}_{10}),$$

where

$$\mathbf{B}_i = (\mathbf{b}_i, \mathbf{0})_{10\times 10}, \qquad \mathbf{b}_i^T = (0, \ldots, 0, \overbrace{1}^{i\text{th}}, 0, \ldots, 0)_{1\times 10},$$

$$\mathbf{A}_i = \begin{pmatrix} 0 & i\delta & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 2i\delta & -1 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & i\delta & -2 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 3i\delta & -1 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2i\delta & -2 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & i\delta & -3 \end{pmatrix},$$

$$\mathbf{C}_i = (C^i_{00}, C^i_{10}, C^i_{01}, C^i_{20}, C^i_{11}, C^i_{02}, C^i_{30}, C^i_{21}, C^i_{12}, C^i_{03}).$$

The measure Π is taken to be uniform density over the region $(0,1] \times (0,1]$. We summarize and compare the results obtained by the three approaches in Table 2.

Again, we observe a remarkable agreement between the CQL and QL approaches, and, in this case, they both agree well with the MLE approach. The information identity holds approximately at the estimate $\widetilde{\theta}$:

$$\left\{E_\theta\left(-\frac{\partial q_3^*}{\partial\theta}\right)\right\}_{\widetilde{\theta}} = \begin{pmatrix}63.6 & 280.1\\ 280.1 & 1400.1\end{pmatrix} \approx \begin{pmatrix}60.6 & 259.9\\ 259.9 & 1301.6\end{pmatrix} = \left\{E_\theta\left(q_3^* q_3^{*T}\right)\right\}_{\widetilde{\theta}}.$$

In Figure 2, we compare the likelihood and the potential function of q_3^*. The parameters for both the contour plots and the perspective plots are on the region $[0,1] \times [0, 0.7]$.

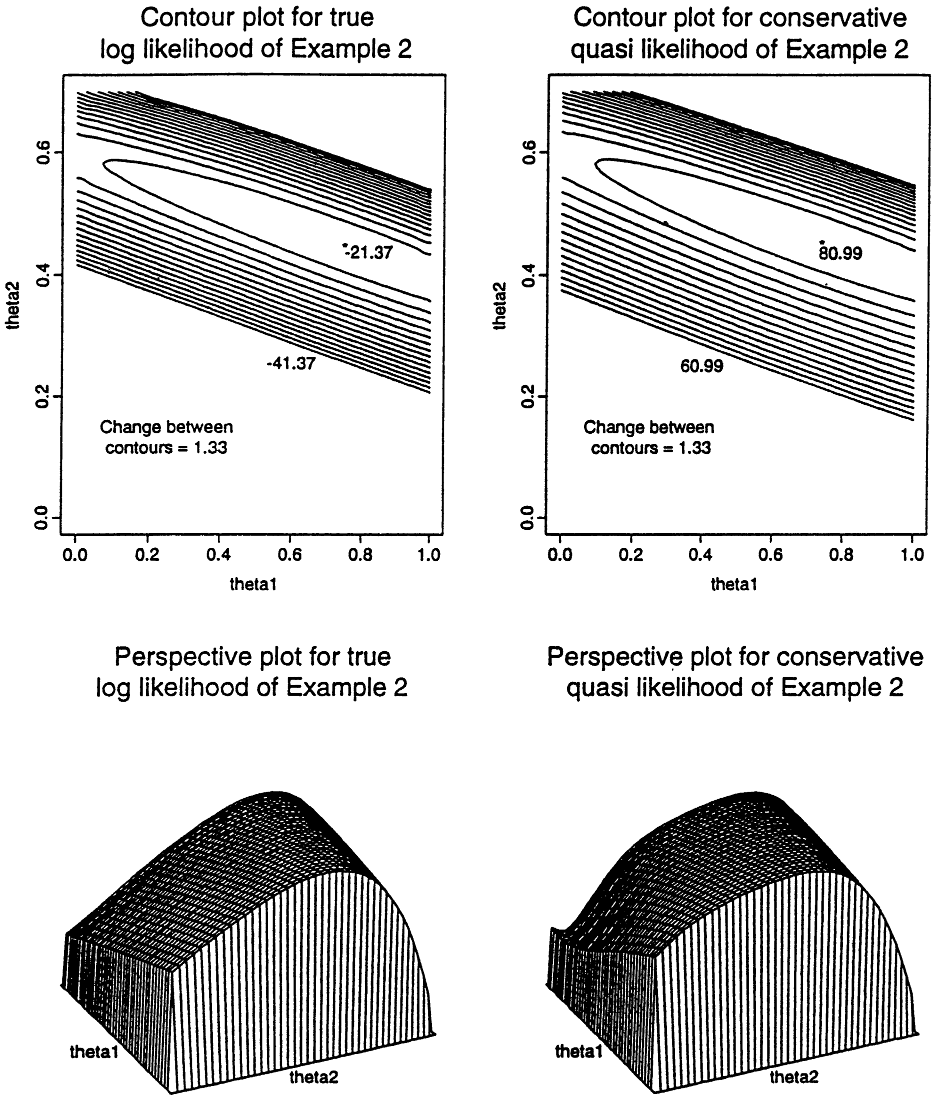

FIG. 2. *Comparison of true likelihood and conservative quasilikelihood for Example 2.*

TABLE 2
Comparison among MLE, QL and CQL for Example 2

Methods	Point estimates	Covariance matrices
MLE	$\widehat{\theta}_1 = 0.751$ $\widehat{\theta}_2 = 0.448$	$\begin{pmatrix} 0.153 & -0.0305 \\ -0.0305 & 0.00674 \end{pmatrix}$
QL	$\bar{\theta}_1 = 0.736$ $\bar{\theta}_2 = 0.452$	$\begin{pmatrix} 0.152 & -0.0304 \\ -0.0304 & 0.00676 \end{pmatrix}$
CQL	$\widetilde{\theta}_1 = 0.737$ $\widetilde{\theta}_2 = 0.452$	$\begin{pmatrix} 0.153 & -0.0306 \\ -0.0306 & 0.00679 \end{pmatrix}$

7. Discussion. In developing the quasilikelihood function presented here we have two purposes in mind. First, the quasilikelihood estimate is now obtained by maximizing a potential function, and confidence sets can be obtained from the contours of the potential functions. This avoids the ambiguity in defining the estimate as the solution to (1) when it has multiple solutions and when its potential function is not defined. Second, with a potential function at hand, it is possible to combine the prior knowledge of θ into the inference procedure. One possibility of doing so is to base inference on the "posterior distribution" $\exp\{Q^*(\theta, y) + \log \pi(\theta)\}$. When the sample size is moderate, $\pi(\theta)$ has a significant contribution to inference; when the sample size is large, the quasilikelihood $Q^*(\theta, y)$ dominates. The prior density $\pi(\theta)$ for Bayesian application need not be the same as the one we use for the projection of the score.

There is a degree of arbitrariness in the choice of the measure Π on the parameter space. In the examples, although only one choice has been presented, we have experimented with various choices of Π. In general, it seems that the effect of the choice of Π is very weak. This is perhaps to be expected since Π has no effect whatsoever in those cases where the quasiscore function (1) happens to be conservative. In this sense, the effect of the choice of Π seems to be less than the effect of the choice of prior in the Bayesian inference.

We now make some comparison between the projected likelihood ratio (PL) introduced by McLeish and Small (1992) and CQL. The PL is a function of two values of the parameter θ, so it needs a reference point to be used as a likelihood function. The CQL does not need a reference point; it does, however, depend on Π. In certain special cases, PL recovers a likelihood function for the mixture of random variables, whereas the CQL recovers a likelihood in the linear exponential family. The function $\mathrm{PL}(\theta_1, \theta_2)$ has the property that $E_{\theta_1}\mathrm{PL}(\theta_1, \theta_2) = 1$, which mimics a likelihood function, and is useful; this does not hold for exp(CQL).

The CQL can also be developed for quadratic or higher-order estimating equations [Jarrett (1984), Crowder (1987) and Godambe and Thompson (1989)] when the functional form of the higher moments of data, such as skewness and kurtosis, is available. The idea of projecting the likelihood score with respect to a prior in the parameter space also finds application in combining likelihood conditioned on different and nonnested sub-σ-fields of sample spaces, which we shall study further in separate research.

Although a conservative quasilikelihood estimate is in the theory not fully efficient among the family of linear and unbiased estimating functions in terms of the Loewner ordering of asymptotic covariance matrices, both examples indicate that the covariance matrices differ from their QL counterparts only by a negligible amount. They also indicate that the projection q^* may be approximated well using polynomials of low degree.

Acknowledgments. We wish to thank the referees and Editor for their helpful comments.

REFERENCES

BHAPKAR, V. P. (1972). On a measure of efficiency in an estimating equation. *Sankhyā Ser. A* **34** 467–472.

COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics* Chapman and Hall, London.

CROWDER, M. (1987). On linear and quadratic estimating functions. *Biometrika* **74** 591–597.

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood estimation from the EM algorithm (with discussion). *J. Roy. Statist. Soc. Ser. B* **39** 1–38.

FIRTH, D. and HARRIS, I. R. (1991). Quasi-likelihood for multiplicative random effects. *Biometrika* **78** 545–555.

GODAMBE, V. P. (1960). An optimum property of regular maximum likelihood estimation. *Ann. Math. Statist.* **31** 1208–1211.

GODAMBE, V. P. and HEYDE, C. C. (1987). Quasi-likelihood and optimal estimation. *Internat. Statist. Rev.* **55** 231–244.

GODAMBE V. P. and THOMPSON, M. E. (1989). An extension of quasi-likelihood estimation (with discussion). *J. Statist. Plann. Inference* **22** 137–152.

GRIFFITHS, P. and HILL, I. D. (1985). *Applied Statistics Algorithms*. Horwood, London.

JARRETT, R. G. (1984). Bounds and expansions for Fisher information when the moments are known. *Biometrika* **71** 101–113.

JOHNSON, N. L. and KOTZ, S. (1970). *Continuous Univariate Distributions* **2**. Wiley, New York.

MCCULLAGH, P. (1983). Quasi-likelihood functions. *Ann. Statist* **11** 59–67.

MCCULLAGH, P. (1991). Quasi-likelihood and estimating functions. In *Statistical Theory and Modelling: In Honour of Sir David Cox* (D. V. Hinkley, N. Reid and E. J. Snell, eds.) 265–268. Chapman and Hall, London.

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

MCLEISH, D. L. (1984). Estimation for aggregate models: the aggregate Markov chain. *Canad. J. Statist.* **12** 265–282.

MCLEISH, D. L. and SMALL, C. G. (1988). *The Theory and Applications of Statistical Inference Functions. Lecture Notes in Statist.* **44**. Springer, New York.

MCLEISH, D. L. and SMALL, C. G. (1992). A projected likelihood function for semiparametric models. *Biometrika* **79** 93–102.

NOETHER, G. E. (1955). On a theory of Pitman. *Ann. Math. Statist.* **26** 64–68.

PITMAN, E. J. G. (1949). Lecture notes on nonparametric statistical inference. Columbia Univ.

SMALL, C. G. and MCLEISH, D. L. (1988). Generalizations of ancillarity, completeness and sufficiency in an inference function space. *Ann. Statist.* **16** 534–551.

SMALL, C. G. and MCLEISH, D. L. (1989). Projection as a method for increasing sensitivity and eliminating nuisance parameters. *Biometrika* **76** 693–703.

WEDDERBURN, R. W. M. (1974). Quasi-likehood functions, generalized linear models, and the Gauss–Newton method. *Biometrika* **61** 439–447.

DEPARTMENT OF STATISTICS
PENNSYLVANIA STATE UNIVERSITY
326 CLASSROOM BUILDING
UNIVERSITY PARK, PENNSYLVANIA 16802

DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
5734 SOUTH UNIVERSITY AVENUE
CHICAGO, ILLINOIS 60637