

Maximum empirical likelihood estimation and related topics

Hanxiang Peng and Anton Schick

*Indiana University Purdue University
at Indianapolis
Department of Mathematical Sciences
Indianapolis, IN 46202-3267
USA
e-mail: hpeng@math.iupui.edu*

*Binghamton University
Department of Mathematical Sciences
Binghamton, NY 13902-6000
USA
e-mail: anton@math.binghamton.edu*

Abstract: This article develops a theory of maximum empirical likelihood estimation and empirical likelihood ratio testing with irregular and estimated constraint functions that parallels the theory for parametric models and is tailored for semiparametric models. The key is a uniform local asymptotic normality condition for the local empirical likelihood ratio. This condition is shown to hold under mild assumptions on the constraint function. Applications of our results are discussed to inference problems about quantiles under possibly additional information on the underlying distribution and to residual-based inference about quantiles.

AMS 2000 subject classifications: Primary 62G05; secondary 62G10, 62G20.

Keywords and phrases: Irregular and estimated constraints, nuisance parameters, empirical likelihood ratio tests, uniform local asymptotic normality condition.

Received February 2017.

1. Introduction

Let $(\mathcal{X}, \mathcal{S})$ be a measurable space, \mathcal{Q} be a family of probability measures on \mathcal{S} , and κ be a function from \mathcal{Q} onto an open subset Θ of \mathbb{R}^k . Let Z_1, \dots, Z_n be independent and identically distributed \mathcal{X} -valued random variables with an unknown distribution Q belonging to the model \mathcal{Q} . We are interested in inference about the characteristic $\theta = \kappa(Q)$ of Q . Let us look at the following case.

(K0) There is a function u from $\mathcal{X} \times \Theta$ into \mathbb{R}^m , with $m \geq k$, such that, for every R in \mathcal{Q} , the identity

$$\int u(z, \kappa(R)) dR(z) = 0$$

holds and the matrix

$$W(R) = \int u(z, \kappa(R))u^\top(z, \kappa(R)) dR(z)$$

is positive definite.

We refer to u as a *constraint function*. To simplify notation we abbreviate $W(Q)$ by W and set

$$U_n = n^{-1/2} \sum_{j=1}^n u(Z_j, \theta).$$

Let \mathcal{P}_n denote the closed probability simplex of dimension n ,

$$\mathcal{P}_n = \left\{ \pi = (\pi_1, \dots, \pi_n)^\top \in [0, 1]^n : \sum_{j=1}^n \pi_j = 1 \right\}.$$

To construct confidence sets for θ , Owen (1988 [11], 1990 [12], 2001 [13]) introduced the empirical likelihood

$$\mathcal{R}_n(\vartheta) = \sup \left\{ \prod_{j=1}^n n\pi_j : \pi \in \mathcal{P}_n, \sum_{j=1}^n \pi_j u(Z_j, \vartheta) = 0 \right\}, \quad \vartheta \in \Theta,$$

and proved the following theorem.

Theorem 1.1. *Suppose (K0) holds. Then $-2 \log \mathcal{R}_n(\theta)$ is asymptotically chi-square with m degrees of freedom.*

This allowed him to show that the set

$$\mathcal{C}_n = \{ \vartheta \in \Theta : -2 \log \mathcal{R}_n(\vartheta) < \chi_{1-\alpha}^2(m) \},$$

where $\chi_{1-\alpha}^2(m)$ denotes the $(1 - \alpha)$ -quantile of the chi-square distribution with m degrees of freedom, is a confidence set for θ of asymptotic size $1 - \alpha$. Indeed, this follows from

$$P(\theta \in \mathcal{C}_n) = P(-2 \log \mathcal{R}_n(\theta) < \chi_{1-\alpha}^2(m)) \rightarrow 1 - \alpha.$$

Of course, this result can also be used to test whether θ equals some specific value, say θ_0 . The corresponding test

$$\delta_n = \mathbf{1}[-2 \log \mathcal{R}_n(\theta_0) \geq \chi_{1-\alpha}^2(m)]$$

rejects the null hypothesis $H_0 : \theta = \theta_0$ if the test statistic $-2 \log \mathcal{R}_n(\theta_0)$ equals or exceeds $\chi_{1-\alpha}^2(m)$. This test has asymptotic size α .

Soon it was realized that the empirical likelihood can also be used to construct point estimators. Qin and Lawless (1994 [16]) studied the maximum empirical likelihood estimator (MELE)

$$\hat{\theta} = \arg \max_{\vartheta \in \Theta} \mathcal{R}_n(\vartheta).$$

Similar to the classical theory for parametric models, where the behavior of the maximum likelihood estimator is tied to the behavior of the local log-likelihood ratio, the behavior of the empirical likelihood analogs is now linked to the behavior of the *local empirical log-likelihood ratio*

$$\mathcal{L}_n(t) = \log \frac{\mathcal{R}_n(\theta + n^{-1/2}t)}{\mathcal{R}_n(\theta)}, \quad t \in \mathbb{R}^k, \theta + n^{-1/2}t \in \Theta.$$

The local empirical log-likelihood is said to satisfy the *uniform local asymptotic normality* (ULAN) condition if the expansion

$$\sup_{|t| \leq C} |\mathcal{L}_n(t) - t^\top \Gamma_n + 1/2 t^\top J t| = o_P(1) \quad (1.1)$$

holds for all finite constants C , some invertible $k \times k$ dispersion matrix J , and random vectors Γ_n satisfying

$$\Gamma_n \implies N(0, J).$$

Qin and Lawless (1994 [16]) obtain this condition under regularity and integrability conditions on the constraint function u and its partial derivatives with respect to the parameter. Their assumptions imply the following conditions with the matrix A equal to $-E[\dot{u}(Z, \theta)]$, where \dot{u} denotes the derivative of u with respect to the parameter.

(Q1) For each sequence C_n of positive numbers satisfying $C_n = o(n^{1/2})$, one has

$$\sup_{|t| \leq C_n} \frac{1}{n} \sum_{j=1}^n |u(Z_j, \theta + n^{-1/2}t) - u(Z_j, \theta)|^2 = o_P(1).$$

(Q2) There is an $m \times k$ matrix A of full rank k such that the expansion

$$\sup_{|t| \leq C_n} \frac{|n^{-1/2} \sum_{j=1}^n [u(Z_j, \theta + n^{-1/2}t) - u(Z_j, \theta)] + At|}{1 + |t|} = o_P(1)$$

holds for each sequence C_n of positive numbers satisfying $C_n = o(n^{1/2})$.

The conditions (Q1) and (Q2) are also implied by Assumptions 2.1 and 2.2 in Parente and Smith (2011 [14]) who allow for irregular u and treat generalized empirical likelihood. Under their assumptions the matrix A equals the derivative of the map $\vartheta \mapsto -E[u(X, \vartheta)]$ at θ .

Here we shall show that the ULAN condition holds under the following weaker conditions which restrict the sequences C_n in (Q1) and (Q2) to be bounded.

(K1) For every finite constant C , one has

$$D_n(C) = \sup_{|t| \leq C} \frac{1}{n} \sum_{j=1}^n |u(Z_j, \theta + n^{-1/2}t) - u(Z_j, \theta)|^2 = o_P(1).$$

(K2) There is an $m \times k$ matrix A of full rank k such that the expansion

$$\sup_{|t| \leq C} \left| n^{-1/2} \sum_{j=1}^n [u(Z_j, \theta + n^{-1/2}t) - u(Z_j, \theta)] + At \right| = o_P(1)$$

holds for every finite constant C .

Theorem 1.2. *Suppose (K0)–(K2) hold. Then the expansion*

$$\sup_{|t| \leq C} | -2 \log \mathcal{R}_n(\theta + n^{-1/2}t) - (U_n - At)^\top W^{-1}(U_n - At) | = o_P(1)$$

holds for every finite C . Thus the local empirical log-likelihood satisfies the ULAN condition with $J = A^\top W^{-1}A$ and $\Gamma_n = A^\top W^{-1}U_n$.

The expansion (1.1) is critical to the study of maximum empirical likelihood estimation. Assume for the moment that the map $\vartheta \mapsto \mathcal{R}_n(\vartheta)$ attains a maximum on each compact subset of Θ . This is the case when the map is upper semi-continuous or if it takes only finitely many values. Note that the function h defined by

$$\begin{aligned} h(t) &= t^\top \Gamma_n - 1/2 t^\top J t \\ &= 1/2 [\Gamma_n^\top J^{-1} \Gamma_n - (t - J^{-1} \Gamma_n)^\top J (t - J^{-1} \Gamma_n)], \quad t \in \mathbb{R}^k, \end{aligned} \quad (1.2)$$

is uniquely maximized by $\hat{t} = J^{-1} \Gamma_n$. This shows that under the ULAN condition the random function $\vartheta \mapsto \mathcal{R}_n(\vartheta)$ has a local maximum $\hat{\theta}$ such that

$$n^{1/2}(\hat{\theta} - \theta) - J^{-1} \Gamma_n = o_P(1). \quad (1.3)$$

In particular, if $\vartheta \mapsto \mathcal{R}_n(\vartheta)$ has one local maximizer with probability tending to 1, then this local maximizer $\hat{\theta}$ will obey the expansion (1.3). The theory becomes more involved if $\vartheta \mapsto \mathcal{R}_n(\vartheta)$ has several local maxima or if maxima do not exist.

For J and Γ_n of Theorem 1.2, the expansion (1.3) can be written as

$$\hat{\theta} = \theta + (A^\top W^{-1}A)^{-1} A^\top W^{-1} \frac{1}{n} \sum_{j=1}^n u(Z_j, \theta) + o_P(n^{-1/2}). \quad (1.4)$$

If $m = k$, then A will be invertible, and (1.4) simplifies to

$$\hat{\theta} = \theta + \frac{1}{n} \sum_{j=1}^n A^{-1} u(Z_j, \theta) + o_P(n^{-1/2}).$$

We call an estimator $\hat{\theta}$ that satisfies (1.4) *central*. Qin and Lawless (1994 [16]) have shown that central estimators possess some optimality properties, i.e., they are efficient in the model defined by the constraint function u which is the largest model for which (K0) holds.

Qin and Lawless (1994 [16]) claim that the MELE is central under their assumptions. Their proof, however, only establishes that a maximizer in a neighborhood of the true parameter of radius $n^{-1/3}$ has this property, see the proof of their Lemma 1. Parente and Smith (2011 [14]) consider a *compact* parameter space, an interior point θ of this parameter space and show that the MELE is central under their assumptions which imply (Q1) and (Q2). The assumption that the parameter space is compact is mathematically convenient, but typically not realistic in applications. We shall obtain central estimators under the weaker conditions (K1) and (K2) and without compactness, but require the availability of a \sqrt{n} -consistent estimator.

In Section 3 we address two methods of constructing central estimators, namely, *one-step maximum empirical likelihood estimation* and *guided maximum empirical likelihood estimation*. These methods yield the following *constructive* existence result for central estimators.

Theorem 1.3. *Suppose (K0)–(K2) hold, and $\tilde{\theta}$ is a \sqrt{n} -consistent estimator in the sense that $n^{1/2}(\tilde{\theta} - \theta) = O_P(1)$. Then one can construct a central estimator.*

It follows from the previous two theorems that, under the assumptions of Theorem 1.3, every central estimator $\hat{\theta}$ satisfies the expansion

$$-2 \log \mathcal{R}_n(\hat{\theta}) = \tilde{U}_n^\top (I - \Pi_A) \tilde{U}_n + o_P(1)$$

with $\tilde{U}_n = W^{-1/2} U_n$ and Π_A the idempotent matrix

$$\Pi_A = W^{-1/2} A (A^\top W^{-1} A)^{-1} A^\top W^{-1/2}.$$

Since the m -dimensional random vector \tilde{U}_n is asymptotically standard normal, we see that $-2 \log \mathcal{R}_n(\hat{\theta})$ is asymptotically chi-square with $m - k$ degrees of freedom provided m is greater than k . If m equals k , then $-2 \log \mathcal{R}_n(\hat{\theta})$ converges to zero in probability. For $m > k$, a similar result has been proved in Corollary 4 by Qin and Lawless (1994 [16]) under their regularity assumptions and more recently for the irregular case under stronger conditions in Theorem 1 by Lopez, Van Keilegom and Veraverbeke (2009 [7]) and in Theorem 3.1 by Parente and Smith (2011 [14]). We avoid some of the difficulties by working with central estimators instead of the maximum empirical likelihood estimator. The latter satisfies the expansion (1.4) only under additional requirements such as consistency. Note the simplicity of our conditions as compared to conditions (C0)–(C6) of Lopez, Van Keilegom and Veraverbeke (2009 [7]).

In the above we have focused on maximum empirical likelihood estimation. The key to this was the ULAN condition. As this condition plays a key role in the theory of likelihood ratio tests for parametric models, it should not be surprising that the theory for likelihood ratio testing for parametric model carries over to the empirical likelihood setting. Indeed, Qin and Lawless (1994 [16]) have already discussed this under their sufficient conditions for ULAN. We shall develop the appropriate theory for empirical likelihood ratio testing in Section 4.

So far we have discussed a simple approach to maximum empirical likelihood estimation which generalizes results of Qin and Lawless (1994 [16]) to allow for

irregular constraint functions and relaxes the conditions in Parente and Smith (2011 [14]). Of great interest are extensions to constraint functions that depend on nuisance parameters. Generalizations of Theorem 1.1 that allow for estimated constraint functions have been developed in Hjort, McKeague and Van Keilegom (2009 [3]) and Peng and Schick (2013 [15]). Here we are interested in developing a theory parallel to Theorems 1.2 and 1.3 that allows for constraint functions with estimated nuisance parameters. The theory will be developed in Section 2.

The remainder of this paper is organized as follows. In Section 2 we discuss the case when the constraint function depends on characteristics of the underlying distribution and is thus unknown. We develop a theory parallel to that given in this introduction based on estimates of the unknown constraint function. The key result is Theorem 2.2 which gives the ULAN condition for the local empirical likelihood based on random constraint functions. In Section 3 we address the construction of central estimators in the more general setting of Section 2. Section 4 treats empirical likelihood ratio testing again for random constraint functions. In Section 5 we treat several inference problems related to quantiles as these provide constraints that are not regular. In particular, we treat maximum empirical likelihood estimation of quantiles with and without additional information, and empirical likelihood ratio testing about quantiles and about the equality of median and mean. The results of a simulation study are reported in Section 6, where we compare the behavior of the various constructions of central estimators in small to moderate sample sizes and present simulated significance levels and powers of empirical likelihood ratio tests. Residual-based inference about a quantile is considered in Section 7 for regression models. We first treat linear regression and then discuss how the results carry over to nonparametric and semiparametric regression models. In Section 8 we present a uniform expansion for an abstract general empirical likelihood process. This result is then used to prove Theorem 2.2 and other related expansions.

2. Maximum empirical likelihood estimation in the presence of nuisance parameters

Our goal is to extend the results discussed in the Introduction beyond the basic assumption (K0). We are interested in extensions that allow for nuisance parameters. This is important for applications to semiparametric models. A formulation that allows for this is given next. Again, let m be an integer satisfying $m \geq k$.

- (L0) For every R in \mathcal{Q} there is a function u_R from $\mathcal{Z} \times \Theta$ into \mathbb{R}^m such that the identity

$$\int u_R(z, \kappa(R)) dR(z) = 0$$

holds and the matrix

$$W(R) = \int u_R(z, \kappa(R)) u_R^\top(z, \kappa(R)) dR(z)$$

is positive definite.

Note that (K0) is the special case of (L0) in which $u_R = u$ for all $R \in \mathcal{Q}$. To simplify notation we abbreviate $W(Q)$ by W and set

$$U_n = n^{-1/2} \sum_{j=1}^n u_Q(Z_j, \theta). \quad (2.1)$$

Since u_Q is not known, we work with the modified empirical likelihood

$$\hat{\mathcal{H}}_n(\vartheta) = \sup \left\{ \prod_{j=1}^n n\pi_j : \pi \in \mathcal{P}_n, \sum_{j=1}^n \pi_j \hat{u}_n(Z_j, \vartheta) = 0 \right\}, \quad \vartheta \in \Theta,$$

where \hat{u}_n is an estimator of u_Q based on the observations Z_1, \dots, Z_n . Generalizations of Theorem 1.1 for the modified empirical likelihood have been discussed by Hjort, McKeague and Van Keilegom (2009 [3]) and Peng and Schick (2013 [15]). Possible limit distributions of $-2 \log \hat{\mathcal{H}}_n(\theta)$ do now include generalized chi-square distributions. But direct analogues of Theorem 1.1 are possible if u_R and \hat{u}_n are chosen carefully, see Peng and Schick (2013 [15]).

Theorem 2.1. *Suppose (L0) holds and \hat{u}_n satisfies*

$$\frac{1}{n} \sum_{j=1}^n |\hat{u}_n(Z_j, \theta) - u_Q(Z_j, \theta)|^2 = o_P(1)$$

and

$$n^{-1/2} \sum_{j=1}^n \left(\hat{u}_n(Z_j, \theta) - u_Q(Z_j, \theta) \right) = o_P(1).$$

Then $-2 \log \hat{\mathcal{H}}_n(\theta)$ is asymptotically chi-square with m degrees of freedom.

Next we are looking for a generalization of Theorem 1.2. The corresponding local empirical log-likelihood ratio is

$$\hat{\mathcal{L}}_n(t) = \log \frac{\hat{\mathcal{H}}_n(\theta + n^{-1/2}t)}{\hat{\mathcal{H}}_n(\theta)}, \quad t \in \mathbb{R}^k, \theta + n^{-1/2}t \in \Theta.$$

Motivated by the conditions (K1) and (K2), we introduce the following conditions.

(L1) For every finite constant C one has

$$\hat{D}_n(C) = \sup_{|t| \leq C} \frac{1}{n} \sum_{j=1}^n |\hat{u}_n(Z_j, \theta + n^{-1/2}t) - u_Q(Z_j, \theta)|^2 = o_P(1).$$

(L2) There is an $m \times k$ matrix A of full rank k such that the expansion

$$\sup_{|t| \leq C} \left| n^{-1/2} \sum_{j=1}^n \left(\hat{u}_n(Z_j, \theta + n^{-1/2}t) - u_Q(Z_j, \theta) \right) + At \right| = o_P(1)$$

holds for each finite constant C .

Theorem 2.2. *Suppose (L0) – (L2) hold. Then the expansions*

$$\sup_{|t| \leq C} | -2 \log \hat{\mathcal{K}}_n(\theta + n^{-1/2}t) - (U_n - At)^\top W^{-1}(U_n - At) | = o_P(1)$$

and

$$\sup_{|t| \leq C} | \hat{\mathcal{L}}_n(t) - t^\top A^\top W^{-1}U_n + 1/2t^\top A^\top W^{-1}At | = o_P(1) \tag{2.2}$$

hold for every finite C .

Theorem 1.2 is a special case of this theorem. To see this take $u_R = u$ for all $R \in \mathcal{Q}$ and $\hat{u}_n = u$. Theorem 2.2 lets us also treat the case when (K0) holds, but we want to work with a slightly perturbed version u_n of u . In this case $\hat{u}_n = u_n$ is non-stochastic. In particular, this allows the treatment of smoothed versions of u . If $\Theta = \mathbb{R}^k$, a possible smoothed version is given by $u_n(z, \vartheta) = \int u(z, \vartheta + b_n u) K(u) du$, where K is a kernel and b_n is a bandwidth.

Having obtained the ULAN property for the modified empirical likelihood, the theory for central estimators based on it can be developed as before. Now a central estimator must satisfy the expansion

$$\hat{\theta} = \theta + (A^\top W^{-1}A)^{-1} \frac{1}{n} \sum_{j=1}^n A^\top W^{-1}u_{Q_j}(Z_j, \theta) + o_P(n^{-1/2}). \tag{2.3}$$

The following theorem is a consequence of the results of Section 3.

Theorem 2.3. *Suppose (L0)–(L2) hold, and $\tilde{\theta}$ is a \sqrt{n} -consistent estimator. Then one can construct a central estimator.*

One has to be careful in selecting the functions $\{u_R : R \in \mathcal{Q}\}$ in order to achieve (L2). This will be explained by means of an example in Section 7.

3. On the construction of central estimators

In this section we address the construction of central estimators. We shall restrict our attention to the more general case when the assumptions (L0)–(L2) are met. Results for this case immediately yield results for the case (K0)–(K2); simply take $u_R = u$ and $\hat{u}_n = u$. All our methods require the availability of a preliminary \sqrt{n} -consistent estimator of θ . Thus throughout this section we shall always assume that the following condition is met.

- (A) The conditions (L0)–(L2) hold and $\tilde{\theta}$ is a \sqrt{n} -consistent estimator, i.e. $n^{1/2}(\tilde{\theta} - \theta) = O_P(1)$.

We abbreviate $W(Q)$ from (L0) by W and let U_n be the random vector defined in (2.1). It follows from Theorem 2.2 that the ULAN condition holds with $\Gamma_n = A^\top W^{-1}U_n$ and $J = A^\top W^{-1}A$.

We begin with a simple observation. A $n^{1/2}$ -consistent (generalized) MELE is central.

Lemma 3.1. *Suppose condition (A) holds and $\tilde{\theta}$ is a generalized MELE, i.e. $\tilde{\theta}$ satisfies*

$$\hat{\mathcal{R}}_n(\tilde{\theta}) \geq e^{-1/n} \sup_{\vartheta \in \Theta} \hat{\mathcal{R}}_n(\vartheta).$$

Then $\tilde{\theta}$ is central.

Proof. We need to show $\hat{\Delta} = n^{1/2}(\tilde{\theta} - \theta) - J^{-1}\Gamma_n = o_P(1)$. Let C be a constant. Then, on the event $A_n = \{n^{1/2}|\tilde{\theta} - \theta| \leq C\} \cap \{|J^{-1}\Gamma_n| \leq C\}$, we derive from the ULAN condition and the identity (1.2) the expansions

$$B_{n1} = \hat{\mathcal{L}}_n(n^{1/2}(\tilde{\theta} - \theta)) = (1/2)[\Gamma_n J^{-1} \Gamma_n - \hat{\Delta}^\top J^{-1} \hat{\Delta}] + o_P(1)$$

and

$$B_{n2} = \sup_{|t| \leq C} \hat{\mathcal{L}}_n(t) = (1/2)\Gamma_n J^{-1} \Gamma_n + o_P(1).$$

On this event, we also have $B_{n1} \geq B_{n2} - 1/n$, and therefore $\mathbf{1}[A_n] \hat{\Delta} = o_P(1)$ by the positive definiteness of J . Since this holds for every C , we obtain the desired result in view of the $n^{1/2}$ -consistency of $\tilde{\theta}$. \square

The previous lemma was formulated for a generalized MELE, which in contrast to a MELE does always exist. The practical value of this lemma is limited, as it does not provide a method of constructing a $n^{1/2}$ -consistent generalized MELE and hence a central estimator. Explicit methods of constructing central estimators are discussed next.

Method 1: One-step maximum empirical likelihood estimation. One-step maximum likelihood estimators were introduced by Le Cam (1960 [6]), who showed that such estimators are asymptotically efficient in parametric LAN families. He actually used a discretized preliminary estimator in his construction. Discretization is not needed here, in view of the more stringent ULAN condition.

A one-step maximum empirical likelihood estimator is of the form

$$\tilde{\theta} + n^{-1/2} \tilde{J}^{-1} \tilde{\Gamma}_n$$

where \tilde{J} is a consistent estimator of $J = A^\top W^{-1} A$, i.e.,

$$\tilde{J} = J + o_P(1) \tag{3.1}$$

and the random vector $\tilde{\Gamma}_n$ obeys the expansion

$$\tilde{\Gamma}_n = \Gamma_n - J\Delta + o_P(1) \tag{3.2}$$

with $\Gamma_n = A^\top W^{-1} U_n$ and $\Delta = n^{1/2}(\tilde{\theta} - \theta)$. It is easy to see that such an estimator is central.

There are several ways to construct the quantities \tilde{J} and $\tilde{\Gamma}_n$. One such method is described in Fabian and Hannan (1982 [1]) who use first and second order differences of the log-likelihood. Here we follow a different approach which uses least squares.

Let us set

$$r(t) = \log \frac{\hat{\mathcal{L}}_n(\tilde{\theta} + n^{-1/2}t)}{\hat{\mathcal{L}}_n(\tilde{\theta})}, \quad t \in \mathbb{R}^k.$$

Then $r(t)$ equals $\hat{\mathcal{L}}_n(\Delta + t) - \hat{\mathcal{L}}_n(\Delta)$ and we obtain the identity

$$r(t) = t^\top (\Gamma_n - J\Delta) - (1/2)t^\top Jt + \eta_n(t)$$

with

$$\sup_{|t| \leq C} |\eta_n(t)| = o_p(1)$$

for every finite C . This identity can be rewritten as

$$r(t) = t^\top (\Gamma_n - J\Delta) + s(t)^\top \text{diag}(-J) + \phi(tt^\top)^\top \phi(-J) + \eta_n(t)$$

with $s(t) = (1/2)(t_1^2, \dots, t_k^2)^\top$, and where, for a $k \times k$ matrix M , $\text{diag}(M)$ denotes the vector formed by the diagonal of M and $\phi(M)$ denotes the vector $(M_{1,2}, \dots, M_{1,k}, M_{2,3}, \dots, M_{2,k}, \dots, M_{k-1,k})^\top$ formed by the upper triangular entries M_{ij} , $1 \leq i < j \leq k$, ordered by row index and then by column index. Now let t_1, \dots, t_L be vectors in \mathbb{R}^k and set

$$\mathbf{r} = \begin{pmatrix} r(t_1) \\ \vdots \\ r(t_L) \end{pmatrix} \quad \mathbf{e} = \begin{pmatrix} \eta_n(t_1) \\ \vdots \\ \eta_n(t_L) \end{pmatrix} \quad V = \begin{pmatrix} \Gamma_n - J\Delta \\ -\text{diag}(J) \\ -\phi(J) \end{pmatrix}$$

and

$$D = \begin{bmatrix} t_1^\top & s^\top(t_1) & \phi^\top(t_1 t_1^\top) \\ \vdots & \vdots & \vdots \\ t_L^\top & s^\top(t_L) & \phi^\top(t_L t_L^\top) \end{bmatrix}.$$

Then we have the identity $\mathbf{r} = DV + \mathbf{e}$. Assume now that the matrix D has full rank $K = k+k+k(k-1)/2 = k(k+3)/2$. Then the minimizer $\hat{b} = (D^\top D)^{-1} D^\top \mathbf{r}$ of $\|\mathbf{r} - Db\|^2$ satisfies $\hat{b} = V + (D^\top D)^{-1} D^\top \mathbf{e} = V + o_p(1)$. Thus the requirements (3.1) and (3.2) are met if we take $\tilde{\Gamma}_n = (\hat{b}_1, \dots, \hat{b}_k)^\top$ and \tilde{J} the symmetric matrix with diagonal $-(\hat{b}_{k+1}, \dots, \hat{b}_{2k})^\top$ and upper triangular part formed by the last $K - 2k$ entries of $-\hat{b}$ in an obvious way. We refer to the one-step MELE with these choices of \tilde{J} and $\tilde{\Gamma}_n$ as the *least squares* one-step MELE, short LSMELE.

Let us summarize our findings in the following theorem.

Theorem 3.1. *Suppose condition (A) holds and the vectors t_1, \dots, t_L are chosen such that the matrix D has full rank K . Then the LSMELE is central.*

In the case $k = 1$, we can take t_1, \dots, t_L to be distinct non-zero numbers that satisfy $t_1^3 + \dots + t_L^3 = 0$. Then the LSMELE takes on the simple form

$$\tilde{\theta} - n^{-1/2} \frac{\sum_{l=1}^L t_l r(t_l) \sum_{l=1}^L t_l^4}{2 \sum_{l=1}^L t_l^2 r(t_l) \sum_{l=1}^L t_l^2}.$$

The above theorem remains true even if the vectors t_1, \dots, t_L are replaced by random vectors which are bounded in probability and for which $D^\top D$ is positive definite. Moreover L , the number of vectors, can be made random as long as we keep it bounded by a fixed number L_0 .

Method 2. Guided maximum empirical likelihood estimation using one-step estimators. Let $\hat{\theta}_*$ denote a one-step MELE such as the LSMELE. Although this estimator is central under condition (A), we might want to slightly modify it to resemble more a MELE. Roughly speaking our second method works with an approximate maximizer of $\hat{\mathcal{H}}_n(\vartheta)$ in a ball of radius $cn^{-1/2}$ centered at the one-step MELE. More precisely, we call an estimator $\hat{\theta}$ that satisfies $n^{1/2}|\hat{\theta} - \hat{\theta}_*| \leq c$ and

$$\hat{\mathcal{H}}_n(\hat{\theta}) \geq e^{-1/n} \sup_{|\vartheta - \hat{\theta}_*| \leq cn^{-1/2}} \hat{\mathcal{H}}_n(\vartheta)$$

for some (small) positive c a (generalized) maximum empirical likelihood estimator guided by a one-step estimator, short GOMELE. If the map $\vartheta \mapsto \hat{\mathcal{H}}_n(\vartheta)$ is upper semi-continuous, then we can take $\hat{\theta}$ to be a maximizer of $\hat{\mathcal{H}}_n$ on the random set $\{\vartheta \in \Theta : |\vartheta - \hat{\theta}_*| \leq c\}$. Such a maximizer may no longer exist if we do not have upper semi-continuity. In this case, we need to work with the more general definition.

Theorem 3.2. *Under condition (A) every GOMELE is central.*

Proof. Let us set $\Delta_* = n^{1/2}(\hat{\theta}_* - \theta)$ and $\Delta = n^{1/2}(\hat{\theta} - \hat{\theta}_*)$. Then we have

$$\hat{\mathcal{L}}_n(\Delta_* + \Delta) \geq \sup_{|t| \leq c} \hat{\mathcal{L}}_n(\Delta_* + t) - 1/n.$$

Since Δ_* is bounded in probability, the ULAN condition implies

$$\sup_{|t| \leq c} |\hat{\mathcal{L}}_n(\Delta_* + t) - h(\Delta_* + t)| = o_P(1)$$

with h as in (1.2). Since $\hat{\theta}_*$ is central, we have $\Delta_* - J^{-1}\Gamma_n = o_P(1)$ and thus

$$\sup_{|t| \leq c} |h(\Delta_* + t) - 1/2\Gamma_n^\top J\Gamma_n + 1/2t^\top Jt| = o_P(1).$$

Thus we have the expansion

$$\sup_{|t| \leq c} |\hat{\mathcal{L}}_n(\Delta_* + t) - 1/2\Gamma_n^\top J\Gamma_n + 1/2t^\top Jt| = o_P(1).$$

From this and the invertibility of J we immediately conclude the desired result $\Delta = o_P(1)$. \square

Method 3: Guided maximum empirical likelihood estimation using a $n^{1/2}$ -consistent estimator. Guided (generalized) maximum empirical likelihood estimation can also be done using the $n^{1/2}$ -consistent estimator $\hat{\theta}$ rather than the one-step estimator. This, however, requires a larger neighborhood and a stronger version of the ULAN condition.

Theorem 3.3. *Let condition (A) hold and let C_n be a sequence of positive numbers tending to infinity and satisfying $C_n = o(n^{1/2})$. Suppose that*

$$\sup_{|t| \leq 2C_n} \frac{|\hat{\mathcal{L}}_n(t) - t^\top \Gamma_n + 1/2t^\top Jt|}{(1 + |t|)^2} = o_P(1) \tag{3.3}$$

holds. Then an estimator $\hat{\theta}$ that satisfies $n^{1/2}|\hat{\theta} - \tilde{\theta}| \leq C_n$ and

$$\hat{\mathcal{H}}_n(\hat{\theta}) \geq e^{-1/n} \sup_{n^{1/2}|\vartheta - \tilde{\theta}| \leq C_n} \hat{\mathcal{H}}_n(\vartheta)$$

is central.

Proof. Set $\tilde{\Delta} = n^{1/2}(\tilde{\theta} - \theta)$. Then, with h as in (1.2) and λ the smallest eigenvalue of J , we have

$$\begin{aligned} \hat{\mathcal{L}}_n(\tilde{\Delta} + t) &= h(\tilde{\Delta} + t) + R_n(t) \\ &\leq \frac{1}{2}[\Gamma_n^\top J^{-1} \Gamma_n - \lambda|\tilde{\Delta} + t - J^{-1} \Gamma_n|^2] + R_n(t) \\ &\leq \frac{1}{2}[\Gamma_n^\top J^{-1} \Gamma_n - \lambda(|t| - |J^{-1} \Gamma_n - \tilde{\Delta}|)^2] + R_n(t) \\ &\leq \frac{1}{2}[-\lambda|t|^2 + 2\lambda|t||J^{-1} \Gamma_n - \tilde{\Delta}| + \Gamma_n^\top J^{-1} \Gamma_n] + R_n(t) \end{aligned}$$

where $\sup_{|t| \leq C_n} |R_n(t)|/(1 + |t|)^2 = o_P(1)$. Let $c_n \leq C_n$ be a sequence that tends to infinity. It is now easy to see that

$$\sup_{c_n \leq |t| \leq C_n} \hat{\mathcal{L}}_n(\tilde{\Delta} + t) \rightarrow -\infty$$

in probability. From this and $\hat{\mathcal{L}}_n(0) = o_P(1)$, we derive $P(n^{1/2}|\hat{\theta} - \tilde{\theta}| > c_n) \rightarrow 0$. Since c_n is arbitrary, we conclude the $n^{1/2}$ -consistency of $\hat{\theta}$. The desired result now follows as in Lemma 3.1. \square

From a practical point it is preferable to work with a very slowly growing C_n , say $C_n = (\log n)^{1/2}$. Sufficient conditions for the strengthened version of ULAN needed in the theorem can be given by strengthening (L0)–(L2). A general result will be given in Section 8. Here we mention the special case for $C_n = (\log n)^{1/2}$.

Lemma 3.2. *Suppose (L0) holds, $E[\log(1 + |u_Q(Z, \theta)|)|u_Q(Z, \theta)|^2]$ is finite, and we have the rates*

$$\sup_{|t| \leq 2(\log n)^{1/2}} \frac{1}{n} \sum_{j=1}^n |\hat{u}_n(Z_j, \theta + n^{-1/2}t) - u_Q(Z_j, \theta)|^2 = o_P((\log n)^{-1}),$$

and

$$\sup_{|t| \leq 2(\log n)^{1/2}} \frac{|n^{-1/2} \sum_{j=1}^n (\hat{u}_n(Z_j, \theta + n^{-1/2}t) - u_Q(Z_j, \theta)) + At|}{1 + |t|} = o_P(1),$$

for an $m \times k$ matrix A of full rank k . Then (3.3) holds with $C_n = (\log n)^{1/2}$.

4. Empirical likelihood ratio testing

In this section we shall discuss empirical likelihood ratio testing. For this we assume again the setting of the introduction and require that (L0)–(L2) hold so that we have the ULAN condition for the likelihood ratio. We do not separately discuss the case for the conditions (K0)–(K2) as this is just the special case with $u_R = u$ for all $R \in \mathcal{Q}$ and $\hat{u}_n = u$.

We begin with a preliminary result. Let us set

$$\tilde{U}_n = n^{-1/2} \sum_{j=1}^n W^{-1/2} u_Q(Z_j, \theta).$$

In view of Theorems 2.2 and 2.3, a central estimator $\hat{\theta}$ satisfies the expansion

$$-2 \log \hat{\mathcal{R}}_n(\hat{\theta}) = \tilde{U}_n^\top (I - \Pi_A) \tilde{U}_n + o_P(1)$$

with Π_A the idempotent matrix

$$\Pi_A = W^{-1/2} A (A^\top W^{-1} A)^{-1} A^\top W^{-1/2}.$$

We are interested in testing the null hypothesis $H_0 : \theta \in \Theta_0$ for some subset Θ_0 of Θ . We assume that Θ_0 is the image $\{\psi(t) : t \in \Delta\}$ of some open subset Δ of \mathbb{R}^l under some injective differentiable function ψ which has derivatives of full rank $l < k$. With Θ_0 we associate the submodel

$$\mathcal{Q}_0 = \{R \in \mathcal{Q} : \kappa(R) \in \Theta_0\}$$

and the functional κ_0 from \mathcal{Q}_0 onto Δ defined by

$$\kappa_0(R) = \psi^{-1}(\kappa(R)), \quad R \in \mathcal{Q}_0,$$

where $\psi^{-1} : \Theta_0 \rightarrow \Delta$ is the inverse map of ψ . Suppose from now on that θ belongs to Θ_0 so that the null hypothesis is true. Then there is a unique τ in Δ such that $\theta = \psi(\tau)$, and the derivative B of ψ at τ has full rank l . We have

$$\psi(\tau + n^{-1/2}s) = \theta + n^{-1/2}Bs + n^{-1/2}t_n$$

with $t_n = n^{1/2}(\psi(\tau + n^{-1/2}s) - \psi(\tau)) - Bs \rightarrow 0$. It is now easy to see that for every finite constant C

$$\sup_{|s| \leq C} \frac{1}{n} \sum_{j=1}^n |\hat{u}_n(Z_j, \psi(\tau + n^{-1/2}s)) - u(Z_j, \psi(\tau))|^2 = o_P(1)$$

and

$$\sup_{|s| \leq C} \left| n^{-1/2} \sum_{j=1}^n (\hat{u}_n(Z_j, \psi(\tau + n^{-1/2}s)) - u(Z_j, \psi(\tau)) + ABs) \right| = o_P(1).$$

Thus the analogues of the conditions (L0)–(L2) hold for the submodel \mathcal{Q}_0 and the functional κ_0 . The roles of $\hat{\mathcal{H}}_n$, θ , and A are now played by $\hat{\mathcal{H}}_n \circ \psi$, $\tau = \kappa_0(Q)$, and AB . Thus Theorem 2.2 yields the expansions

$$\sup_{|s| \leq C} | -2 \log \hat{\mathcal{H}}_n(\psi(\tau + n^{-1/2}s)) - (U_n - ABs)^\top W^{-1}(U_n - ABs) | = o_P(1)$$

and

$$\sup_{|s| \leq C} | \hat{\mathcal{L}}_n(\phi(\tau + n^{-1/2}s)) - s^\top B^\top \Gamma_n + \frac{1}{2} s^\top B^\top J B s | = o_P(1)$$

for every finite C with $\Gamma_n = A^\top W^{-1}U_n$ and $J = A^\top W^{-1}A$. Hence a central estimator $\hat{\tau}$ of τ for the submodel satisfies the expansion

$$\hat{\tau} = \tau + \frac{1}{n} \sum_{j=1}^n M u(Z_j, \theta) + o_P(n^{-1/2}) \tag{4.1}$$

with $M = (B^\top A^\top W^{-1}AB)^{-1} B^\top A^\top W^{-1}$. The delta-method yields the expansion

$$\psi(\hat{\tau}) = \theta + \frac{1}{n} \sum_{j=1}^n B M u(Z_j, \theta) + o_P(n^{-1/2}).$$

Thus we find

$$-2 \log \hat{\mathcal{H}}_n(\psi(\hat{\tau})) = \tilde{U}_n^\top (I - \Pi_{AB}) \tilde{U}_n + o_P(1)$$

with Π_{AB} the idempotent matrix defined by

$$\Pi_{AB} = W^{-1/2} AB (B^\top A^\top W^{-1} AB)^{-1} B^\top A^\top W^{-1/2}.$$

Analogous to the classical likelihood ratio, the *empirical likelihood ratio test* rejects the null hypothesis for small values of the test statistic

$$\frac{\sup_{\vartheta \in \Theta_0} \hat{\mathcal{H}}_n(\vartheta)}{\sup_{\vartheta \in \Theta} \hat{\mathcal{H}}_n(\vartheta)} = \frac{\sup_{t \in \Delta} \hat{\mathcal{H}}_n(\psi(t))}{\sup_{\vartheta \in \Theta} \hat{\mathcal{H}}_n(\vartheta)}.$$

It will be more convenient to work instead with the test statistic

$$T_n = \frac{\hat{\mathcal{H}}_n(\psi(\hat{\tau}))}{\hat{\mathcal{H}}_n(\hat{\theta})},$$

where $\hat{\theta}$ is a central estimator in the full model and $\hat{\tau}$ is a central estimator in the submodel \mathcal{Q}_0 with functional κ_0 . In view of the previous results, we have the expansion

$$-2 \log T_n = \tilde{U}_n^\top (\Pi_A - \Pi_{AB}) \tilde{U}_n + o_P(1).$$

The matrix $\Pi_A - \Pi_{AB}$ is idempotent with trace $k - l$. Thus $-2 \log T_n$ has a limiting chi-square distribution with $k - l$ degrees of freedom. Consequently, the test $\mathbf{1}[-2 \log T_n > \chi_{1-\alpha}^2(k - l)]$ has asymptotic size α . The above shows that the empirical likelihood ratio test behaves like the usual parametric likelihood ratio test.

5. Inference about quantiles

Throughout this section X_1, \dots, X_n are independent copies of a random variable X with distribution function F . We shall focus on inference problems related to quantiles as these provide constraints that are not regular. We let F^{-1} denote the left inverse of F defined by

$$F^{-1}(\gamma) = \inf\{x \in \mathbb{R} : F(x) \geq \gamma\}, \quad 0 < \gamma < 1.$$

We say F is γ -regular, if γ belongs to the interval $(0, 1)$ and F has a positive derivative at $F^{-1}(\gamma)$. If F is γ -regular, then $F^{-1}(\gamma)$ is the unique γ -quantile and the sample γ -quantile \hat{q}_γ obeys the expansion

$$\hat{q}_\gamma = F^{-1}(\gamma) - \frac{1}{n} \sum_{j=1}^n \frac{\mathbf{1}[X_j \leq F^{-1}(\gamma)] - \gamma}{F'(F^{-1}(\gamma))} + o_P(n^{-1/2}).$$

In the following examples the verification of the conditions (K1) and (K2) will rely on the following well known result.

Lemma 5.1. *Let C be a finite constant and q be a real number. Then we have*

$$\sup_{|t| \leq C} \frac{1}{n} \sum_{j=1}^n |\mathbf{1}[X_j \leq q + n^{-1/2}t] - \mathbf{1}[X_j \leq q]|^2 = o_P(1)$$

if F is continuous at q , and have

$$\sup_{|t| \leq C} \left| n^{-1/2} \sum_{j=1}^n \mathbf{1}[X_j \leq q + n^{-1/2}t] - \mathbf{1}[X_j \leq q] - tF'(q) \right| = o_P(1)$$

if F is differentiable at q .

Example 5.1. Let us assume that F is γ -regular. We want to estimate the γ -quantile $\theta = F^{-1}(\gamma)$ of F using the empirical likelihood approach. Since F is continuous, θ satisfies $E[\mathbf{1}[X_1 \leq \theta]] = \gamma$. This suggests to look at the empirical likelihood

$$\mathcal{R}_n(q) = \sup \left\{ \prod_{j=1}^n n\pi_j : \pi \in \mathcal{P}_n, \sum_{j=1}^n \pi_j (\mathbf{1}[X_j \leq q] - \gamma) = 0 \right\}, \quad q \in \mathbb{R}.$$

In view of Lemma 5.1 and the γ -regularity of F , the conditions (K0)–(K2) hold with $u(z, \vartheta) = \mathbf{1}[z \leq \vartheta] - \gamma$, $W = \gamma(1 - \gamma)$ and $A = -F'(\theta)$. Thus we obtain from Theorem 1.3 that a GOMELE $\hat{\theta}$ satisfies

$$\hat{\theta} = \theta - \frac{1}{n} \sum_{j=1}^n \frac{\mathbf{1}[X_j \leq \theta] - \gamma}{F'(\theta)} + o_P(n^{-1/2})$$

and hence is asymptotically equivalent to the sample quantile.

Here we have an explicit formula for the empirical likelihood,

$$\mathcal{R}_n(q) = \mathbf{1}[0 < N(q) < n] \left(\frac{n\gamma}{N(q)}\right)^{N(q)} \left(\frac{n(1-\gamma)}{n-N(q)}\right)^{n-N(q)}, \quad q \in \mathbb{R},$$

where $N(q) = \sum_{j=1}^n \mathbf{1}[X_j \leq q]$. This follows as in Owen (2001 [13]), page 43, who considered a slightly modified version. From the formula we derive the identity

$$\mathcal{R}_n(\vartheta) = \sum_{j=1}^{n-1} \mathbf{1}[X_{(j)} \leq \vartheta < X_{(j+1)}] g_\gamma(j, n),$$

where $X_{(1)}, \dots, X_{(n)}$ are the order statistics and

$$g_\gamma(x, y) = y^y \left(\frac{1-\gamma}{y-x}\right)^{y-x} \left(\frac{\gamma}{x}\right)^x, \quad 0 < x < y.$$

It is easy to check that the function $x \mapsto g_\gamma(x, n)$ is increasing on the interval $(0, n\gamma]$ and decreasing on the interval $[n\gamma, n)$. This shows that, almost surely, the function $\vartheta \mapsto \mathcal{R}_n(\vartheta)$ is piecewise constant, non-decreasing on $(-\infty, X_{(k_n+1)})$ and non-increasing on $[X_{(k_n+1)}, \infty)$, where k_n is the integer part of $n\gamma$. Thus a MELE is given by

$$\mathbf{1}[g_\gamma(k_n, n) \geq g_\gamma(k_n + 1, n)]X_{(k_n)} + \mathbf{1}[g_\gamma(k_n, n) < g_\gamma(k_n + 1, n)]X_{(k_n+1)}.$$

Remark 5.1. For $\gamma = 1/2$, we can also use the empirical likelihood

$$\mathcal{R}_n(q) = \sup \left\{ \prod_{j=1}^n n\pi_j : \pi \in \mathcal{P}_n, \sum_{j=1}^n \pi_j \text{sign}(X_j - q) = 0 \right\}, \quad q \in \mathbb{R},$$

with $\text{sign}(x) = \mathbf{1}[x > 0] - \mathbf{1}[x < 0]$. Then the conditions (K0)–(K2) hold with $u(z, \vartheta) = \text{sign}(z - \vartheta)$, $W = 1$ and $A = 2F'(\theta)$. Thus a GOMELE obeys the expansion

$$\hat{\theta} = \theta + \frac{1}{n} \sum_{j=1}^n \frac{\text{sign}(X_j - \theta)}{2F'(\theta)} + o_P(n^{-1/2}).$$

It is easy to show that the sample median is a MELE.

The above example is easily extended to cover the simultaneous estimation of several quantiles. Let us sketch this briefly.

Example 5.2. Let $0 < \gamma_1 < \dots < \gamma_m < 1$ and assume that F is γ_i -regular for $i = 1, \dots, m$. We set $\theta = (\theta_1, \dots, \theta_m)^\top$ with $\theta_i = F^{-1}(\gamma_i)$ for $i = 1, \dots, m$. Note that θ belongs to $\Theta = \{\vartheta \in \mathbb{R}^m : \vartheta_1 < \dots < \vartheta_m\}$. We can work with the empirical likelihood

$$\mathcal{R}_n(\vartheta) = \sup \left\{ \prod_{j=1}^n n\pi_j : \pi \in \mathcal{P}_n, \sum_{j=1}^n \pi_j \begin{pmatrix} \mathbf{1}[X_j \leq \vartheta_1] - \gamma_1 \\ \vdots \\ \mathbf{1}[X_j \leq \vartheta_m] - \gamma_m \end{pmatrix} = 0 \right\}, \quad \vartheta \in \Theta.$$

Here (K0)–(K2) hold with $A = -\text{diag}(F'(\theta_1), \dots, F'(\theta_m))$ and W the matrix with entries $W_{ij} = \gamma_i - \gamma_i \gamma_j$ for $1 \leq i \leq j \leq m$. From this and Theorem 1.3 we find that the i -th component $\hat{\theta}_i$ of a GOMELE $\hat{\theta}$ satisfies the expansion

$$\hat{\theta}_i = \theta_i - \frac{1}{n} \sum_{j=1}^n \frac{\mathbf{1}[X_j \leq \theta_i] - \gamma_i}{F'(\theta_i)} + o_P(n^{-1/2}).$$

The above empirical likelihoods can be used to test composite hypothesis about quantiles. We explain this in a concrete example.

Example 5.3. Suppose that F is $i/4$ -regular for $i = 1, 2, 3$. We want to test the null hypothesis $H_0 : \theta_1 + \theta_3 = 2\theta_2$ about the quartiles $\theta_i = F^{-1}(i/4)$, $i = 1, 2, 3$. The empirical likelihood associated with the quartiles is

$$\mathcal{R}_n(\vartheta) = \sup \left\{ \prod_{j=1}^n n\pi_j : \pi \in \mathcal{P}_n, \sum_{j=1}^n \pi_j (\mathbf{1}[X_j \leq \vartheta_i] - i/4) = 0, i = 1, 2, 3 \right\},$$

where ϑ is such that $\vartheta_1 < \vartheta_2 < \vartheta_3$. We take

$$\Theta_0 = \{\psi(t_1, t_2) = (t_2 - t_1, t_2, t_2 + t_1)^\top : t_1 > 0, t_2 \in \mathbb{R}\}.$$

Let now $\hat{\tau}$ be a GOMELE for the empirical likelihood $\mathcal{R}_n(\psi(t))$. Then the test statistic $-2 \log \mathcal{R}_n(\psi(\hat{\tau}))$ has a limiting chi-square distribution with 1 degree of freedom. Consequently, the test $\mathbf{1}[-2 \log \mathcal{R}_n(\psi(\hat{\tau})) \geq \chi_{1-\alpha}^2(1)]$ has asymptotic size α .

Example 5.4. We assume that X has a finite variance σ^2 and its distribution function F has a positive derivative $F'(m_F)$ at its (unique) median m_F . We want to test whether the mean μ_F of F equals the median m_F . For this we look at

$$\mathcal{R}_n(q, r) = \sup \left\{ \prod_{j=1}^n n\pi_j : \pi \in \mathcal{P}_n, \sum_{j=1}^n \pi_j \begin{pmatrix} \text{sign}(X_j - q) \\ X_j - r \end{pmatrix} = 0 \right\}, \quad q, r \in \mathbb{R}.$$

The assumptions (K0)–(K2) hold with

$$u(z, q, r) = \begin{pmatrix} \text{sign}(z - q) \\ z - r \end{pmatrix}, \quad A = \begin{bmatrix} 2F'(m_F) & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad W = \begin{bmatrix} 1 & \rho \\ \rho & \sigma^2 \end{bmatrix},$$

where ρ is the covariance of $\varepsilon = X - \mu_F$ and $\text{sign}(X - m_F)$. This follows from Lemma 5.1 and simple calculations. The map ψ can be taken to be $\psi(t) = (t, t)^\top$ and has derivative $(1, 1)^\top$ of rank 1. It is easy to see that $\mathcal{R}_n(q, r)$ is maximized by \hat{q}, \hat{r} , where \hat{q} is the sample median and \hat{r} is the sample mean and that $\mathcal{R}_n(\hat{q}, \hat{r}) = 1$. The empirical likelihood ratio statistic T_n simplifies to $T_n = \mathcal{R}_n(\hat{\tau}, \hat{\tau})$, where $\hat{\tau}$ is a GOMELE under the null hypothesis of the common value τ of μ_F and m_F , and $-2 \log T_n$ has a limiting chi-square distribution with one degree of freedom. From this we conclude that the test $\mathbf{1}[-2 \log T_n \geq \chi_{1-\alpha}^2(1)]$ has asymptotic size α .

In the next examples we address estimation of a quantile under additional assumptions on the underlying distribution function F .

Example 5.5. Suppose F is γ -regular and has zero mean and finite variance σ^2 . We estimate $\theta = F^{-1}(\gamma)$ using the empirical likelihood

$$\mathcal{R}_n(\vartheta) = \sup \left\{ \prod_{j=1}^n n\pi_j : \pi \in \mathcal{P}_n, \sum_{j=1}^n \pi_j \left(\frac{\mathbf{1}[X_j \leq \vartheta]}{X_j} - \gamma \right) = 0 \right\}, \quad \vartheta \in \mathbb{R}.$$

It is easy to check that (K0)–(K2) hold in this case with

$$u(z, \vartheta) = \begin{pmatrix} \mathbf{1}[z \leq \vartheta] - \gamma \\ z \end{pmatrix}, \quad W = \begin{bmatrix} \gamma(1-\gamma) & \rho \\ \rho & \sigma^2 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} -F'(\theta) \\ 0 \end{bmatrix}$$

where ρ is the covariance between X and $\mathbf{1}[X \leq \theta]$. Thus a GOMELE $\hat{\theta}$ of θ satisfies

$$\hat{\theta} = \theta - \frac{1}{n} \sum_{j=1}^n \frac{\mathbf{1}[X_j \leq \theta] - \gamma - \rho/\sigma^2 X_j}{F'(\theta)} + o_P(n^{-1/2})$$

and has asymptotic variance $(\gamma(1-\gamma) - \rho^2/\sigma^2)/(F'(\theta))^2$.

Example 5.6. Suppose F is γ -regular for some $\gamma \neq 1/2$ and has known median 0. To estimate $\theta = F^{-1}(\gamma)$, we rely on the empirical likelihood

$$\mathcal{R}_n(\vartheta) = \sup \left\{ \prod_{j=1}^n n\pi_j : \pi \in \mathcal{P}_n, \sum_{j=1}^n \pi_j \left(\frac{\mathbf{1}[X_j \leq \vartheta]}{\text{sign}(X_j)} - \gamma \right) = 0 \right\}, \quad \vartheta \in \mathbb{R}.$$

It is easy to check that (K0)–(K2) hold in this case with

$$u(z, \vartheta) = \begin{pmatrix} \mathbf{1}[z \leq \vartheta] - \gamma \\ \text{sign}(z) \end{pmatrix}, \quad W = \begin{bmatrix} \gamma(1-\gamma) & \rho \\ \rho & 1 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} -F'(\theta) \\ 0 \end{bmatrix}$$

where ρ is the covariance between $\text{sign}(X)$ and $\mathbf{1}[X \leq \theta]$. Thus a GOMELE $\hat{\theta}$ of θ satisfies

$$\hat{\theta} = \theta - \frac{1}{n} \sum_{j=1}^n \frac{\mathbf{1}[X_j \leq \theta] - \gamma - \rho \text{sign}(X_j)}{F'(\theta)} + o_P(n^{-1/2})$$

and has asymptotic variance $(\gamma(1-\gamma) - \rho^2)/(F'(\theta))^2$.

6. Simulations

To study the performance of one-step and guided maximum likelihood estimation and of likelihood ratio tests in small to moderate sample sizes we carried out a small simulation study. This was done with the aid of the R package [17]. We used the function `e1m` provided by Art Owen to calculate the log-empirical likelihood.

6.1. Simulations for Example 5.5

We first looked at estimating the median θ when the distribution has known mean zero utilizing the empirical likelihood

$$\mathcal{R}_n(\vartheta) = \sup \left\{ \prod_{j=1}^n n\pi_j : \pi \in \mathcal{P}_n, \sum_{j=1}^n \pi_j \left(\mathbf{1}[X_j \leq \vartheta] - .5 \right) = 0 \right\}, \quad \vartheta \in \mathbb{R}.$$

The theory for this was treated in Example 5.5 where we considered the more general problem of estimating a quantile when the mean is known. We chose this problem as the criterion function is irregular.

Table 1 reports n times the simulated mean square errors of five estimators of θ , sample median $\tilde{\theta}$ (SMED), LSMELE, GOMELE (guided by LSMELE), GMELE (guided by the sample median), and MELE, for four different distributions and four sample sizes ($n = 30, 60, 90, 120$). The distributions chosen were

- (1) the standard normal distribution with density

$$f_1(x) = (2\pi)^{-1/2} \exp(-x^2/2),$$

- (2) the logistic distribution with density

$$f_2(x) = \exp(-x)/(1 + \exp(-x))^2,$$

- (3) the double exponential distribution with density

$$f_3(x) = \exp(-|x|)/2,$$

- (4) a shifted exponential distribution with density

$$f_4(x) = \exp(-x - 1)\mathbf{1}[x + 1 > 0].$$

The respective asymptotic variances of the sample median for these four distributions are $\pi/2 \simeq 1.5708$, 4, 1, and 1, while those of a central estimator are $\pi/2 - 1 = .5708$, $4 - 48(\log(2))^2/\pi^2 \simeq 1.6634$, .5, and $1 - (\log(2))^2 \simeq .5195$. These numbers show that much can be gained from the knowledge that the mean is zero.

The present empirical likelihood is piecewise constant between neighboring order statistics and thus only needs to be calculated for the midpoints between neighboring order statistics. We used this observation to calculate LSMELE, GOMELE, GMELE and MELE. The GOMELE was chosen as a maximizer over a neighborhood of the LSMELE of radius $2.5\hat{\sigma}n^{-1/2}$ and the GMELE over a neighborhood of the sample median of radius $\hat{\sigma} \log(n)n^{-1/2}$, with $\hat{\sigma} = n/(2 \sum_{i=1}^n \mathbf{1}[|X_j - \tilde{\theta}| < .5])$ an estimator of the asymptotic standard deviation of the sample median. Since the maximizers are not unique, we used the midpoint of the maximizing interval. For the LSMELE, we first computed $\tilde{\theta}_i = \tilde{\theta} + n^{-1/2}\hat{\sigma}(-2.55 + i/10)$, $i = 0, \dots, 51$, then associated with $\tilde{\theta}_i$ the average $\tilde{\theta}_i^*$ of the largest observation less than or equal to $\tilde{\theta}_i$ and the smallest observation larger than $\tilde{\theta}_i$, computed $T_i = n^{1/2}(\tilde{\theta}_i^* - \tilde{\theta})$, eliminated duplicate values among

TABLE 1
ESTIMATING MEDIAN WHEN MEAN IS ZERO

density	n	SMED	LSMELE	GOMELE	GMELE	MELE
f_1	30	1.5610	0.3622	0.5626	0.5603	0.5638
	60	1.5574	0.4086	0.5729	0.5533	0.5533
	90	1.5332	0.4576	0.5756	0.5687	0.5687
	120	1.5765	0.4373	0.5723	0.5723	0.5723
f_2	30	3.8309	1.1881	1.6196	1.6197	1.6251
	60	3.9929	1.2723	1.6850	1.6836	1.6857
	90	3.9978	1.3218	1.6533	1.6526	1.6526
	120	3.9320	1.3021	1.6462	1.6462	1.6462
f_3	30	1.2524	0.5445	0.6189	0.6189	0.6198
	60	1.2046	0.5400	0.5727	0.5706	0.5706
	90	1.1891	0.5306	0.5568	0.5557	0.5557
	120	1.1405	0.5202	0.5633	0.5633	0.5633
f_4	30	0.9926	0.3496	0.5240	0.5120	0.5549
	60	0.9635	0.3815	0.5219	0.5125	0.5169
	90	0.9872	0.4013	0.5100	0.5100	0.5104
	120	0.9747	0.4370	0.5353	0.5229	0.5229

Each entry is the sample size times the simulated mean square error for the corresponding estimator, sample size and error density. The results are based on 4000 repetitions.

T_0, \dots, T_{51} resulting in t_1, \dots, t_L distinct values with $L \leq 52$. Here L and the values t_1, \dots, t_L are random.

From Table 1 we see that the LSMELE performs best in all cases considered and that GOMELE, GMELE and MELE perform about the same. The performance of the LSMELE is better than suggested by the asymptotic theory for three of the four densities.

6.2. Simulations for Example 5.6

We also looked at an example with a smooth constraint function, namely estimating the mean when the median is known to be zero using the empirical likelihood

$$\mathcal{R}_n(\vartheta) = \sup \left\{ \prod_{j=1}^n n\pi_j : \pi \in \mathcal{P}_n, \sum_{j=1}^n \pi_j \left(\frac{X_j - \vartheta}{\text{sign}(X_j)} \right) = 0 \right\}, \quad \vartheta \in \mathbb{R}.$$

A \sqrt{n} -consistent estimator is given by the sample mean \bar{X} . Table 2 reports again n times the simulated mean square errors of the sample mean (SM), the LSMELE (guided by the sample mean), and versions of the LSMELE associated with different design points and the corresponding GOMELE for the first three densities which are symmetric about zero and thus have median zero. The respective asymptotic variances of the sample mean for these three distributions are 1 , $\pi^2/3 \simeq 3.290$ and 2 , while those of a central estimator are $1 - 2/\pi \simeq .3634$, 1.3681 and 1 . These numbers show that much can be gained from the knowledge that the median is zero.

TABLE 2
ESTIMATING MEAN WHEN MEDIAN IS ZERO

density	n	SM	L(1)	G(1)	L(1.5)	G(1.5)	L(2)	G(2)	GMELE
f_1	50	1.018	0.398	0.384	0.397	0.384	0.400	0.384	0.384
	100	1.016	0.375	0.366	0.375	0.366	0.378	0.365	0.366
f_2	50	3.239	1.459	1.437	1.504	1.438	1.580	1.439	1.439
	100	3.268	1.454	1.372	1.465	1.373	1.504	1.374	1.374
f_3	50	1.983	1.000	1.014	1.058	1.014	1.138	1.014	1.014
	100	2.006	1.012	0.999	1.042	0.999	1.097	0.999	0.999

Each entry is the sample size times the simulated mean square error for the corresponding estimator, sample size and error density. The results are based on 4000 repetitions.

We ran the simulations for the sample sizes $n = 50$ and $n = 100$ and used 4000 iterations in each case. The GMELE was found via a grid search using the grid $\{\bar{X} + i\hat{\sigma}C_n/(100\sqrt{n}) : i = -100, \dots, 100\}$ with $\hat{\sigma}$ the sample standard deviation and $C_n = 4 + \sqrt{\log n}$. For $a = 1, 1.5, 2$, we used the design points $\{t_i = i\hat{\sigma}C_n/100, i = -a_n, \dots, a_n\}$ where a_n is the integer closest to $100a/C_n$ to compute the LSMELE $L(a)$ and used a grid search with the above grid points within $2\hat{\sigma}/\sqrt{n}$ units of $L(a)$ to find the GOMELE $G(a)$.

From Table 2 we see that the performance of the GOMELE's and the GMELE are the same, and that the performance of the LSMELE's is slightly worse and seems to be better for smaller a .

6.3. Simulations for Example 5.3

Here we report simulation results for the empirical likelihood ratio test described in Example 5.3 which addresses testing the null hypothesis $F^{-1}(1/4) + F^{-1}(3/4) - 2F^{-1}(1/2) = 0$. The data were generated from $F = F_0$ for computing the significance level and from the mixture distribution $F = .65F_0 + .35G_\beta$ for computing the power. The distribution F_0 was taken to be symmetric about zero so that the null hypothesis was met by $F = F_0$. The distribution G_β was taken from a parametric family with parameter β . The parameter β was selected so that the difference $\delta = F^{-1}(1/4) + F^{-1}(3/4) - 2F^{-1}(1/2)$ took the values .4, .6 and .8. We worked with three choices for F_0 , the standard normal distribution $N(0, 1)$, the Cauchy distribution $\text{Cau}(0)$, and the Laplace distribution $\text{Lap}(0)$. We picked six choices for G_β , namely, the Cauchy distribution $\text{Cau}(\beta)$ with location parameter β , the exponential distribution $\text{Exp}(\beta)$ with rate parameter β , the Laplace distribution $\text{Lap}(\beta)$ with location parameter β , the Logistic distribution $\text{Logis}(\beta)$ with location parameter β , the normal distribution $N(\beta, 1)$ with mean β and variance 1, and the uniform distribution $\text{Unif}(\beta, 2\beta)$.

The R function `constrOptim` was used to compute the two-dimensional GMELE of the parameter $\tau = (\tau_1, \tau_2)$. Here the arguments `ui`, `ci` (the constraint matrix and constraint vector) of the R function `constrOptim` were set to ensure that the search region is $0 < \tau_1 \leq \tilde{\tau}_1 + 3\sqrt{\log(n)/n}$ and $\tilde{\tau}_2 - 3\sqrt{\log(n)/n} \leq \tau_2 \leq$

TABLE 3

Simulated significance levels and powers of the EL test about $H_0 : \delta = F^{-1}(1/4) + F^{-1}(3/4) - 2F^{-1}(1/2) = 0$ at the nominal level .05 for the sample sizes $n = 120, 140, 160$. Data were generated from $F = F_0$ and $F = 0.65F_0 + 0.35G_\beta$.

G_β	δ	$F_0 = N(0, 1)$			$F_0 = \text{Cau}(0)$			$F_0 = \text{Lap}(0)$		
		120	140	160	120	140	160	120	140	160
	0.0	.046	.030	.020	.101	.063	.046	.039	.028	.027
Cau(β)	0.4	.169	.198	.222	.088	.101	.103	.275	.301	.354
	0.6	.284	.348	.397	.172	.175	.191	.460	.560	.615
	0.8	.436	.509	.560	.234	.292	.337	.626	.723	.790
Exp(β)	0.4	.305	.283	.282	.180	.174	.188	.240	.264	.309
	0.6	.400	.416	.452	.261	.281	.313	.391	.430	.514
	0.8	.536	.535	.565	.380	.404	.439	.487	.574	.620
Lap(β)	0.4	.137	.151	.194	.166	.145	.124	.129	.171	.184
	0.6	.264	.319	.356	.222	.228	.201	.230	.274	.329
	0.8	.429	.504	.553	.290	.311	.327	.369	.419	.483
Logis(β)	0.4	.224	.221	.274	.125	.114	.131	.187	.203	.242
	0.6	.357	.406	.451	.174	.199	.238	.305	.349	.420
	0.8	.467	.542	.599	.260	.308	.352	.435	.504	.556
N($\beta, 1$)	0.4	.153	.184	.197	.106	.093	.105	.139	.129	.156
	0.6	.286	.314	.348	.186	.175	.199	.197	.240	.263
	0.8	.436	.487	.559	.296	.278	.285	.313	.331	.409
Unif($\beta, 2\beta$)	0.4	.522	.513	.510	.621	.646	.675	.390	.355	.334
	0.6	.532	.533	.574	.655	.655	.694	.456	.449	.445
	0.8	.613	.629	.657	.636	.671	.668	.535	.569	.575

$\tilde{\tau}_2 + 3\sqrt{\log(n)/n}$, where $\tilde{\tau}_1$ was chosen to be the difference of the third and first sample quartile, and $\tilde{\tau}_2$ was taken to be the sample median.

Table 3 reports the simulated significance level and power of the test given in Example 5.3. The results are based on 2000 repetitions and the sample sizes $n = 120, 140$ and 160.

6.4. Simulations for Example 5.4

Here we report simulation results for the empirical likelihood ratio test described in Example 5.4 to test for the equality of mean and median. The data were generated from F_0 for computing the significance level and from the mixture distribution $F = 0.95F_0 + 0.05G_\beta$ for computing the power. Here again F_0 is a symmetric distribution and G_β comes from a parametric model. We selected F_0 to be the standard normal distribution $N(0, 1)$, the standard logistic distribution $\text{Logis}(0, 1)$, the t -distribution $t(4)$ with 4 degrees of freedom, and the slash t -distribution $\text{SLt}(4, 5)$ with 4 degrees of freedom and tail index 5 (stochastically equivalently, $\text{SLt}(4, 5) = t(4)/\text{Unif}(0, 1)^{1/5}$). We used two choices for G_β , the exponential distribution $\text{Exp}(\beta)$ with rate β and the Gamma distribution $\text{Gam}(\beta, 5)$ with rate β and shape equal to 5. The parameter β was selected so that the difference $\delta = \mu_F - m_F$ took the values .4, .6 and .8.

The R function `optimize` was used to find the one-dimensional GMELE with the argument interval having endpoints $\tilde{\tau} \pm \tilde{\sigma}\sqrt{\log(n)}$, where $\tilde{\tau}$ is the average

of the sample mean and sample median, and $\tilde{\sigma}$ is the jackknife estimator of the standard error of $\tilde{\tau}$.

Table 4 reports the simulated significance level and power of the empirical likelihood test from Example 5.4. Also listed in the table are the values of β which correspond to the values of δ . The results are based on 2000 repetitions and for the sample sizes $n = 40, 80$ and 120 .

TABLE 4
Simulated significance level and power of the EL test about $H_0 : \delta = \mu_F - \text{med}_F = 0$ at the nominal level .05 for several sample sizes n . Data were generated from F_0 and from the contaminated symmetric distribution $F = 0.95F_0 + 0.05F_1$.

F_0	δ	$F_1 = \text{Exp}(\beta)$				$F_1 = \text{Gam}(\beta, 5)$			
		40	80	120	β	40	80	120	β
$N(0, 1)$	0.0	.046	.044	.049		.040	.050	.047	
	0.2	.150	.261	.380	0.19	.170	.308	.420	0.94
	0.4	.310	.509	.672	0.11	.353	.580	.758	0.54
	0.6	.404	.642	.806	0.08	.492	.741	.878	0.36
$\text{Logis}(0, 1)$	0.0	.036	.051	.048		.039	.040	.043	
	0.2	.096	.145	.196	0.14	.088	.158	.218	0.70
	0.4	.194	.340	.464	0.09	.230	.391	.550	0.45
	0.6	.298	.476	.654	0.07	.329	.570	.731	0.33
$\text{SLt}(4, 5)$	0.0	.056	.045	.048		.052	.043	.044	
	0.2	.130	.200	.251	0.17	.114	.222	.314	0.85
	0.4	.241	.381	.520	0.10	.278	.464	.612	0.50
	0.6	.358	.564	.688	0.07	.400	.638	.789	0.40
$t(4)$	0.0	.049	.044	.048		.034	.040	.044	
	0.2	.156	.228	.314	0.19	.149	.260	.357	0.92
	0.4	.278	.475	.610	0.11	.308	.540	.722	0.53
	0.6	.408	.612	.768	0.07	.445	.704	.852	0.37

7. Residual-based inference about a quantile

Let Z_1, \dots, Z_n be independent replicas of the random vector $Z = (X^\top, Y)^\top$ which forms the linear regression model

$$Y = \beta_0 + \beta_1^\top X + \varepsilon,$$

where ε and X are independent, X has a positive definite dispersion matrix, and ε has mean zero, a finite variance σ^2 , and a uniformly continuous density f with $\{f > 0\}$ an interval. We are interested in estimating the γ -quantile θ of ε for some $0 < \gamma < 1$. If the error variables $\varepsilon_1, \dots, \varepsilon_n$ were observable, we could work with the empirical likelihood from Example 5.5

$$\mathcal{R}_n(\vartheta) = \sup \left\{ \prod_{j=1}^n n\pi_j : \pi \in \mathcal{P}_n, \sum_{j=1}^n \pi_j \begin{pmatrix} \mathbf{1}[\varepsilon_j \leq \vartheta] - \gamma \\ \varepsilon_j \end{pmatrix} = 0 \right\}$$

which takes into account the fact that the errors are centered. A naive approach would now be to replace the unobservable error variables by the residuals $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$, based on the least squares approach. While this choice yields

the desired (L1), it does not produce (L2). This follows from the fact that the sum of the residuals is zero. We should also point out the following additional properties of the residuals,

$$M_n = \max_{1 \leq j \leq n} |\hat{\varepsilon}_j - \varepsilon_j| = o_P(1) \quad \text{and} \quad D_n = \frac{1}{n} \sum_{j=1}^n |\hat{\varepsilon}_j - \varepsilon_j|^2 = O_P(n^{-1}).$$

To find an appropriate choice of u_Q , we start with the fact that the least squares residuals satisfy the property

$$\sup_{t \in \mathbb{R}} \left| \frac{1}{n} \sum_{j=1}^n \left(\mathbf{1}[\hat{\varepsilon}_j \leq t] - \mathbf{1}[\varepsilon_j \leq t] - f(t)\varepsilon_j \right) \right| = o_P(n^{-1/2}). \quad (7.1)$$

This can be derived from results of Koul (1969 [4]) for the fixed design case and Müller, Schick and Wefelmeyer (2007 [8]) for the random design case used here, see also Remark 2 in Müller, Schick and Wefelmeyer (2009 [9]). The expansion (7.1) and the fact that the residuals sum to zero imply that

$$\sup_{t \in \mathbb{R}} \left| n^{-1/2} \sum_{j=1}^n \left[\left(\mathbf{1}[\hat{\varepsilon}_j \leq t] - F(t) + \hat{f}(t)\hat{\varepsilon}_j \right) - \left(\mathbf{1}[\varepsilon_j \leq t] - F(t) + f(t)\varepsilon_j \right) \right] \right|$$

converges to zero in probability for any estimator \hat{f} of f . This suggests to work with the empirical likelihood

$$\hat{\mathcal{H}}_n(\vartheta) = \sup \left\{ \prod_{j=1}^n n\pi_j : \pi \in \mathcal{P}_n, \sum_{j=1}^n \pi_j \left(\mathbf{1}[\hat{\varepsilon}_j \leq \vartheta] - \gamma + \hat{f}(\vartheta)\hat{\varepsilon}_j \right) = 0 \right\},$$

where \hat{f} is a residual based kernel density estimator of f ,

$$\hat{f}(y) = \frac{1}{nb} \sum_{j=1}^n K\left(\frac{y - \hat{\varepsilon}_j}{b}\right), \quad y \in \mathbb{R},$$

with K a symmetric Lipschitz-continuous density and b a bandwidth satisfying $nb^4 \rightarrow \infty$ and $b \rightarrow 0$. Then the residual-based density estimator \hat{f} is uniformly consistent,

$$\|\hat{f} - f\|_\infty = \sup_{y \in \mathbb{R}} |\hat{f}(y) - f(y)| = o_P(1).$$

This follows from uniform consistency of the error-based kernel estimator \tilde{f} (defined as \hat{f} but with ε_j in place of $\hat{\varepsilon}_j$) and the inequality

$$|\hat{f}(y) - \tilde{f}(y)| \leq \frac{L}{b^2} \frac{1}{n} \sum_{j=1}^n |\hat{\varepsilon}_j - \varepsilon_j| \leq \frac{LD_n^{1/2}}{b^2}$$

with L the Lipschitz constant for K .

Here we have

$$u_Q(Z, \vartheta) = \mathbf{1}[\varepsilon \leq \vartheta] - \gamma + f(\vartheta)\varepsilon$$

and

$$W = \gamma(1 - \gamma) + f^2(\theta)\sigma^2 + 2f(\theta)E[\varepsilon\mathbf{1}[\varepsilon \leq \theta]].$$

Let us now show that (L1) and (L2) hold, the latter with $A = -f(\theta)$. We conclude (L1) from the inequality

$$\begin{aligned} & \sup_{|t| \leq C} \frac{1}{n} \sum_{j=1}^n \left| \mathbf{1}[\hat{\varepsilon}_j \leq \theta + n^{-1/2}t] + \hat{f}(\theta + n^{-1/2}t)\hat{\varepsilon}_j - \mathbf{1}[\varepsilon_j \leq \theta] - f(\theta)\varepsilon_j \right|^2 \\ & \leq \frac{1}{n} \sum_{j=1}^n 3\mathbf{1}[|\varepsilon_j - \theta| \leq M_n + Cn^{-1/2}] + 3D_n \|\hat{f}\|_\infty^2 \\ & \quad + \frac{1}{n} \sum_{j=1}^n 3\varepsilon_j^2 \left(\|\hat{f} - f\|_\infty + \sup_{|t| \leq C} |f(\theta + n^{-1/2}t) - f(\theta)| \right)^2. \end{aligned}$$

Since the residuals sum to zero, the desired (L2) follows from the inequality

$$\begin{aligned} & \left| n^{-1/2} \sum_{j=1}^n \left(\mathbf{1}[\hat{\varepsilon}_j \leq \theta + n^{-1/2}t] - \mathbf{1}[\varepsilon_j \leq \theta] - f(\theta)\varepsilon_j \right) - f(\theta)t \right| \\ & \leq \left| n^{-1/2} \sum_{j=1}^n \left(\mathbf{1}[\hat{\varepsilon}_j \leq \theta + n^{-1/2}t] - \mathbf{1}[\varepsilon_j \leq \theta + n^{-1/2}t] - f(\theta + n^{-1/2}t)\varepsilon_j \right) \right| \\ & \quad + \left| n^{-1/2} \sum_{j=1}^n \left(\mathbf{1}[\varepsilon_j \leq \theta + n^{-1/2}t] - \mathbf{1}[\varepsilon_j \leq \theta] - (F(\theta + n^{-1/2}t) - F(\theta)) \right) \right| \\ & \quad + |n^{1/2}(F(\theta + n^{-1/2}t) - F(\theta)) - tf(\theta)| \\ & \quad + |f(\theta + n^{-1/2}t) - f(\theta)| \left| n^{-1/2} \sum_{j=1}^n \varepsilon_j \right|, \end{aligned}$$

equation (7.1), properties of the empirical process and the uniform continuity of f .

We obtain from Theorem 2.2 that a GOMELE $\hat{\theta}$ obeys the expansion

$$\hat{\theta} = \theta - \frac{1}{n} \sum_{j=1}^n \left(\frac{\mathbf{1}[\varepsilon_j \leq \theta] - \gamma}{f(\theta)} + \varepsilon_j \right) + o_P(1). \quad (7.2)$$

As pointed out in Müller, Schick and Wefelmeyer (2012 [10]), the uniform expansion (7.1) and results of Gill (1989 [2]) yield the expansion

$$\hat{F}^{-1}(\gamma) = \theta - \frac{1}{n} \sum_{j=1}^n \left(\frac{\mathbf{1}[\varepsilon_j \leq \theta] - \gamma}{f(\theta)} + \varepsilon_j \right) + o_P(1)$$

with \hat{F} the residual-based empirical distribution function. This shows that residual-based sample quantiles are central and hence equivalent to the GOMELE.

This implies that a residual-based sample quantile can be used instead of a one-step MELE in the construction of the GOMELE.

The asymptotic variance of the residual-based quantile or a GOMELE is

$$\frac{\gamma(1-\gamma)}{f^2(\theta)} + \sigma^2 + 2\frac{E[\varepsilon\mathbf{1}[\varepsilon \leq \theta]]}{f(\theta)}.$$

The first summand is the asymptotic variance of the sample quantile based on the actual errors. The sum of the last two terms may be negative. Indeed, if f is a centered normal density, then the sum of the last two terms equals $-\sigma^2$. Thus the residual-based sample quantile can have a smaller variance than the sample quantile based on the actual errors.

TABLE 5
COMPARISON OF ESTIMATORS OF THE MEDIAN

n	EBSM	RBSM	G(0.1)	G(0.2)	G(0.3)	G(0.4)	G(0.5)
N(0, 1)							
31	1.5761	0.5406	0.5493	0.5569	0.5631	0.5686	0.5725
61	1.6405	0.5662	0.5723	0.5778	0.5827	0.5864	0.5900
91	1.5767	0.5667	0.5714	0.5757	0.5794	0.5826	0.5847
$t(4, 0, 1/\sqrt{2})$							
31	0.8777	0.4995	0.5108	0.5217	0.5323	0.5420	0.5515
61	0.8927	0.5439	0.5530	0.5619	0.5702	0.5783	0.5862
91	0.9205	0.5543	0.5619	0.5695	0.5768	0.5839	0.5907
Lap(0, 1/ $\sqrt{2}$)							
31	0.6515	0.4804	0.4921	0.5034	0.5142	0.5246	0.5344
61	0.6104	0.4971	0.5054	0.5138	0.5218	0.5296	0.5371
91	0.5768	0.4876	0.4940	0.5005	0.5066	0.5126	0.5187
Logis(0, $\sqrt{3}/\pi$)							
31	1.2148	0.5028	0.5141	0.5242	0.5337	0.5421	0.5488
61	1.2205	0.5256	0.5343	0.5427	0.5507	0.5580	0.5649
91	1.1920	0.5092	0.5166	0.5236	0.5301	0.5365	0.5422

Each entry is the sample size times the simulated mean square error for the corresponding estimator, sample size and error distribution. The results are based on 4000 repetitions.

We performed a simulation study to illustrate the above for the case $\gamma = 1/2$, i.e., for estimating the median. We restricted ourselves to the simple linear regression model $Y = 1 + 2X + \varepsilon$ with X having a normal distribution with mean 1 and variance 1 and considered four symmetric error distributions, namely, $N(0, 1)$, $\text{Lap}(0, 1/\sqrt{2})$, $\text{Logis}(0, \sqrt{3}/\pi)$ and $t(4, 0, 1/\sqrt{2})$ which is the Student t -distribution with 4 degrees of freedom, location parameter 0 and scale parameter $1/\sqrt{2}$. Each of these distributions has variance 1 and median $\theta = F^{-1}(1/2)$ equal to 0. We compared the error-based sample median (EBSM), the residual-based sample median (RBSM), and the modified GOMELE (with the RBSM in place of a one-step MELE)

$$G(c) = \arg \max_{|\vartheta - \tilde{\vartheta}| \leq c\tilde{\sigma}} \hat{\mathcal{R}}_n(\vartheta)$$

for several choices of c . Here $\tilde{\vartheta}$ denotes the RBSM and $\tilde{\sigma}$ denotes the jackknife estimate of the standard error of $\tilde{\vartheta}$. Reported in Table 5 are the simulated mean squared errors multiplied by the sample size n of the estimators EBSM, RBSM and $G(c)$ with $c = .1, .2, .3, .4, .5$. We looked at odd sample sizes since for an even sample size $n = 2m$ the RBSM $(\hat{\varepsilon}_{(m)} + \hat{\varepsilon}_{(m+1)})/2$ is also a MELE. The table shows that the RBSM has smallest simulated mean square error in all cases considered and the mean square errors of the estimators $G(c)$ increase with c . Based on this we recommend the use of the RBSM.

The above results for linear regression carry over to nonparametric regression. Let us explain this in the simplest case when $Z = (X, Y)^\top$ and $Y = r(X) + \varepsilon$, with r a twice continuously differentiable function and X and ε are independent, with ε having mean zero and finite variance, and X is quasi-uniform on the unit interval $[0, 1]$. The latter means X has a density g that is bounded and bounded away from zero on its support $[0, 1]$. Under the additional assumption that the density f is Hölder of order $1/3$ and has a finite moment of order greater than order $8/3$, Müller et al (2007) have shown that there are estimators \hat{r}_n of r such that (7.1) also hold for the nonparametric residuals $\hat{\varepsilon}_j = Y_j - \hat{r}_n(X_j)$. These nonparametric residuals satisfy, for some $\rho > 1/2$,

$$M_n = o_P(1), \quad \sum_{j=1}^n \hat{\varepsilon}_j = o_P(n^{1/2}), \quad \text{and} \quad D_n = o_P(n^{-\rho}). \quad (7.3)$$

It is now easy to check that the kernel density estimator \hat{f} based on these nonparametric residuals is uniformly consistent for f if also $n^\rho b^4 \rightarrow \infty$. One verifies (L1) and (L2) with $A = -f(\theta)$ and again obtains the expansion (7.2) for the corresponding GOMELE.

Expansions to nonparametric regression models with multivariate covariates are possible using the results of Müller et al (2009 [9]). The results in Müller et al (2007 [8], 2012 [10]) can be used to obtain extensions to the partly linear regression model and to the additive nonparametric regression model, while those of Koul, Müller and Schick (2017 [5]) can be used for extensions to single-index models. In all these models one can construct residuals so that (7.1) and (7.3) hold and then obtains the expansion (7.2). We should mention that we do not get the second part of (7.3) for all regression models. This is already so in linear regression *without* an intercept.

Remark 7.1. A key point of the section is that one has to be careful in selecting the constraint function in order to be able to achieve (L1) and (L2). We are not the first to observe this. Zhu and Xue (2006 [20]) have pointed this out in the context of a single index model. Here we look at a more general single-index random-coefficient model, namely

$$Y = \beta^\top (V + \theta^\top U)X + \varepsilon$$

where V is a random variable, U is a k -dimensional random vector, X is a q -dimensional random vector, the error variable is independent of the covariates

(U, V, X) with mean zero and finite positive variance, β is an unknown smooth function from \mathbb{R} into \mathbb{R}^q , and the k -dimensional vector θ is the parameter of interest.

If $q = 1$ and $X = 1$, then this model reduces to the single index model

$$Y = \beta(V + \theta^\top U) + \varepsilon.$$

For this model, Xue and Zhu (2006 [19]) used the constraint function

$$u_R(U, V, Y, t) = (Y - \beta_R(I_t))\beta'_R(I_t)U,$$

with $I_t = V + t^\top U$ and were unable to verify (L1) and (L2). Zhu and Xue (2006 [20]) considered instead the constraint function

$$u_R(U, V, Y, t) = (Y - \beta_R(I_t))\beta'_R(I_t)[U - E_R(U|I_t)]$$

and were able to verify (L1) and (L2).

In the general case, the constraint function

$$u_R(U, V, X, Y, t) = (Y - \beta_R(I_t))\beta'_R(I_t)^\top XU$$

used by Xue and Wang (2012 [18]) is not suitable for obtaining (L1) and (L2). Instead, one should work with the constraint function

$$u_R(U, V, X, Y, t) = (Y - \beta_R(I_t))[\beta'_R(I_t)^\top XU - M_R(I_t)X]$$

where the matrix

$$M_R(I_t) = E_R(\beta'_R(I_t)^\top XU X^\top | I_t) E_R(XX^\top | I_t)^{-1} X$$

is the projection of $\beta'_R(I_t)^\top XU$ onto the space of functions of the form $A(I_t)^\top X$ with A a function into $\mathbb{R}^{k \times q}$ satisfying $E_R[|A(I_t)|^2] < \infty$. Here one needs to assume that the matrix $E_R(XX^\top | I_t)$ is invertible.

8. A general result

Let $\mathbb{T}_{n1}(t), \dots, \mathbb{T}_{nn}(t)$ be m -dimensional random vectors indexed by $t \in \mathbb{R}^k$, where $k \leq m$. Let C_n be a sequence of positive numbers such that $\inf_n C_n > 0$ and $C_n = o(n^{1/2})$. We are interested in the asymptotic behavior of the empirical likelihood process

$$\mathcal{R}_n(t) = \sup \left\{ \prod_{j=1}^n n\pi_j : \pi \in \mathcal{P}_n, \sum_{j=1}^n \pi_j \mathbb{T}_{nj}(t) = 0 \right\}, \quad |t| \leq C_n.$$

To this end we shall use the following result which is a special case of Lemma 5.2 of Peng and Schick (2013 [15]).

Lemma 8.1. *Let x_1, \dots, x_n be m -dimensional vectors. Set*

$$x^* = \max_{1 \leq j \leq n} |x_j|, \quad \bar{x} = \frac{1}{n} \sum_{j=1}^n x_j, \quad S = \frac{1}{n} \sum_{j=1}^n x_j x_j^\top,$$

and let λ denote the smallest and Λ the largest eigenvalue of the matrix S . Then the inequality $\lambda > 5|\bar{x}|x^$ implies*

$$\left| -2 \log \mathcal{R} - n\bar{x}^\top S^{-1} \bar{x} \right| \leq \left(\Lambda + \frac{\Lambda^3}{4\lambda^2} \right) \frac{2n|\bar{x}|^3 x^*}{(\lambda - |\bar{x}|x^*)^3} \tag{8.1}$$

where

$$\mathcal{R} = \sup \left\{ \prod_{j=1}^n n\pi_j : \pi \in \mathcal{P}_n, \sum_{i=1}^n \pi_i x_i = 0 \right\}.$$

Motivated by this we introduce the quantities

$$\mathbb{T}_n^*(t) = \max_{1 \leq j \leq n} |\mathbb{T}_{nj}(t)|, \quad \bar{\mathbb{T}}_n(t) = \frac{1}{n} \sum_{j=1}^n \mathbb{T}_{nj}(t),$$

and

$$\mathbb{S}_n(t) = \frac{1}{n} \sum_{j=1}^n \mathbb{T}_{nj}(t) \mathbb{T}_{nj}^\top(t).$$

We impose the following conditions.

- (B1) $\sup_{|t| \leq C_n} (1 + |t|) \mathbb{T}_n^*(t) = o_P(n^{1/2})$.
- (B2) There is a positive definite $m \times m$ matrix S such that

$$\sup_{|t| \leq C_n} |\mathbb{S}_n(t) - S| = o_P(1).$$

- (B3) There exist k -dimensional random vectors U_n and an $m \times k$ matrix A of full rank k such that $U_n = O_P(1)$ and

$$\sup_{|t| \leq C_n} \frac{|\sqrt{n} \bar{\mathbb{T}}_n(t) - U_n + At|}{1 + |t|} = o_P(1).$$

We have the following result.

Theorem 8.1. *Suppose (B1)–(B3) hold. Then we have*

$$\sup_{|t| \leq C_n} \frac{|-2 \log \mathcal{R}_n(t) - (U_n - At)^\top S^{-1} (U_n - At)|}{(1 + |t|)^2} = o_P(1)$$

and therefore

$$\sup_{|t| \leq C_n} \frac{|\log(\mathcal{R}_n(t)/\mathcal{R}_n(0)) - t^\top A^\top S^{-1} U_n + \frac{1}{2} t^\top A^\top S^{-1} A t|}{(1 + |t|)^2} = o_P(1).$$

Proof. Let $\lambda_n(t)$ and $\Lambda_n(t)$ denote the smallest and largest eigenvalues of $\mathbb{S}_n(t)$. It follows from (B2) that there are constants $0 < \eta < K < \infty$ such that

$$P\left(\sup_{|t| \leq C_n} \Lambda_n(t) > K\right) \rightarrow 0 \quad \text{and} \quad P\left(\inf_{|t| \leq C_n} \lambda_n(t) < \eta\right) \rightarrow 0. \tag{8.2}$$

It follows from (B3) that

$$\sup_{|t| \leq C_n} \frac{|\bar{\mathbb{T}}_n(t)|}{1 + |t|} = O_P(n^{-1/2}).$$

This and (B1) yield

$$\sup_{|t| \leq C_n} \mathbb{T}_n^*(t) |\bar{\mathbb{T}}_n(t)| = o_P(1) \tag{8.3}$$

and

$$\sup_{|t| \leq C_n} \frac{n \mathbb{T}_n^*(t) |\bar{\mathbb{T}}_n(t)|^3}{(1 + |t|)^2} = o_P(1). \tag{8.4}$$

From (8.1) – (8.4) it follows that

$$\sup_{|t| \leq C_n} \frac{|-2 \log \mathcal{R}_n(t) - n \bar{\mathbb{T}}_n(t)^\top \mathbb{S}_n(t)^{-1} \bar{\mathbb{T}}_n(t)|}{(1 + |t|)^2} = o_P(1). \tag{8.5}$$

From (B2) we derive

$$\sup_{|t| \leq C_n} |\mathbb{S}_n(t)^{-1} - S^{-1}| = o_P(1)$$

and thus obtain the expansion

$$\sup_{|t| \leq C_n} \frac{|n \bar{\mathbb{T}}_n(t)^\top \mathbb{S}_n(t)^{-1} \bar{\mathbb{T}}_n(t) - n \bar{\mathbb{T}}_n(t)^\top S^{-1} \bar{\mathbb{T}}_n(t)|}{(1 + |t|)^2} = o_P(1). \tag{8.6}$$

The first conclusion in the theorem follows from (8.6), (8.5) and (B3). The second conclusion is a simple consequence of the first one. \square

From the above we immediately derive the following result which gives sufficient conditions for (3.3). The assumptions used in this result imply (L0)–(L2). We use the notation of Section 2. We need the following stronger version of (L2).

(SL2) For an $m \times k$ matrix A of full rank k we have the expansion

$$\sup_{|t| \leq 2C_n} \frac{\left| \frac{1}{\sqrt{n}} \sum_{j=1}^n \left(\hat{u}_n(Z_j, \theta + n^{-1/2}t) - u_Q(Z_j, \theta) \right) + At \right|}{1 + |t|} = o_P(1).$$

Lemma 8.2. *In addition to (L0) and (SL2) assume the rates*

$$\sup_{|t| \leq 2C_n} \max_{1 \leq j \leq n} C_n |\hat{u}_n(Z_j, \theta + n^{-1/2}t)| = o_P(n^{1/2}), \tag{8.7}$$

$$\sup_{|t| \leq 2C_n} \frac{1}{n} \sum_{j=1}^n |\hat{u}_n(Z_j, \theta + n^{-1/2}t) - u_Q(Z_j, \theta)|^2 = o_P(1). \tag{8.8}$$

Then the expansion

$$\sup_{|t| \leq 2C_n} \frac{|-2 \log \hat{\mathcal{K}}_n(\theta + n^{-1/2}t) - (U_n - At)^\top W^{-1}(U_n - At)|}{(1 + |t|)^2} = o_P(1)$$

holds, and this implies (3.3).

Proof. The desired result follows if we verify the assumptions of Theorem 8.1 with $\mathbb{T}_{n_j}(t) = \hat{u}_n(Z_j, \theta + n^{-1/2}t)$, $S = W$, and U_n as in (2.1) and with C_n replaced by $2C_n$. Note that (B1) follows from (8.7) and (B3) from (SL2) and the central limit theorem. We are left to verify (B2). To simplify notation we set $w(z) = u_Q(z, \theta)$ and $\bar{W}_n = \frac{1}{n} \sum_{j=1}^n w(Z_j)w^\top(Z_j)$. Since $\int |w|^2 dQ$ is finite, we obtain

$$|\bar{W}_n - W_n| = o_P(1). \tag{8.9}$$

Then (B2) follows from (8.8), (8.9) and the bound

$$\begin{aligned} |a^\top (\mathbb{S}_n(t) - \bar{W}_n)a| &= \left| \frac{1}{n} \sum_{j=1}^n (a^\top \hat{u}_n(Z_j, \theta + n^{-1/2}t))^2 - \frac{1}{n} \sum_{j=1}^n (a^\top w(Z_j))^2 \right| \\ &\leq \hat{D}_n + 2 \left(\frac{1}{n} \sum_{j=1}^n |w(Z_j)|^2 \hat{D}_n \right)^{1/2} \end{aligned}$$

valid for every unit vector a in \mathbb{R}^k , every t with $|t| \leq 2C_n$, and with \hat{D}_n the left-hand side of (8.8). The above inequality was already used in Peng and Schick (2013 [15]), see their (4.3). \square

Remark 8.1. Note that we have the bound

$$\sup_{|t| \leq 2C_n} \max_{1 \leq j \leq n} |\hat{u}_n(Z_j, \theta + n^{-1/2}t)| \leq \max_{1 \leq j \leq n} |u_Q(Z_j, \theta)| + n^{1/2} \hat{D}_n^{1/2}.$$

Thus we can replace (8.7) and (8.8) by the conditions

$$U_n^* = \max_{1 \leq j \leq n} |u_Q(Z_j, \theta)| = o_p(n^{1/2}/C_n) \tag{8.10}$$

and

$$\sup_{|t| \leq 2C_n} \frac{1}{n} \sum_{j=1}^n |\hat{u}_n(Z_j, \theta + n^{-1/2}t) - u_Q(Z_j, \theta)|^2 = o_p(C_n^{-2}). \tag{8.11}$$

The first condition can typically be verified under additional moment conditions on $T = |u_Q(Z, \theta)|$. For example, if $E[[T^2 \log(1 + T)]]$ is finite, then we have

$$U_n^* = o_p((n/\log n)^{1/2}).$$

Indeed, with $\eta_n = s(n/\log n)^{1/2}$ and $s > 0$, we have the bound

$$P(U_n^* > \eta_n) \leq nP(T > \eta_n) \leq \frac{nE[T^2 \log(1+T)\mathbf{1}[T > \eta_n]]}{\eta_n^2 \log(1+\eta_n)}$$

which tends to zero by the Lebesgue dominated convergence theorem and the fact that $n/(\eta_n^2 \log(1+\eta_n))$ is bounded. This verifies Lemma 3.2.

PROOF OF THEOREM 2.2. Fix a finite C . The desired result follows from Lemma 8.2 with $C_n = C$. In view of the previous remark it suffices to verify (8.10), (8.11) and (SL2) with $C_n = C$. Since $\int |u_Q(z, \theta)|^2 dQ$ is finite, we obtain

$$\max_{1 \leq j \leq n} |u_Q(Z_j, \theta)| = o_P(n^{1/2}), \quad (8.12)$$

which is (8.10) for $C_n = C$. Of course, for $C_n = C$, (L1) is equivalent to (8.11) and (L2) is equivalent to (SL2).

Acknowledgment

Anton Schick was supported by NSF Grant DMS 0906551. Hanxiang Peng was supported in part by Lilly Endowment, Inc., through its support for the Indiana University Pervasive Technology Institute, and in part by the Indiana METACyt Initiative. The Indiana METACyt Initiative at IU was also supported in part by Lilly Endowment, Inc.

References

- [1] FABIAN, V. and HANNAN, J. (1982). On estimation and adaptive estimation for locally asymptotically normal families. *Z. Wahrsch. Verw. Gebiete* **59** 459–478. [MR0656510](#)
- [2] GILL, R. D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method. I. With a discussion by J. A. Wellner and J. Praestgaard and a reply by the author. *Scand. J. Statist.* **16** 97–128. [MR1028971](#)
- [3] HJORT, N. L., MCKEAGUE, I. W. and VAN KEILEGOM, I. (2009). Extending the scope of empirical likelihood. *Ann. Statist.* **37** 1079–1111. [MR2509068](#)
- [4] KOUL, H. L. (1969). Asymptotic behavior of Wilcoxon type confidence regions in multiple linear regression. *Ann. Statist.* **40** 1950–1979. [MR0260126](#)
- [5] KOUL, H. L., MÜLLER, U.U. and SCHICK, A. (2017). Estimating the error distribution in a single-index model. *From Statistics to Mathematical Finance*, Festschrift in Honour of Winfried Stute (D. Ferger, W. González Manteiga, T. Schmidt, J.-L. Wang, eds.), 209–233, Springer, Cham. [MR3752334](#)
- [6] LE CAM, L. (1960). Locally asymptotically normal families of distributions. *Univ. California Publ. Statist.* **3** 37–98. [MR0126903](#)

- [7] MOLANES LOPEZ, E. M., VAN KEILEGOM, I. and VERAVERBEKE, N. (2009). Empirical likelihood for non-smooth criterion functions. *Scand. J. Statist.* **36** 413–432. [MR2549702](#)
- [8] MÜLLER, U. U., SCHICK A. and WEFELMEYER, W. (2007). Estimating the error distribution function in semiparametric regression. *Statist. Decisions* **25** 1–18. [MR2370101](#)
- [9] MÜLLER, U. U., SCHICK A. and WEFELMEYER, W. (2009). Estimating the error distribution function in nonparametric regression with multivariate covariates. *Statist. Probab. Lett.* **79** 957–964. [MR2509488](#)
- [10] MÜLLER, U. U., SCHICK A. and WEFELMEYER, W. (2012). Estimating the error distribution function in semiparametric additive regression models. *J. Statist. Plann. Inference* **142** 552–566. [MR2843057](#)
- [11] OWEN, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75** 237–249. [MR0946049](#)
- [12] OWEN, A. B. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18** 90–120. [MR1041387](#)
- [13] OWEN, A. B. (2001). *Empirical Likelihood*. Chapman & Hall/CRC, London.
- [14] PARENTE, P. M. D. C. and SMITH, R. J. (2011). GEL methods for non-smooth moment indicators. *Econometric Theory* **27** 74–113. [MR2771012](#)
- [15] PENG, H. and SCHICK, A. (2013). Empirical likelihood approach to goodness of fit testing. *Bernoulli* **19** 954–981. [MR3079302](#)
- [16] QIN, J. and LAWLESS, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22** 300–325. [MR1272085](#)
- [17] R CORE TEAM (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- [18] XUE, L. and WANG, Q. (2012). Empirical likelihood for single-index varying-coefficient models. *Bernoulli* **18** 836–856. [MR2948904](#)
- [19] XUE, L. and ZHU, L. (2006) Empirical likelihood for single-index models. *J. Multivariate Anal.* **97** 1295–1312. [MR2279674](#)
- [20] ZHU, L. and XUE, L. (2006). Empirical likelihood confidence regions in a partially linear single-index model. *J. Roy. Statist. Soc., Series B* **68** 549–570. [MR2278341](#)