

ON MAXIMUM EMPIRICAL LIKELIHOOD ESTIMATION AND RELATED TOPICS

BY HANXIANG PENG AND ANTON SCHICK*

*Indiana University Purdue University at Indianapolis and Binghamton
University*

This article studies maximum empirical likelihood estimation in the case of constraint functions that may be discontinuous and/or depend on additional parameters. The later is the case in applications to semiparametric models where the constraint functions may depend on the nuisance parameter. Our results are thus formulated for empirical likelihoods based on estimated constraint functions that may also be irregular. The key to our analysis is a uniform local asymptotic normality condition for the local empirical likelihood ratio. This condition holds under mild assumptions on the estimated constraint functions and allows for a study of maximum empirical likelihood estimation and empirical likelihood ratio testing similar to that for parametric models with the uniform local asymptotic normality condition. Applications of our results are discussed to inference problems about quantiles under possibly additional information on the underlying distribution, to residual-based inference about quantiles, and to partial adaption.

1. Introduction. Let $(\mathcal{Z}, \mathcal{S})$ 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{Z} -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 $\mathcal{Z} \times \Theta$ into \mathbb{R}^m , with $m \geq k$, such that for every R in \mathcal{Q} ,

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

and the matrix

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

*Supported by NSF Grant DMS 0906551

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

Keywords and phrases: irregular constraints; nuisance parameters; empirical likelihood ratio tests; uniform local asymptotic normality condition

is positive definite.

We refer to u as the *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 ,

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

To construct confidence sets for θ , Owen (1988, 1990, 2001) 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 $-2 \log \mathcal{R}_n(\theta)$ has as limiting distribution the chi-square distribution with m degrees of freedom.*

This allowed him to show that the set

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

is a confidence set for θ of asymptotic size $1-\alpha$. Indeed, we have $P(\theta \in \mathcal{C}_n) = P(-2 \log \mathcal{R}_n(\theta) < \chi_{1-\alpha}^2(m)) \rightarrow 1-\alpha$. 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) 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 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

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

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) obtain this condition under regularity and integrability conditions on the constraint function u and its partial derivatives with respect to the parameter. We shall show that the ULAN condition holds under the following weaker conditions, which allow for irregular u .

(K1) For every finite constant C ,

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

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

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

for every constant C .

Here A plays the same role as the quantity $-E[\dot{u}(Z, \theta)]$ does in Qin and Lawless (1994), where \dot{u} denotes the derivative of u with respect to the parameter.

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. In the ensuing discussion we assume 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

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

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

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

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

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

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) have shown that central estimators possess some optimality properties.

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) under their regularity assumptions and more recently in Theorem 1 by Lopez, Van Keilegom and Veraverbeke (2009) for the irregular case, using stronger conditions. We avoid some of the difficulties by working with central estimators instead of the maximum empirical likelihood estimator. The later 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).

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 empirical likelihood setting. Indeed, Qin and Lawless (1994) 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) to allow for irregular constraint functions. 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) and Peng and Scick (2010). 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 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. In Section 6 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. 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. Finally, in Section 8 we treat the case of partial adaptation. This is done by means of two examples, sample mean and sample median in symmetric location model, and least squares and least absolute deviation estimation in regression with symmetric errors.

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

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

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

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

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

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

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) and Peng and Schick (2010). Possible limit distributions of $-2 \log \hat{\mathcal{R}}_n(\theta)$ 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 (2010).

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 $-2 \log \hat{\mathcal{R}}_n(\theta)$ has a limiting chi-square distribution with m degrees of freedom.

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

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

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

(L1) For every finite 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

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

for each finite constant C .

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

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

and

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

hold for every finite C .

Note that 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

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

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 in the sense that $n^{1/2}(\tilde{\theta} - \theta) = O_P(1)$. 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 always assume that the following condition is met.

- (A) The conditions (L0)–(L2) hold, $\tilde{\theta}$ is a \sqrt{n} -consistent estimator, i.e. $n^{1/2}(\tilde{\theta} - \theta) = O_P(1)$, and \tilde{M} are positive definite $k \times k$ random dispersion matrices converging in probability to a positive definite dispersion matrix M .

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. Every $n^{1/2}$ -consistent (generalized) MELE is central.

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

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

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) that

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

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 $\mathbf{1}[A_n]\hat{\Delta}_n = o_P(1)$ 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 likelihood estimation. One-step maximum likelihood estimators were introduced by Le Cam (1960), 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.

We base the construction of the one-step MELE on the following consequences of the ULAN condition. For every finite constant C , the expansions

$$(3.1) \quad \sup_{|s|, |t| \leq C} |\hat{\mathcal{L}}_n(s+t) - \hat{\mathcal{L}}_n(s-t) - 2t^\top \Gamma_n + 2t^\top J s| = o_P(1)$$

and

$$(3.2) \quad \sup_{|r|, |s|, |t| \leq C} |\hat{\mathcal{L}}_n(r+s+t) - \hat{\mathcal{L}}_n(r+s-t) - \hat{\mathcal{L}}_n(r-s+t) + \hat{\mathcal{L}}_n(r-s-t) + 4t^\top J s| = o_P(1)$$

hold. Let e_1, \dots, e_k denote the standard unit vectors in \mathbb{R}^k . Let $\tilde{\gamma}$ denote the vector with components

$$\tilde{\gamma}_i = [\log \hat{\mathcal{R}}_n(\tilde{\theta} + n^{-1/2} \tilde{M} e_i) - \hat{\mathcal{R}}_n(\tilde{\theta} - n^{-1/2} \tilde{M} e_i)]/2$$

and \tilde{V} be the matrix with entries

$$\begin{aligned} \tilde{V}_{ij} = & - \left[\log \hat{\mathcal{R}}_n(\tilde{\theta} + n^{-1/2} \tilde{M}(e_i + e_j)) - \log \hat{\mathcal{R}}_n(\tilde{\theta} + n^{-1/2} \tilde{M}(e_j - e_i)) \right. \\ & \left. - \log \hat{\mathcal{R}}_n(\tilde{\theta} + n^{-1/2} \tilde{M}(e_i - e_j)) + \log \hat{\mathcal{R}}_n(\tilde{\theta} - n^{-1/2} \tilde{M}(e_j + e_i)) \right] / 4. \end{aligned}$$

It follows from (3.1) and (3.2) that

$$\tilde{\gamma}_i = e_i^\top \tilde{M} (\Gamma_n - J^{-1} n^{1/2} (\tilde{\theta} - \theta)) + o_P(1)$$

and

$$\tilde{V}_{ij} = e_i^\top \tilde{M} J \tilde{M} e_j + o_P(1).$$

This shows that

$$(3.3) \quad \tilde{\Gamma} = \tilde{M}^{-1} \tilde{\gamma} = \Gamma_n - J n^{1/2} (\tilde{\theta} - \theta) + o_P(1)$$

and

$$(3.4) \quad \tilde{J} = \tilde{M}^{-1} \tilde{V} \tilde{M}^{-1} = J + o_P(1).$$

Now we define the one-step MELE by

$$\hat{\theta}_* = \tilde{\theta} + n^{-1/2} \tilde{J}^{-1} \tilde{\Gamma}.$$

It follows from (3.3) and (3.4) and the $n^{1/2}$ -consistency of $\tilde{\theta}$ that the one-step MELE $\hat{\theta}_*$ is central. Let us summarize this in the following theorem.

THEOREM 3.1. *Suppose condition (A) holds. Then the one-step MELE $\hat{\theta}_*$ is central.*

Note that we can always take $\tilde{M} = M = aI_k$ for some positive a . Suppose that $n^{1/2}(\tilde{\theta} - \theta)$ is asymptotically normal with mean vector zero and positive definite dispersion matrix D . Then we can take $\tilde{M} = a\tilde{D}_n$ and $M = aD$, for some positive a and some consistent positive definite estimator \tilde{D}_n of D .

Method 2. Guided maximum empirical likelihood estimation using one-step estimators. Let $\hat{\theta}_*$ denote a one-step MELE. 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 $\hat{\mathcal{R}}_n(\vartheta)$ 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

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

for some (small) positive c a (generalised) maximum empirical likelihood estimator guided by a one-step estimator, short GOMELE. If the map $\vartheta \mapsto \hat{\mathcal{R}}_n(\vartheta)$ is upper semi-continuous, then we can take $\hat{\theta}$ to be a maximizer of $\hat{\mathcal{R}}_n$ 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 (generalised) maximum empirical likelihood estimation can also be done using the $n^{1/2}$ -consistent estimator $\tilde{\theta}$ 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*

$$(3.5) \quad \sup_{|t| \leq 2C_n} \frac{|\hat{\mathcal{L}}_n(t) - t^\top \Gamma_n + 1/2 t^\top J t|}{(1 + |t|)^2} = o_P(1).$$

Then an estimator $\hat{\theta}$ that satisfies $|\hat{\theta} - \tilde{\theta}| \leq C_n$ and

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

is central.

PROOF. Set $\tilde{\Delta} = n^{1/2}(\tilde{\theta} - \theta)$. Then, with h as in (1.2) and λ the smallest eigen value 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 5. 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*

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

and

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

for an $m \times k$ matrix A of full rank k . Then (3.5) 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 $\hat{\mathcal{R}}_n$, θ , and A are now played by $\hat{\mathcal{R}}_n \circ \psi$, $\tau = \kappa_0(Q)$, and AB . Thus Theorem 2.2 yields the expansions

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

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

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

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

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

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

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

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

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

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. A general result. Let $\mathbb{T}_{n1}(t), \dots, \mathbb{T}_{nm}(t)$ 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 (2010).

LEMMA 5.1. *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 eigen value of the matrix S . Then the inequality $\lambda > 5|\bar{x}|x^$ implies*

$$(5.1) \quad \left| -2 \log \mathcal{R} - n \bar{x}^\top S^{-1} \bar{x} \right| \leq \left(\Lambda + \frac{\Lambda^3}{4\lambda^2} \right) \frac{2n|\bar{x}|^3 x^*}{(\lambda - |\bar{x}|x^*)^3}$$

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

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

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

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

We have the following result.

THEOREM 5.1. *Suppose (B1)–(B3) hold. Then*

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

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 eigen values of $\mathbb{S}_n(t)$. It follows from (B2) that there are constants $0 < \eta < K < \infty$ such that

$$(5.2) \quad P\left(\sup_{|t| \leq C_n} \Lambda_n(t) > K\right) \rightarrow 0 \quad \text{and} \quad P\left(\inf_{|t| \leq C_n} \lambda_n(t) > \eta\right) \rightarrow 0.$$

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

$$(5.3) \quad \sup_{|t| \leq C_n} \mathbb{T}_n^*(t) |\bar{\mathbb{T}}_n(t)| = o_P(1)$$

and

$$(5.4) \quad \sup_{|t| \leq C} \frac{n \mathbb{T}_n^*(t) |\bar{\mathbb{T}}_n(t)|^3}{(1 + |t|)^2} = o_P(1).$$

From (5.1) – (5.4) it follows that

$$(5.5) \quad \sup_{|t| \leq C_n} \frac{|-2 \log \mathcal{R}_n(t) - n \bar{\mathbb{T}}_n(t)^\top \mathbb{S}_n(t)^{-1} \bar{\mathbb{T}}_n(t)|}{(1 + |t|)^2} = o_P(1).$$

From (B2) we derive

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

and thus obtain the expansion

$$(5.6) \quad \sup_{|t| \leq C_n} \frac{|n \bar{\mathbb{T}}_n(t)^\top \mathbb{S}_n(t)^{-1} \bar{\mathbb{T}}_n(t) - n \bar{\mathbb{T}}_n(t)^\top S^{-1} \bar{\mathbb{T}}_n(t)|}{(1 + |t|)^2} = o_P(1).$$

The first conclusion in the theorem follows from (5.6), (5.5) 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 (3.5). The assumption used in this result imply (L0)–(L2), and 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 5.2. *In addition to (L0) and (SL2) assume that we have the rates*

$$(5.7) \quad \sup_{|t| \leq C_n} \max_{1 \leq j \leq n} C_n |\hat{u}_n(Z_j, \theta + n^{-1/2}t)| = o_P(n^{1/2}),$$

$$(5.8) \quad \sup_{|t| \leq 2C_n} \frac{1}{n} \sum_{j=1}^n |\hat{u}_n(Z_j, \theta + n^{-1/2}t) - u_Q(Z_j, \theta)|^2 = o_P(1).$$

Then we have the expansion

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

and this implies (3.5).

PROOF. The desired result follows if we verify the assumptions of Theorem 5.1 with $\mathbb{T}_{nj}(t) = \hat{u}_n(Z_j, \theta + n^{-1/2}t)$, $S = W$, and U_n as in (2.1). Note that (B1) follows (5.7) 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

$$(5.9) \quad |\bar{W}_n - W_n| = o_P(1).$$

Then (B2) follows from (5.8), (5.9) 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 $|t| \leq C_n$, and with \hat{D}_n the left-hand side of 5.8. The above inequality was already used in Peng and Schick (2010), see their (4.3). \square

REMARK 5.1. 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 (5.7) and (5.8) by the conditions

$$(5.10) \quad U^* = \max_{1 \leq j \leq n} |u_Q(Z_j, \theta)| = o_p(n^{1/2}/C_n)$$

and

$$(5.11) \quad \sup_{|t| \leq 2C_n} \frac{1}{n} \sum_{j=1}^n |\hat{u}_n(Z_j, \theta + n^{-1/2}t) - u_Q(Z_j, \theta)|^2 = o_p(C_n^{-2}).$$

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

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

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

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

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 5.2 with $C_n = C$. In view of the previous remark it suffices to verify (5.10), (5.11) and (SL2) with $C_n = C$. Since $\int |u_Q(z, \theta)| dQ$ is finite, we obtain

$$(5.12) \quad \max_{1 \leq j \leq n} |u_Q(Z_j, \theta)| = o_P(n^{1/2})$$

which is (5.10) for the $C_n = C$. Of course, for $C_n = C$, (L1) is equivalent to (5.11) and (L2) is equivalent to (SL2).

6. Inference about quantiles. Throughout this section X_1, \dots, X_n are independent copies of a random variable X with *continuous* 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

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

The latter is equivalent to

$$\hat{q}_\gamma = F^{-1}(\gamma) + \frac{1}{n} \sum_{j=1}^n \frac{\phi(X_j - F^{-1}(\gamma), \gamma)}{F'(F^{-1}(\gamma))} + o_P(n^{-1/2}),$$

where

$$(6.1) \quad \phi(x, \gamma) = \gamma \mathbf{1}[x > 0] - (1 - \gamma) \mathbf{1}[x < 0].$$

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

LEMMA 6.1. *Suppose that F is differentiable at some q . Then the following hold for every finite constant C ,*

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

and

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

The above hold also if \leq is replaced by $<$, and if \leq and $F'(q)$ are replaced by $>$ and $-F'(q)$.

EXAMPLE 6.1. 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 $E[X_1 \leq \theta] = \gamma$. This suggests to look at the empirical likelihood

$$\mathcal{R}_{n1}(q) = \sup \left\{ \prod_{j=1}^n n\pi_j : \pi \in \mathcal{P}_n, \sum_{j=1}^n \pi_j (\mathbf{1}[X_j \leq q] - \gamma) = 0 \right\}, \quad q \in \mathbb{R}.$$

In view of Lemma 6.1 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 every gMELE $\hat{\theta}$ satisfies

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

Consequently the gMELE is asymptotically equivalent to the sample quantile.

We have $E[\phi(X - \theta, \gamma)] = 0$, where ϕ is defined in (6.1). Thus, we can also work with the empirical likelihood

$$\mathcal{R}_{n2}(q) = \sup \left\{ \prod_{j=1}^n n\pi_j : \pi \in \mathcal{P}_n, \sum_{j=1}^n \pi_j \phi(X_j - q, \gamma) = 0 \right\}, \quad q \in \mathbb{R}.$$

For this choice the conditions (K0)–(K2) hold with $u(z, \vartheta) = \phi(z - \vartheta, \gamma)$, $W = \gamma(1 - \gamma)$ and $A = F'(\theta)$. Thus every gMELE obeys the expansion

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

If $\gamma = 1/2$, we could also work with the empirical likelihood

$$\mathcal{R}_{n3}(q) = \sup \left\{ \prod_{j=1}^n n\pi_j : \pi \in \mathcal{P}_n, \sum_{j=1}^n \pi_j \text{sign}(X_j - q) = 0 \right\}, \quad q \in \mathbb{R},$$

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

REMARK 6.1. We should point out that the previous example is one of the few cases where there is an explicit formula for the empirical likelihood. For example, the first empirical likelihood is given by

$$(6.2) \quad \mathcal{R}_{n1}(q) = \mathbf{1}[0 < N(q) < n] \left(\frac{n\gamma}{N(q)} \right)^{N(q)} \left(\frac{n(1-\gamma)}{n-N(q)} \right)^{n-N(q)},$$

where $N(q) = \sum_{j=1}^n \mathbf{1}[X_j \leq q]$. Indeed, as shown by Owen, see Owen (2001), we have

$$\mathcal{R}_{n1}(q) = \prod_{j=1}^n \frac{1}{1 + \zeta(\mathbf{1}[X_j \leq q] - \gamma)}$$

if ζ is a solution to the equation

$$\sum_{j=1}^n \frac{\mathbf{1}[X_j \leq q] - \gamma}{1 + \zeta(\mathbf{1}[X_j \leq q] - \gamma)} = 0,$$

subject to $1 + \zeta(\mathbf{1}[X_j \leq q] - \gamma) > 0$, and $\mathcal{R}_{n1}(q) = 0$ if no such ζ exists. No such ζ exists for $N(q) = 0$ and $N(q) = n$. The left-hand side of the equation can be expressed as

$$N(q) \frac{1 - \gamma}{1 + \zeta(1 - \gamma)} - (n - N(q)) \frac{\gamma}{1 - \zeta\gamma}.$$

A unique solution of the equation is then given by

$$\zeta = \frac{N(q) - n\gamma}{n\gamma(1 - \gamma)}$$

on the event $\{0 < N(q) < n\}$. On this event the quantities $1 + \zeta(\mathbf{1}[X_j \leq q] - \gamma)$ are positive. Simple calculations show that $\mathcal{R}_{n1}(q)$ has the desired form (6.2).

From (6.2) we derive the identity

$$\mathcal{R}_{n1}(\vartheta) = \sum_{j=1}^{n-1} \mathbf{1}[X_{(j)} \leq \vartheta < X_{(j+1)}] g_\gamma(j, n),$$

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 $\vartheta \mapsto \mathcal{R}_{n1}(\vartheta)$ 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 the gMELE based on the preliminary estimator $X_{(k_n)}$ is given by

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

One can also show that

$$\mathcal{R}_{n2}(\vartheta) = \sum_{j=1}^{n-1} \mathbf{1}[X_{(j)} < \vartheta < X_{(j+1)}] g_\gamma(j, n) + \sum_{j=2}^{n-1} \mathbf{1}[X_{(j)} = \vartheta] g_\gamma(j, n-1).$$

The function $y \mapsto g_\gamma(x, y)$ is increasing on the interval $(x, x/\gamma]$ and decreasing on $(x/\gamma, \infty)$. This shows that $g_\gamma(i, n-1) > g_\gamma(i, n)$ for $i = 1, \dots, k_n$ and $g_\gamma(i, n-1) < g_\gamma(i, n)$ for $i = k_n + 1, \dots, n$. Thus a MELE exists in this case.

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

EXAMPLE 6.2. 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

$$\mathcal{R}_{n4}(\vartheta) = \sup \left\{ \prod_{j=1}^n n\pi_j : \pi \in \mathcal{P}_n, \sum_{j=1}^n \pi_j \begin{pmatrix} \mathbf{1}[X_j \leq \vartheta_1] - \gamma_1 \\ \dots \\ \mathbf{1}[X_j \leq \vartheta_m] - \gamma_m \end{pmatrix} = 0 \right\},$$

defined for $\vartheta \in \Theta$. 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 gMELE $\hat{\theta}$ satisfies the expansion

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

Of course, we could also work with the empirical likelihood

$$\mathcal{R}_{n5}(\vartheta) = \sup \left\{ \prod_{j=1}^n n\pi_j : \pi \in \mathcal{P}_n, \sum_{j=1}^n \pi_j \begin{pmatrix} \phi(X_j - \vartheta_1, \gamma_1) \\ \dots \\ \phi(X_j - \theta_m, \gamma_m) \end{pmatrix} = 0 \right\},$$

for $\vartheta \in \Theta$. Now $A = \text{diag}(F'(\theta_1), \dots, F'(\theta_m))$ and W is as before. From this and Theorem 1.3 we find that the i -th component $\hat{\theta}_i$ of a gMELE $\hat{\theta}$ satisfies the expansion

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

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

EXAMPLE 6.3. 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

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

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 gMELE 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 6.4. 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 6.1 and the calculations

$$\frac{1}{n} \sum_{j=1}^n |\varepsilon_j - n^{-1/2}t - \varepsilon_j|^2 \leq t^2/n,$$

and

$$n^{-1/2} \sum_{j=1}^n (\varepsilon_j - n^{-1/2}t - \varepsilon_j) + t = 0.$$

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 $T_n = \mathcal{R}_n(\hat{q}, \hat{r})$, where $\hat{\tau}$ is a gMELE (guided by the average of the sample mean and sample median) 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 6.5. Suppose F is γ -regular and has zero mean and standard deviation σ . We estimate $\theta = F^{-1/2}(\gamma)$ using the empirical likelihood

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

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 every gMELE $\hat{\theta}$ 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 6.6. Suppose F is γ -regular for some $\gamma \neq 1/2$ and has known median 0. To estimate $\theta = F^{-1/2}(\gamma)$, we rely on the empirical likelihood

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

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

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

where ρ is the covariance between $\text{sign}(X)$ and $\mathbf{1}[X \leq \theta]$. Thus every gMELE $\hat{\theta}$ 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 $(\gamma(1-\gamma) - \rho^2)/(F'(\theta))^2$.

EXAMPLE 6.7. Suppose F is γ -regular for some $\gamma > 1/2$ and $2F - 1$ is odd. The latter implies that X and $-X$ have the same distribution. To estimate $\theta = F^{-1/2}(\gamma)$, we rely on the empirical likelihood

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

defined for $\vartheta \in \mathbb{R}$. The conditions (K0)–(K2) are met by $u(z, \vartheta) = \mathbf{1}[z \leq \vartheta] + \mathbf{1}[-z \leq \vartheta] - 2\gamma$, $W = 6\gamma - 4\gamma^2 - 2$, and $A = -2F'(\theta)$. Thus every gMELE $\hat{\theta}$ of θ satisfies

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

and has asymptotic variance $(6\gamma - 4\gamma^2 - 2)/(2F'(\theta))^2$.

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 6.5

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

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.

$$(7.1) \quad M_n = \max_{1 \leq j \leq n} |\hat{\varepsilon}_j - \varepsilon_j| = o_P(1)$$

and

$$(7.2) \quad D_n = \frac{1}{n} \sum_{j=1}^n |\hat{\varepsilon}_j - \varepsilon_j|^2 = O_P(n^{-1}).$$

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

$$(7.3) \quad \sup_{x \in \mathbb{R}} \left| \frac{1}{n} \sum_{j=1}^n \left(\mathbf{1}[\hat{\varepsilon}_j \leq x] - \mathbf{1}[\varepsilon_j \leq x] - f(x)\varepsilon_j \right) \right| = o_P(n^{-1/2}).$$

This can be derived from results of Koul (1969) for the fixed design case and MSW (2007) for the random design case used here, see also Remark 2 in MSW (2008). The above expansion suggests to work with the empirical likelihood

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

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 density. If K is Lipschitz and the bandwidth b satisfies $nb^4 \rightarrow \infty$ 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

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

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 . To have uniform consistency, we assume from now that K is Lipschitz and that the bandwidth b satisfies $nb^4 \rightarrow \infty$ and $b \rightarrow 0$.

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

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

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

We obtain from Theorem 2.2 that a gMELE $\hat{\theta}$ (guided say by the sample quantile of the residuals) obeys the expansion

$$(7.4) \quad \hat{\theta} = \theta - \frac{1}{n} \sum_{j=1}^n \left(\frac{\mathbf{1}[\varepsilon_j \leq \theta] - \gamma}{f(\theta)} + \varepsilon_j \right) + o_P(1).$$

The asymptotic variance of a gMELE is thus

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

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 gMELE can have a smaller variance than the sample quantile based on the actual errors.

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 (7.3) also hold for the

nonparametric residuals $\hat{\varepsilon}_j = Y_j - \hat{r}_n(X_j)$. These nonparametric residuals also satisfy (7.1),

$$(7.5) \quad \frac{1}{n} \sum_{j=1}^n \hat{\varepsilon}_j = o_P(n^{-1/2})$$

and

$$(7.6) \quad \frac{1}{n} \sum_{j=1}^n (\hat{\varepsilon}_j - \varepsilon_j)^2 = o_P(n^{-\rho})$$

for some $\rho > 1/2$. 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 $n^\rho b^4 \rightarrow \infty$. Using (7.3), (7.5) and (7.6) one verifies (L1) and (L2) with $A = -f(\theta)$ and again obtains the expansion (7.4) for the corresponding gMELE.

Expansions to nonparametric regression models with multivariate covariates are possible using the results of Müller et al (2009). The results in Müller et al (2007) and Müller et al (2010) can be used to obtain extensions to the partly linear regression model and to the additive nonparametric regression model. In all these models one can construct residuals so that (7.3), (7.5) and (7.6) hold and then obtains the expansion (7.4). We should mention that we do not get (7.5) for all regression models. This is already so in linear regression *without* an intercept.

8. Partial adaptation. In this section we look at the case when there are several candidates for the function u . This happens often in semiparametric models. For example, we can identify the location parameter in the symmetric location model as the mean or as the median. Either case provides a constraint function. Similarly, in a regression model with symmetric errors we can estimate the regression parameters by either the least squares approach or the least absolute deviation approach. Either approach provides a constraint function. The empirical likelihood approach lets us use both constraint functions at the same time by stacking them into a new constraint function. It is easy to check that if the constraint functions u_1 and u_2 satisfy conditions (K0)–(K2), so does the stacked constraint function $u = (u_1^\top, u_2^\top)^\top$ provided the matrices $W(R)$ for this choice of u are positive definite. In the next two examples we shall see that the gMELE associated with the stacked constraint function u is asymptotically equivalent to the linear combination $s\hat{\theta}_1 + (1-s)\hat{\theta}_2$ of the gMELEs $\hat{\theta}_1$ and $\hat{\theta}_2$ associated with u_1 and u_2 respectively, with the smallest asymptotic variance. In particular,

if the gMELE associated with u_1 provides an efficient estimator for some distribution and the gMELE associated with u_2 provides an efficient estimator for another distribution, then gMELE associated with the stacked constraint function u is efficient for each distribution.

Brown and Chen (1998) addressed these issues in the location problem using the constraint functions for mean and median arriving at the same conclusions. They derived these results heuristically and for a smoothed version of empirical likelihood, the least squares empirical likelihood as referred by them (or the euclidean likelihood by Owen (1991)).

EXAMPLE 8.1. Let X_1, \dots, X_n be independent copies of a random variable X , where $X = \theta + \varepsilon$ and ε has mean zero and a finite variance σ^2 , ε and $-\varepsilon$ have the same distribution, and their common distribution function F has a positive derivative $f(0)$ at 0. In this case we can estimate θ by either the sample mean or the sample median. Note that the sample mean \bar{X} is a solution to the estimating equation

$$\frac{1}{n} \sum_{j=1}^n X_j - \vartheta = 0,$$

while the sample median M_n solves the equation

$$\frac{1}{n} \sum_{j=1}^n \text{sign}(X_j - \vartheta) = 0$$

and obeys the expansion

$$M_n = \theta + \frac{1}{n} \sum_{j=1}^n \frac{\text{sign}(\varepsilon_j)}{2f(0)} + o_P(n^{-1/2}).$$

Instead of working with either the sample mean or sample median, we could work with a linear combination $\tilde{\theta}(s) = s\bar{X} + (1-s)M_n$ of the two and select the coefficient s which minimizes the asymptotic variance. It is easy to see that the asymptotic variance of $\tilde{\theta}(s)$ equals $s^2\sigma^2 + 2s(1-s)\nu + (1-s)^2/(4f^2(0))$ with $\nu = E[|\varepsilon|]$ and is minimized by $s = a/c$ where $a = 1 - 2\nu f(0)$ and $c = 1 - 4\nu f(0) + 4\sigma^2 f^2(0)$. One could now use the estimator $\tilde{\theta}(\hat{a}/\hat{c})$ with \hat{a} and \hat{c} estimators of a and c . If \hat{a} and \hat{c} are consistent for a and c , then $\tilde{\theta}(\hat{a}/\hat{c})$ will be asymptotically equivalent to $\tilde{\theta}(a/c)$.

We shall now show that every gMELE of the empirical likelihood

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

is asymptotically equivalent to $\tilde{\theta}(a/c)$. We are in the setting (K0) with $u(z, \vartheta) = (z - \vartheta, \text{sign}(z - \vartheta))^\top$ and

$$W = \begin{bmatrix} E[\varepsilon^2] & E[|\varepsilon|] \\ E[|\varepsilon|] & 1 \end{bmatrix} = \begin{bmatrix} \sigma^2 & \nu \\ \nu & 1 \end{bmatrix}.$$

Here (K1) and (K2) hold with $A = [1, 2f(0)]^\top$, see Example 6.4. Doing the necessary calculations we see that every gMELE satisfies the expansion

$$(8.1) \quad \hat{\theta} = \theta + \frac{1}{n} \sum_{j=1}^n \frac{a\varepsilon_j + b\text{sign}(\varepsilon_j)}{c} + o_P(n^{-1/2}),$$

where a and c are as above and $b = 2\sigma^2 f(0) - \nu$. If f is a normal density, then $b = 0$ and $a = c = 1 - 2/\pi$, and every gMELE is asymptotically equivalent to the sample mean. If f is a double exponential density, then $a = 0$ and $b = c = \sigma/\sqrt{2}$ and every gMELE is equivalent to the sample median. Since $a + 2f(0)b = c$, we can write (8.1) as

$$\hat{\theta} = \theta + \frac{1}{n} \sum_{j=1}^n \left(\frac{a}{c} \varepsilon_j + \left(1 - \frac{a}{c}\right) \frac{\text{sign}(\varepsilon_j)}{2f(0)} \right) + o_P(n^{-1/2}),$$

This shows that every gMELE is asymptotically equivalent to $\tilde{\theta}(a/c)$. Thus its asymptotic variance $(\sigma^2 - \nu^2)/c$ is not larger than that of either the sample mean or the sample median. The gMELE avoids estimation of the unknown a and c .

Suppose that F has finite Fisher information for location. This means that F has an absolutely continuous density f and the score function $\ell_f = -f/f$ for location belongs to $L_2(F)$. Under this assumption one has

$$\int x \ell_f(x) f(x) dx = 1 \quad \text{and} \quad \int \text{sign}(x) \ell_f(x) f(x) dx = 2f(0).$$

Let ψ be the function from \mathbb{R} to \mathbb{R}^2 defined by $\psi(x) = [x, \text{sign}(x)]^\top$. Then the function $h_f = A^\top W^{-1} \psi$ equals $\int \ell_f \psi^\top dF (\int \psi \psi^\top dF)^{-1} \psi$ and is therefore the projection of the score function ℓ_f onto the linear span $\{a^\top \psi : a \in \mathbb{R}^2\}$ of the components of ψ . Moreover, $A^\top W^{-1} A$ equals $\int h_f^2 dF$. Thus the expansion (8.1) can be written as

$$\hat{\theta} = \theta + \frac{1}{n} \sum_{j=1}^n \frac{h_f(\varepsilon_j)}{\int h_f^2 dF} + o_P(n^{-1/2}).$$

EXAMPLE 8.2. Assume that Z_1, \dots, Z_n are independent replicas of the observation $Z = (X^\top, Y)^\top$ from the regression model $Y = \theta^\top X + \varepsilon$, where ε and X are independent, the matrix $M = E[XX^\top]$ is well defined and positive definite, ε has finite variance σ^2 , ε and $-\varepsilon$ have the same distribution, and their common distribution function F has a positive derivative $f(0)$ at 0. Here we could use both the least squares approach as well as the least absolute deviation approach to estimate the regression parameter θ . Motivated by the previous example we instead look at the gMELE of the empirical likelihood

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

Here we again are in the setting (K0) with

$$u(Z, \vartheta) = \begin{pmatrix} X(Y - \vartheta X) \\ X \text{sign}(Y - \vartheta^\top X) \end{pmatrix} \quad \text{and} \quad W = \begin{bmatrix} \sigma^2 M & \nu M \\ \nu M & M \end{bmatrix}.$$

The conditions (K1) and (K2) are now be verified with $A = [M, 2f(0)M]^\top$. Let $\tau = \sigma^2 - \nu^2$ and a, b and c be as in the previous example. Then we have

$$W^{-1} = \frac{1}{\sigma^2 - \nu^2} \begin{bmatrix} M^{-1} & -\nu M^{-1} \\ -\nu M^{-1} & \sigma^2 M^{-1} \end{bmatrix}$$

and thus obtain that $A^\top W^{-1} = (1/\tau)[aI, bI]$ and $A^\top W^{-1}A = c/\tau M$. Thus every gMELE obeys the expansion

$$\hat{\theta} = \theta + \frac{1}{n} \sum_{j=1}^n M^{-1} X_j [(a/c)\varepsilon_j + (b/c)\text{sign}(\varepsilon_j)] + o_P(n^{-1/2})$$

and has asymptotic variance τ/cW^{-1} . As in the previous example we find that this estimator is asymptotically equivalent to the least squares estimator for a normal density f and asymptotically equivalent to the least absolute deviation estimator for a double exponential density f . In all cases the asymptotic variance of a gMELE is no larger than those of the least squares or least absolute deviation estimators.

References.

- Brown, B.M. and Chen, S. X. (1998). Combined and least squares empirical likelihood. *Ann. Inst. Statist. Math.*, **50**, 697–714.
- Hjort, N.L., McKeague, I.W. and Van Keilegom, I. (2009). Extending the scope of empirical likelihood. *Ann. Statist.* **37**, 1079–1111.

- Koul, H.L. (1969). Asymptotic behavior of Wilcoxon type confidence regions in multiple linear regression. *Ann. Statist.*, **40**, 1950–1979.
- Le Cam, L. (1960). Locally asymptotically normal families of distributions. *Univ. California Publ. Statist.*, **3**, 37–98.
- Lopez, E.M.M., Van Keilegom, I. and Veraverbeke (2009). Empirical likelihood for non-smooth criterion functions. *Scand. J. Statist.*, **36**, 413–432.
- Müller, U.U., Schick A. and Wefelmeyer, W. (2007). Estimating the error distribution function in semiparametric regression. *Statist. Decisions*, **25**, 1–18.
- 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.
- Müller, U.U., Schick A. and Wefelmeyer, W. (2010). Estimating the error distribution function in semiparametric additive regression models. Preprint.
- Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237–249.
- Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.*, **18**, 90–120.
- Owen, A. B. (2001). *Empirical Likelihood*. Chapman & Hall/CRC, London.
- Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.*, **22**, 300–325.

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