The Annals of Statistics
1994, Vol. 22, No. 1, 300–325

EMPIRICAL LIKELIHOOD AND GENERAL ESTIMATING EQUATIONS

BY JING QIN AND JERRY LAWLESS[1]

University of Waterloo

For some time, so-called empirical likelihoods have been used heuristically for purposes of nonparametric estimation. Owen showed that empirical likelihood ratio statistics for various parameters $\theta(F)$ of an unknown distribution F have limiting chi-square distributions and may be used to obtain tests or confidence intervals in a way that is completely analogous to that used with parameteric likelihoods. Our objective in this paper is twofold: first, to link estimating functions or equations and empirical likelihood; second, to develop methods of combining information about parameters. We do this by assuming that information about F and θ is available in the form of unbiased estimating functions. Empirical likelihoods for parameters are developed and shown to have properties similar to those for parameteric likelihood. Efficiency results for estimates of both θ and F are obtained. The methods are illustrated on several problems, and areas for future investigation are noted.

1. Introduction. Likelihood is arguably the most important concept for inference in parameteric models. Recently it has also been shown to be useful in nonparametric contexts. For some time it has been used to obtain nonparametric estimates of distribution functions [e.g., Kaplan and Meier (1958), Vardi (1985). Recently Owen (1988, 1990, 1991), building on an earlier suggestion of Thomas and Grunkemeier (1975), has introduced an "empirical" likelihood ratio statistic for nonparametric problems. Owen has shown that the statistics have limiting chi-square distributions in certain situations, and has shown how to obtain tests and confidence limits for parameters, expressed as functionals $\theta(F)$ of an unknown distribution function F. Other asymptotic properties—and the possibility of correcting likelihood ratio statistics or their signed roots—have been studied by DiCiccio and Romano (1989), Hall (1990), DiCiccio, Hall and Romano (1989, 1991) and others.

Empirical likelihood, described in Section 2, provides likelihood ratio statistics for parameters by profiling a nonparametric likelihood; the approach is analoguous to that used for parameteric models, although it is computationally more complex. Owen (1990) showed that for independent and identically distributed (i.i.d.) data the approach applies to quite general parameters $\theta(F)$. Owen (1991) made extensions to linear regression problems, and Kolaczyk

Received September 1991; revised December 1992.

[1]This research partially supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

AMS 1991 *subject classification*. Primary 62E20.

Key words and phrases. Asymptotic efficiency, auxiliary information, empirical likelihood, estimating equations, parameteric likelihood, semiparametric models, testing hypotheses, Wilks' theorem.

(1992) and Owen (1992) have made further extensions to generalized linear and projection pursuit regression. Although further investigation of this methodology is needed, especially in small to moderate size samples, it appears to provide a valuable approach to tests and interval estimation in nonparametric or distribution-free contexts.

Our objective in this paper is twofold: first, to link estimating equations and empirical likelihood; second, to develop methods of combining information about parameters. We achieve both for i.i.d. data as follows. Consider d-variate i.i.d. random variables $x_1, \dots, x_n$ with unknown distribution function F, and a p-dimensional parameter θ associated with F. We assume that information about θ and F is available in the form of $r \geq p$ functionally independent unbiased estimating functions, that is functions $g_j(x, \theta)$, $j = 1, 2, \dots, r$, such that $E_F\{g_j(x, \theta)\} = 0$. In vector form, we have

$$g(x, \theta) = \left(g_1(x, \theta), \dots, g_r(x, \theta)\right)^{\tau}, \tag{1.1}$$

where

$$E_F\{g(x, \theta)\} = 0. \tag{1.2}$$

We will show how to use such information to estimate θ and F, in conjunction with empirical likelihood.

When $r = p$, our methods are the same as those of Owen (1988, 1990) and provide (empirical) likelihood-based methods of interval estimation for parameters with $\theta(F)$. Our main interest, however, is in the case where $r > p$. This allows us to deal with the combination of pieces of information about a distribution. For illustration we introduce some examples that we will return to later.

EXAMPLE 1. Sometimes we have information relating the first and second moments of a variable [e.g., Godamble and Thompson (1989) and McCullagh and Nelder (1989)]. For example, let $y_1, \dots, y_n$ be i.i.d., univariate observations with mean θ, and suppose that it is known that $E(y^2) = m(\theta)$, where $m(\cdot)$ is a known function. Our aim is to estimate θ. The information about F can be expressed in the form (1.1), (1.2) by taking

$$g(y, \theta) = \left(y - \theta, y^2 - m(\theta)\right)^{\tau}.$$

EXAMPLE 2. Let $(x_1, y_1), \dots, (x_n, y_n)$ be bivariate i.i.d. observations with $E(x_i) = E(y_i) = \theta$. In this case we can take $g((x_i, y_i), \theta) = (x_i - \theta, y_i - \theta)$. A somewhat similar problem is when $E(x_i) = c$ is known and $E(y_i) = \theta$ is to be estimated, in which case we would have $g((x_i, y_i), \theta) = (x_i - c, y_i - \theta)$. Such problems are common in survey sampling [e.g., Kuk and Mak (1989) and Chen and Qin (1993)].

EXAMPLE 3. Several authors have considered nonparametric estimation of a distribution F when information about certain functionals of F is available.

For example, Haberman (1984) and Sheehy (1988) consider estimation of $F(x)$ based on an i.i.d. sample $x_1,\dots,x_n$ when it is known that $E_F\{T(x)\} = a$, for some specified function $T(\cdot)$. Our methods deal with this by taking $g(x) = T(x) - a$; that is, $r = 1$ and the dimension p of θ is 0.

We show in this paper that empirical likelihood may be brought to bear on problems such as these. The basic idea is to maximize an empirical likelihood (see Section 2) subject to constraints provided by (1.2). We show how estimators both of parameters θ and the underlying distribution F may be obtained and determine asymptotic normal distributions for the estimators. We also demonstrate that empirical likelihood ratio statistics for parameters have asymptotic χ^2 distributions. All of these results parallel closely similar results for parametric likelihood inference. Section 2 reviews Owen's (1988, 1990) definition of empirical likelihood and the concept of optimal estimating functions. Section 3 presents our methods and associated asymptotic results; it is also shown that our method combines information in the form of estimating functions in an optimal way. Section 4 gives two other asymptotic results. Section 5 presents several examples, and Section 6 discusses some additional points. Outlines of proofs of the results in Sections 3 and 4 are provided in the Appendix. Further details are given in a technical report available from the authors.

2. Definition of empirical likelihood and optimal estimating functions. We first outline empirical likelihood as discussed by Owen (1988, 1990). Let $x_1,x_2,\dots,x_n$ be i.i.d. observations from a d-variate distribution F having mean μ and nonsingular covariance matrix. The empirical likelihood function is

$$L(F) = \prod_{i=1}^{n} dF(x_i) = \prod_{i=1}^{n} p_i, \tag{2.1}$$

where $p_i = dF(x_i) = \Pr(X = x_i)$. Only distributions with an atom of probability on each x_i have nonzero likelihood, and (2.1) is maximized by the empirical distribution function $F_n(x) = n^{-1}\sum_{i=1}^{n} I(x_i < x)$. The empirical likelihood ratio is then defined as $R(F) = L(F)/L(F_n)$, and it is easily shown that this may be written as

$$R(F) = \prod_{i=1}^{n} np_i. \tag{2.2}$$

We remark that formulas here and elsewhere in this paper do not require that the x_i's be distinct.

Suppose now that we want to estimate a parameter $\theta = T(F)$: For simplicity we consider the mean μ of F. To obtain confidence regions for μ, we define the profile empirical likelihood ratio function

$$R_E(\mu) = \sup\left\{ \prod_{i=1}^{n} np_i \,\middle|\, p_i \geq 0,\ \sum_{i=1}^{n} p_i = 1,\ \sum_{i=1}^{n} p_i x_i = \mu \right\}. \tag{2.3}$$

As noted by Owen (1988, 1990), a unique value for the right-hand side of (2.3) exists, provided that μ is inside the convex hull of the points $x_1, \dots, x_n$. An explicit expression for $R_E(\mu)$ can be derived by a Lagrange multiplier argument: the maximum of $\Pi_{i=1}^n np_i$ subject to the constraints $p_i \geq 0, \Sigma_{i=1}^n p_i = 1$ and $\Sigma_{i=1}^n p_i x_i = \mu$ is attained when

$$p_i = p_i(\mu) = n^{-1}\left\{1 + t^\tau(x_i - \mu)\right\}^{-1}, \tag{2.4}$$

where $t = t(\mu)$ is a $d \times 1$ vector given as the solution to

$$\sum_{i=1}^{n} \left\{1 + t^\tau(x_i - \mu)\right\}^{-1}(x_i - \mu) = 0. \tag{2.5}$$

Since $\Pi_{i=1}^n p_i$ is maximized unconditionally by F_n, it follows that $R_E(\mu)$ is maximized with respect to μ at $\widehat{\mu} = \overline{x}$ and that

$$R_E(\mu) = \prod_{i=1}^{n} \left\{1 + t^\tau(x_i - \mu)\right\}^{-1}. \tag{2.6}$$

The empirical likelihood ratio statistic is $W_E(\mu) = -2\log R_E(\mu)$, that is,

$$W_E(\mu) = 2\sum_{i=1}^{n} \log\left\{1 + t^\tau(x_i - \mu)\right\}. \tag{2.7}$$

Owen (1988, 1990) has proved under mild conditions that if $\mu = \mu_0$, then $W_E(\mu_0)$ converges in distribution to $\chi^2_{(d)}$ as $n \to \infty$. Approximate α-level confidence regions for μ may therefore be obtained as the set of points μ such that $W_E(\mu) \leq c_\alpha$, where c_α is defined such that $\Pr(\chi^2_{(d)} \leq c_\alpha) = \alpha$. Profile empirical likelihood ratio statistics for subsets of $\mu = (\mu_1, \dots, \mu_d)$ can also be used to obtain confidence regions for subsets of the parameters, in the usual way. Owen (1990) has shown that the preceding approach applies to quite general parameters $\theta(F)$, including multidimensional M-estimates. Owen (1991, 1992) and Kolaczyk (1992) have extended the methodology to a broad range of regression problems involving linear, generalized linear and projection pursuit models.

Let $g_1(x, \theta), \dots, g_r(x, \theta)$ be a set of functionally independent estimating functions, as in (1.1) and (1.2), where θ is a p-dimensional parameter. If $r = p$, estimates $\widehat{\theta}(x)$ may be obtained as roots of the corresponding estimating equation $g(x, \theta) = 0$. More generally, if the $g_j(x, \theta)$'s are r specified functions, we may consider the class of p-dimensional estimating functions

$$\Psi = \left\{\psi(x, \theta) \middle| \psi(x, \theta) = A(\theta)g(x, \theta)\right\}, \tag{2.8}$$

where $A(\theta)$ is a $p \times r$ matrix of real functions. In estimating function theory [e.g., see Goldambe and Heyde (1987)] an estimating function $\psi^*(x,\theta) \in \Psi$ is called optimum in Ψ if the estimator $\widehat{\theta}$ from $\psi^*(x,\theta) = 0$ has minimum asymptotic variance.

3. Main results. We assume that $x_1, \ldots, x_n$ are i.i.d. observations from an unknown distribution F, that there is a p-dimensional parameter θ associated with F and that information about θ and F is available in the form of $r \geq p$ functionally independent unbiased estimating functions, as described by (1.1) and (1.2). We apply empirical likelihood to this framework by maximizing (2.1) subject to restrictions

$$p_i \geq 0, \qquad \sum_i p_i = 1, \qquad \sum_i p_i g(x_i, \theta) = 0. \tag{3.1}$$

For a given θ, a unique maximum exists, provided that 0 is inside the convex hull of the points $g(x_1,\theta), \ldots, g(x_n,\theta)$. The maximum may be found via Lagrange multipliers. Let

$$H = \sum_i \log p_i + \lambda\left(1 - \sum_i p_i\right) - nt^\tau \sum_i p_i g(x_i, \theta),$$

where λ and $t = (t_1, t_2, \ldots, t_r)^\tau$ are Lagrange multipliers. Taking derivatives with respect to p_i, we have

$$\frac{\partial H}{\partial p_i} = \frac{1}{p_i} - \lambda - nt^\tau g(x_i, \theta) = 0,$$
$$\sum_i p_i \frac{\partial H}{\partial p_i} = n - \lambda = 0 \quad \Rightarrow \quad \lambda = n$$

and

$$p_i = \left(\frac{1}{n}\right) \frac{1}{1 + t^\tau g(x_i, \theta)}, \tag{3.2}$$

with the restriction from the third part of (3.1) that

$$0 = \sum_i p_i g(x_i, \theta) = \frac{1}{n} \sum_i \frac{1}{1 + t^\tau g(x_i, \theta)} g(x_i, \theta), \tag{3.3}$$

from which (see below) t can be determined in terms of θ.

Note that it is necessary that $0 \leq p_i \leq 1$, which implies that t and θ must satisfy $1 + t^\tau g(x_i, \theta) \geq 1/n$ for each i. For fixed θ, let $D_\theta = \{t\colon 1 + t^\tau g(x_i, \theta) \geq 1/n\}$; D_θ is convex and closed, and it is bounded if 0 is inside the convex hull of the $g(x_i, \theta)$'s. Moreover,

$$\frac{\partial}{\partial t}\left\{\frac{1}{n} \sum_{i=1}^n \frac{1}{1 + t^\tau g(x_i, \theta)} g(x_i, \theta)\right\} = -\frac{1}{n} \sum_{i=1}^n \frac{g(x_i, \theta) g^\tau(x_i, \theta)}{\{1 + t^\tau g(x_i, \theta)\}^2}$$

is negative definite for t in D_θ, provided that $\sum_{i=1}^n g(x_i,\theta)g^\tau(x_i,\theta)$ is positive definite. By the inverse function theorem, $t = t(\theta)$ is thus a continuous differentiable function of θ.

The (profile) empirical likelihood function for θ is now defined as

$$L_E(\theta) = \prod_{i=1}^n \left\{ \left(\frac{1}{n}\right) \frac{1}{1 + t^\tau(\theta) g(x_i,\theta)} \right\}.$$

Since $\Pi_{i=1}^n p_i$ is maximized for $p_i = n^{-1}$ in the absence of the parameteric constraints we define, analogous to (2.6), the empirical log-likelihood ratio

$$l_E(\theta) = \sum_{i=1}^n \log\left[1 + t^\tau(\theta) g(x_i,\theta)\right]. \tag{3.4}$$

Obviously (2.4) and (2.5) are special cases of (3.2) and (3.3), given by $g(x_i,\mu) = x_i - \mu$, and $2^{-1}W_E(\mu)$ from (2.7) is a special case of (3.4).

We may minimize $l_E(\theta)$ to obtain an estimate $\tilde{\theta}$ of the parameter θ, called the maximum empirical likelihood estimate (MELE). In addition, this yields estimates $\tilde{p}_i$, from (3.2), and an estimate for the distribution function F, as

$$\tilde{F}_n(x) = \sum_{i=1}^n \tilde{p}_i I(x_i < x). \tag{3.5}$$

When $r = p$ it is easily seen that $\tilde{\theta} = \hat{\theta}$ maximizes $l_E(\theta)$, where $\hat{\theta}$ is the solution to the estimating equations $\sum_{i=1}^n g(x_i,\theta) = 0$. In addition, $\tilde{p}_i = n^{-1}$ and (3.5) is the empirical cumulative distribution function. The empirical log-likelihood ratio $l_E(\theta)$ covers as special cases (2.7) and similar statistics for other problems considered by Owen (1990, 1991, 1992) and Kolaczyk (1992).

When $r > p$ and when profile empirical likelihoods are wanted, computational issues arise as to the best ways to obtain $\tilde{\theta}$ and profiles of $l_E(\theta)$. We discuss this in Section 5, where we consider specific examples. The remainder of this section presents first-order asymptotic properties of $\tilde{\theta}$, $\tilde{F}_n(x)$ and the empirical log likelihood ratio statistics. Proofs for the various propositions are given in the Appendix.

In the following, we use $\|\cdot\|$ to denote Euclidean norm.

LEMMA 1. *Assume that* $E[g(x,\theta_0)g^\tau(x,\theta_0)]$ *is positive definite,* $\partial g(x,\theta)/\partial\theta$ *is continuous in a neighborhood of the true value* θ_0, $\|\partial g(x,\theta)/\partial\theta\|$ *and* $\|g(x,\theta)\|^3$ *are bounded by some integrable function* $G(x)$ *in this neighborhood, and the rank of* $E[\partial g(x,\theta_0)/\partial\theta]$ *is* p. *Then, as* $n \to \infty$, *with probability* 1 $l_E(\theta)$ *attains its minimum value at some point* $\tilde{\theta}$ *in the interior of the ball* $\|\theta - \theta_0\| \le n^{-1/3}$, *and* $\tilde{\theta}$ *and* $\tilde{t} = t(\tilde{\theta})$ *satisfy*

$$Q_{1n}(\tilde{\theta},\tilde{t}) = 0, \qquad Q_{2n}(\tilde{\theta},\tilde{t}) = 0, \tag{3.6}$$

where

$$Q_{1n}(\theta,t) = \frac{1}{n}\sum_i \frac{1}{1+t^\tau g(x_i,\theta)} g(x_i,\theta), \tag{3.7}$$

$$Q_{2n}(\theta,t) = \frac{1}{n}\sum_i \frac{1}{1+t^\tau g(x_i,\theta)} \left(\frac{\partial g(x_i,\theta)}{\partial\theta}\right)^\tau t. \tag{3.8}$$

THEOREM 1. *In addition to the conditions of Lemma* 1, *we assume that* $\partial^2 g(x,\theta)/\partial\theta\,\partial\theta^\tau$ *is continuous in* θ *in a neighborhood of the true value* θ_0. *Then if* $\|\partial^2 g(x,\theta)/\partial\theta\,\partial\theta^\tau\|$ *can be bounded by some integrable function* $G(x)$ *in the neighborhood, then*

$$\sqrt{n}(\tilde\theta - \theta_0) \to N(0,V), \qquad \sqrt{n}(\tilde t - 0) \to N(0,U),$$
$$\sqrt{n}(\tilde F_n(x) - F(x)) \to N(0,W(x)),$$

where

$$\tilde F_n(x) = \sum_{i=1}^n \tilde p_i 1(x_i < x),$$

$$\tilde p_i = \left(\frac{1}{n}\right)\frac{1}{1+\tilde t^\tau g(x_i,\tilde\theta)},$$

$$V = \left[E\left(\frac{\partial g}{\partial\theta}\right)^\tau (Egg^\tau)^{-1} E\left(\frac{\partial g}{\partial\theta}\right)\right]^{-1},$$

$$W(x) = F(x)(1-F(x)) - B(x)UB^\tau(x),$$

$$B(x) = E\{g(x_i,\theta_0)I(x_i < x)\},$$

$$U = [E(gg^\tau)]^{-1}\left\{I - E\left(\frac{\partial g}{\partial\theta}\right) V E\left(\frac{\partial g}{\partial\theta}\right)^\tau [E(gg^\tau)]^{-1}\right\}$$

and $\tilde\theta$ *and* $\tilde t$ *are asymptotically uncorrelated.*

We can use Theorem 1 to get approximate confidence limits for θ or $F(x)$. The asymptotic variance V of $\sqrt{n}(\tilde\theta - \theta_0)$ is consistently estimated by

$$\left[\left\{\sum_{i=1}^n \tilde p_i \frac{\partial g(x_i,\tilde\theta)}{\partial\theta}\right\}^\tau \left\{\sum_{i=1}^n \tilde p_i g(x_i,\tilde\theta) g^\tau(x_i,\tilde\theta)\right\}^{-1} \left\{\sum_{i=1}^n \tilde p_i \frac{\partial g(x_i,\tilde\theta)}{\partial\theta}\right\}\right]^{-1}$$

or by the same expression with the $\tilde p_i$'s, replaced by n^{-1}.

We give some addition properties of empirical likelihood methods in the following corollaries to Theorem 1. We assume throughout that the conditions in Theorem 1 hold.

COROLLARY 1. *When $r > p$, the asymptotic variance $V = V_r$ of $\sqrt{n}(\tilde{\theta} - \theta)$ cannot decrease if an estimating equation is dropped.*

COROLLARY 2. *The MELE $\tilde{\theta}$ based on $g_1(x, \theta), \dots, g_r(x, \theta)$ is fully efficient in the sense that it has the same asymptotic variance as the optimal estimator obtained from the class of $p \times 1$ estimating equations that are linear combinations of $g_1(x, \theta), \dots, g_r(x, \theta)$; see (2.8).*

The proofs of Corollaries 1 and 3 are sketched in the Appendix. Corollary 2 is obtained by direct comparison of V in Theorem 1 with the asymptotic covariance matrix of the estimator obtained from the class of $p \times 1$ estimating equations based on $g_1(x, \theta), \dots, g_r(x, \theta)$ [see, e.g., McCullagh and Nelder (1989), page 341] or by noting the equivalence of MELE's based on equivalent sets of estimating functions (i.e., sets which are in 1–1 correspondence) and applying Corollary 1. See Section 6 for additional information.

We know that when the number of estimating equations and parameters are equal, the score equations are optimal [see Godambe and Heyde (1987)]. Corollary 3 shows that if p of the r estimating functions $g_j(x, \theta)$ are actually the score functions, then, as seems obvious, the covariance matrix V_r for $\sqrt{n}(\tilde{\theta} - \theta)$ is the same as that for the MLE, $\sqrt{n}(\hat{\theta} - \theta)$. However, $\tilde{F}_n(x)$ of (3.5) is generally more efficient than the empirical c.d.f. $F_n(x)$.

COROLLARY 3. *If we know the distribution of x up to parameter, let $g = (h_1^\tau, h_2^\tau)$, where*

$$h_1 = \big(g_1(x, \theta), \dots, g_p(x, \theta)\big)^\tau = \left(\frac{\partial \log f(x, \theta)}{\partial \theta_1}, \dots, \frac{\partial \log f(x, \theta)}{\partial \theta_p}\right)^\tau,$$
$$h_2 = \big(g_{p+1}(x, \theta), \dots, g_r(x, \theta)\big)^\tau$$

and x is assumed to have density $f(x, \theta)$, so that h_1 is the score. Then

$$V_r = V_p, \qquad W_r \geq E\big(h_1(X, \theta) 1(X < x)\big)^\tau V_p E\big(h_1(X, \theta) 1(X < x)\big),$$

where V_p and V_r are the asymptotic covariance matrices of the MELE's $\tilde{\theta}$ based on h_1 and on (h_1, h_2), respectively.

Empirical likelihood provides a way to find efficient estimates in semiparametric models which are specified in terms of $r > p$ estimating functions. It also parallels likelihood in full parametric models with respect to the likelihood ratio statistic, as the next theorem shows.

THEOREM 2. *The empirical likelihood ratio statistic for testing $H_0\colon \theta = \theta_0$ is*

$$W_E(\theta_0) = 2l_E(\theta_0) - 2l_E(\tilde{\theta}), \tag{3.9}$$

where $l_E(\theta)$ is given by (3.4). Under the assumptions of Theorem 1, $W_E(\theta_0) \to \chi^2_{(p)}$ as $n \to \infty$, when H_0 is true.

Similarly we can prove the following corollary.

COROLLARY 4. *In order to test model* (1.2), *we may consider the empirical likelihood ratio statistic*

$$W_1 = 2\sum_{i=1}^{n} \log\left[1 + \tilde{t}^{\tau} g(x_i, \tilde{\theta})\right].$$

Under the assumptions of Theorem 1, W_1 *is asymptotically* $\chi^2_{(r-p)}$ *if* (1.2) *is correct.*

COROLLARY 5. *Let* $\theta^{\tau} = (\theta_1, \theta_2)^{\tau}$, *where* θ_1 *and* θ_2 *are* $q \times 1$ *and* $(p-q) \times 1$ *vectors, respectively. For* H_0: $\theta_1 = \theta_1^0$, *the profile empirical likelihood ratio test statistic is*

$$W_2 = 2l_E(\theta_1^0, \tilde{\theta}_2^0) - 2l_E(\tilde{\theta}_1, \tilde{\theta}_2), \tag{3.10}$$

where $\tilde{\theta}_2^0$ *minimizes* $l_E(\theta_1^0, \theta_2)$ *with respect to* θ_2. *Under* H_0, $W_2 \to \chi^2_{(q)}$ *as* $n \to \infty$.

Theorem 2 and Corollary 5 allow us to use the empirical likelihood ratio statistic for testing or obtaining confidence limits for parameters in a completely analogous way to that for parametric likelihoods. With full parametric models, there are several likelihood-based statistics equivalent to the first order for testing H_0: $\theta = \theta_0$, including the likelihood ratio statistic, score statistic and Wald's maximum likelihood estimate statistic. A similar equivalence exists here, but we will not explore this topic now.

4. Other results. In this section we mention two other results. The first is that we may use results of Van der Vaart (1988) and Bickel, Klaassen, Ritov and Wellner (1993) to derive a convolution theorem for "regular estimators" of P_0 and $\theta(P_0)$, where P_0 is the probability measure corresponding to F_0, and to show that our maximum empirical likelihood estimates are asymptotically efficient in the sense of those authors and of Sheehy (1988). This gives the following theorem.

THEOREM 3. *Under the conditions of Theorem* 1 *and Lemma* 2 *in the Appendix, the MELE's for both the parameters and the distribution function are asymptotically efficient in the sense of Van der Vaart* (1988) *and Bickel, Klaassen, Ritov and Wellner* (1993).

The second result is about local asymptotic normality of the empirical likelihood ratio statistic.

THEOREM 4. *Under the conditions of Theorem* 1, *let*

$$Z_{n,\theta}(u) = \prod_{i=1}^{n}\left[\frac{1 + t^{\tau}(\theta + un^{-1/2})g(x_i, \theta + un^{-1/2})}{1 + t^{\tau}(\theta)g(x_i, \theta)}\right]$$

be the normalized empirical likelihood ratio. Then we have a LeCam–type representation,

$$\begin{aligned} Z_{n,\theta}(u) &= \exp\left\{ \sum_{i=1}^{n} \log\left[1 + t^{\tau}(\theta + un^{-1/2}) g(x_i, \theta + un^{-1/2})\right] \right. \\ &\qquad \left. - \sum_{i=1}^{n} \log\left[1 + t^{\tau}(\theta) g(x_i, \theta)\right] \right\} \\ &= \exp\left\{ u^{\tau} E_\theta\left(\frac{\partial g}{\partial \theta}\right)^{\tau} \left[E_\theta(gg^{\tau})\right]^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(x_i, \theta) \right. \\ &\qquad \left. + \frac{1}{2} u^{\tau} E_\theta\left(\frac{\partial g}{\partial \theta}\right)^{\tau} \left[E_\theta(gg^{\tau})\right]^{-1} E_\theta\left(\frac{\partial g}{\partial \theta}\right) u + o_p(1) \right\} \end{aligned}$$

and

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} E_\theta\left(\frac{\partial g}{\partial \theta}\right)^{\tau} \left[E_\theta(gg^{\tau})\right]^{-1} g(x_i, \theta) \to N\left(0, E_\theta\left(\frac{\partial g}{\partial \theta}\right)^{\tau} \left[E_\theta(gg^{\tau})\right]^{-1} E_\theta\left(\frac{\partial g}{\partial \theta}\right)\right).$$

The proof is straightforward by Taylor expansion.

Note that $E_\theta(\partial g/\partial\theta)^{\tau}[E_\theta(gg^{\tau})]^{-1}E_\theta(\partial g/\partial\theta)$ is the inverse of the asymptotic variance of $\sqrt{n}(\widetilde{\theta} - \theta)$, so this representation is similar to the representation of the normalized parametric likelihood ratio [see Ibragimov and Has'minskii (1981), page 114].

5. Examples. We consider several illustrations of the estimation procedures. We primarily consider large-sample aspects, but for the first example we also present some numerical results.

Computational issues are discussed by Owen (1990, 1992) and by Qin and Lawless (1992a). Additional experience with empirical likelihood methods is needed before specific recommendations can be given, but a few points may be mentioned. In order to evaluate $l_E(\theta)$ for a given θ, we have to solve (3.3) for $t(\theta)$; this is often handled well by Newton's method, bearing in mind, however, the remarks preceding Lemma 1. To obtain $\widetilde{\theta}$ we may proceed in two stages, essentially obtaining $l_E(\theta)$ and then maximizing it. Alternatively, we may attempt to solve (3.5) simultaneously for $\widetilde{\theta}$ and $\widetilde{t}$; some care is needed because the solution sought is one of many saddlepoints of the function $h(\theta, t) = \sum_{i=1}^{n} \log\{1 + t^{\tau} g(x_i, \theta)\}$ and, in particular, must satisfy $1 + t^{\tau} g(x_i, \theta) \geq n^{-1}$ for each i.

EXAMPLE 1 (Continued). Recall that $y_1, y_2, \ldots, y_n$ are i.i.d., with unknown univariate distribution F and first and second moments $\mu_1 = \theta$, and $\mu_2 = m(\theta)$, respectively, where $m(\cdot)$ is a known function. We can apply the approach used in Section 3 leading to equations (3.6), (3.7) and (3.8). This yields

$$n^{-1}\sum_i \frac{y_i-\theta}{1+t_1(y_i-\theta)+t_2[y_i^2-m(\theta)]}=0, \tag{5.1}$$

$$n^{-1}\sum_i \frac{y_i^2-m(\theta)}{1+t_1(y_i-\theta)+t_2[y_i^2-m(\theta)]}=0, \tag{5.2}$$

$$n^{-1}\sum_i \frac{-t_1-t_2m'(\theta)}{1+t_1(y_i-\theta)+t_2[y_i^2-m(\theta)]}=0. \tag{5.3}$$

The third equation implies $t_1 = -t_2m'(\theta)$ and by substituting this into (5.1) and (5.2) we may get two equations in t_2 and θ to solve. Recalling the discussion proceeding Lemma 1 in Section 3, we note that the desired solution $\tilde{\theta}, \tilde{t}_2$ must satisfy the conditions $1-\tilde{t}_2m'(\theta)(y_i-\tilde{\theta})+\tilde{t}_2[y_i^2-m(\tilde{\theta})] \geq 1/n$, for each $i=1,\dots,n$. In moderately large samples Newton iteration starting from the initial value $(\theta, t_2) = (\bar{y}, 0)$ often works; alternatively, we can obtain $l_E(\theta)$ by finding $t(\theta)$ (see the comments before Lemma 1) as an "inner" iteration and can iterate on θ to find $\tilde{\theta}$.

The results of Theorem 1 show that $\sqrt{n}(\tilde{\theta}-\theta_0) \to N(0,V)$ where V is given in Theorem 1. Some algebra shows that

$$V = \operatorname{var}(y) - \Delta^{-1}[m'(\theta_0)\operatorname{var}(y) + \theta_0 m(\theta_0) - E(y^3)]^2,$$

where $\Delta = E[m'(\theta_0)(y-\theta_0)+m(\theta_0)-y^2]^2$. Thus $V \leq \operatorname{var}(y)$, which is the variance of $\sqrt{n}(\bar{y}-\theta_0)$ and so $\tilde{\theta}$ is asymptotically at least as efficient as $\bar{y}$. In practice, any higher efficiency of $\tilde{\theta}$ when the second moment relationship $E(y^2) = m(\theta)$ holds will of course have to be balanced against a lack of robustness under departures from the relationship.

We consider for illustration a model with first and second moments satisfying $Ex = \theta$ and $E(x^2) = 2\theta^2 + 1$. We generated 1000 pseudorandom samples of sizes 15, 20, 30 and 40 from $N(\theta, \theta^2+1)$, for two values of θ. For each sample we obtained three estimates of θ: the sample mean, the MELE based on the additional knowledge that $E(x^2) = 2\theta^2+1$ and the parametric MLE based on the normal distribution. Table 1 shows that estimated mean and variance of each estimator, obtained from the simulation. We see that the variance of the MELE lies between that of the sample mean and the parametric maximum likelihood estimator. In Table 2 we compare three methods of obtaining confidence intervals for θ. The first two are based on our empirical likelihood methods: one (ELR) obtains confidence intervals from the empirical likelihood ratio statistic (3.9) and the $\chi^2_{(1)}$ approximation of Theorem 2; the other (NCI) is based on the limiting normal distribution for the MELE $\tilde{\theta}$ given in Theorem 1 and the variance estimator following Theorem 1. The third method (PLR) is based on the parametric likelihood ratio statistic from the normal distribution $N(\theta, \theta^2+1)$ from which the data were generated, with $\chi^2_{(1)}$ as the approximating distribution. Table 2 shows, for 1000 samples, of sizes $n = 30$ and 60, two

TABLE 1
Estimated mean and variance of three estimators of θ, from 1000 simulations

n	Sample mean Mean	Sample mean Var	MELE Mean	MELE Var	MLE Mean	MLE Var
			$N(0,1)$, true value of $\theta = 0.0$			
15	0.004484	0.067624	0.006848	0.061824	0.006482	0.058516
20	0.000956	0.049740	0.001945	0.048108	−0.002313	0.045455
30	−0.005714	0.031004	−0.005119	0.030921	−0.004360	0.029835
40	0.000956	0.024572	0.002931	0.024221	−0.000947	0.023431
			$N(1,2)$, true value of $\theta = 1.0$			
15	1.004317	0.128445	0.946416	0.086383	0.966406	0.083193
20	0.995677	0.106569	0.952668	0.062353	0.972931	0.059177
30	1.006338	0.068629	0.968523	0.035759	0.984540	0.034930
40	1.015897	0.044045	0.984512	0.021883	0.994275	0.020584

θ values and nominal 90% and 95% confidence intervals, the average length (Avl) and empirical coverage (Ecv) for each type of interval. It is interesting to note that the two empirical likelihood methods agree closely and that for smaller samples their coverage probability is substantially less than the nominal 90 and 95%. By comparison, the parameteric likelihood yields intervals with close to the nominal coverage. Further investigation of this is needed, but these results raise the question of whether, even for small samples, likelihood ratio intervals are similar to ones based on normality of $\tilde{\theta}$ for empirical likelihood and whether higher-order corrections for both methods are needed.

Finally, as an example, Figure 1 depicts an empirical likelihood ratio curve $W_E(\theta)$ (solid line) and a parametric likelihood ratio curve (dotted line), for a particular sample with $n = 30$ and $\theta = 1$. We can see that for this example the curves are very close.

EXAMPLE 2 [(Continued) Two-sample problem with common mean]. In this case observations (x_i, y_i), $i = 1, 2, \ldots, n$, occur in independent pairs and $E(x_i) = E(y_i) = \theta$. To estimate θ, we consider the estimating equations based on $g_1 = x - \theta$ and $g_2 = y - \theta$ and we associate the empirical likelihood probability p_i with (x_i, y_i). After some simplification, from estimating equations (3.6)–(3.8) we have

$$\tilde{\theta} = \sum_{i=1}^{n} \left[\frac{x_i}{1 + \tilde{t}(x_i - y_i)}\right] \Big/ \sum_{i=1}^{n} \left[\frac{1}{1 + \tilde{t}(x_i - y_i)}\right],$$

where $\tilde{t}$ is determined by

$$\sum_{i=1}^{n} \frac{x_i - y_i}{1 + \tilde{t}(x_i - y_i)} = 0. \tag{5.4}$$

TABLE 2
Average length and coverage for three confidence interval methods, from 1000 simulations

		90%		95%	
		Avl	Ecv	Avl	Ecv
		$n = 30$			
$N(0,1)$	ELR	0.55064	85.8%	0.65714	92.4%
	NCI	0.56965	86.0%	0.67889	92.0%
	PLR	0.60197	89.6%	0.72441	94.3%
$N(1,2)$	ELR	0.56698	83.3%	0.67737	89.2%
	NCI	0.56863	84.3%	0.67767	90.1%
	PLR	0.61489	88.3%	0.73900	93.6%
		$n = 60$			
$N(0,1)$	ELR	0.41535	89.5%	0.49611	95.4%
	NCI	0.41291	89.2%	0.49210	94.9%
	PLR	0.42549	90.8%	0.50950	96.1%
$N(1,2)$	ELR	0.41200	88.6%	0.49267	94.1%
	NCI	0.40933	89.0%	0.48782	93.2%
	PLR	0.42845	91.3%	0.51265	96.1%

Moreover, we seek the solution $\tilde{t}$ to (5.4) that satisfies $1 + \tilde{t}(x_i - y_i) \geq 1/n$ for each $i = 1, 2, \ldots, n$. Such a solution exists if and only if the $x_i - y_i$'s are not all of the same sign. In that case there is exactly one such value $\tilde{t}$, which lies in the interval (t_L, t_U), where

$$t_L = \frac{(1/n - 1)}{\max[0, \max(x_i - y_i)]}, \qquad t_U = \frac{(1 - 1/n)}{\min[0, \max(y_i - x_i)]}.$$

The asymptotic covariance matrix for $\sqrt{n}(\tilde{\theta} - \theta_0)$ is given by V of Theorem 1 as

$$V = \left\{ E\left(\frac{\partial g}{\partial \theta}\right)^{\tau} (E(gg^{\tau}))^{-1} E\left(\frac{\partial g}{\partial \theta}\right) \right\}^{-1} = \frac{\sigma_x^2 \sigma_y^2 - \sigma_{xy}^2}{\sigma_x^2 + \sigma_y^2 - 2\sigma_{xy}},$$

where $\sigma_x^2 = \mathrm{var}(x_i)$, $\sigma_y^2 = \mathrm{var}(y_i)$ and $\sigma_{xy} = \mathrm{cov}(x_i, y_i)$; V may be estimated using the sample covariance matrix entries.

It is easily shown directly that $\tilde{\theta}$ is asymptotically equivalent to the optimal (minimum asymptotic variance) linear combination of $\bar{x}$ and $\bar{y}$. In particular, note that in the case where $\sigma_{xy} = \mathrm{cov}(x_i, y_i) = 0$ that $V = \sigma_x^2 \sigma_y^2 / (\sigma_x^2 + \sigma_y^2)$, which is the same as the variance of the optimal linear combination estimator

$$\hat{\theta} = \left(\frac{\sigma_y^2}{\sigma_x^2 + \sigma_y^2}\right)\bar{x} + \left(\frac{\sigma_x^2}{\sigma_x^2 + \sigma_y^2}\right)\bar{y}.$$

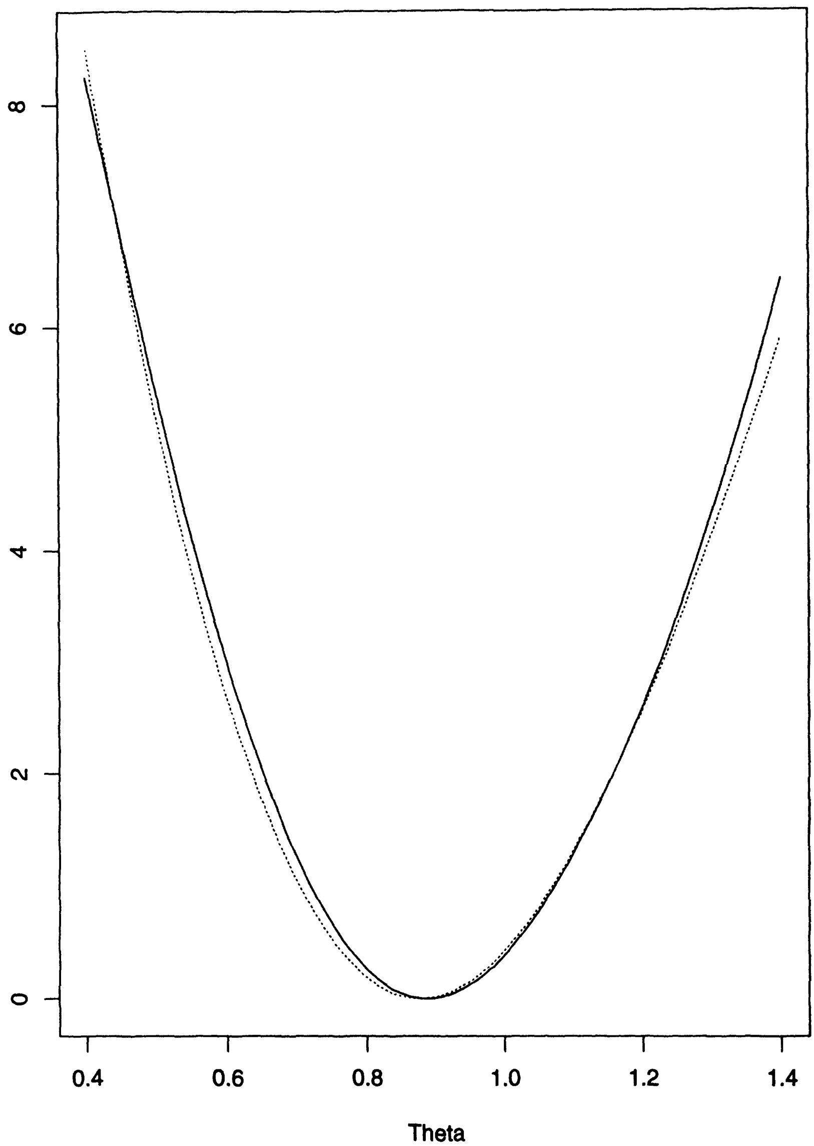

FIG. 1. *Empirical likelihood ratio and parametric likelihood ratio curves.*

EXAMPLE 3 (Continued). Haberman (1984) and Sheehy (1988) have considered constrained estimation of probability measures based on i.i.d. sample $x_1, x_2, \ldots, x_n$ from a distribution P_0, where it is assumed known that $E_{P_0}T(x) = a$, for some specified function T. The estimators they considered are based on the minimization of Kullback–Leibler divergence from certain collections of probability measures to the empirical measure of the x_i's and Sheehy (1988) has proved that the estimate is asymptotically efficient. We may apply empirical likelihood to this problem, utilizing the fixed constraint $\sum_{i=1}^n p_i T(x_i) - a = 0$. In this case earlier results about the c.d.f. still apply, and we obtain the estimate $\widetilde{F}_n(t)$ of the distribution function. It is easy to check that for this estimate $\sqrt{n}(\widetilde{F}_n(t) - F_0(t))$ is asymptotically equivalent to

$$\int \left[1(x < t) - F_0(t) - \operatorname{cov}\big(1(x < t), T\big)(\operatorname{var} T)^{-1}(T - a)\right]$$
$$\times d\sqrt{n}\big(F_n(x) - F_0(x)\big) + o_p(1).$$

This matches with Sheehy's Theorem 2′.

EXAMPLE 4 (An example of "semi-empirical" likelihood). The methods developed in this paper are also useful for dealing with incomplete information in parametric or semiparametric models. As an example we consider a problem arising in field studies of equipment failure [Kalbfleisch and Lawless (1988)]. In this situation, N items are in use and associated with item i are a time to failure $y_i \geq 0$ and a vector of covariates x_i; a regression model with density $f(y|x;\theta)$ specifies the distribution of y_i given x_i.

An incomplete data problem arises because only items that fail by some time T are inspected: each item that fails at time $y_i \leq T$ is inspected, and covariate values x_i are determined. For items with $y_i > T$, the x_i values are unknown. One approach is to base inferences about θ on the likelihood function for y_i's such that $y_i \leq T$:

$$L(\theta) = \prod_{i: y_i \leq T} \frac{f(y_i|x_i;\theta)}{F(T|x_i;\theta)},$$

where $F(y|x;\theta)$ is the c.d.f. of y given x. However, this does not use the information that the remaining y_i's exceed T. This information cannot be used in a parametric likelihood framework without specifying a distribution for the covariates, but it is possible to use empirical likelihood, as follows.

Consider y and x to be jointly distributed, and define estimating functions

$$g_1(y,x,\theta) = I(y \leq T)\frac{\partial}{\partial\theta}\left[\log f(y|x;\theta) - \log F(T|x;\theta)\right],$$
$$g_2(y,x,\theta) = \frac{I(y \leq T)}{F(T|x;\theta)} - 1.$$

Note that $Eg_1 = Eg_2 = 0$, and $\sum_{i=1}^n g_1(x_i, t_i, \theta)$ is the score function from $L(\theta)$. We now associate p_i in the empirical likelihood formulation of Section 3 with

$Pr(y_i, x_i)$, $i = 1, 2, \ldots, N$, and use (g_1, g_2) as our estimating functions. This leads to an MELE $\tilde{\theta}$ where, by Theorem 1, $\sqrt{N}(\tilde{\theta} - \theta) \to N(0, V)$ as $N \to \infty$.

To illustrate, we consider the special case where $f(y|x;\theta) = \theta \exp(-\theta y)$, that is, there are no covariates. We find that $V = \theta^2/\{1 - \exp(-\theta T)\}$, which is precisely the asymptotic variance of the MLE $\hat{\theta}$ obtain from the censored data likelihood

$$L_1 = \prod_{i: y_i \le T} \theta \exp\bigl(-\theta y_i\bigr) \prod_{i: y_i > T} \exp\bigl(-\theta T\bigr).$$

Further investigation is needed for cases where covariates have an effect, but it is interesting that in this special case the MELE is equivalent to the fully efficient MLE. Qin and Lawless (1992b) consider similar applications of empirical likelihood.

6. Additional remarks. Other approaches may be taken to combine estimating functions. First, we remark that likelihood is not the only distance in the simplex for $(p_1, p_2, \ldots, p_n)$ that can be used to generate confidence sets for θ with a chi-square calibration. Efron (1981) and DiCiccio and Romano (1990) consider the $(n-2)$-dimensional subfamily of multinomials generated by minimizing the Kullback–Leibler distance $D(F, F_n) = \sum_i p_i \log(np_i)$ subject to $\sum_i p_i x_i = \mu$ and $\sum_i p_i = 1$. Owen (1991) has considered log Euclidean likelihood, defined as $l_{\mathrm{EU}} = -\frac{1}{2}\sum_i (np_i - 1)^2$, as an alternative to $\sum_{i=1}^n \log p_i$. This is quite tractable and leads to methods asymptotically equivalent to the ones in this paper.

Another approach to combining estimating functions is to consider the optimal (minimum asymptotic variance) linear combination of the r estimating functions $g_j(x_i, \theta)$, $j = 1, 2, \ldots, r$, as mentioned in Corollary 2. This leads to [e.g., see McCullagh and Nelder (1989), page 341] the estimating equations

$$\sum_{i=1}^n D_i^\tau v_i^{-1} g_i = 0, \tag{6.1}$$

where $g_i = \bigl(g_1(x_i, \theta), \ldots, g_r(x_i, \theta)\bigr)^\tau$, $D_i = \partial g_i / \partial \theta$ and $v_i = \mathrm{var}(g_i)$, which is assumed nonsingular. The v_i's are unknown, but when the g_i's are i.i.d. we have $v_i = v$, which may be consistently estimated, for example, by

$$\hat{v} = \frac{1}{n} \sum_{i=1}^n \hat{g}_i \hat{g}_i^\tau,$$

where $\hat{g}_i$ is g_i evaluated at any consistent estimate $\hat{\theta}$ of θ. We then obtain $\tilde{\theta}$ by solving

$$\sum_{i=1}^n D_i^\tau \hat{v}^{-1} g_i = 0. \tag{6.2}$$

The empirical likelihood approach has the potential advantage of providing likelihood ratio statistics, upon which tests and confidence intervals may be based. One might hope that these possess better small-sample properties than methods based on approximate normality of the estimates. This was not apparent from our simulation in Example 1, however, and needs investigation. The empirical likelihoods also appear to be Bartlett or signed square root correctable and may be generalized to handle independent but not identically distributed data.

A good deal of work is needed to apply and assess the methods in practical situations. Experience is needed to determine how easily estimates can be obtained in small- to moderate-size samples and what the properties of the estimators and the empirical likelihood ratio statistics are in these situations. Higher-order asymptotic properties and comparisons with resampling methods are also of interest. We hope to consider some of these topics in future communications.

APPENDIX

Here we give proofs for the results in Sections 3 and 4.

PROOF OF LEMMA 1. Denote $\theta = \theta_0 + un^{-1/3}$, for $\theta \in \{\theta \mid \|\theta - \theta_0\| = n^{-1/3}\}$, where $\|u\| = 1$. First, we give a lower bound for $l_E(\theta)$ on the surface of the ball. Similar to the proof of Owen (1990), when $E\|g(x,\theta)\|^3 < \infty$ and $\|\theta - \theta_0\| \le n^{-1/3}$ we have

$$\text{(A.1)}\quad \begin{aligned} t(\theta) &= \left[\frac{1}{n}\sum_{i=1}^{n} g(x_i,\theta)g^{\tau}(x_i,\theta)\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^{n} g(x_i,\theta)\right] + o(n^{-1/3}) \quad \text{(a.s.)} \\ &= O(n^{-1/3}) \quad \text{(a.s.)}, \end{aligned}$$

uniformly about $\theta \in \{\theta \mid \|\theta - \theta_0\| \le n^{-1/3}\}$.

By this and Taylor expansion, we have (uniformly for u),

$$\begin{aligned} l_E(\theta) &= \sum_i t^{\tau}(\theta)g(x_i,\theta) - \frac{1}{2}\sum_i \left[t^{\tau}(\theta)g(x_i,\theta)\right]^2 + o(n^{1/3}) \quad \text{(a.s.)} \\ &= \frac{n}{2}\left[\frac{1}{n}\sum_i g(x_i,\theta)\right]^{\tau}\left[\frac{1}{n}\sum_i g(x_i,\theta)g^{\tau}(x_i,\theta)\right]^{-1}\left[\frac{1}{n}\sum_i g(x_i,\theta)\right] \\ &\quad + o(n^{1/3}) \quad \text{(a.s.)} \\ &= \frac{n}{2}\left[\frac{1}{n}\sum_i g(x_i,\theta_0) + \frac{1}{n}\sum_i \frac{\partial g(x_i,\theta_0)}{\partial\theta}un^{-1/3}\right]^{\tau}\left[\frac{1}{n}\sum_i g(x_i,\theta)g^{\tau}(x_i,\theta)\right]^{-1} \\ &\quad \times \left[\frac{1}{n}\sum_i g(x_i,\theta_0) + \frac{1}{n}\sum_i \frac{\partial g(x_i,\theta_0)}{\partial\theta}un^{-1/3}\right] + o(n^{1/3}) \quad \text{(a.s.)} \end{aligned}$$

$$= \frac{n}{2}\left[O\left(n^{-1/2}(\log\log n)^{1/2}\right) + E\left(\frac{\partial g(x,\theta_0)}{\partial\theta}\right)un^{-1/3}\right]^{\tau}$$
$$\times\left[E\left(g(x,\theta_0)g^{\tau}(x,\theta_0)\right)\right]^{-1}$$
$$\times\left[O\left(n^{-1/2}(\log\log n)^{1/2}\right) + E\left(\frac{\partial g(x,\theta_0)}{\partial\theta}\right)un^{-1/3}\right] + o\left(n^{1/3}\right) \quad \text{(a.s.)}$$
$$\geq (c-\varepsilon)n^{1/3}, \quad \text{a.s.},$$

where $c-\varepsilon > 0$ and c is the smallest eigenvalue of

$$E\left(\frac{\partial g(x,\theta_0)}{\partial\theta}\right)^{\tau}\left[E\left(g(x,\theta_0)g^{\tau}(x,\theta_0)\right)\right]^{-1}E\left(\frac{\partial g(x,\theta_0)}{\partial\theta}\right).$$

Similarly,

$$l_E(\theta_0) = \frac{n}{2}\left[\frac{1}{n}\sum_i g(x_i,\theta_0)\right]^{\tau}\left[\frac{1}{n}\sum_i g(x_i,\theta_0)g^{\tau}(x_i,\theta_0)\right]^{-1}$$
$$\times\left[\frac{1}{n}\sum_i g(x_i,\theta_0)\right] + o(1) \quad \text{(a.s.)}$$
$$= O(\log\log n), \quad \text{(a.s.)}.$$

Since $l_E(\theta)$ is a continuous function about θ as θ belongs to the ball $\|\theta-\theta_0\| \leq n^{-1/3}$, $l_E(\theta)$ has minimum value in the interior of this ball, and $\widetilde{\theta}$ satisfies

$$\left.\frac{\partial l_E(\theta)}{\partial\theta}\right|_{\theta=\widetilde{\theta}} = \sum_i \left.\frac{(\partial t^{\tau}(\theta)/\partial\theta)g(x_i,\theta) + (\partial g(x_i,\theta)/\partial\theta)^{\tau}t(\theta)}{1+t^{\tau}(\theta)g(x_i,\theta)}\right|_{\theta=\widetilde{\theta}}$$
$$= \sum_i \frac{1}{1+t^{\tau}(\theta)g(x_i,\theta)}\left(\frac{\partial g(x_i,\theta)}{\partial\theta}\right)^{\tau}t(\theta)\Big|_{\theta=\widetilde{\theta}}$$
$$= 0. \qquad \square$$

PROOF OF THEOREM 1. Taking derivatives about θ and t^{τ}, we have

$$\frac{\partial Q_{1n}(\theta,0)}{\partial\theta} = \frac{1}{n}\sum_i\frac{\partial g(x_i,\theta)}{\partial\theta}, \qquad \frac{\partial Q_{1n}(\theta,0)}{\partial t^{\tau}} = -\frac{1}{n}\sum_i g(x_i,\theta)g(x_i,\theta)^{\tau},$$
$$\frac{\partial Q_{2n}(\theta,0)}{\partial\theta} = 0, \qquad \frac{\partial Q_{2n}(\theta,0)}{\partial t^{\tau}} = \frac{1}{n}\sum_i\left(\frac{\partial g(x_i,\theta)}{\partial\theta}\right)^{\tau}.$$

Expanding $Q_{1n}(\widetilde{\theta},\widetilde{t})$, $Q_{2n}(\widetilde{\theta},\widetilde{t})$ at $(\theta_0,0)$, by the conditions of the theorem and

Lemma 1, we have

$$
\begin{aligned}
&0=Q_{1n}(\tilde{\theta},\tilde{t})\\
\text{(A.2)}\quad &=Q_{1n}(\theta_0,0)+\frac{\partial Q_{1n}(\theta_0,0)}{\partial\theta}(\tilde{\theta}-\theta_0)+\frac{\partial Q_{1n}(\theta_0,0)}{\partial t^\tau}(\tilde{t}-0)+o_p(\delta_n),\\
&0=Q_{2n}(\tilde{\theta},\tilde{t})\\
\text{(A.3)}\quad &=Q_{2n}(\theta_0,0)+\frac{\partial Q_{2n}(\theta_0,0)}{\partial\theta}(\tilde{\theta}-\theta_0)+\frac{\partial Q_{2n}(\theta_0,0)}{\partial t^\tau}(\tilde{t}-0)+o_p(\delta_n),
\end{aligned}
$$

where $\delta_n = \|\tilde{\theta}-\theta_0\| + \|\tilde{t}\|$. We have

$$\begin{pmatrix} \tilde{t} \\ \tilde{\theta}-\theta_0 \end{pmatrix} = S_n^{-1}\begin{pmatrix} -Q_{1n}(\theta_0,0)+o_p(\delta_n) \\ o_p(\delta_n) \end{pmatrix},$$

where

$$\text{(A.4)}\quad S_n = \begin{pmatrix} \dfrac{\partial Q_{1n}}{\partial t^\tau} & \dfrac{\partial Q_{1n}}{\partial\theta} \\ \dfrac{\partial Q_{2n}}{\partial t^\tau} & 0 \end{pmatrix}_{(\theta_0,0)} \to \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & 0 \end{pmatrix} = \begin{pmatrix} -E(gg^\tau) & E\left(\dfrac{\partial g}{\partial\theta}\right) \\ E\left(\dfrac{\partial g}{\partial\theta}\right)^\tau & 0 \end{pmatrix}.$$

From this and $Q_{1n}(\theta_0,0) = (1/n)\sum_{i=1}^n g(x_i,\theta_0) = O_p(n^{-1/2})$, we know that $\delta_n = O_p(n^{-1/2})$. Easily we have

$$\sqrt{n}(\tilde{\theta}-\theta_0) = S_{22.1}^{-1}S_{21}S_{11}^{-1}\sqrt{n}Q_{1n}(\theta_0,0)+o_p(1)\to N(0,V),$$

where

$$V = S_{22.1}^{-1} = \left\{E\left(\frac{\partial g}{\partial\theta}\right)^\tau (Egg^\tau)^{-1}E\left(\frac{\partial g}{\partial\theta}\right)\right\}^{-1}.$$

The rest of the proof is similarly straightforward. □

PROOF OF COROLLARY 1. Write

$$
\begin{aligned}
D_r(\theta) &= \left(\left(\frac{\partial g_1}{\partial\theta}\right)^\tau,\ldots,\left(\frac{\partial g_{r-1}}{\partial\theta}\right)^\tau,\left(\frac{\partial g_r}{\partial\theta}\right)^\tau\right) = \left(D_{r-1}^\tau(\theta),\left(\frac{\partial g_r}{\partial\theta}\right)^\tau\right),\\
C_r(\theta) &= E(gg^\tau) = \begin{pmatrix} C_{11}(\theta) & C_{12}(\theta) \\ C_{21}(\theta) & C_{22}(\theta) \end{pmatrix},
\end{aligned}
$$

where $C_{11}(\theta)$ is an $(r-1)\times(r-1)$ matrix. For square matrices A and B of the same order, let $A \geq B$ denote that $A-B$ is positive semidefinite. Then

$$
\begin{aligned}
V_r^{-1} &= E\left(\frac{\partial g}{\partial\theta}\right)^\tau E(gg^\tau)^{-1}E\left(\frac{\partial g}{\partial\theta}\right)\\
&= D_r^\tau(\theta)C_r^{-1}(\theta)D_r(\theta)\\
&\geq \left(D_{r-1}^\tau(\theta),\left(\frac{\partial g_r}{\partial\theta}\right)^\tau\right)\begin{pmatrix} C_{11}^{-1}(\theta) & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} D_{r-1}^\tau(\theta) \\ \frac{\partial g_r}{\partial\theta} \end{pmatrix}\\
&= V_{r-1}^{-1}.
\end{aligned}
$$

□

PROOF OF COROLLARY 3. We know that

$$V_r^{-1} = \left(\left(E\frac{\partial h_1}{\partial\theta}\right)^\tau, \left(E\frac{\partial h_2}{\partial\theta}\right)^\tau\right)\begin{pmatrix} E(h_1h_1^\tau) & E(h_1h_2^\tau) \\ E(h_2h_1^\tau) & E(h_2h_2^\tau)\end{pmatrix}^{-1}\begin{pmatrix} E\left(\frac{\partial h_1}{\partial\theta}\right) \\ E\left(\frac{\partial h_2}{\partial\theta}\right)\end{pmatrix}$$

$$= \left(\left(E\frac{\partial h_1}{\partial\theta}\right)^\tau, -E\left(\frac{\partial h_1}{\partial\theta}\right)^\tau (Eh_1h_1^\tau)^{-1}E(h_1h_2^\tau) + E\left(\frac{\partial h_2}{\partial\theta}\right)^\tau\right)$$

$$\times\begin{pmatrix} (Eh_1h_1^\tau)^{-1} & 0 \\ 0 & A_{22.1}^{-1}\end{pmatrix}\begin{pmatrix} E\left(\frac{\partial h_1}{\partial\theta}\right) \\ -(Eh_2h_1^\tau)(Eh_1h_1^\tau)^{-1}E\left(\frac{\partial h_1}{\partial\theta}\right) + E\left(\frac{\partial h_2}{\partial\theta}\right)\end{pmatrix},$$

where

$$A_{22.1} = E(h_2h_2^\tau) - E(h_2h_1^\tau)(Eh_1h_1^\tau)^{-1}E(h_1h_2^\tau).$$

Since h_1 is the score, $(Eh_1h_1^\tau) = -E(\partial h_1/\partial\theta)$ and, taking derivatives about θ in the equation $E(h_2) = 0$, we have

$$\int \frac{\partial h_2}{\partial\theta} f(x,\theta)\,dx + \int h_2 \frac{\partial f(x,\theta)}{\partial\theta}\,dx = 0,$$

that is

$$E\left(\frac{\partial h_2}{\partial\theta}\right) + E(h_2h_1^\tau) = 0.$$

Thus

$$V_r^{-1} = \left(E\left(\frac{\partial h_1}{\partial\theta}\right)^\tau, 0\right)\begin{pmatrix} (Eh_1h_1^\tau)^{-1} & 0 \\ 0 & A_{22.1}^{-1}\end{pmatrix}\begin{pmatrix} E\left(\frac{\partial h_1}{\partial\theta}\right) \\ 0\end{pmatrix}$$

$$= E\left(\frac{\partial h_1}{\partial\theta}\right)^\tau (Eh_1h_1^\tau)^{-1}E\left(\frac{\partial h_1}{\partial\theta}\right) = (Eh_1h_1^\tau) = V_p^{-1}.$$

If $\tilde\theta$ is the MLE for θ, then $F(x,\tilde\theta)$ is the MLE for $F(x,\theta)$; so the asymptotic variance W_r of $\sqrt{n}(\tilde F_n(x) - F(x))$ is no less than the asymptotic variance of $\sqrt{n}(F(x,\tilde\theta) - F(x))$. We only need to calculate the asymptotic variance of the latter, which is easily found to be

$$E\{h_1(X,\theta)1(X<x)\}^\tau V_p E\{h_1(X,\theta)1(X<x)\},$$

and the corollary is proved. □

PROOF OF THEOREM 2. The log-empirical likelihood ratio test statistic is

$$W_E(\theta_0) = 2\left\{ \sum_i \log\left[1 + t_0^\tau g(x_i,\theta_0)\right] - \sum_i \log\left[1 + \tilde{t}^\tau g(x_i,\tilde{\theta})\right] \right\}.$$

Note that

$$l_E(\tilde{\theta},\tilde{t}) = \sum_{i=1}^n \log\left[1 + \tilde{t}^\tau g(x_i,\tilde{\theta})\right] = -\frac{n}{2} Q_{1n}^\tau(\theta_0,0) A Q_{1n}(\theta_0,0) + o_p(1)$$

where $A = S_{11}^{-1}\{I + S_{12}S_{22.1}^{-1}S_{21}S_{11}^{-1}\}$. Also under H_0,

$$\frac{1}{n}\sum_i \frac{1}{1 + t_0^\tau g(x_i,\theta_0)} g(x_i,\theta_0) = 0 \quad \Rightarrow \quad t_0 = -S_{11}^{-1}Q_{1n}(\theta_0,0) + o_p(1)$$

and

$$\sum_i \log\left[1 + t_0^\tau g(x_i,\theta_0)\right] = -\frac{n}{2} Q_{1n}^\tau(\theta_0,0) S_{11}^{-1} Q_{1n}(\theta_0,0) + o_p(1).$$

Thus

$$\begin{aligned} W_E(\theta_0) &= nQ_{1n}^\tau(\theta_0,0)\left(A - S_{11}^{-1}\right)Q_{1n}(\theta_0,0) + o_p(1) \\ &= nQ_{1n}^\tau(\theta_0,0)S_{11}^{-1}S_{12}S_{22.1}^{-1}S_{21}S_{11}^{-1}Q_{1n}(\theta_0,0) + o_p(1) \\ &= \left[(-S_{11})^{-1/2}\sqrt{n}Q_{1n}(\theta_0,0)\right]^\tau \left[(-S_{11})^{-1/2}S_{12}S_{22.1}^{-1}S_{21}(-S_{11})^{-1/2}\right] \\ &\quad \times \left[(-S_{11})^{-1/2}\sqrt{n}Q_{1n}(\theta_0,0)\right] + o_p(1). \end{aligned}$$

Note that $(-S_{11})^{-1/2}\sqrt{n}Q_{1n}(\theta_0,0)$ converges to a standard multivariate normal distribution and that $(-S_{11})^{-1/2}S_{12}S_{22.1}^{-1}S_{21}(-S_{11})^{-1/2}$ is symmetric and idempotent, with trace equal to p. Hence the empirical likelihood ratio statistic $W_E(\theta_0)$ converges to χ_p^2. □

PROOF OF COROLLARY 5. By Taylor expansion we have

$$\begin{aligned} W_2 &= 2l_E\left(\theta_1^0,\tilde{\theta}_2^0\right) - 2l_E\left(\tilde{\theta}_1,\tilde{\theta}_2\right) \\ &= \left[(-S_{11})^{-1/2}\sqrt{n}Q_{1n}(\theta_0,0)\right]^\tau (Egg^\tau)^{-1/2} \\ &\quad \times\left\{ \left(E\frac{\partial g}{\partial \theta}\right)\left[\left(E\frac{\partial g}{\partial \theta}\right)^\tau (Egg^\tau)^{-1}\left(E\frac{\partial g}{\partial \theta}\right)\right]^{-1}\left(E\frac{\partial g}{\partial \theta}\right)^\tau \right. \\ &\quad \left. - \left(E\frac{\partial g}{\partial \theta_1}\right)\left[\left(E\frac{\partial g}{\partial \theta_1}\right)^\tau (Egg^\tau)^{-1}\left(E\frac{\partial g}{\partial \theta_1}\right)\right]^{-1}\left(E\frac{\partial g}{\partial \theta_1}\right)^\tau \right\} \\ &\quad \times (Egg^\tau)^{-1/2}\left[(-S_{11})^{-1/2}\sqrt{n}Q_{1n}(\theta_0,0)\right] + o_p(1). \end{aligned}$$

By a result in Rao [(1973), page 187], we only need to show that

$$\Delta := \left(E\frac{\partial g}{\partial \theta}\right)\left[\left(E\frac{\partial g}{\partial \theta}\right)^{\tau}(Egg^{\tau})^{-1}\left(E\frac{\partial g}{\partial \theta}\right)\right]^{-1}\left(E\frac{\partial g}{\partial \theta}\right)^{\tau}$$
$$\geq \left(E\frac{\partial g}{\partial \theta_1}\right)\left[\left(E\frac{\partial g}{\partial \theta_1}\right)^{\tau}(Egg^{\tau})^{-1}\left(E\frac{\partial g}{\partial \theta_1}\right)\right]^{-1}\left(E\frac{\partial g}{\partial \theta_1}\right)^{\tau}.$$

In fact,

$$\Delta = \left(E\frac{\partial g}{\partial \theta}\right)\left[\left(E\frac{\partial g}{\partial \theta}\right)^{\tau}(Egg^{\tau})^{-1}\left(E\frac{\partial g}{\partial \theta}\right)\right]^{-1}\left(E\frac{\partial g}{\partial \theta}\right)^{\tau}$$
$$\geq \left(E\frac{\partial g}{\partial \theta_1}, E\frac{\partial g}{\partial \theta_2}\right)\begin{pmatrix}\left[\left(E\frac{\partial g}{\partial \theta_1}\right)^{\tau}(Egg^{\tau})^{-1}\left(E\frac{\partial g}{\partial \theta_1}\right)\right]^{-1}\left(E\frac{\partial g}{\partial \theta_1}\right)^{\tau} & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}E\left(\frac{\partial g}{\partial \theta_1}\right)^{\tau}\\ E\left(\frac{\partial g}{\partial \theta_2}\right)\end{pmatrix}$$
$$= \left(E\frac{\partial g}{\partial \theta_1}\right)\left[\left(E\frac{\partial g}{\partial \theta_1}\right)^{\tau}(Egg^{\tau})^{-1}\left(E\frac{\partial g}{\partial \theta_1}\right)\right]^{-1}\left(E\frac{\partial g}{\partial \theta_1}\right)^{\tau}.$$

Thus $W_2 \to \chi^2_{[r-(p-q)-(r-p)]} = \chi^2_q$. $\square$

Now we consider the proof of Theorem 3. For this we need the concept of a tangent space $T(P_0)$ of $L_2(P_0)$; see Sheehy (1988) for a concise description. First we give a lemma; let

$$g_A(x,\theta) = \left(g_1(x,\theta),\dots,g_p(x,\theta)\right)^{\tau}, \qquad g_B(x,\theta) = \left(g_{p+1}(x,\theta),\dots,g_r(x,\theta)\right)^{\tau},$$
$$Eg(x,\theta) = 0 \quad \Rightarrow \quad Eg_A = 0, \quad \text{and } Eg_B = 0.$$

LEMMA 2. *Assume that* $(E_{p_0}\,\partial g_A/\partial\theta)^{-1}$ *exists. Define*

$$T^*(P_0) = \Big\{h\colon h \in L_2(P_0), |h| < K, \text{ some } K > 0, E_{P_0}h = 0,$$
$$E_{P_0}\left(\left[g_B - E_{P_0}\left(\frac{\partial g_B}{\partial \theta}\right)\left(E_{P_0}\frac{\partial g_A}{\partial \theta}\right)^{-1}g_A\right]h\right) = 0\Big\}. \tag{A.5}$$

Then $T^*(P_0)$ *is a maximal tangent space at* $P_0 \in \mathcal{P}$.

Since $T^*(P_0)$ *is maximal we must have*

$$\overline{T_m(P_0)} = \Big\{h\colon h \in L_2(P_0), E_{P_0}h = 0,$$
$$E_{P_0}\left(\left[g_B - E_{P_0}\left(\frac{\partial g_B}{\partial \theta}\right)\left(E_{P_0}\frac{\partial g_A}{\partial \theta}\right)^{-1}g_A\right]h\right) = 0\Big\},$$

for any maximal tangent space $T_m(P_0)$.

PROOF. Consider a parametric submodel F_η, with density f_η. Since $\int g_A\left(x, \theta(F_\eta)\right) dF_\eta = 0$, we have

$$\int \frac{\partial g_A\left(x, \theta(F_\eta)\right)}{\partial \theta} \frac{\partial \theta(F_\eta)}{\partial \eta} dF_\eta + \int g_A\left(x, \theta(F_\eta)\right) \frac{\partial \log f_\eta}{\partial \eta} dF_\eta = 0,$$

that is,

$$\int \frac{\partial g_A(x, \theta_0)}{\partial \theta} \frac{\partial \theta(F_\eta)}{\partial \eta}\bigg|_{\eta=0} dP_0 + \int g_A(x, \theta_0) h \, dP_0 = 0$$

or

$$\frac{\partial \theta(F_\eta)}{\partial \eta}\bigg|_{\eta=0} = -\left[E_{P_0}\left(\frac{\partial g_A(x, \theta_0)}{\partial \theta}\right)\right]^{-1} E_{P_0}\left[g_A(x, \theta_0) h\right]. \tag{A.6}$$

Similarly,

$$\int \frac{\partial g_B(x, \theta_0)}{\partial \theta} \frac{\partial \theta(F_\eta)}{\partial \eta}\bigg|_{\eta=0} dP_0 + \int g_B(x, \theta_0) h \, dP_0 = 0,$$

that is,

$$E_{P_0}\left(\frac{\partial g_B(x, \theta_0)}{\partial \theta}\right)\left\{-\left[E_{P_0}\left(\frac{\partial g_A(x, \theta_0)}{\partial \theta}\right)\right]^{-1} E_{P_0}\left[g_A(x, \theta_0) h\right]\right\} + E_{P_0}\left[g_B(x, \theta_0) h\right] = 0$$

or

$$E_{P_0}\left(\left[g_B - E_{P_0}\left(\frac{\partial g_B}{\partial \theta}\right)\left(E_{P_0}\frac{\partial g_A}{\partial \theta}\right)^{-1} g_A\right] h\right) = 0,$$

so

$$T(P_0) \subset \overline{T^*(P_0)} = \left\{h\colon h \in L_2(P_0), E_{P_0} h = 0, \; E_{P_0}\left(\left[g_B - E_{P_0}\left(\frac{\partial g_B}{\partial \theta}\right)\left(E_{P_0}\frac{\partial g_A}{\partial \theta}\right)^{-1} g_A\right] h\right) = 0\right\}.$$

In order to show that $T^*(P_0)$ is itself a tangent space, we need to show that if $h \in T^*(P_0)$ there exists $\{P_t\colon 0 \le t \le 1\} \subset \mathcal{P}$ so that

$$\int \left[t^{-1}\left((dP_t)^{1/2} - (dP_0)^{1/2}\right) - \frac{1}{2} h (dP_0)^{1/2}\right]^2 \to 0 \quad \text{as } t \to 0.$$

It is easy to show that the sequence $\{P_t\}$ defined by $dP_t/dP_0 = 1 + th$ will do, and we are done. □

PROOF OF THEOREM 3. We will show that the MELE $\tilde{\theta}$ is fully efficient. Since the tangent space $T_m(P_0)$ is linear, it is sufficient to show that the influence function of $\tilde{\theta}$ is the projection of the (pathwise) derivative of the functional $\theta(P_0)$ onto the tangent space. Since

$$\frac{\partial\theta(F_\eta)}{\partial\eta}\bigg|_{\eta=0} = -\left[E_{P_0}\left(\frac{\partial g_A(x,\theta_0)}{\partial\theta}\right)\right]^{-1} E_{P_0}\left[g_A(x,\theta_0)h\right],$$

the derivative $\dot{\theta}$ of θ is

$$\dot{\theta} = -\left[E_{P_0}\left(\frac{\partial g_A}{\partial\theta}\right)\right]^{-1} g_A.$$

Noting that

$$\dot{\theta} = S_{22.1}^{-1}S_{21}S_{11}^{-1}g(x,\theta_0) - \left\{\left[E_{P_0}\left(\frac{\partial g_A}{\partial\theta}\right)\right]^{-1} g_A + S_{22.1}^{-1}S_{21}S_{11}^{-1}g(x,\theta_0)\right\}$$

and

$$\sqrt{n}(\tilde{\theta} - \theta_0) = \frac{1}{\sqrt{n}}\sum_i S_{22.1}^{-1}S_{21}S_{11}^{-1}g(x_i,\theta_0) + o_p(1),$$

we want to show that

$$S_{22.1}^{-1}S_{21}S_{11}^{-1}g(x,\theta_0) \in T_m(P_0), \tag{A.7}$$

$$\alpha := \left\{\left[E_{P_0}\left(\frac{\partial g_A}{\partial\theta}\right)\right]^{-1} g_A + S_{22.1}^{-1}S_{21}S_{11}^{-1}g(x,\theta_0)\right\} \in T_m^{\perp}(P_0). \tag{A.8}$$

By Bickel, Klaassen, Ritov and Wellner [(1993), Section 3], our empirical likelihood estimate for θ is then efficient.

Easily we can show

$$E_{P_0}\left\{\left[g_B - E_{P_0}\left(\frac{\partial g_B}{\partial\theta}\right)\left(E_{P_0}\frac{\partial g_A}{\partial\theta}\right)^{-1} g_A\right]\left[g^\tau S_{11}^{-1}S_{12}S_{22.1}^{-1}\right]\right\} = 0.$$

Next we show $\alpha \in T_m^{\perp}(P_0)$.

Note that $S_{22.1} = E_{P_0}(\partial g/\partial\theta)^\tau E(gg^\tau)E_{P_0}(\partial g/\partial\theta)$, $S_{12} = S_{21}^\tau = E_{P_0}(\partial g/\partial\theta)$ and $S_{11} = -E_{P_0}(gg^\tau)$, so

$$\begin{aligned}\alpha &= S_{22.1}^{-1}\left\{S_{22.1}\left[E_{P_0}\left(\frac{\partial g_A}{\partial\theta}\right)\right]^{-1} g_A + S_{21}S_{11}^{-1}g\right\} \\ &= S_{22.1}^{-1}S_{21}S_{11}^{-1}\begin{pmatrix} 0 \\ g_B - E_{P_0}\left(\frac{\partial g_B}{\partial\theta}\right)\left(E_{P_0}\frac{\partial g_A}{\partial\theta}\right)^{-1} g_A\end{pmatrix} \perp T_m(P_0).\end{aligned}$$

In a similar way, we can establish the efficiency of the distribution function estimate $\tilde{F}_n$. □

Acknowledgments. The authors wish to thank Professors J. Chen, V. P. Godambe, J. D. Kalbfleisch, D. L. McLeish and A. B. Owen and two anonymous referees for many helpful comments.

REFERENCES

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins Univ. Press.

CHEN, J. and QIN, J. (1993). Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika* **80** 107–116.

DICICCIO, T. J., HALL, P. and ROMANO, J. P. (1989). Comparison of parametric and empirical likelihood functions. *Biometrika*. **76** 465–476.

DICICCIO, T. J., HALL, P. and ROMANO, J. P. (1991). Empirical likelihood is Bartlett-correctable. *Ann. Statist.* **19** 1053–1061.

DICICCIO, T. J. and ROMANO, J. (1989). On adjustments to the signed root of the empirical likelihood ratio statistic. *Biometrika* **76** 447–456.

DICICCIO, T. J. and ROMANO, J. (1990). Nonparametric confidence limits by resampling methods and least favorable families. *Internat. Statist. Rev.* **58** 59–76.

EFRON, B. (1981). Nonparametric standard errors and confidence intervals (with discussion). *Canad. J. Statist.* **9** 139–172.

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** 137–172.

HABERMAN, S. J. (1984). Adjustment by minimum discriminant information. *Ann. Statist.* **12** 971–988.

HALL, P. (1990). Pseudo-likelihood theory for empirical likelihood. *Ann. Statist.* **18** 121–140.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1981). *Statistical Estimation, Asymptotic Theory*. Springer, New York.

KALBFLEISCH, J. D. and LAWLESS, J. F. (1988). Estimation of reliability in field-performance studies. *Technometrics* **30** 365–388.

KAPLAN, E. L. and MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481.

KOLACZYK, E. D. (1992). Empirical likelihood for generalized linear models. Technical Report 389, Dept. Statistics, Stanford Univ.

KUK, A. Y. C. and MAK, T. K. (1989). Median estimation in the presence of auxiliary information. *J. Roy. Statist. Soc. Ser. B* **51** 261–269.

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

OWEN, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75** 237–249.

OWEN, A. B. (1990). Empirical likelihood confidence regions. *Ann. Statist.* **18** 90–120.

OWEN, A. B. (1991). Empirical likelihood for linear models. *Ann. Statist.* **19** 1725–1747.

OWEN, A. B. (1992). Empirical likelihood and generalized projection pursuit. Technical Report 393, Dept. Statistics, Stanford Univ.

QIN, J. and LAWLESS, J. F. (1992a). Empirical likelihood estimation and tests of parameters subject to constraints. Technical Report 92-13, Dept. Statistics, Univ. Waterloo.

QIN, J. and LAWLESS, J. F. (1992b). Some applications of semi-empirical likelihood in incomplete data problems. Unpublished manuscript.

RAO, C. R. (1973). *Linear Statistical Inference and Its Applications*. Wiley, New York.

SHEEHY, A. (1988). Kullback–Leibler constrained estimation of probability measures, Technical Report 137, Dept. Statistics, Stanford Univ.

THOMAS, D. R. and GRUNKEMEIER, G. L. (1975). Confidence interval estimation of survival probabilities for censored data. *J. Amer. Statist. Assoc.* **70** 865–871.

VAN DER VAART, A. W. (1988). *Statistical Estimation in Large Parameter Spaces*. Centrum voor Wiskunde en Informatica, Amsterdam.

VARDI, Y. (1985). Empirical distributions in selection bias models. *Ann. Statist.* **13** 178–203.

DEPARTMENT OF STATISTICS
AND ACTUARIAL SCIENCE
UNIVERSITY OF WATERLOO
WATERLOO, ONTARIO
CANADA N2L 3G1