HODGES-LEHMANN OPTIMALITY FOR TESTING MOMENT CONDITIONS

By

Ivan Canay and Taisuke Otsu

March 2011

COWLES FOUNDATION DISCUSSION PAPER NO. 1789

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Hodges-Lehmann Optimality for Testing Moment Conditions*

IVAN CANAY[1] and TAISUKE OTSU[2,†]

[1] *Department of Economics, Northwestern University, 2001 Sheridan Road, Evanston, Illinois 60208, USA. E-mail:* iacanay@northwestern.edu*. url:*
http://faculty.wcas.northwestern.edu/~iac879/

[2] *Department of Economics, Yale University, 30 Hillhouse Ave., Rm. 18, New Haven, CT 06520, USA. E-mail:* taisuke.otsu@yale.edu*. url:*
http://cowles.econ.yale.edu/faculty/otsu.htm

Abstract. This paper studies the Hodges and Lehmann (1956) optimality of tests in a general setup. The tests are compared by the exponential rates of growth to one of the power functions evaluated at a fixed alternative while keeping the asymptotic sizes bounded by some constant. We present two sets of sufficient conditions for a test to be Hodges-Lehmann optimal. These new conditions extend the scope of the Hodges-Lehmann optimality analysis to setups that cannot be covered by other conditions in the literature. The general result is illustrated by our applications of interest: testing for moment conditions and overidentifying restrictions. In particular, we show that (i) the empirical likelihood test does not necessarily satisfy existing conditions for optimality but does satisfy our new conditions; and (ii) the generalized method of moments (GMM) test and the generalized empirical likelihood (GEL) tests are Hodges-Lehmann optimal under mild primitive conditions. These results support the belief that the Hodges-Lehmann optimality is a weak asymptotic requirement.

Keywords: asymptotic optimality, large deviations, moment condition, generalized method of moments, generalized empirical likelihood.

1. Introduction

There are numerous testing problems in statistics and econometrics where alternative tests under consideration have the same asymptotic properties under the null hypothesis and local alternatives. As asymptotic comparisons are intended to approximate finite sample behaviors, it is important to assess whether such equivalence is preserved in different asymptotic frameworks. This paper studies an alternative notion of asymptotic

*First version: October 2009. This version: January 11, 2011. We would like to thank Andrew Barron, Patrik Guggenberger, Joel Horowitz, Azeem Shaikh, and seminar participants at Brown, Columbia, Cornell, Indiana, Northwestern, NYU, Princeton, Stanford, Texas-Austin, UCLA, UCSD, Virginia and the Econometrics Jamboree at Duke for helpful comments.

†Financial support from the National Science Foundation (SES-0720961) is gratefully acknowledged.

comparison of tests due to Hodges and Lehmann (1956) in a general setup. More specifically, we focus on global properties and compare the tests in terms of the exponential rate of growth to one of the power functions evaluated at a fixed alternative while keeping the asymptotic sizes bounded by some constant. We present two sets of sufficient conditions for a test to be Hodges-Lehmann optimal. These new conditions extend the scope of the Hodges-Lehmann optimality analysis to setups that cannot be covered by other conditions in the literature (e.g. Kallenberg and Kourouklis, 1992). This point is illustrated by our applications of interest: testing for moment conditions and overidentifying restrictions (or generalized estimating equations). In particular, we show that the empirical likelihood test (Owen, 1988; Qin and Lawless, 1994) does not necessarily satisfy the existing conditions but does satisfy the new conditions we propose, and that the generalized method of moments (GMM) test of Hansen (1982) and the generalized empirical likelihood (GEL) tests of Smith (1997) and Newey and Smith (2004) (including empirical likelihood, continuous updating GMM, and exponential tilting as special cases) are Hodges-Lehmann optimal for testing overidentifying restrictions under mild primitive conditions.

The dominant approach to approximate finite sample power functions in statistics and econometrics is based on sequences of local (or Pitman) alternatives. There are yet some reasons to go beyond the local analysis. First, although the local analysis might provide a good approximation of the power function for alternatives close to the null hypothesis, there are risks in extrapolating whatever lessons we learn locally to alternatives that are far from the null. This is particularly true, for example, when the finite sample power function is non-monotone. Second, there are cases where different tests, with different exact power functions, have the same asymptotic behavior under local alternatives. Then it is important to look for approximations that are pertinent for the regions of high power (as it is the case for the Hodges-Lehmann approach) and see if such equivalence is preserved in those regions.

Our Hodges-Lehmann optimality analysis contributes to the literature in several ways. First, we show that the existing general sufficient conditions by Kallenberg and Kourouklis (1992) for a test to be Hodges-Lehmann optimal are too strong for our applications of interest. We provide an example where the empirical likelihood test does not satisfy an even weaker version of those sufficient conditions. Second, we provide novel sets of sufficient conditions for the Hodges-Lehmann optimality. One set is similar to the conditions in Kallenberg and Kourouklis (1992), although we require lower semicontinuity in the weak topology instead of continuity in the τ-topology. The other set involves a localized version of semicontinuity and turns out to be very useful to analyze discontinuous cases. In our applications of interest, this new condition allows us to establish the Hodges-Lehmann optimality of the GEL tests. Our conditions and results are presented in a general hypothesis testing framework. Thus, they have wide applicability as a starting point for studying the Hodges-Lehmann optimality in other applications. Third, we apply our sufficient conditions to the problems of testing moment conditions and overidentifying restrictions. For testing moment conditions, we show that the Hotelling's T and GEL tests are Hodges-Lehmann optimal. For testing overidentifying restrictions (i.e., testing the validity of estimating equations whose dimension is higher than that of parameters), we show that the GMM and GEL tests are Hodges-Lehmann optimal. These findings together with the mildness of the new sufficient conditions provide further evidence to the belief that the Hodges-Lehmann optimality seems to be a weak asymptotic requirement.

Our application to overidentified moment condition models (or generalized estimating equations) is of extreme importance particularly in econometrics. It is known that the GMM and GEL tests have the same asymptotic properties under the null hypothesis and local alternatives. Several papers study statistical properties of the GMM and GEL methods beyond their first-order local asymptotic properties (e.g. Imbens, Spady and Johnson, 1998; Newey and Smith, 2004; Schennach, 2007). In terms of global analysis based on large deviation theory, Kitamura (2001) and Kitamura, Santos and Shaikh (2009) provide conditions under which the empirical likelihood test is uniformly most powerful in a generalized Neyman-Pearson sense for testing overidentifying restrictions. Additional global optimality results include those in Canay (2010) and Otsu (2010). We deviate from these papers in two important dimensions. First, the type I error probability in our Hodges-Lehmann analysis converges to a positive constant, as opposed to converging to zero. Although the approximations that take the type I error probability to zero may help understanding the differences between different tests, they do not resemble the standard situation where a test statistic is compared to a fixed asymptotic critical value. Second, we provide Hodges-Lehmann optimality results for commonly used tests of moment conditions. On the other hand, the previous papers prove that the empirical likelihood test achieves some form of global optimality, but do not address the possibility that other competing tests are optimal as well.

The remainder of the paper is organized as follows. Section 2 introduces the basic notation and concepts, and presents the general Hodges-Lehmann optimality results. Section 3 applies the general optimality results to moment condition tests and overidentifying restriction tests.

We use the following notation. Let $\bar{\mathbb{R}} \equiv \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$ be the extended real line, A^c be the complement of a set A, $A \setminus B \equiv A \cap B^c$ be the set subtraction of a set B from a set A, $1\{A\}$ be the indicator function for an event A, $\Pr\{A : P\}$ be the probability of an event A evaluated under a probability measure P, $E_P[\cdot]$ be the mathematical expectation under a probability measure P, and "$\Rightarrow$" denote the weak convergence.

2. General Results

Consider a random sample $\{x_i : i = 1, \dots, n\}$ generated from a probability measure P_0 with support $\mathcal{X}$. Let $\mathcal{M}$ be the set of all probability measures on $\mathcal{X}$. For subsets $\mathcal{P}$ and $\mathcal{Q}$ of $\mathcal{M}$ with $\mathcal{P} \subset \mathcal{Q}$, we consider the hypothesis testing problem

$$H_0 : P_0 \in \mathcal{P}, \text{ versus } H_1 : P_0 \in \mathcal{Q} \setminus \mathcal{P}.$$

A test ϕ_n is defined as a binary function of the sample, where $\phi_n = 0$ means acceptance and $\phi_n = 1$ means rejection. Performance of ϕ_n is evaluated by two kinds of error probabilities: $\alpha_n(P) = E_P[\phi_n]$ for $P \in \mathcal{P}$ (type I) and $\beta_n(P) = E_P[1 - \phi_n]$ for $P \in \mathcal{Q} \setminus \mathcal{P}$

(type II). The Hodges-Lehmann optimality analysis focuses on the convergence rate of the type II error probability $\beta_n(P_1)$ (or power) of the test under a fixed alternative $P_1 \in \mathcal{Q} \setminus \mathcal{P}$, while fixing the limit of the type I error probability $\alpha_n(P)$ over $P \in \mathcal{P}$. Our definition of the Hodges-Lehmann optimality is given below.

***Definition* 2.1** (Hodges-Lehmann optimality)**.** *A test $\phi_{HL,n}$ is called* Hodges-Lehmann optimal *at $P_1 \in \mathcal{Q} \setminus \mathcal{P}$ if*

(i) $\phi_{HL,n}$ is pointwise asymptotically level $\alpha \in (0,1)$, i.e.,

$$\limsup_{n\to\infty} E_P[\phi_{HL,n}] \le \alpha \quad \text{for each } P \in \mathcal{P},$$

(ii) for any pointwise asymptotically level α test ϕ_n, it holds

$$\limsup_{n\to\infty} \frac{1}{n} \log E_{P_1}[1-\phi_{HL,n}] \le \liminf_{n\to\infty} \frac{1}{n} \log E_{P_1}[1-\phi_n].$$

This is, given a restriction on the type I error probability, a test is called Hodges-Lehmann optimal at the fixed alternative measure P_1 if the rate of exponential convergence of the type II error probability evaluated at P_1 is faster than that of any alternative test. Although this definition for optimality is intuitive, the set of alternative tests is potentially very large and therefore it might be infeasible to explore the second inequality in Definition 2.1 for every possible alternative test. The approach we take here divides the analysis in two parts. First, we show that there exists an optimal convergence rate for the type II error probability (or equivalently, a lower bound for $\liminf_{n\to\infty} n^{-1} \log E_{P_1}[1-\phi_n]$). Then we investigate sufficient conditions to achieve the optimal rate.

We first derive the optimal convergence rate of the type II error probability. Let $Q \ll P$ denote that Q is absolutely continuous with respect to P, and

$$K(Q,P) \equiv \begin{cases} \int_{\mathcal{X}} \log(dQ/dP)dQ & \text{if } Q \ll P \\ \infty & \text{otherwise} \end{cases}$$

denote the Kullback-Leibler divergence (or relative entropy) for probability measures Q and P. Define $K(\mathcal{A},P) \equiv \inf_{Q\in\mathcal{A}} K(Q,P)$ for a subset $\mathcal{A} \subseteq \mathcal{M}$. The following lemma presents the best possible exponential rate of decay to zero of the type II error probability of a test.

Lemma 2.1. *For any pointwise asymptotically level α test ϕ_n, it holds*

$$\liminf_{n\to\infty} \frac{1}{n} \log E_{P_1}[1-\phi_n] \ge -K(\mathcal{P},P_1),$$

for each $P_1 \in \mathcal{Q} \setminus \mathcal{P}$.

This lemma, an adaptation of Stein's lemma to our setup, shows that the best exponential growth rate of power depends on the Kullback-Leibler divergence between the set $\mathcal{P}$ for the null hypothesis and the fixed alternative measure P_1. The bound $-K(\mathcal{P}, P_1)$ is informative for alternatives with $0 < K(\mathcal{P}, P_1) < \infty$. It is interesting to note that in the Bahadur optimality analysis (e.g. Bahadur, 1960), where the roles of the type I and type II error probabilities are interchanged, the best possible rate of the type I error is typically obtained as $-K(P_1, \mathcal{P})$. Since $K(P, Q) \neq K(Q, P)$ in general, this Bahadur bound is different from the Hodges-Lehmann bound obtained here.

To achieve the bound in Lemma 2.1, we concentrate on tests that take the form of

$$\phi_n = 1\{T(\hat{P}_n) > c_n\}, \tag{1}$$

where $T(\hat{P}_n)$ is a test statistic based on a mapping $T : \mathcal{M} \to \bar{\mathbb{R}}$ and the empirical measure $\hat{P}_n$, and $\{c_n : n \in \mathbb{N}\}$ is a sequence of positive real numbers monotonically decreasing to zero. Given this form, our task reduces to explore sufficient conditions for the mapping T to attain the bound in Lemma 2.1.

There are results in the literature which indicate that several tests can be Hodges-Lehmann optimal in standard testing problems, such as parameter hypothesis and goodness-of-fit testing problems (see, Kallenberg and Kourouklis, 1992; Tusnády, 1977). In particular, Kallenberg and Kourouklis (1992) show that the Hodges-Lehmann optimality emerges in general when the acceptance region of a test converges to the set of measures for the null hypothesis in a coarse way, provided the mapping T is continuous in the τ-topology. We show that their continuity assumption in the τ-topology can be replaced with a lower semicontinuity assumption or its localized version in the weak topology. Our conditions are presented as follows. Condition 2.1 is fundamental, and either Condition 2.2 or 2.3 is required for the optimality.

Condition 2.1. $\mathcal{P} = \{Q \in \mathcal{Q} : T(Q) \leq 0\}$.

Condition 2.2. *T is lower semicontinuous in the weak topology on $\{Q \in \mathcal{Q} : K(Q, P_1) < \infty\}$.*

Condition 2.3. *$\mathcal{P}$ and $\mathcal{Q}$ are compact in the weak topology. Furthermore, T is such that $T(Q) \leq 0$ whenever a sequence of measures $\{Q_m : m \in \mathbb{N}\}$ in $\mathcal{Q}$ and a positive sequence $\{\eta_m : m \in \mathbb{N}\}$ decreasing to zero satisfy $Q_m \Rightarrow Q \in \mathcal{Q}$ and $T(Q_m) \leq \eta_m$ for all $m \in \mathbb{N}$.*

Condition 2.1, imposed by Kallenberg and Kourouklis (1992) as well, says that the set of measures for the null hypothesis should coincide with the level set by the mapping T at zero (or the acceptance region in the limit). Condition 2.2 is on the continuity of T. Relative to Kallenberg and Kourouklis (1992) this condition uses a weaker notion of continuity and a weaker topology, meaning that it is neither stronger nor weaker than theirs. That is, lower semicontinuity in the weak topology implies lower semicontinuity in the τ-topology, but does not imply continuity in the τ-topology as required by Kallenberg and Kourouklis (1992). Although Condition 2.2 seems intuitive and mild, this condition may be too restrictive to accommodate the GEL tests for testing moment conditions or overidentifying restrictions, which will be discussed in the next section. Example 3.1 below demonstrates that the mapping to define the empirical likelihood test is not lower semicontinuous in the weak (or τ) topology. Motivated by this problem, we propose an alternative requirement in Condition 2.3. Note that Condition 2.3 is neither weaker nor stronger than Condition 2.2. Condition 2.3 requires that the sets $\mathcal{P}$ and $\mathcal{Q}$ are compact in the weak topology, which is not imposed in Condition 2.2. On the other hand, the continuity requirement on T of Condition 2.3, which is a localized version of the lower semicontinuity, is weaker than that of Condition 2.2 and can accommodate the mappings for the GEL tests discussed in the next section. In our applications, these conditions are verified under some primitive conditions.

Based on these conditions, our general Hodges-Lehmann optimality results are presented as follows.

Theorem 2.1. *Suppose that a test ϕ_n taking the form of (1) is pointwise asymptotically level α, and Condition 2.1 is satisfied. Then under either Condition 2.2 or 2.3, ϕ_n is Hodges-Lehmann optimal at each $P_1 \in \mathcal{Q} \setminus \mathcal{P}$ satisfying $0 < K(\mathcal{P}, P_1) < \infty$.*

The first part (the statement under Condition 2.2) is a generalization of Theorem 2.1 in Kallenberg and Kourouklis (1992). This part is useful to show the Hodges-Lehmann optimality of the Hotelling's T, two-step GMM, and continuous updating GMM tests. The second part (the statement under Condition 2.3) is applied to establish the Hodges-Lehmann optimality of the GEL test.

Note that Definition 2.1 and Theorem 2.1 apply to tests that are *pointwise* asymptotically level α. However, it is worth mentioning that we can alternatively define and present the results for *uniformly* asymptotically level α tests (i.e., $\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} E_P[\phi_n] \leq \alpha$), which is stronger than the pointwise asymptotic requirement. This would require additional restrictions in the applications of the next section.[1]

3. Applications

3.1. Test for Moment Conditions

We now apply the general Hodges-Lehmann optimality results obtained in the last section. In this subsection, we consider the testing problem for moment conditions $E_{P_0}[m(x)] = 0$, where $m : \mathcal{X} \to \mathbb{R}^q$ is a vector of known functions. Pick any $\epsilon > 0$, and define $\Sigma(P) \equiv E_P[(m(x) - E_P[m(x)])(m(x) - E_P[m(x)])']$ and

$$\mathcal{Q}_\epsilon \equiv \{P \in \mathcal{M} : \det(\Sigma(P)) \geq \epsilon\}, \quad \mathcal{P}_\epsilon \equiv \{P \in \mathcal{Q}_\epsilon : E_P[m(x)] = 0\}.$$

[1] For example, in order to control the size uniformly in the application of section 3.1, the set $\mathcal{Q}_\epsilon$ should impose bounded $2+\delta$ moments or a uniform integrability condition in addition to a restriction on the determinant.

The testing problem of interest is $H_0 : P_0 \in \mathcal{P}_\epsilon$ versus $H_1 : P_0 \in \mathcal{Q}_\epsilon \setminus \mathcal{P}_\epsilon$. The requirement in $\mathcal{Q}_\epsilon$ for the determinant is used to control the asymptotic size of tests. Note that we do not make parametric assumptions on the distributional form of P_0. For this problem, we consider the following setup.

Condition 3.1. *$\mathcal{X}$ is compact and m is continuous on $\mathcal{X}$.*

This condition guarantees that the sets $\mathcal{M}$, $\mathcal{P}_\epsilon$, and $\mathcal{Q}_\epsilon$ are compact in the weak topology (see, Theorem D.8 of Dembo and Zeitouni (1998) and Lemma B.3), and simplifies the technical argument below.

One way to test H_0 is to employ Hotelling's T-test statistic $T_H(\hat{P}_n)$, where

$$T_H(Q) \equiv E_Q[m(x)]'\Sigma(Q)^{-1}E_Q[m(x)].$$

Since $nT_H(\hat{P}_n) \Rightarrow \chi^2_q$ under H_0, the T-test is written as $\phi_{H,n} \equiv 1\{T_H(\hat{P}_n) > \chi^2_{q,1-\alpha}/n\}$, where $\chi^2_{q,1-\alpha}$ is the $(1-\alpha)$-th quantile of the χ^2_q distribution. Note that $\phi_{H,n}$ takes the form of (1).

An alternative way to test H_0 is to employ the GEL approach. For example, consider the Cressie and Read (1984) family of criterion functions

$$\rho_a(v) \equiv -(1+av)^{(a+1)/a}/(a+1),$$

for $a \in \mathbb{R}$. The GEL test statistic is defined as $T_a(\hat{P}_n)$, where

$$T_a(Q) \equiv \sup_{\gamma \in \Gamma_Q} E_Q[\rho_a(\gamma' m(x)) - \rho_a(0)],$$

$\Gamma_Q \equiv \{\gamma \in \mathbb{R}^q : \Pr\{\gamma' m(x) \in \mathcal{V} : Q\} = 1\}$, and $\mathcal{V}$ is the domain of $\rho_a(v)$. This GEL test statistic covers several existing statistics, such as empirical likelihood ($a = -1$), Hellinger distance ($a = -1/2$), exponential tilting ($a = 0$), and Hotelling's T-statistic ($a = 1$) discussed above. By Newey and Smith (2004), we can see that $2nT_a(\hat{P}_n) \Rightarrow \chi^2_q$ under H_0. Thus, the GEL test is written as $\phi_{a,n} \equiv 1\{T_a(\hat{P}_n) > \chi^2_{q,1-\alpha}/(2n)\}$ taking the form of (1).

By applying the general result in Theorem 2.1, we can show the Hodges-Lehmann optimality of the Hotelling's T and GEL tests.

Theorem 3.1. *Assume that Condition 3.1 holds, and pick any $\epsilon > 0$ and $a \in \mathbb{R}$. Then*

- *(i) the Hotelling's T-test $\phi_{H,n}$ is pointwise asymptotically level α and T_H satisfies Conditions 2.1 and 2.2, i.e., $\phi_{H,n}$ is Hodges-Lehmann optimal at each $P_1 \in \mathcal{Q}_\epsilon \setminus \mathcal{P}_\epsilon$ satisfying $0 < K(\mathcal{P}_\epsilon, P_1) < \infty$.*
- *(ii) the GEL test $\phi_{a,n}$ is pointwise asymptotically level α and T_a satisfies Conditions 2.1 and 2.3, i.e., $\phi_{a,n}$ is Hodges-Lehmann optimal at each $P_1 \in \mathcal{Q}_\epsilon \setminus \mathcal{P}_\epsilon$ satisfying $0 < K(\mathcal{P}_\epsilon, P_1) < \infty$.*

It is interesting to note that T_a is not necessarily continuous in the τ-topology, as required by Kallenberg and Kourouklis (1992). In fact, T_a does not necessarily satisfy our Condition 2.2, lower semicontinuity in the weak topology. Indeed, this lack of lower semicontinuity becomes our motivation to develop the alternative requirement in Condition 2.3. To illustrate the discontinuity of T_a, let us consider the case of empirical likelihood, where the mapping T_{EL} is defined by $\rho_a(v) = \log(1-v)$ with $a=-1$ and $\mathcal{V} = (-\infty, 1)$. The following example shows that T_{EL} is not lower semicontinuous both in the weak and τ-topology.

***Example* 3.1** (T_{EL} is not lower semicontinuous). *Suppose $m(x) = x$ and $\mathcal{X} = [-x_L, x_H]$ for some $x_L > 0$ and $x_H > 0$. Note that Condition 3.1 is satisfied. For a probability measure Q, let $\mathcal{X}_Q$ denote the support of Q and $-x_{L_Q}$ and x_{H_Q} denote the lower and upper bounds of $\mathcal{X}_Q$. If $\{Q_m : m \in \mathbb{N}\}$ is a sequence of measures, we use $-x_{L_m}$ and x_{H_m}. In this setup,*

$$T_{EL}(Q) \equiv \sup_{\gamma \in \Gamma_Q} \int_{\mathcal{X}} \log(1+\gamma x) dQ,$$

and $\Gamma_Q = (-1/x_{H_Q}, 1/x_{L_Q})$ (if $x_{H_Q} \le 0$ or $x_{L_Q} \le 0$, the reciprocals are set to ∞). This example shows that $\Lambda_\eta \equiv \{Q \in \mathcal{M} : T_{EL}(Q) \le \eta\}$ is not closed in the weak topology for any $\eta > 0$, meaning that T_{EL} is not lower semicontinuous in the weak topology.

Consider the following sequence of probability measures,

$$Q_m(X = -x_L) = \frac{1}{m}, \quad Q_m(X=0) = 1-p-\frac{1}{m}, \quad Q_m(X = x^*) = p,$$

for some $x^* \in (0, x_H)$. Note that $\Gamma_{Q_m} = (-1/x^*, 1/x_L)$ for all $m \in \mathbb{N}$. This sequence weakly converges to the probability measure Q satisfying $Q(X=0) = 1-p$ and $Q(X = x^*) = p$, where $x_{L_Q} = 0$ and then $\Gamma_Q = (-1/x^*, \infty)$. Note that x_{L_m} does not converge to $x_{L_Q} = 0$ since $x_{L_m} > 0$ for all $m \in \mathbb{N}$, so that $\liminf_{m\to\infty}(x_{L_n} - x_{L_Q}) > 0$. To show that Λ_η is not closed it is sufficient to prove that $Q_m \in \Lambda_\eta$ for all $m \in \mathbb{N}$ but $Q \notin \Lambda_\eta$. Since $\int_{\mathcal{X}} \log(1+\gamma x) dQ_m = \log(1 - x_L\gamma)/m + \log(1+\gamma x^*)p$, the value $\gamma_m^* \in \Gamma_{Q_m}$ that maximizes this integral is

$$\gamma_m^* = \frac{px^* - x_L/m}{(p+1/m)x_L x^*}.$$

As $\gamma_m^* \to 1/x_L$ as $m \to \infty$, it follows that $T_{EL}(Q_m) \nearrow \log(1 + x^*/x_L)p$. Therefore, if we pick the value x^* so that $x^* \le x_L(\exp(\eta/p) - 1)$, then it holds $T_{EL}(Q_m) \le \log(1 + x^*/x_L)p \le \eta$ for all $m \in \mathbb{N}$, i.e., $Q_m \in \Lambda_\eta$ for all $m \in \mathbb{N}$. However,

$$T_{EL}(Q) = \sup_{\gamma \in (-1/x^*, \infty)} \int_{\mathcal{X}} \log(1+\gamma x) dQ = \sup_{\gamma \in (-1/x^*, \infty)} \log(1+\gamma x^*)p = \infty.$$

Note that $T_{EL}(Q) = \infty$ regardless of how small x^* or p might be, as long as both are positive. Therefore, for any given $\eta > 0$ we can always find a sequence $Q_m \in \Lambda_\eta$ that weakly converges to $Q \notin \Lambda_\eta$, and so the mapping T_{EL} is not lower semicontinuous in the weak topology. Since it is also true that Q_m converges to Q in the τ-topology, it follows that the mapping T_{EL} is not lower semicontinuous in the τ-topology either. ■

3.2. Overidentifying Restriction Test

In this subsection, we consider the testing problem for overidentifying restrictions, which are common particularly in econometrics. Consider the (generalized) estimating functions $m : \mathcal{X} \times \Theta \to \mathbb{R}^q$, where $\Theta \subset \mathbb{R}^k$ is the parameter space. It is assumed that $q > k$, i.e., the parameter is overidentified. Let $\Sigma(P, \theta) \equiv E_P[(m(x,\theta) - E_P[m(x,\theta)])(m(x,\theta) - E_P[m(x,\theta)])']$ and $\mathcal{Q}_{\epsilon,\theta} \equiv \{P \in \mathcal{M} : \det(\Sigma(P,\theta)) \geq \epsilon\}$. We redefine

$$\mathcal{P}_\epsilon \equiv \cup_{\theta\in\Theta}\{P \in \mathcal{Q}_{\epsilon,\theta} : E_P[m(x,\theta)] = 0\}, \quad \mathcal{Q}_\epsilon \equiv \cup_{\theta\in\Theta}\mathcal{Q}_{\epsilon,\theta}.$$

The testing problem of interest is $H_0 : P_0 \in \mathcal{P}_\epsilon$ versus $H_1 : P_0 \in \mathcal{Q}_\epsilon \setminus \mathcal{P}_\epsilon$, i.e., the estimating equations are valid and the restriction $E_{P_0}[m(x,\theta_0)] = 0$ is satisfied at some $\theta_0 \in \Theta$. For this problem, we consider the following setup.

Condition 3.2. *$\mathcal{X}$ and Θ are compact, and m is continuous in both of its arguments.*

One common test for H_0 is based on the GMM of Hansen (1982). The two-step GMM test statistic is defined as $T_{GMM}(\hat{P}_n)$, where

$$T_{GMM}(Q) \equiv \inf_{\theta\in\Theta} E_Q[m(x,\theta)]'\Sigma(Q,\tilde{\theta}(Q))^{-1}E_Q[m(x,\theta)],$$

and $\tilde{\theta}(Q) \equiv \arg\min_{\theta\in\Theta} E_Q[m(x,\theta)]'WE_Q[m(x,\theta)]$ with a $q \times q$ fixed weight matrix W (i.e., $\tilde{\theta}(\hat{P}_n)$ is a preliminary estimator for θ_0). Here we consider the GMM test in the form of $\phi_{GMM,n} \equiv 1\{T_{GMM}(\hat{P}_n) > \chi^2_{q,1-\alpha}/n\}$.[2]

Alternatively, we can apply the GEL approach. Let $\Gamma_Q(\theta) \equiv \{\gamma \in \mathbb{R}^q : \Pr\{\gamma' m(x,\theta) \in \mathcal{V} : Q\} = 1)\}$. By using the criterion function ρ_a defined in the last subsection, the GEL test statistic for H_0 is given by $T_a(\hat{P}_n)$, where

$$T_a(Q) \equiv \inf_{\theta\in\Theta} \sup_{\gamma\in\Gamma_Q(\theta)} E_Q[\rho_a(\gamma' m(x,\theta)) - \rho_a(0)].$$

Here we consider the GEL test in the form of $\phi_{a,n} \equiv 1\{T_a(\hat{P}_n) > \chi^2_{q,1-\alpha}/(2n)\}$. Again, the GEL test includes several existing tests, such as the empirical likelihood, exponential tilting, and continuous updating GMM tests.

By applying the general result in Theorem 2.1, we can show the Hodges-Lehmann optimality of the GMM and GEL tests.

Theorem 3.2. *Assume that Condition 3.2 holds, and pick any $\epsilon > 0$ and $a \in \mathbb{R}$. Then*

[2]Under additional regularity conditions (such as uniqueness of θ_0 and a rank condition for $E_P[\partial m(x,\theta_0)/\partial\theta]$), we can see that $nT_{GMM}(\hat{P}_n) \Rightarrow \chi^2_{q-k}$ (see, Hansen, 1982). Since we do not impose such additional requirements in the space $\mathcal{Q}_\epsilon$, we employ the critical value $\chi^2_{q,1-\alpha}/n$ instead of $\chi^2_{q-k,1-\alpha}/n$ to guarantee that $\phi_{GMM,n}$ is pointwise asymptotically level α (see, Lemma B.4). The same comment applies to the critical value of the GEL test below.

(i) *the GMM test* $\phi_{GMM,n}$ *with a continuous mapping* $\tilde{\theta}(\cdot)$ *in the weak topology is pointwise asymptotically level* α *and* T_{GMM} *satisfies Conditions 2.1 and 2.2, i.e.,* $\phi_{GMM,n}$ *is Hodges-Lehmann optimal at each* $P_1 \in \mathcal{Q}_\epsilon \backslash \mathcal{P}_\epsilon$ *satisfying* $0 < K(\mathcal{P}_\epsilon, P_1) < \infty$.

(ii) *the GEL test* $\phi_{a,n}$ *is pointwise asymptotically level* α *and* T_a *satisfies Conditions 2.1 and 2.3, i.e.,* $\phi_{a,n}$ *is Hodges-Lehmann optimal at each* $P_1 \in \mathcal{Q}_\epsilon \setminus \mathcal{P}_\epsilon$ *satisfying* $0 < K(\mathcal{P}_\epsilon, P_1) < \infty$.

As the proof of this theorem shows, the mapping T_{GMM} to define the two-step GMM test (and also for the mapping to define the continuous updating GMM test) is lower semicontinuous in the weak topology. Thus, we can apply the first part of Theorem 2.1. On the other hand, as Example 3.1 shows, the mapping T_a to define the GEL test is not lower semicontinuous in general. Thus, we verify Condition 2.3 as an alternative route to derive the Hodges-Lehmann optimality.

Our analysis can be also applied to parameter hypothesis tests in estimating equations, i.e., $H_0 : P_0 \in \mathcal{P}_\epsilon \equiv \cup_{\theta \in \Theta_0}\{P \in \mathcal{Q}_{\epsilon,\theta} : E_P[m(x,\theta)] = 0\}$ versus $H_1 : P_0 \in \mathcal{Q}_\epsilon \setminus \mathcal{P}_\epsilon$ for a subset $\Theta_0 \subset \Theta$. It is also worth mentioning that the results in Theorem 2.1 can be applied to a variety of alternative testing problems, including setups where the parameter of interest is partially identified and the statistical model involves moment inequality conditions.

Appendix A: Proof of the main results

In the Appendices, let $\bar{\mathcal{A}}$ be the closure of a set $\mathcal{A} \subseteq \mathcal{M}$ with respect to the weak topology, and denote $\Omega_\eta \equiv \{Q \in \mathcal{Q} : T(Q) \leq \eta\}$ and $\kappa(\eta) \equiv K(\Omega_\eta, P_1)$. We also use $\mu(Q,\theta) \equiv E_Q[m(z,\theta)]$.

To analyze the large deviation behavior of the empirical measure $\hat{P}_n$, we use Sanov's Theorem (see, Theorem 6.2.10 of Dembo and Zeitouni, 1998) i.e.,

$$\limsup_{n\to\infty} \frac{1}{n} \log E_P[1\{\hat{P}_n \in \mathcal{A}\}] \leq -K(\mathcal{A}, P),$$

for any closed sets $\mathcal{A} \subseteq \mathcal{M}$ in the weak topology, and

$$\liminf_{n\to\infty} \frac{1}{n} \log E_P[1\{\hat{P}_n \in \mathcal{B}\}] \geq -K(\mathcal{B}, P),$$

for any open sets $\mathcal{B} \subseteq \mathcal{M}$ in the weak topology.

A.1. Proof of Lemma 2.1

Pick any $P_1 \in \mathcal{Q} \setminus \mathcal{P}$. If $K(\mathcal{P}, P_1) = \infty$ (i.e., $P \not\ll P_1$ for all $P \in \mathcal{P}$), the conclusion is trivially satisfied. So, we concentrate on the case of $K(\mathcal{P}, P_1) < \infty$. Pick any $\epsilon >$

0. There exists $P_0^* \in \mathcal{P}$ such that $K(P_0^*, P_1) < K(\mathcal{P}, P_1) + \epsilon < \infty$ and the Radon-Nykodym derivative $r(x) \equiv \frac{dP_0^*}{dP_1}$ exists. Since $\{x_i : i = 1, \dots, n\}$ is an i.i.d. sample and $E_{P_0^*}[\log r(x)] = K(P_0^*, P_1) < \infty$, Kolmogorov's strong law of large numbers (see, p. 125 of Chow and Teicher, 1997) implies

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \log r(x_i) = E_{P_0^*}[\log r(x)] < \infty, \quad P_0^* - a.s. \tag{2}$$

Let P^n be the n-fold product measure of P and define the event $E_n \equiv \{\prod_{i=1}^n r(x_i) < \exp(n[K(P_0^*, P_1) + \epsilon])\}$. Observe that

$$\begin{aligned} E_{P_1}[1 - \phi_n] &\geq \int_{E_n} 1\{\phi_n = 0\} dP_1^n \\ &\geq \exp(-n[K(P_0^*, P_1) + \epsilon]) \int_{E_n} 1\{\phi_n = 0\} \prod_{i=1}^{n} r(X_i) dP_1^n \\ &= \exp(-n[K(P_0^*, P_1) + \epsilon]) \int_{E_n} 1\{\phi_n = 0\} dP_0^{*n} \\ &\geq \exp(-n[K(P_0^*, P_1) + \epsilon])(\Pr\{\phi_n = 0 : P_0^{*n}\} - \Pr\{E_n^c : P_0^{*n}\}), \end{aligned}$$

where the first inequality follows from the set inclusion relation, the second inequality follows from the definition of E_n, the equality follows from the change of measures, and the last inequality follows from the set inclusion relation. Since $\liminf_{n\to\infty} \Pr\{\phi_n = 0 : P_0^{*n}\} = 1 - \limsup_{n\to\infty} \Pr\{\phi_n = 1 : P_0^{*n}\} \geq 1 - \alpha \in (0, 1)$ (because ϕ_n is pointwise asymptotically level α) and $\lim_{n\to\infty} \Pr\{E_n^c : P_0^{*n}\} = 0$ (by (2)), it follows that

$$\liminf_{n\to\infty} \frac{1}{n} \log E_{P_1}[1 - \phi_n] \geq -K(P_0^*, P_1) - \epsilon > -K(\mathcal{P}, P_1) - 2\epsilon,$$

where the second inequality follows from the definition of P_0^*. Since ϵ is arbitrary, the conclusion is obtained.

A.2. Proof of Theorem 2.1

Proof under Condition 2.2. Pick any $\epsilon > 0$. Note that the function $\kappa(\eta)$ is non-increasing (by definition) and right continuous in $\eta \in [0, \infty)$ (by Lemma B.1). Thus, there exists $\delta > 0$ such that

$$-\kappa(\delta) < -\kappa(0) + \epsilon. \tag{3}$$

For this δ, we have

$$\begin{aligned} \limsup_{n\to\infty} \frac{1}{n} \log E_{P_1}[1\{T(\hat{P}_n) \leq c_n\}] &\leq \limsup_{n\to\infty} \frac{1}{n} \log E_{P_1}[1\{\hat{P}_n \in \Omega_\delta\}] \\ &\leq -\kappa(\delta) < -\kappa(0) + \epsilon = -K(\mathcal{P}, P_1) + \epsilon, \end{aligned}$$

where the first inequality follows from $c_n \searrow 0$, the second inequality follows by Sanov's Theorem based on the fact that Ω_δ is closed in the weak topology (by Condition 2.2), the third inequality follows from (3), and the equality follows from $\Omega_0 = \mathcal{P}$ (by Condition 2.1). Since ϵ is arbitrary, the conclusion is obtained.

Proof under Condition 2.3. The first step involves proving that $\bar{\Omega}_\eta \subseteq \bar{\Omega}_{\eta'}$ for $\eta \leq \eta'$. Note that $\bar{\Omega}_\eta = \Omega_\eta \cup \partial^* \Omega_\eta$, where the set for boundary points is defined as

$$\partial^* \Omega_\eta \equiv \{Q \notin \Omega_\eta : \exists \text{ a sequence } \{Q_k : k \in \mathbb{N}\} \subseteq \Omega_\eta \text{ such that } Q_k \Rightarrow Q\}.$$

If $Q \in \Omega_\eta$ then $Q \in \Omega_{\eta'}$ by definition. Now suppose $Q \in \partial^* \Omega_\eta$. By definition there exists a sequence $\{Q_k : k \in \mathbb{N}\} \subseteq \Omega_\eta$ such that $Q_k \Rightarrow Q$. It then follows that $\{Q_k : k \in \mathbb{N}\} \subseteq \Omega_{\eta'}$, which implies $Q \in \bar{\Omega}_{\eta'}$. Thus, we obtain $\bar{\Omega}_\eta \subseteq \bar{\Omega}_{\eta'}$.

The second step is to prove that $\bar{\kappa}(\eta) \equiv K(\bar{\Omega}_\eta, P_1)$ is right continuous at $\eta = 0$. Pick any sequence of positive numbers $\{\eta_m : m \in \mathbb{N}\}$ with $\eta_m \searrow 0$. Note that by Condition 2.1, closedness of $\mathcal{P}$, and $0 < K(\mathcal{P}, P_1) < \infty$, we have $\bar{\kappa}(0) < \infty$. Since $\bar{\Omega}_\eta \subseteq \bar{\Omega}_{\eta'}$ for $\eta \leq \eta'$, the function $\bar{\kappa}(\cdot)$ is non-increasing. Thus, the limit $\lim_{m\to\infty} \bar{\kappa}(\eta_m)$ exists and it holds $\lim_{m\to\infty} \bar{\kappa}(\eta_m) \leq \bar{\kappa}(0) < \infty$. Since $\bar{\Omega}_\eta$ is closed in the weak topology by definition and $K(Q, P_1)$ is lower semicontinuous under the weak topology in Q (see, Lemma 1.4.3 of Dupuis and Ellis, 1997), there exists $Q_m \in \bar{\Omega}_{\eta_m}$ for all $m \in \mathbb{N}$ such that $K(Q_m, P_1) = \bar{\kappa}(\eta_m) < \infty$. Since the sequence $\{Q_m : m \in \mathbb{N}\}$ is on the compact set $\mathcal{Q}$, there exists a subsequence $\{Q_{m_j} : j \in \mathbb{N}\}$ such that $Q_{m_j} \Rightarrow Q^*$ for some $Q^* \in \mathcal{Q}$. Since $K(Q, P_1)$ is lower semicontinuous in Q,

$$K(Q^*, P_1) \leq \liminf_{j\to\infty} K(Q_{m_j}, P_1) < \infty.$$

There are two possibilities. First, if there exists a further subsequence $\{Q_{m_k} : k \in \mathbb{N}\}$ of $\{Q_{m_j} : j \in \mathbb{N}\}$ such that $Q_{m_k} \in \Omega_{\eta_{m_k}}$ for all $k \in \mathbb{N}$, then $T(Q_{m_k}) \leq \eta_{m_k}$ for each $k \in \mathbb{N}$ and Condition 2.3 implies $T(Q^*) = 0$ meaning that $Q^* \in \Omega_0$. As a result,

$$\bar{\kappa}(0) \geq \lim_{k\to\infty} \bar{\kappa}(\eta_{m_k}) = \liminf_{k\to\infty} K(Q_{m_k}, P_1) \geq K(Q^*, P_1) \geq \bar{\kappa}(0), \tag{4}$$

and it follows that $\lim_{k\to\infty} \bar{\kappa}(\eta_{m_k}) = \bar{\kappa}(0)$. Second, if such a subsequence does not exist, then it must be the case that $Q_{m_j} \in \partial^* \Omega_{\eta_{m_j}}$ for all j large enough. Since $Q_{m_j} \Rightarrow Q^*$ and $\eta_{m_j} \searrow 0$, it follows from Lemma B.2 that $T(Q^*) = 0$ and (4) follows. Therefore, $\bar{\kappa}(\eta)$ is right continuous at $\eta = 0$, i.e., for any $\epsilon > 0$ there exists $\delta > 0$ such that $\bar{\kappa}(0) - \bar{\kappa}(\delta) < \epsilon$.

The third step is to derive the conclusion by using Sanov's theorem and the results in the previous steps. Now, pick an arbitrary $\epsilon > 0$. Then

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n} \log E_{P_1}[1\{T(\hat{P}_n) \leq c_n\}] &\leq \limsup_{n\to\infty} \frac{1}{n} \log E_{P_1}[1\{\hat{P}_n \in \bar{\Omega}_\delta\}] \\ &\leq -\bar{\kappa}(\delta) < -\bar{\kappa}(0) + \epsilon = -K(\mathcal{P}, P_1) + \epsilon,\end{aligned}$$

for some $\delta > 0$, where the first inequality follows from $c_n \searrow 0$ and $\Omega_\delta \subset \bar{\Omega}_\delta$, the second inequality follows by Sanov's Theorem based on the fact that $\bar{\Omega}_\delta$ is closed in the weak topology, the third inequality follows from the right continuity of $\bar{\kappa}(\eta)$ at $\eta = 0$, and the equality follows from $\bar{\Omega}_0 = \mathcal{P}$ (by Condition 2.1 and closedness of $\mathcal{P}$). Since ϵ is arbitrary, we obtain the conclusion.

A.3. Proof of Theorem 3.1

Proof of (i). The proof is a special case of that of Theorem 3.2 (i) with replacements of $m(x,\theta)$ with $m(x)$.

Proof of (ii). Pick any $\epsilon > 0$ to define $\mathcal{P}_\epsilon$ and $\mathcal{Q}_\epsilon$. Also pick any $a \in \mathbb{R}$ to define ρ_a. First, by applying Lemma B.4 (with replacements of $m(x,\theta)$ with $m(x)$), $\phi_{a,n}$ is pointwise asymptotically level α.

Second, we present some properties of T_a. Let $\mathcal{P}_0 \equiv \{P \in \mathcal{M} : E_P[m(x)] = 0\}$ and $\mathcal{P}_0(Q) \equiv \{P \in \mathcal{P}_0 : P \ll Q, Q \ll P\}$. From theory of convex duality (see, e.g., Borwein and Lewis, 1993), we have

$$T_a(Q) \equiv \sup_{\gamma \in \Gamma_Q} E_Q[\rho_a(\gamma' m(x)) - \rho_a(0)] = \inf_{P \in \mathcal{P}_0(Q)} D_a(Q,P), \tag{5}$$

for each $Q \in \mathcal{M}$, where

$$D_a(Q,P) \equiv \begin{cases} \int \frac{1}{a(a+1)} \left(\left(\frac{dP}{dQ} \right)^{a+1} - 1 \right) dQ & \text{if } P \ll Q \\ \infty & \text{otherwise} \end{cases}.$$

Note that the mapping $D_a : \mathcal{M} \times \mathcal{M} \to [0, \infty]$ is a special case of the so-called f-divergence (see Liese and Vajda, 1987). It is known that

(D1) $D_a(Q,P) = 0$ if and only if $Q = P$;

(D2) $D_a(Q,P)$ is lower semicontinuous under the product topology for $(Q,P) \in \mathcal{M} \times \mathcal{M}$ induced by the weak topology for $\mathcal{M}$ and $\mathcal{M}$ (Liese and Vajda, 1987, Theorem 1.47).

Third, we check Condition 2.1 for T_a, i.e., $\mathcal{P}_\epsilon = \{R \in \mathcal{Q}_\epsilon : T_a(R) = 0\}$ in this case. Suppose $Q \in \mathcal{P}_\epsilon \subset \mathcal{P}_0$. Then the definition of $\mathcal{P}_0(Q)$ implies $Q \in \mathcal{P}_0(Q)$. Also, (5) and the set inclusion relation imply $0 \le T_a(Q) = \inf_{P \in \mathcal{P}_0(Q)} D(Q,P) \le D(Q,Q) = 0$. Therefore, from $Q \in \mathcal{P}_\epsilon \subset \mathcal{Q}_\epsilon$, we have $Q \in \{R \in \mathcal{Q}_\epsilon : T_a(R) = 0\}$. On the other hand, suppose $Q \in \{R \in \mathcal{Q}_\epsilon : T_a(R) = 0\}$. Then since $Q \in \mathcal{Q}_\epsilon$ and $Q \in \mathcal{P}_0(Q) \subset \mathcal{P}_0$ (by (D1)), we obtain $Q \in \mathcal{P}_\epsilon$. Combining these results, Condition 2.1 is verified.

Finally, we check Condition 2.3. Pick any sequence $\{Q_m : m \in \mathbb{N}\} \subseteq \mathcal{Q}_\epsilon$ such that $Q_m \Rightarrow Q \in \mathcal{Q}_\epsilon$ and $T_a(Q_m) \le \eta_m$ for all $m \in \mathbb{N}$. Since the set $\mathcal{P}_0$ is compact in the weak topology (by applying Lemma B.3 with replacements of $m(x,\theta)$ with $m(x)$) and $D_a(Q,P)$ is lower semicontinuous in the weak topology for $P \in \mathcal{M}$ (by (D2)), there exists a sequence $P_m^* \in \mathcal{P}_0$ such that $D_a(Q_m, P_m^*) = \inf_{P \in \mathcal{P}_0} D_a(Q_m, P) \le T_a(Q_m)$ for each $m \in \mathbb{N}$. Since $\{P_m^* : m \in \mathbb{N}\}$ is a sequence on the compact set $\mathcal{P}_0$, there exists a subsequence $\{P_{m_j}^* : j \in \mathbb{N}\}$ such that $P_{m_j}^* \Rightarrow P^* \in \mathcal{P}_0$. Now, from (D2), it follows that

$$0 = \liminf_{j \to \infty} \eta_{m_j} \ge \liminf_{j \to \infty} T_a(Q_{m_j}) \ge \liminf_{j \to \infty} D_a(Q_{m_j}, P_{m_j}^*) \ge D_a(Q, P^*).$$

which means $Q = P^*$ (by (D1)). Therefore, it holds $P^* \in \mathcal{P}_0(Q)$ and $T_a(Q) = \inf_{P \in \mathcal{P}_0(Q)} D_a(Q,P) \le D_a(Q,P^*) = 0$, which completes the proof.

A.4. Proof of Theorem 3.2

Proof of (i). Pick any $\epsilon > 0$ to define $\mathcal{P}_\epsilon$ and $\mathcal{Q}_\epsilon$. From Lemma B.4, $\phi_{GMM,n}$ is pointwise asymptotically level α. Also, Condition 2.1 is trivially satisfied. So we concentrate on showing that T_{GMM} is lower semicontinuous in $\mathcal{M}$ under the weak topology.

First, re-write the mapping as $T_{GMM}(Q) \equiv \inf_{\theta\in\Theta} T_{GMM}(Q,\theta)$, where

$$T_{GMM}(Q,\theta) \equiv \mu(Q,\theta)'\Sigma(Q,\tilde{\theta}(Q))^{-1}\mu(Q,\theta),$$

and $\mu(Q,\theta) \equiv E_Q[m(x,\theta)]$. When $\Sigma(Q,\tilde{\theta}(Q))$ is singular, we define $T_{GMM}(Q,\theta)$ to be infinity if $||\mu(Q,\theta)|| \neq 0$ and to be zero if $||\mu(Q,\theta)|| = 0$. By Condition 3.2 and the Portmanteau Lemma (see van der Vaart, 1998, Lemma 2.2), it follows that both $\mu(Q,\theta)$ and $\Sigma(Q,\theta)$ are uniformly continuous in $(Q,\theta) \in \mathcal{M}\times\Theta$, as both $\mathcal{M}$ and Θ are compact. Thus, since $\tilde{\theta}(Q)$ is continuous in Q, $\Sigma(Q,\tilde{\theta}(Q))$ is continuous in $Q \in \mathcal{M}$.

Pick any sequence $\{(Q_m,\theta_m) : m \in \mathbb{N}\}$ such that $Q_m \Rightarrow Q^* \in \mathcal{M}$ and $\theta_m \to \theta^* \in \Theta$. We split into three cases. First, suppose $\det(\Sigma(Q^*,\tilde{\theta}(Q^*))) > 0$. Then since $\det(\Sigma(Q_m,\tilde{\theta}(Q_m))) > 0$ for all m large enough, we obtain $T_{GMM}(Q^*,\theta^*) = \lim_{m\to\infty} T_{GMM}(Q_m,\theta_m)$. Second, suppose $\det(\Sigma(Q^*,\tilde{\theta}(Q^*))) = 0$ and $||\mu(Q^*,\theta^*)|| = 0$. Then $T_{GMM}(Q^*,\theta^*) = 0 \leq \liminf_{m\to\infty} T_{GMM}(Q_m,\theta_m)$, since $T_{GMM}(Q_m,\theta_m) \geq 0$ for all $m \in \mathbb{N}$ by definition. Third, suppose $\det(\Sigma(Q^*,\tilde{\theta}(Q^*))) = 0$ and $||\mu(Q^*,\theta^*)|| \neq 0$, so that $T_{GMM}(Q^*,\theta^*) = \infty$. In this case, if $\det(\Sigma(Q_m,\tilde{\theta}(Q_m))) = 0$ for all m large enough, then $T_{GMM}(Q_m,\theta_m) = \infty$ for all m large enough, as $||\mu(Q_m,\theta_m)|| \neq 0$ for all m large enough. If $\det(\Sigma(Q_m,\tilde{\theta}(Q_m)) > 0$ for all $m \in \mathbb{N}$ large enough, then $T_{GMM}(Q_m,\theta_m)$ is a continuous transformation of $\mu(Q_m,\theta_m)$ and $\Sigma(Q_m,\tilde{\theta}(Q_m))$ and therefore continuous in (Q,θ), which implies $T_{GMM}(Q^*,\theta^*) = \lim_{m\to\infty} T_{GMM}(Q_m,\theta_m)$. Combining these results,

$$T_{GMM}(Q^*,\theta^*) \leq \liminf_{m\to\infty} T_{GMM}(Q_m,\theta_m), \tag{6}$$

for any sequence $\{(Q_m,\theta_m) : m \in \mathbb{N}\}$ such that $Q_m \Rightarrow Q^* \in \mathcal{M}$ and $\theta_m \to \theta^* \in \Theta$.

Now pick any sequence $\{R_m : m \in \mathbb{N}\}$ such that $R_m \Rightarrow R^*$. By compactness of Θ and continuity of $T_{GMM}(Q,\theta)$ in $\theta \in \Theta$ for each $Q \in \mathcal{M}$, there exists a sequence $\{\theta^*_m : m \in \mathbb{N}\}$ such that $T_{GMM}(R_m) = T_{GMM}(R_m,\theta^*_m)$ for each $m \in \mathbb{N}$. Since $\{\theta^*_m : m \in \mathbb{N}\}$ is a sequence on the compact set Θ, there exists a subsequence $\{\theta^*_{m_j} : j \in \mathbb{N}\}$ such that $\theta^*_{m_j} \to \theta^*$ for some $\theta^* \in \Theta$. Then

$$\begin{aligned} T_{GMM}(R^*) &\leq T_{GMM}(R^*,\theta^*) \\ &\leq \liminf_{j\to\infty} T_{GMM}(R_{m_j},\theta^*_{m_j}) = \liminf_{j\to\infty} T_{GMM}(R_{m_j}), \end{aligned}$$

where the first inequality follows by the definition of $T_{GMM}(R^*)$ and the second one by (6). We can conclude that $T_{GMM}(\cdot)$ is lower semicontinuous on $\mathcal{M}$ in the weak topology.

Proof of (ii). The proof is similar to that of Theorem 3.1 (ii) by noting that

$$T_a(Q) \equiv \inf_{\theta\in\Theta} \sup_{\gamma\in\Gamma_Q(\theta)} E_Q[\rho_a(\gamma' m(x,\theta)) - \rho_a(0)] = \inf_{P\in\mathcal{P}_0(Q)} D_\alpha(Q,P),$$

where $\mathcal{P}_0(Q) \equiv \{P \in \mathcal{P}_0 : P \ll Q, Q \ll P\}$ and $\mathcal{P}_0 \equiv \cup_{\theta\in\Theta}\{P \in \mathcal{M} : E_P[m(x,\theta)] = 0\}$.

Appendix B: Additional Lemmas

Lemma B.1. *Under Condition 2.2, the function $\kappa(\eta)$ is right continuous in $\eta \in [0, \infty)$.*

Proof. First, note that $\Omega_{\eta_1} \subseteq \Omega_{\eta_2}$ if $\eta_2 > \eta_1$ meaning that $K(\Omega_{\eta_2}, P_1) \leq K(\Omega_{\eta_1}, P_1)$. Thus, $\kappa(\cdot)$ is a non-increasing function. Second, let $\{\eta_m : m \in \mathbb{N}\}$ be a sequence in $\mathbb{R}$ monotonically decreasing to some $\eta \in [0, \infty)$ such that $\kappa(\eta) < \infty$. Since $\kappa(\cdot)$ is non-increasing, $\{\kappa(\eta_m) : m \in \mathbb{N}\}$ is a non-decreasing sequence bounded by $\kappa(\eta)$ from above, and $\lim_{m\to\infty} \kappa(\eta_m)$ exists. By the lower semicontinuity of T and $\kappa(\eta) < \infty$, the set Ω_η is closed in the weak topology meaning that there exists $Q \in \Omega_\eta$ such that $K(Q, P_1) = K(\Omega_\eta, P_1)$. Therefore, for each $m \in \mathbb{N}$ there exists $Q_m \in \Omega_{\eta_m}$ such that $K(Q_m, P_1) = \kappa(\eta_m) \leq \kappa(\eta)$. Since $K(\cdot, P_1)$ has compact level sets for each $P_1 \in \mathcal{M}$ (see Dupuis and Ellis, 1997, Lemma 1.4.3), $\{Q_m : m \in \mathbb{N}\}$ has a subsequence $\{Q_{m_j} : j \in \mathbb{N}\}$ such that $Q_{m_j} \Rightarrow Q \in \mathcal{M}$ and $K(Q, P_1) \leq \liminf_{j\to\infty} K(Q_{m_j}, P_1) < \infty$ (by the lower semicontinuity of $K(\cdot, P_1)$ in the weak topology). Since $T(Q_{m_j}) \leq \eta_{m_j}$ for each $j \in \mathbb{N}$ and T is lower semicontinuous (by Condition 2.2), it follows that $T(Q) \leq \liminf_{j\to\infty} T(Q_{m_j}) \leq \liminf_{j\to\infty} \eta_{m_j} = \eta$. Therefore, $Q \in \Omega_\eta$ (meaning that $K(Q, P_1) \geq \kappa(\eta)$) and $\kappa(\eta) \geq \lim_{j\to\infty} \kappa(\eta_{m_j}) \geq \liminf_{j\to\infty} K(Q_{m_j}, P_1) \geq K(Q, P_1) \geq \kappa(\eta)$, which means $\lim_{j\to\infty} \kappa(\eta_{m_j}) = \kappa(\eta)$. Note that the result also holds for $\eta \in [0, \infty)$ such that $\kappa(\eta) = \infty$. To see this suppose by contradiction that $\kappa(\eta) = \infty$ but $\lim_{m\to\infty} \kappa(\eta_m)$ exists for a sequence $\{\eta_m : m \in \mathbb{N}\}$ with $\eta_m \searrow \eta$. By applying the previous argument, there exists $Q \in \Omega_\eta$ such that $K(Q, P_1) < \infty$, which violates $\kappa(\eta) = \infty$. □

Lemma B.2. *Let*

$$\partial^* \Omega_\eta \equiv \{Q \notin \Omega_\eta : \exists \text{ a sequence } \{Q_k : k \in \mathbb{N}\} \subseteq \Omega_\eta \text{ such that } Q_k \Rightarrow Q\}$$

be the set of boundary points of Ω_η in the weak topology. Under Condition 2.3, if $Q_m \in \partial^* \Omega_{\eta_m}$ for all $m \in \mathbb{N}$ with a sequence $\eta_m \searrow 0$ and $Q_m \Rightarrow Q^* \in \mathcal{M}$, then it holds $T(Q^*) = 0$.

Proof. Pick any sequence $\{Q_m : m \in \mathbb{N}\}$ such that $Q_m \in \partial^* \Omega_{\eta_m}$ for all $m \in \mathbb{N}$ with some sequence $\eta_m \searrow 0$ and $Q_m \Rightarrow Q^*$ for some $Q^* \in \mathcal{M}$. For this Q^*, suppose that

$$\exists \{Q'_m : m \in \mathbb{N}\} \text{ such that } Q'_m \in \Omega_{\eta_m} \text{ for all } m \in \mathbb{N} \text{ and } Q'_m \Rightarrow Q^*. \tag{7}$$

Then Condition 2.3 implies $T(Q^*) = 0$. So it is sufficient to show (7). Define the Lévy metric

$$d_L(P, Q) \equiv \inf\{\epsilon > 0 : F_P(x - \epsilon \mathbf{1}) - \epsilon \leq F_Q(x) \leq F_P(x - \epsilon \mathbf{1}) + \epsilon \text{ for all } x \in \mathcal{X}\},$$

for measures $P, Q \in \mathcal{M}$, where F_P and F_Q are distribution functions associated with P and Q, respectively, and $\mathbf{1} \equiv \{1, \dots, 1\}$ is the vector of ones with the same dimension as x. Note that the Lévy metric is compatible with the weak topology. Pick any $\epsilon > 0$. From $Q_m \in \partial^* \Omega_{\eta_m}$, there exists $\{Q'_m : m \in \mathbb{N}\}$ such that $Q'_m \in \Omega_{\eta_m}$ and $d_L(Q'_m, Q_m) \leq \epsilon/2$

for all $m \in \mathbb{N}$. Also from $Q_m \Rightarrow Q^*$, there exists $M \in \mathbb{N}$ such that $d_L(Q_m, Q^*) \leq \epsilon/2$ for all $m \geq M$. Combining these results, we have $d_L(Q'_m, Q^*) \leq \epsilon$ for all $m \geq M$. Since ϵ is arbitrary, we obtain $Q'_m \Rightarrow Q^*$ and (7) holds true, which completes the proof. □

Lemma B.3. *Let $\mathcal{Q}_{\epsilon,\theta} \equiv \{P \in \mathcal{M} : \det(\Sigma(P, \theta)) \geq \epsilon\}$, $\mathcal{Q}_\epsilon \equiv \cup_{\theta\in\Theta}\mathcal{Q}_{\epsilon,\theta}$, and $\mathcal{P}_\epsilon \equiv \cup_{\theta\in\Theta}\{P \in \mathcal{Q}_{\epsilon,\theta} : E_P[m(x, \theta)] = 0\}$. Under Condition 3.2, $\mathcal{Q}_\epsilon$ and $\mathcal{P}_\epsilon$ are compact in the weak topology for every $\epsilon > 0$.*

Proof. Pick any $\epsilon > 0$. From Theorem D.8 of Dembo and Zeitouni (1998), the set $\mathcal{M}$ is compact in the weak topology if the support $\mathcal{X}$ is compact (assumed in Condition 3.2). Thus, it is sufficient to show that $\mathcal{Q}_\epsilon$ and $\mathcal{P}_\epsilon$ are closed in the weak topology. We first show that $\mathcal{Q}_\epsilon$ is closed. To this end, take a sequence $\{Q_m \in \mathcal{Q}_\epsilon : m \in \mathbb{N}\}$ such that $Q_m \Rightarrow Q^* \in \mathcal{M}$. Note that for every $m \in \mathbb{N}$ there exits θ_m such that $\det(\Sigma(Q_m, \theta_m)) \geq \epsilon$. Also, by compactness of Θ there exists a subsequence θ_{m_k} of θ_m such that $\theta_{m_k} \to \theta^* \in \Theta$. Let $g(x, \theta, Q) = (m(x, \theta) - \mu(Q, \theta))(m(x, \theta) - \mu(Q, \theta))'$. By Condition 3.2, $g(x, \theta, Q)$ is uniformly continuous on $\mathcal{X} \times \Theta \times \mathcal{M}$. Then

$$\begin{aligned}
||\Sigma(Q_m, \theta_m) - \Sigma(Q^*, \theta^*)|| &\leq ||\int (g(x, \theta_m, Q_m) - g(x, \theta^*, Q^*))dQ_m|| \\
&\quad + ||\int g(x, \theta^*, Q^*)(dQ^* - dQ_m)|| \\
&\leq \sup_{x\in\mathcal{X}} ||g(x, \theta_m, Q_m) - g(x, \theta^*, Q^*)|| \\
&\quad + ||\int g(x, \theta^*, Q^*)(dQ^* - dQ_m)|| \\
&= o(1) \qquad (8)
\end{aligned}$$

as $m \to \infty$, where the convergence follows from the Portmanteau Lemma (van der Vaart, 1998, Lemma 2.2) and the uniform continuity of $g(x, \theta, Q)$. Since the determinant is a continuous function, it follows that $\mathcal{Q}_\epsilon$ is closed.

We next show the closedness of $\mathcal{P}_\epsilon$. Take a sequence $\{P_m \in \mathcal{P}_\epsilon : m \in \mathbb{N}\}$ such that $P_m \Rightarrow P^* \in \mathcal{M}$. Then there exists a sequence $\{\theta_m : m \in \mathbb{N}\}$ such that $\int_{\mathcal{X}} m(x, \theta_m)dP_m = 0$. Since Θ is compact, there exists a subsequence $\{\theta_{m_j} : j \in \mathbb{N}\}$ such that $\theta_{m_j} \to \theta^{**}$ for some $\theta^{**} \in \Theta$. Therefore, it is sufficient to show that $E_{P^*}[m(x, \theta^{**})] = 0$. To prove this, note that

$$\begin{aligned}
\left\|\int_{\mathcal{X}} m(x, \theta)dP\right\| &\leq \lim_{j\to\infty} \left\|\int_{\mathcal{X}} m(x, \theta)(dP - dP_{m_j})\right\| \\
&\quad + \lim_{j\to\infty} \left\|\int_{\mathcal{X}} [m(x, \theta) - m(x, \theta_{m_j})]dP_{m_j}\right\| \\
&\leq \lim_{j\to\infty} \sup_{x\in\mathcal{X}} \left\|m(x, \theta) - m(x, \theta_{m_j})\right\| = 0,
\end{aligned}$$

where the first inequality follows from the definition of θ_m, the second inequality follows by the Portmanteau Lemma (van der Vaart, 1998, Lemma 2.2) as $m(\cdot, \theta)$ is bounded and continuous for all $\theta \in \Theta$, and the equality follows by the uniform continuity of $m(x, \theta)$ on $\mathcal{X} \times \Theta$. $\square$

Lemma B.4. *Pick any $\epsilon > 0$ and $a \in \mathbb{R}$. Under Condition 3.2, the two-step GMM test $\phi_{GMM,n}$ and the GEL test $\phi_{a,n}$ defined in Section 3.2 are pointwise asymptotically level α.*

Proof. First, consider the continuous updating GMM test statistic (i.e., the case of $a = 1$). In this case, the supremum for γ has an explicit solution and the test statistic is written as $T_{CU}(\hat{P}_n) \equiv \inf_{\theta \in \Theta} \ell_{CU}(\theta)$, where $\ell_{CU}(\theta) \equiv (1/2)\bar{m}_n(\theta)'\Sigma(\hat{P}_n, \theta)^{-1}\bar{m}_n(\theta)$ and $\bar{m}_n(\theta) \equiv n^{-1}\sum_{i=1}^n m(x_i, \theta)$. Take any $P \in \mathcal{P}_\epsilon$. There exists $\theta^* \in \Theta$ such that $E_P[m(x, \theta^*)] = 0$ and $\Sigma(P, \theta^*)$ is positive definite. Let $\phi_{CU,n} \equiv 1\{T_{CU}(\hat{P}_n) > \chi^2_{q,1-\alpha}/(2n)\}$. By the central limit theorem, $2n\ell_{CU}(\theta^*) \Rightarrow \chi^2_q$ and therefore,

$$\begin{aligned}\limsup_{n\to\infty} E_P[\phi_{CU,n}] &= \limsup_{n\to\infty} P\left(\inf_{\theta\in\Theta} 2n\ell_{CU}(\theta) > \chi^2_{q,1-\alpha}\right) \\ &\leq \limsup_{n\to\infty} P\left(2n\ell_{CU}(\theta^*) > \chi^2_{q,1-\alpha}\right) = \alpha.\end{aligned}$$

Similarly, we can define the objective functions $\ell_{GMM}(\theta)$ and $\ell_a(\theta)$ for the two-step GMM and GEL tests, respectively. Since $\ell_{GMM}(\theta^*)$ and $\ell_a(\theta^*)$ are asymptotically equivalent to $\ell_{CU}(\theta^*)$ under $P \in \mathcal{P}_\epsilon$ (see, Newey and Smith, 2004), we obtain the conclusion. $\square$

References

Bahadur, R. R. (1960). Stochastic Comparison of Test. *Annals of Mathematical Statistics* **31** 276-295.

Borwein, J. M. and Lewis, A. S. (1993). Partially-finite Programing in L_1 and the Existence of Maximum Entropy Estimates. *SIAM Journal of Optimization* **3** 248-267.

Canay, I. A. (2010). EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity. *Journal of Econometrics* **156** 408-425.

Chow, Y. S. and Teicher, H. (1997). *Probability Theory*. Springer.

Cressie, N. and Read, R. C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society* **46** 440-464.

Dembo, A. and Zeitouni, O. (1998). *Large Deviations Techniques and Applications*. Second Edition, Springer: New York.

Dupuis, P. and Ellis, R. S. (1997). *A Weak Convergence Approach to the Theory of Large Deviations*. Wiley Series in Probability and Statistics: New York.

Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica* **50** 1029-1054.

Hodges, J. L. and Lehmann, E. L. (1956). The Efficiency of Some Nonparametric Competitors of the t-test. *Annals of Mathematical Statistics* **27** 324-335.

Imbens, G., Spady, R. H. and Johnson, P. (1998). Information Theoretic Approaches to Inference in Moment Condition Models. *Econometrica* **66** 333-357.

KALLENBERG, W. and KOUROUKLIS, S. (1992). Hodges-Lehmann Optimality of Tests. *Statistics & Probability Letters* **14**.

KITAMURA, Y. (2001). Asymptotic Optimality of Empirical Likelihood for Testing Moment Restrictions. *Econometrica* **69** 1661-1672.

KITAMURA, Y., SANTOS, A. and SHAIKH, A. (2009). On the Asymptotic Optimality of Empirical Likelihood for Testing Moment Restrictions. manuscript, University of Chicago.

LIESE, F. and VAJDA, I. (1987). *Convex Statistical Distances.* Teubner Leipzig.

NEWEY, W. K. and SMITH, R. J. (2004). Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators. *Econometrica* **72** 219-255.

OTSU, T. (2010). On Bahadur Efficiency of Empirical Likelihood. *Journal of Econometrics* **157** 248-256.

OWEN, A. (1988). Empirical Likelihood Ratio Confidence Intervals for a Single Functional. *Biometrika* **75** 237-249.

QIN, B. J. and LAWLESS, J. (1994). Empirical Likelihood and General Estimating Equations. *Annals of Statistics* **22** 300-325.

SCHENNACH, S. M. (2007). Point Estimation with Exponentially Tilted Empirical Likelihood. *Annals of Statistics* **32** 634-672.

SMITH, R. J. (1997). Alternative Semi-parametric Likelihood Approaches to Generalised Method of Moments Estimation. *The Economic Journal* **107** 503–519.

TUSNÁDY, G. (1977). On Asymptotic Optimality of Tests. *Annals of Statistics* **5** 385-393.

VAN DER VAART, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press, Cambridge.