ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

An International Society for the Advancement of Economic Theory in its Relation to Statistics and Mathematics

http://www.econometricsociety.org/

Econometrica, Vol. 80, No. 1 (January, 2012), 413–423

ON THE ASYMPTOTIC OPTIMALITY OF EMPIRICAL LIKELIHOOD FOR TESTING MOMENT RESTRICTIONS

YUICHI KITAMURA
Yale University, New Haven, CT 06520-8281, U.S.A.

ANDRES SANTOS
University of California–San Diego, La Jolla, CA 92093-0508, U.S.A.

AZEEM M. SHAIKH
University of Chicago, Chicago, IL 60637, U.S.A.

Econometrica, Vol. 80, No. 1 (January, 2012), 413–423

NOTES AND COMMENTS

ON THE ASYMPTOTIC OPTIMALITY OF EMPIRICAL LIKELIHOOD FOR TESTING MOMENT RESTRICTIONS

BY YUICHI KITAMURA, ANDRES SANTOS, AND AZEEM M. SHAIKH[1]

We show by example that empirical likelihood and other commonly used tests for moment restrictions are unable to control the (exponential) rate at which the probability of a Type I error tends to zero unless the possible distributions for the observed data are restricted appropriately. From this, it follows that for the optimality claim for empirical likelihood in Kitamura (2001) to hold, additional assumptions and qualifications are required. Under stronger assumptions than those in Kitamura (2001), we establish the following optimality result: (i) empirical likelihood controls the rate at which the probability of a Type I error tends to zero and (ii) among all procedures for which the probability of a Type I error tends to zero at least as fast, empirical likelihood maximizes the rate at which the probability of a Type II error tends to zero for *most* alternatives. This result further implies that empirical likelihood maximizes the rate at which the probability of a Type II error tends to zero for all alternatives among a class of tests that satisfy a weaker criterion for their Type I error probabilities.

KEYWORDS: Empirical likelihood, large deviations, Hoeffding optimality, moment restrictions.

1. INTRODUCTION

LET $X_i, i = 1, \ldots, n$, be an independent and identically distributed (i.i.d.) sequence of random variables with distribution $P \in \mathbf{P} \subseteq \mathbf{M}$, where $\mathbf{M}$ denotes the set of probability distributions on $\mathcal{X} \subseteq \mathbf{R}^d$ (with the Borel σ-algebra). Let $\Theta \subseteq \mathbf{R}^r$ with $m > r$ and $g : \mathbf{R}^d \times \Theta \to \mathbf{R}^m$ be given, and define

$$\mathbf{P}_0 = \{P \in \mathbf{P} : E_P[g(X, \theta)] = 0 \text{ for some } \theta \in \Theta\}. \quad (1)$$

In this paper, we consider testing

$$H_0 : P \in \mathbf{P}_0 \quad \text{versus} \quad H_1 : P \in \mathbf{P}_1, \quad (2)$$

where $\mathbf{P}_1 = \mathbf{P} \setminus \mathbf{P}_0$. Note that any nonrandomized test of (2) can be identified with a partition $(\Omega_{0,n}, \Omega_{1,n})$ of $\mathbf{M}$, wherein the test rejects the null hypothesis if the empirical distribution of the observations, denoted by $\hat{P}_n$, falls in $\Omega_{1,n}$ and fails to reject it otherwise.

[1]We thank J. Stock and two anonymous referees for valuable suggestions that greatly improved the paper. The research of the first author has been supported by National Science Foundation Grants SES-0551271 and SES-0851759. The research of the third author has been supported by National Science Foundation Grant DMS-0820310.

 DOI: 10.3982/ECTA8773

Following Owen (1988), Qin and Lawless (1994) proposed a test of (2) based on the empirical likelihood ratio. To describe their test, recall that the Kullback–Leibler divergence of P from Q is defined as

$$I(Q|P) = \begin{cases} \int \log\left(\frac{dQ}{dP}\right) dQ, & \text{if } Q \ll P, \\ \infty, & \text{otherwise,} \end{cases}$$

and let

$$\mathbf{M}_0(Q) = \left\{P \in \mathbf{M} : P \ll Q, Q \ll P, E_P[g(X, \theta)] = 0 \text{ for some } \theta \in \Theta\right\}. \tag{3}$$

Using this notation, their test can be described as rejecting the null hypothesis for large values of

$$\inf_{P \in \mathbf{M}_0(\hat{P}_n)} I(\hat{P}_n|P). \tag{4}$$

It therefore corresponds, for some sequence of critical values $\{\eta_n > 0 : n \geq 1\}$, to the partition $(\Lambda_0(\eta_n), \Lambda_1(\eta_n))$ of $\mathbf{M}$, where

$$\Lambda_0(\eta_n) = \left\{Q \in \mathbf{M} : \inf_{P \in \mathbf{M}_0(Q)} I(Q|P) < \eta_n\right\}, \tag{5}$$
$$\Lambda_1(\eta_n) = \mathbf{M} \setminus \Lambda_0(\eta_n).$$

Here, the infimum over the empty set is understood to be infinity.

Kitamura (2001) claimed that the empirical likelihood ratio test is optimal for testing (2) in the large deviations sense of Hoeffding (1965) provided that the following conditions hold:

$$P\left\{\sup_{\theta \in \Theta} \|g(X, \theta)\| = \infty\right\} = 0 \quad \text{for all } P \in \mathbf{P}, \tag{6}$$

$$g(x, \cdot) \text{ is continuous at every } \theta \in \Theta \text{ for each } x \in \mathbf{R}^d. \tag{7}$$

Specifically, part (a) of Theorem 2 in Kitamura (2001) asserts that for any $\eta > 0$,

$$\sup_{P \in \mathbf{P}_0} \limsup_{n \to \infty} \frac{1}{n} \log P^n\{\hat{P}_n \in \Lambda_1(\eta)\} \leq -\eta, \tag{8}$$

whereas part (b) of the same theorem asserts that if another test $(\Omega_{0,n}, \Omega_{1,n})$ satisfies

$$\sup_{P \in \mathbf{P}_0} \limsup_{n \to \infty} \frac{1}{n} \log P^n\{\hat{P}_n \in \Omega^{\delta}_{1,n}\} \leq -\eta$$

for some $\delta > 0$, then

$$\limsup_{n\to\infty} \frac{1}{n} \log P^n\{\hat{P}_n \in \Omega_{0,n}\} \geq \limsup_{n\to\infty} \frac{1}{n} \log P^n\{\hat{P}_n \in \Lambda_0(\eta)\}$$

for any $P \in \mathbf{P}$. Here, P^n denotes the n-fold product measure $\bigotimes_{i=1}^n P$ and A^δ denotes the (open) δ "blowup" of a set $A \subseteq \mathbf{M}$ with respect to the Prokhorov–Lévy metric, that is, $A^\delta = \bigcup_{P\in A} B(P, \delta)$, where $B(P, \delta)$ denotes an open ball with center P and radius δ in the Prokhorov–Lévy metric.

In Section 2, we provide two examples that demonstrate that the empirical likelihood ratio test of Qin and Lawless (1994) in fact fails to control size as required in (8) without additional restrictions on $\mathbf{P}$. More specifically, we show that given the assumptions in Kitamura (2001), (8) is not satisfied for any $\eta > 0$. Interestingly, these examples also reveal that most commonly used tests of (2) have the same feature. In Section 3, we use these examples to motivate restrictions on $\mathbf{P}$ under which we establish a more limited version of Theorem 2 in Kitamura (2001). Proofs of all results are collected in the Supplemental Material (Kitamura, Santos, and Shaikh (2012)).

REMARK 1.1: Strictly speaking, Theorem 2 of Kitamura (2001) considers a test different from that of Qin and Lawless (1994) in that it omits the requirements $P \ll Q$ and $Q \ll P$ in (3).

2. TWO EXAMPLES

We now provide two examples that illustrate the need for further restrictions on $\mathbf{P}$ so that (8) holds. The family of distributions used in the first example is the same as that used in Romano (1989) to show that the bootstrap does not behave well uniformly over certain classes of distributions and used in Romano (2004) to show that the t-test has size equal to 1 in finite samples over certain classes of distributions.

EXAMPLE 2.1: Suppose $d = 1$, $m = 1$, and $g(x, \theta) = x$ for all $\theta \in \Theta$. Let $\mathbf{P}$ be any set of probability distributions that satisfy (6) and (7) and contain

$$\mathbf{C}_0 = \{P_c : 0 < c < 1\},$$

where P_c is the measure with mass $1 - c$ on c and mass c on $-(1 - c)$. Define the event

$$E_n = \{X_i = c \text{ for all } 1 \leq i \leq n\}$$

and notice that E_n implies that $\mathbf{M}_0(\hat{P}_n) = \emptyset$. Since the empirical likelihood ratio test rejects whenever $\mathbf{M}_0(\hat{P}_n) = \emptyset$, we obtain for any $\eta > 0$ that

$$P_c^n\{\hat{P}_n \in \Lambda_1(\eta)\} \geq P_c^n\{E_n\} = (1 - c)^n. \tag{9}$$

Moreover, since $\mathbf{C}_0 \subseteq \mathbf{P}_0$, it follows from (9) that

$$\sup_{P\in\mathbf{P}_0} \limsup_{n\to\infty} \frac{1}{n} \log P^n\{\hat{P}_n \in \Lambda_1(\eta)\} \geq \sup_{c\in(0,1)} \log(1-c) = 0. \tag{10}$$

We conclude, in particular, that (8) cannot be satisfied for any $\mathbf{P}$ that includes $\mathbf{C}_0$.

The above example suggests that if $\mathbf{P}$ is "too rich," then the empirical likelihood ratio test cannot satisfy (8) for any value of $\eta > 0$. It is important to note that this would remain true even if we were to allow the critical value η to depend on n. Furthermore, this shortcoming is not unique to empirical likelihood in that it is shared by many commonly used tests. In particular, Example 2.1 also applies to the t-test. To see this, let P_c and E_n be defined as in Example 2.1 and simply note that the sample mean, $\bar{X}_n$, equals c and the standard error, $\hat{\sigma}_n$, defined by

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{1\leq i\leq n} (X_i - \bar{X}_n)^2,$$

equals zero whenever E_n occurs. Hence, for any critical value $\eta > 0$,

$$P_c^n\{\sqrt{n}|\bar{X}_n| \geq \eta\hat{\sigma}_n\} \geq P_c^n\{E_n\} = (1-c)^n. \tag{11}$$

By arguing as in (10), it follows that the t-test also fails to control size in the sense of Hoeffding (1965) whenever $\mathbf{C}_0 \subseteq \mathbf{P}$. Note that this failure persists even if we estimate the standard error without centering, because (11) continues to hold provided $\sqrt{n} \geq \eta$. For this reason, Example 2.1 applies to the generalized method of moments (GMM) based J-test proposed in Hansen (1982). A similar argument shows that Example 2.1 also applies to tests based on generalized empirical likelihood.

The simplicity of Example 2.1 is illustrative but potentially misleading. First, it is not enough to simply rule out discrete distributions. Indeed, one could modify Example 2.1 by "smoothing" out the mass on either side of zero and it would still apply. Second, it is also not sufficient to rule out distributions that have "too little" mass on one side of zero, as shown by Example 2.2 below.

EXAMPLE 2.2: As in the previous example, assume $d = 1$, $m = 1$, and $g(x, \theta) = x$ for all $\theta \in \Theta$. Let $\mathbf{P}$ be any set of probability distributions that satisfy (6) and (7) and contain

$$\mathbf{K}_0 = \left\{P_{K,c} = cD_{-1} + (1-c)R_{K,c} : 0 < c < \frac{1}{2}, K \geq 2\right\},$$

where D_{-1} is the degenerate distribution at -1 and $R_{K,c}$ is the distribution that satisfies

$$R_{K,c}\left\{X = \frac{-2c}{(1-c)(K-1)}\right\} = \frac{1}{2}, \quad R_{K,c}\left\{X = \frac{2Kc}{(1-c)(K-1)}\right\} = \frac{1}{2}.$$

As in Example 2.1, we show that empirical likelihood fails to satisfy (8) by constructing an event under which it rejects too often. To this end, define

$$A_n = \{X_i \neq -1 \text{ for all } 1 \leq i \leq n\}.$$

Note that there is a unique measure $F_{K,c}$ in $\mathbf{M}_0(R_{K,c})$ given by

$$F_{K,c}\left\{X = \frac{-2c}{(1-c)(K-1)}\right\} = \frac{K}{1+K},$$
$$F_{K,c}\left\{X = \frac{2Kc}{(1-c)(K-1)}\right\} = \frac{1}{1+K}.$$

We can therefore conclude by direct calculation that there is a K_η sufficiently large for which

$$\begin{aligned}\inf_{P\in\mathbf{M}_0(R_{K_\eta,c})} I\big(R_{K_\eta,c}|P\big) &= I\big(R_{K_\eta,c}|F_{K_\eta,c}\big)\\ &= \frac{1}{2}\log\left(\frac{1+K_\eta}{2K_\eta}\right) + \frac{1}{2}\log\left(\frac{1+K_\eta}{2}\right) > \eta.\end{aligned}$$

Next note that under $R^n_{K_\eta,c}$, either $\mathbf{M}_0(\hat{P}_n) = \emptyset$ or $\mathbf{M}_0(\hat{P}_n) = \{F_{K_\eta,c}\}$. It follows that

$$\begin{aligned}&\liminf_{n\to\infty}\frac{1}{n}\log R^n_{K_\eta,c}\{\hat{P}_n \in \Lambda_1(\eta)\}\\ &\quad = \liminf_{n\to\infty}\frac{1}{n}\log R^n_{K_\eta,c}\left\{\inf_{P\in\mathbf{M}_0(\hat{P}_n)} I(\hat{P}_n|P) \geq \eta\right\}\\ &\quad \geq \liminf_{n\to\infty}\frac{1}{n}\log R^n_{K_\eta,c}\big\{I\big(\hat{P}_n|F_{K_\eta,c}\big) \geq \eta \cap \mathbf{M}_0(\hat{P}_n) \neq \emptyset\big\}\\ &\quad = 0,\end{aligned}$$

where the final equality follows from observing that $I(\hat{P}_n|F_{K_\eta,c})$ converges in probability to $I(R_{K_\eta,c}|F_{K_\eta,c})$ and $\mathbf{M}_0(\hat{P}_n) \neq \emptyset$ with probability approaching 1 under $R^n_{K_\eta,c}$. We conclude that

$$
\begin{aligned}
&\sup_{P\in\mathbf{K}_0}\limsup_{n\to\infty}\frac{1}{n}\log P^n\{\hat{P}_n\in\Lambda_1(\eta)\}\\
&\quad\geq\sup_{c\in(0,1/2)}\Big\{\liminf_{n\to\infty}\frac{1}{n}\log P^n_{K_\eta,c}\{\hat{P}_n\in\Lambda_1(\eta)|A_n\}\\
&\qquad+\liminf_{n\to\infty}\frac{1}{n}\log P^n_{K_\eta,c}\{A_n\}\Big\}\\
&\quad=\sup_{c\in(0,1/2)}\Big\{\liminf_{n\to\infty}\frac{1}{n}\log R^n_{K_\eta,c}\{\hat{P}_n\in\Lambda_1(\eta)\}+\liminf_{n\to\infty}\frac{1}{n}\log P^n_{K_\eta,c}\{A_n\}\Big\}\\
&\quad=\sup_{c\in(0,1/2)}\log(1-c)\\
&\quad=0.
\end{aligned}
$$

Therefore, (8) cannot be satisfied for any $\mathbf{P}$ that includes $\mathbf{K}_0$.

Both Examples 2.1 and 2.2 rely on a sequence of distributions for which the rate at which the probability of a Type I error tends to zero itself tends to zero. In Example 2.1, this sequence is denoted by P_c and in Example 2.2, this sequence is denoted by $P_{K_\eta,c}$. While they appear different in nature, these sequences are, in fact, linked together by the observation that

$$\lim_{c\to 0}P_c=\lim_{c\to 0}P_{K_\eta,c}=D_0, \tag{12}$$

where D_0 is the degenerate distribution at 0 and the limits should be interpreted in the weak topology. In the special case where $\mathcal{X}=[-1,1]$, $d=1$, $m=1$, and $g(x,\theta)=x$ for all $\theta\in\Theta$, it follows from our general results in the next section that D_0 is a limit point of any sequence of distributions for which the (exponential) rate at which the probability of a Type I error tends to zero itself tends to zero. Therefore, by requiring that D_0 not be a limit point of $\mathbf{P}$, we show that the test of Qin and Lawless (1994) in this setting satisfies (8) at least for all sufficiently small $\eta>0$.

REMARK 2.1: It is worth noting the implications of Example 2.1 for the ability of the empirical likelihood ratio test to control size in finite samples when only requirements (6) and (7) are imposed on $\mathbf{P}$. To this end, recall the setup of Example 2.1 and note that (9) implies

$$\sup_{P\in\mathbf{P}_0}P\{\hat{P}_n\in\Lambda_1(\eta)\}\geq\sup_{c\in(0,1)}P_c\{\hat{P}_n\in\Lambda_1(\eta)\}=1.$$

Indeed, when only (6) and (7) are imposed on $\mathbf{P}$, it follows from Bahadur and Savage (1956) that there is no "reasonable" test in this setting in the sense that no test can have power against any alternative greater than size. On the other hand, Example 2.1 applies even if we impose the additional requirement that $\mathcal{X} = [-1, 1]$. With this additional requirement, the results of Bahadur and Savage (1956) no longer apply and reasonable tests exist. See, for example, Romano and Wolf (2000).

REMARK 2.2: In parametric models where $\mathbf{P}$ consists of distributions with common finite support, the likelihood ratio test for $H_0 : E_P[X] = 0$ versus $H_1 : E_P[X] \neq 0$ rejects for large values of

$$\inf_{P \in \mathbf{P}_0} I(\hat{P}_n | P), \tag{13}$$

which equals (4) with probability approaching 1 under any $P \in \mathbf{P}$ as n tends to infinity; see Newey and Smith (2004). In more general settings, however, it is important that the empirical likelihood ratio test takes the infimum over $\mathbf{M}_0(\hat{P}_n)$ rather than over $\mathbf{P}_0$. For example, if $\mathbf{P}$ is only required to satisfy (6) and (7), it is straightforward to see that (13) equals zero while the test statistic of Qin and Lawless (1994) converges to a chi-squared random variable.

3. THE MAIN RESULT

Our main result requires the following assumptions.

ASSUMPTION 3.1: *$\mathcal{X}$ and Θ are compact subsets of $\mathbf{R}^d$ and $\mathbf{R}^r$, respectively.*

ASSUMPTION 3.2: *$g : \mathcal{X} \times \Theta \to \mathbf{R}^m$ is continuous in both of its arguments.*

The compactness of $\mathcal{X}$, imposed in Assumption 3.1, ensures that $\mathbf{M}$ is itself compact in the weak topology, a crucial point in the proof of our main result. As implied by Examples 2.1 and 2.2, however, additional requirements must be imposed on the set of probability measures $\mathbf{P}$ so that empirical likelihood is able to control size as in (8). To this end, let

$$\Theta_0(P) = \{\theta \in \Theta : E_P[g(X, \theta)] = 0\}, \tag{14}$$

$$\mathbf{M}_0 = \{P \in \mathbf{M} : E_P[g(X, \theta)] = 0 \text{ for some } \theta \in \Theta\}. \tag{15}$$

Additionally, for a convex set $C \subseteq \mathbf{R}^m$, define its dimension to be the smallest integer $\dim(C)$ such that there exists an affine subspace of dimension $\dim(C)$ containing C. Letting $s(P, \theta)$ be the dimension of the convex hull of the support of $g(X, \theta)$ under P, we then define

$$\mathbf{D}_0 = \{P \in \mathbf{M} : \exists \theta \in \Theta_0(P) \text{ with } s(P, \theta) < m\}, \tag{16}$$

which we note is equal to $\{D_0\}$ when $d=1$, $0\in\mathcal{X}$, and $g(x,\theta)=x$ for all $\theta\in\Theta$.

The set $\mathbf{P}$ is then restricted by requiring that it satisfy the following additional assumptions.

ASSUMPTION 3.3: $\mathbf{P}\subseteq\mathbf{M}$ *is closed in the weak topology*.

ASSUMPTION 3.4: *For each* $P\in\mathbf{P}$, $\Theta_0(P)$ *is either empty or a singleton denoted* $\theta_0(P)$.

ASSUMPTION 3.5: *For some* $\varepsilon>0$, $\mathbf{P}\cap\mathbf{D}_0^\varepsilon=\emptyset$.

Given the compactness of $\mathbf{M}$, Assumption 3.3 implies $\mathbf{P}$ is compact as well. It is left as a high level assumption, but we note closed sets in the weak topology are easily constructed by imposing moment restrictions on bounded continuous functions. Assumption 3.4 is employed to show that $\theta_0(P)$ is continuous in P under the weak topology. Since we are typically interested in cases where $m>r$, we feel that Assumption 3.4 is not particularly restrictive. Assumption 3.5 is made precisely to avoid Examples 2.1 and 2.2 as it ensures $\mathbf{D}_0$ contains no limit points of $\mathbf{P}$. We note that, provided Assumption 3.4 is met, we can let $\mathbf{P}=\mathbf{M}\setminus\mathbf{D}_0^\varepsilon$.

We are now in a position to state our main result:

THEOREM 3.1: *Let* X_i, $i=1,\dots,n$, *be an i.i.d. sequence of random variables with distribution* $P\in\mathbf{P}$. *Let* $(\Lambda_0(\eta),\Lambda_1(\eta))$ *be defined as in* (5). *Suppose Assumptions 3.1–3.5 hold. Then we observe the following results*:

(a) *There exists* $\bar\eta>0$ *such that for all* $0<\eta\le\bar\eta$, *we have that*

$$\sup_{P\in\mathbf{P}_0}\limsup_{n\to\infty}\frac{1}{n}\log P^n\{\hat P_n\in\Lambda_1(\eta)\}\le-\eta.$$

(b) *If a test* $(\Omega_{0,n},\Omega_{1,n})$ *satisfies*

$$\sup_{P\in\mathbf{P}_0}\limsup_{n\to\infty}\frac{1}{n}\log P^n\{\hat P_n\in\Omega_{1,n}^\delta\}\le-\eta \tag{17}$$

for some $\delta>0$, *then*

$$\limsup_{n\to\infty}\frac{1}{n}\log Q^n\{\hat P_n\in\Omega_{0,n}\}\ge\limsup_{n\to\infty}\frac{1}{n}\log Q^n\{\hat P_n\in\Lambda_0(\eta)\}$$

for any $Q\in\mathbf{P}$ *satisfying*

$$\inf_{P\in\overline{(\mathbf{M}_0\setminus\mathbf{P}_0)}^{\sqrt{\eta/2}}}I(Q|P)\ge\inf_{P\in\mathbf{P}_0\setminus\overline{(\mathbf{M}_0\setminus\mathbf{P}_0)}^{\sqrt{\eta/2}}}I(Q|P), \tag{18}$$

where, for $A\subseteq\mathbf{M}$, $\overline{A}^\delta$ *denotes the closure of* A^δ *with respect to the weak topology*.

For the proof, see the Supplemental Material.

Theorem 3.1 establishes the desired optimality property of empirical likelihood for testing (2). First, the probability of a Type I error when using empirical likelihood tends to zero at a (exponential) rate that is bounded away from zero on $\mathbf{P}_0$. Second, for any distribution Q satisfying (18), the probability of a Type II error when using empirical likelihood vanishes at a (exponential) rate at least as fast as that of any nonrandomized test satisfying the requirement (17). We emphasize the nonlocal nature of the optimality property, in that it holds for all distributions Q satisfying (18). Heuristically, condition (18) demands that Q be "closer" to $\mathbf{P}_0$ (the subset of $\mathbf{M}_0$ over which we demand control of the rate at which the probability of a Type I error tends to zero) than to $\mathbf{M}_0 \setminus \mathbf{P}_0$ (the subset of $\mathbf{M}_0$ over which we do not impose this requirement). Note in the special case where $\mathbf{M}_0 = \mathbf{P}_0$ that $\mathbf{M}_0 \setminus \mathbf{P}_0$ is empty, implying that the left hand side of (18) is infinite, so (18) is satisfied for all $P \in \mathbf{P}$.

As illustrated by Examples 2.1 and 2.2, commonly used tests for (2) also fail to control the (exponential) rate at which the probability of a Type I error tends to zero if we allow for distributions in neighborhoods of $\mathbf{D}_0$. Rather than examining the performance of tests on restricted sets $\mathbf{P}$ that exclude $\mathbf{D}_0$, an alternative method of comparison among such procedures is to require

$$\sup_{P\in\mathbf{P}_0} \limsup_{n\to\infty} \frac{1}{n} \log P^n\{\hat{P}_n \in (\Omega_{1,n} \setminus \mathbf{D}_0^{\varepsilon})^{\delta}\} \leq -\eta \tag{19}$$

instead of (17). Requirement (19) should not be interpreted as "size" control, but instead as a benchmark for tests that have difficulty controlling the rate at which the Type I error tends to zero in neighborhoods of $\mathbf{D}_0$. Given the weaker criterion (19), it is clear that any optimal test must satisfy $\mathbf{D}_0^{\varepsilon} \subseteq \Omega_{1,n}$. For this reason, consider

$$\begin{aligned} \tilde{\Lambda}_0(\eta) &= \Lambda_0(\eta) \setminus \mathbf{D}_0^{\varepsilon}, \\ \tilde{\Lambda}_1(\eta) &= \Lambda_1(\eta) \cup \mathbf{D}_0^{\varepsilon}, \end{aligned} \tag{20}$$

where the dependence on ε is omitted in the notation. Note that the tests $(\Lambda_0(\eta), \Lambda_1(\eta))$ and $(\tilde{\Lambda}_0(\eta), \tilde{\Lambda}_1(\eta))$ differ only on the event $\hat{P}_n \in \mathbf{D}_0^{\varepsilon}$. We can use Theorem 3.1 to show that the optimal test in this framework is given by the modified empirical likelihood test defined by (20).

COROLLARY 3.1: *Let X_i, $i = 1, \ldots, n$, be an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$. Let $(\tilde{\Lambda}_0(\eta), \tilde{\Lambda}_1(\eta))$ be defined as in* (20). *Suppose Assumptions* 3.1–3.4 *hold. Then we observe the following results*:

(a) *There exists $\bar{\eta}(\varepsilon) > 0$ such that for all $0 < \eta \leq \bar{\eta}(\varepsilon)$, we have that*

$$\sup_{P\in\mathbf{P}_0} \limsup_{n\to\infty} \frac{1}{n} \log P^n\{\hat{P}_n \in \tilde{\Lambda}_1(\eta) \setminus \mathbf{D}_0^{\varepsilon}\} \leq -\eta.$$

(b) *If a test* $(\Omega_{0,n}, \Omega_{1,n})$ *satisfies*

$$\sup_{P\in\mathbf{P}_0}\limsup_{n\to\infty}\frac{1}{n}\log P^n\{\hat{P}_n\in(\Omega_{1,n}\setminus\mathbf{D}_0^{\varepsilon})^{\delta}\}\leq-\eta \tag{21}$$

for some $\delta>0$, *then for every probability measure* $Q\in\mathbf{P}$,

$$\limsup_{n\to\infty}\frac{1}{n}\log Q^n\{\hat{P}_n\in\Omega_{0,n}\}\geq\limsup_{n\to\infty}\frac{1}{n}\log Q^n\{\hat{P}_n\in\tilde{\Lambda}_0(\eta)\}. \tag{22}$$

See the Supplemental Material for the proof.

We reiterate that (19) differs from (17) only in how the former treats distributions that are close to the set $\mathbf{D}_0$. Remarkably, as a result of this rather simple modification, it is possible to remove Assumption 3.5 entirely. Moreover, in contrast to Theorem 3.1, (22) holds without qualifications on Q. This result may therefore provide some guidance when choosing among tests that have difficulty controlling the rate at which the Type I error tends to zero in neighborhoods of $\mathbf{D}_0$, such as tests based on (generalized) empirical likelihood or the GMM-based J-test.

REFERENCES

BAHADUR, R. R., AND L. J. SAVAGE (1956): "The Nonexistence of Certain Statistical Procedures in Nonparametric Problems," *Annals of Mathematical Statistics*, 27, 1115–1122. [419]

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054. [416]

HOEFFDING, W. (1965): "Asymptotically Optimal Tests for Multinomial Distributions," *Annals of Mathematical Statistics*, 36, 369–401. [414,416]

KITAMURA, Y. (2001): "Asymptotic Optimality of Empirical Likelihood for Testing Moment Restrictions," *Econometrica*, 69, 1661–1672. [413-415]

KITAMURA, Y., A. SANTOS, AND A. M. SHAIKH (2012): "Supplement to 'On the Asymptotic Optimality of Empirical Likelihood for Testing Moment Restrictions'," *Econometrica Supplemental Material*, 80, http://www.econometricsociety.org/ecta/Supmat/8773_Proofs.pdf. [415]

NEWEY, W. K., AND R. J. SMITH (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72, 219–255. [419]

OWEN, A. (1988): "Empirical Likelihood Ratio Confidence Intervals for a Single Functional," *Biometrika*, 75, 237–249. [414]

QIN, J., AND J. LAWLESS (1994): "Empirical Likelihood and Generalized Estimating Equations," *Annals of Statistics*, 22, 300–325. [414,415,418,419]

ROMANO, J. P. (1989): "Do Bootstrap Confidence Intervals Behave Well Uniformly in P?" *Canadian Journal of Statistics*, 17, 75–80. [415]

——— (2004): "On Nonparametric Testing, the Uniform Behavior of the t-test, and Related Problems," *Scandinavian Journal of Statistics*, 31, 567–584. [415]

ROMANO, J. P., AND M. WOLF (2000): "Finite Sample Nonparametric Inference and Large Sample Efficiency," *The Annals of Statistics*, 28, 756–778. [419]

Dept. of Economics, Yale University, New Haven, CT 06520-8281, U.S.A.; yuichi.kitamura@yale.edu, *Dept. of Economics, University of California–San Diego, 9500 Gilman Drive 0508, La Jolla, CA 92093-0508, U.S.A.; a2santos@ucsd.edu,*

and

Dept. of Economics, University of Chicago, Chicago, IL 60637, U.S.A.; amshaikh@uchicago.edu.

Manuscript received August, 2009; final revision received April, 2011.