Journal of Econometrics 171 (2012) 45–53

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Hodges–Lehmann optimality for testing moment conditions[☆]

Ivan A. Canay[a,*], Taisuke Otsu[b]

[a] Department of Economics, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208, USA
[b] Department of Economics, Yale University, 30 Hillhouse Avenue, New Haven, CT 06511, USA

ARTICLE INFO

Article history:
Received 24 January 2011
Received in revised form
15 November 2011
Accepted 5 June 2012
Available online 20 July 2012

Keywords:
Asymptotic optimality
Large deviations
Moment condition
Generalized method of moments
Generalized empirical likelihood

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.

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 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 for optimality 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 properties of tests in statistics and econometrics is based on sequences of local (or Pitman) alternatives. There are still 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 (see Nelson and Savin, 1990). 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.

☆ We would like to thank two anonymous referees, 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. Canay and Otsu thank the National Science Foundation for research support via grants SES-1123586 and SES-0720961, respectively.

* Corresponding author. Tel.: +1 847 491 2929; fax: +1 847 491 7001.
E-mail addresses: iacanay@northwestern.edu (I.A. Canay), taisuke.otsu@yale.edu (T. Otsu).
URLs: http://faculty.wcas.northwestern.edu/~iac879/ (I.A. Canay), http://cowles.econ.yale.edu/faculty/otsu.htm (T. Otsu).

0304-4076/$ – see front matter
doi:10.1016/j.jeconom.2012.06.014

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 for 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 et al., 1998; Newey and Smith, 2004; Schennach, 2007). In terms of global analysis based on large deviation theory, Kitamura (2001) and Kitamura et al. (2012) 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).

There are two key features of our Hodges–Lehmann analysis relative to the other global analyses cited above. First, the type I error probability in the Hodges–Lehmann analysis converges to a positive constant, as opposed to converging to zero. This intends to resemble the situation where a test statistic is compared to a fixed asymptotic critical value. On the other hand, the above cited papers consider the situation where a test statistic is compared to a critical value drifting to zero. Thus, our Hodges–Lehmann analysis complements the existing global optimality analyses for tests of overidentifying restrictions by introducing a different asymptotic framework. Second, the papers cited above 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. We provide Hodges–Lehmann optimality results for several commonly used tests of overidentifying restrictions.

The remainder of the paper is organized as follows. Section 2 introduces basic notation and concepts, and presents 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

Let $\mathcal{X}$ be a Polish space. Consider a random sample $\{x_i : i = 1, \ldots, n\}$ 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

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

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. The 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.,

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

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

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

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 or equal to 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. For probability measures P and Q, let $Q \ll P$ denote that Q is absolutely continuous with respect 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) from Q to 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*

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

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. If $K(\mathcal{P}, P_1) = \infty$, the above inequality trivially holds true. If $0 < K(\mathcal{P}, P_1) < \infty$, this lemma provides the best possible exponential decay rate of the type II error probability. If $K(\mathcal{P}, P_1) = 0$, this lemma says that there is no test which attains an exponential decay rate of the type II error probability.

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 decay 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

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

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 Condition 2.3 is required for the optimality.

Condition 2.1. (a) *$\mathcal{Q}$ is closed in the weak topology and* (b) $\mathcal{P} = \{Q \in \mathcal{Q} : T(Q) \leq 0\}$.

Condition 2.2. *T is lower semicontinuous in the weak topology at all $Q \in \{Q \in \mathcal{Q} : K(Q, P_1) < \infty\}$: this is, for all $Q \in \{Q \in \mathcal{Q} : K(Q, P_1) < \infty\}$ and all sequence $\{Q_m : m \in \mathbb{N}\}$ in $\mathcal{Q}$ such that $Q_m \Rightarrow Q$, it holds $T(Q) \leq \liminf_{m\to\infty} T(Q_m)$.*

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 sequence of positive real numbers $\{\eta_m : m \in \mathbb{N}\}$ 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(a) imposes a weak regularity on the relevant subset of measures $\mathcal{Q}$. Condition 2.1(b), imposed by Kallenberg and Kourouklis (1992) as well, says that the set of measures $\mathcal{P}$ 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 Condition 2.3, ϕ_n is Hodges–Lehmann optimal at each $P_1 \in int(\mathcal{Q}) \setminus \mathcal{P}$ 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 tests.

Note that this theorem establishes optimality for alternatives such that $0 < K(\mathcal{P}, P_1) < \infty$ and that are not at the boundary of $\mathcal{Q}$. For example, if $K(\mathcal{P}, P_1) = 0$, Lemma 2.1 implies that there is no test which attains an exponential decay rate of the type II error probability. Thus the Hodges–Lehmann analysis in such a case, which perhaps compares polynomial decay rates of the type II error probabilities, will be significantly different from ours and is beyond the scope of this paper. Also, for the case of $K(\mathcal{P}, P_1) = \infty$, although the conclusion of Lemma 2.1 trivially holds, the lemma does not provide an optimal decay rate and, to the best of our knowledge, it is not clear how to conduct the Hodges–Lehmann analysis in such a situation.

Also 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 change can be done typically by imposing more restrictions on $\mathcal{Q}$ relative to the pointwise requirement.[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

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

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 $\mathcal{X}$.*

[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.

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.4), 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 Condition 2.1(a) holds true and*

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

Theorem 3.1 shows that several existing tests to test moment conditions are Hodges–Lehmann optimal. This suggests, similarly to previous findings on parametric and nonparametric tests (Kallenberg and Kourouklis, 1992; Tusnády, 1977), that Hodges–Lehmann optimality is a weak asymptotic requirement. We are not aware of any example of a reasonable test which is not Hodges–Lehmann optimal in this setting.[2]

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 ∞). Consider a measure Q^* such that $Q^*(X=0) = 1-p$ and $Q^*(X=x^*) = p$ for some $x^* \in (0, x_H)$. We can always choose $p \in (0,1)$ so that $Q^* \in \mathcal{Q}_\epsilon$. Now consider the following sequence of probability measures,

$$Q_m(X = -x_L) = \frac{1}{m},$$

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

Clearly $Q_m \Rightarrow Q^*$. Note that $\Gamma_{Q_m} = (-1/x^*, 1/x_L)$ for all $m \in \mathbb{N}$, while $x_{L_Q} = 0$ and then $\Gamma_Q = (-1/x^*, \infty)$. This is, x_{L_m} does not converge to $x_{L_Q} = 0$ since $x_{L_m} > 0$ for all $m \in \mathbb{N}$, and so $\liminf_{m\to\infty}(x_{L_n} - x_{L_Q}) > 0$.

Now note that 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 $T_{EL}(Q_m) \nearrow \log(1 + x^*/x_L)p < \infty$. However,

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

Note that $T_{EL}(Q) = \infty$ regardless of how small x^* or p might be, as long as both are positive. Therefore, for a measure $Q^* \in \mathcal{Q}_\epsilon$ we constructed a sequence $\{Q_m : m \in \mathbb{N}\}$ such that $Q_m \Rightarrow Q^*$ and $T_{EL}(Q^*) > \liminf_{m\to\infty} T_{EL}(Q_m)$, which violates Condition 2.2. 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 $\mathcal{Q}_{\epsilon,\theta} \equiv \{P \in \mathcal{M} : \det(\Sigma(P,\theta)) \ge \epsilon\}$. 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 $\theta_0 \in \Theta$.

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

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

[2] If we restrict our attention to the class of distributions having symmetric densities (i.e., $\mathcal{Q}_\epsilon \equiv \{P \in \mathcal{M} : \det(\Sigma(P)) \ge \epsilon, P$ has a symmetric pdf$\}$), then the analysis of Hodges and Lehmann (1956) can be applied to the case of $m(x) = x$ (i.e., testing for location) and, for example, the sign test is typically not Hodges–Lehmann optimal.

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 $\phi_{GMM,n} \equiv 1\{T_{GMM}(\hat{P}_n) > \chi^2_{q,1-\alpha}/n\}$.[3]

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 $\phi_{a,n} \equiv 1\{T_a(\hat{P}_n) > \chi^2_{q,1-\alpha}/(2n)\}$. 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 Condition 2.1(a) holds true and*

(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(b) and 2.2, i.e., $\phi_{GMM,n}$ is Hodges–Lehmann optimal at each $P_1 \in int(\mathcal{Q}_\epsilon)\backslash\mathcal{P}_\epsilon$ 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(b) and 2.3, i.e., $\phi_{a,n}$ is Hodges–Lehmann optimal at each $P_1 \in int(\mathcal{Q}_\epsilon) \setminus \mathcal{P}_\epsilon$ satisfying $0 < K(\mathcal{P}_\epsilon, P_1) < \infty$.*

Theorem 3.2 shows again that all tests under consideration are Hodges–Lehmann optimal, suggesting that Hodges–Lehmann optimality is a weak asymptotic requirement for the problem of testing overidentifying restrictions.

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 what follows, let $\bar{\mathcal{A}}$ be the closure of a set $\mathcal{A} \subseteq \mathcal{M}$ with respect to the weak topology, and denote $\Omega_\eta \equiv \{Q \in \mathcal{Q} : T(Q) \le \eta\}$ and $\kappa(\eta) \equiv K(\Omega_\eta, P_1)$ for $\eta \in [0,\infty)$. We also use $\mu(Q,\theta) \equiv E_Q[m(z,\theta)]$.

Define the Lévy–Prohorov metric for measures $P, Q \in \mathcal{M}$ as

$$d_L(P,Q) \equiv \inf\{\epsilon > 0 : P(A) \le Q(A^\epsilon) + \epsilon, Q(A) \le P(A^\epsilon) + \epsilon \text{ for all Borel sets } A\},$$

where $A^\epsilon \equiv \{x \in \mathcal{X} : \inf_{y\in A} d(x,y) \le \epsilon\}$ for a metric d on $\mathcal{X}$. The Lévy–Prohorov metric is compatible with the weak topology (Billingsley, 1999, Theorem 6.8). Let $B_L(P,r) \equiv \{Q \in \mathcal{M} : d_L(Q,P) \le r\}$ be the ball with respect to the Lévy–Prohorov metric centered at P with radius $r > 0$.

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.,

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

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

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

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 $K(\mathcal{P}, P_1) = \infty$, 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. Now let $t^- \equiv -\min\{t, 0\}$. Since P_0^* is absolutely continuous with respect to P_1 and $s(\log s)^-$ is bounded for all $s \in [0,\infty)$, we have

$$\int_{\mathcal{X}} (\log r(x))^- \, dP_0^* = \int_{\mathcal{X}} r(x) \, (\log r(x))^- \, dP_1 < \infty.$$

Combining this result with $E_{P_0^*}[\log r(x)] = K(P_0^*, P_1) < \infty$ implies $E_{P_0^*}[|\log r(x)|] < \infty$. As $\{x_i : i = 1,\ldots,n\}$ is an i.i.d. sample from P_0^*, the strong law of large numbers (see, Theorem 22.1 of Billingsley, 1995) implies

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

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

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

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 $\liminf_{n\to\infty} \Pr\{\phi_n = 0 : P_0^{*n}\} = 1 - \limsup_{n\to\infty} \Pr\{\phi_n = 1 : P_0^{*n}\} \ge 1 - \alpha \in (0, 1)$ (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

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

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 $P_1 \in int(\mathcal{Q}) \setminus \mathcal{P}$. Since $P_1 \in int(\mathcal{Q})$, there exists $r > 0$ such that $B_L(P_1, r) \subseteq \mathcal{Q}$. The weak convergence $\hat{P}_n \Rightarrow P_1$ implies that $\hat{P}_n \in B_L(P_1, r) \subseteq \mathcal{Q}$ for all n

[3] 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.5). The same comment applies to the critical value of the GEL test.

large enough. Thus, for all n large enough, it holds

$$E_{P_1}[1\{T(\hat{P}_n) \le c_n\}] = E_{P_1}[1\{\hat{P}_n \in \{Q \in \mathcal{Q} : T(Q) \le c_n\}\}]. \tag{3}$$

Now 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.2). Thus, there exists $\delta > 0$ such that

$$-\kappa(\delta) < -\kappa(0) + \epsilon = -K(\mathcal{P}, P_1) + \epsilon, \tag{4}$$

where the equality follows from Condition 2.1(b). For this δ, it holds

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

where the first inequality follows from (3), $c_n \searrow 0$, and $\Omega_\delta \subseteq \bar{\Omega}_\delta$, the second inequality follows by Sanov's Theorem based on the fact that $\bar{\Omega}_\delta$ is closed in the weak topology, the equality follows from Lemma B.1, and the last inequality follows from (4). 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 $\eta \le \eta'$. Note that $\bar{\Omega}_\eta = \Omega_\eta \cup \partial^*\Omega_\eta$, where the set for boundary points is defined as

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

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, the 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 $\eta \le \eta'$, the function $\bar{\kappa}(\cdot)$ is non-increasing. Thus, the limit $\lim_{m\to\infty} \bar{\kappa}(\eta_m)$ exists and it holds $\lim_{m\to\infty} \bar{\kappa}(\eta_m) \le \bar{\kappa}(0) < \infty$. 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,

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

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 $T(Q_{m_k}) \le \eta_{m_k}$ for each $k \in \mathbb{N}$ and Condition 2.3 implies $T(Q^*) = 0$ meaning that $Q^* \in \Omega_0$. As a result,

$$\begin{aligned} \bar{\kappa}(0) &\ge \lim_{k\to\infty} \bar{\kappa}(\eta_{m_k}) \\ &= \liminf_{k\to\infty} K(Q_{m_k}, P_1) \ge K(Q^*, P_1) \ge \bar{\kappa}(0), \end{aligned} \tag{5}$$

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.3 that $T(Q^*) = 0$ and (5) 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

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

for some $\delta > 0$, where the first inequality follows from (3), $c_n \searrow 0$, and $\Omega_\delta \subseteq \bar{\Omega}_\delta$, 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(b) and the closedness of $\mathcal{P}$). Since ϵ is arbitrary, we obtain the conclusion. □

A.3. Proof of Theorem 3.1

Pick any $\epsilon > 0$ to define $\mathcal{P}_\epsilon$ and $\mathcal{Q}_\epsilon$. Condition 2.1(a) follows from Lemma B.4 by replacing $m(x, \theta)$ with $m(x)$.

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 $a \in \mathbb{R}$ to define ρ_a. First, from Lemma B.5 (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 $\mathcal{P}_0(Q) \equiv \{P \in \mathcal{P}_0 : P \ll Q, Q \ll P\}$. Under Condition 3.1, we can apply Theorem 3.4 of (Borwein and Lewis, 1993): if $\mathcal{P}_0(Q)$ is not empty (i.e., the primal constraint qualification of (Borwein and Lewis, 1993) is satisfied), then

$$T_a(Q) \equiv \sup_{\gamma\in\Gamma_Q} E_Q[\rho_a(\gamma' m(x)) - \rho_a(0)] = \inf_{P\in\mathcal{P}_0(Q)} D_a(Q, P), \tag{6}$$

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

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

If $\mathcal{P}_0(Q)$ is empty, then we have $T_a(Q) = \infty$ (because we can take λ so that $\lambda' m(x)$ have the same sign for almost every x under Q) and $\inf_{P\in\mathcal{P}_0(Q)} D_a(Q, P) = \infty$ (by convention).

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(b) 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, (6) and the set inclusion relation imply $0 \le T_a(Q) = \inf_{P\in\mathcal{P}_0(Q)} D_a(Q, P) \le D_a(Q, Q) = 0$. 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 $Q \in \{R \in \mathcal{Q}_\epsilon : T_a(R) = 0\}$. From $T_a(Q) = 0$ and $\mathcal{P}_0(Q) \subset \mathcal{P}_0$, we have $\inf_{P\in\mathcal{P}_0} D_a(Q, P) = 0$. Since $\mathcal{P}_0$ is compact (by applying Lemma B.4 for the case of $\epsilon = 0$ 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 $P^* \in \mathcal{P}_0$ such that $\inf_{P\in\mathcal{P}_0} D_a(Q, P) = D_a(Q, P^*) = 0$. Now (D1) implies $Q = P^* \in \mathcal{P}_0$ and thus $Q \in \mathcal{Q}_\epsilon$ implies $Q \in \mathcal{P}_\epsilon$. Combining these results, Condition 2.1(b) 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.4 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

Pick any $\epsilon > 0$ to define $\mathcal{P}_\epsilon$ and $\mathcal{Q}_\epsilon$. Condition 2.1(a) follows from Lemma B.4.

Proof of (i). From Lemma B.5, $\phi_{GMM,n}$ is pointwise asymptotically level α. Also Condition 2.1(b) follows immediately. So we concentrate on showing that T_{GMM} satisfies Condition 2.2.

Pick any $Q^* \in \{Q \in \mathcal{Q}_\epsilon : K(Q,P_1) < \infty\}$. We 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 $\|\mu(Q,\theta)\| \ne 0$ and to be zero if $\|\mu(Q,\theta)\| = 0$. By Condition 3.2 and the Portmanteau Lemma (see, Lemma 2.2 of van der Vaart, 1998), 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 $Q_m \Rightarrow Q^* \in \{Q \in \mathcal{Q}_\epsilon : K(Q,P_1) < \infty\}$ 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 that $\det(\Sigma(Q^*,\tilde\theta(Q^*))) = 0$ and $\|\mu(Q^*,\theta^*)\| = 0$. Then $T_{GMM}(Q^*,\theta^*) = 0 \le \liminf_{m\to\infty} T_{GMM}(Q_m,\theta_m)$, since $T_{GMM}(Q_m,\theta_m) \ge 0$ for all $m \in \mathbb{N}$ by definition. Third, suppose $\det(\Sigma(Q^*,\tilde\theta(Q^*))) = 0$ and $\|\mu(Q^*,\theta^*)\| \ne 0$, so that $T_{GMM}(Q^*,\theta^*) = \infty$. We can partition the sequence $\{Q_m : m \in \mathbb{N}\}$ into two subsequences $\{Q_{m_j}\}$ and $\{Q_{m_k}\}$ such that

(a) $\det(\Sigma(Q_{m_j},\tilde\theta(Q_{m_j}))) = 0$ along the subsequence,
(b) $\det(\Sigma(Q_{m_k},\tilde\theta(Q_{m_k}))) > 0$ along the subsequence.

We concentrate on the case where both $\{Q_{m_j}\}$ and $\{Q_{m_k}\}$ have infinitely many elements (the case where $\{Q_{m_j}\}$ or $\{Q_{m_k}\}$ has a finite number of elements can be handled in the same manner). Since $\|\mu(Q_m,\theta_m)\| > 0$ for all m large enough, we can construct the above subsequences such that $\|\mu(Q_{m_j},\theta_{m_j})\| > 0$ and $\|\mu(Q_{m_k},\theta_{m_k})\| > 0$ for all j and k large enough. By the construction of the subsequences $\{Q_{m_j}\}$ and $\{Q_{m_k}\}$,

$$\begin{aligned}\liminf_{m\to\infty} T_{GMM}(Q_m,\theta_m) &= \min\left\{\liminf_{j\to\infty} T_{GMM}(Q_{m_j},\theta_{m_j}), \liminf_{k\to\infty} T_{GMM}(Q_{m_k},\theta_{m_k})\right\} \\ &= \liminf_{k\to\infty} T_{GMM}(Q_{m_k},\theta_{m_k}) = T_{GMM}(Q^*,\theta^*),\end{aligned}$$

where the second equality follows from $T_{GMM}(Q_{m_j},\theta_{m_j}) = \infty$ for all j large enough, and the third equality follows by $T_{GMM}(Q_{m_k},\theta_{m_k})$ being a continuous transformation of $\mu(Q_{m_k},\theta_{m_k})$ and $\Sigma(Q_{m_k},\tilde\theta(Q_{m_k}))$ and therefore continuous in (Q_{m_k},θ_{m_k}). Combining all three cases,

$$T_{GMM}(Q^*,\theta^*) \le \liminf_{m\to\infty} T_{GMM}(Q_m,\theta_m), \tag{7}$$

for any sequence $\{(Q_m,\theta_m) : m \in \mathbb{N}\}$ such that $Q_m \Rightarrow Q^* \in \{Q \in \mathcal{Q}_\epsilon : K(Q,P_1) < \infty\}$ and $\theta_m \to \theta^* \in \Theta$.

Now pick any sequence $\{Q_m : m \in \mathbb{N}\}$ in $\mathcal{Q}_\epsilon$ such that $Q_m \Rightarrow Q^* \in \{Q \in \mathcal{Q}_\epsilon : K(Q,P_1) < \infty\}$. By the compactness of Θ and the continuity of $T_{GMM}(Q,\theta)$ in $\theta \in \Theta$ for each $Q \in \mathcal{Q}_\epsilon$, there exists a sequence $\{\theta_m : m \in \mathbb{N}\}$ in Θ such that $T_{GMM}(Q_m) = T_{GMM}(Q_m,\theta_m)$ for each $m \in \mathbb{N}$. From the definition of the limit inferior, we can always take a subsequence $\{(Q_{m_j},\theta_{m_j}) : j \in \mathbb{N}\}$ such that

$$\liminf_{m\to\infty} T_{GMM}(Q_m) = \liminf_{m\to\infty} T_{GMM}(Q_m,\theta_m) = \lim_{j\to\infty} T_{GMM}(Q_{m_j},\theta_{m_j}).$$

If $\liminf_{m\to\infty} T_{GMM}(Q_m) = \infty$, then Condition 2.2 trivially holds for T_{GMM}. Thus consider the case of $\liminf_{m\to\infty} T_{GMM}(Q_m) < \infty$. Since $\{\theta_{m_j} : j \in \mathbb{N}\}$ is a sequence on a compact set Θ, we can take a further subsequence $\{\tilde\theta_{m_k} : k \in \mathbb{N}\}$ which converges to some $\tilde\theta \in \Theta$. It then follows that

$$\liminf_{m\to\infty} T_{GMM}(Q_m) = \lim_{k\to\infty} T_{GMM}(Q_{m_k},\tilde\theta_{m_k}) \ge T_{GMM}(Q^*,\tilde\theta) \ge T_{GMM}(Q^*),$$

where the first inequality follows from (7) and the second inequality follows by the definition of $T_{GMM}(Q^*)$. Therefore, T_{GMM} satisfies Condition 2.2.

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 Conditions 2.1(a) and 2.2, for each $P_1 \in \mathcal{Q}$ and each $\eta \in [0,\infty)$ with $\kappa(\eta) < \infty$, there exists $Q^* \in \Omega_\eta$ such that $K(Q^*,P_1) = K(\bar\Omega_\eta,P_1) = \kappa(\eta)$.*

Proof. Pick any $P_1 \in \mathcal{Q}$ and $\eta \in [0,\infty)$ with $\kappa(\eta) < \infty$. Define $\Omega'_\eta \equiv \{Q \in \mathcal{M} : K(Q,P_1) \le \kappa(\eta)+1\}$. Since Ω'_η is compact in the weak topology (Dupuis and Ellis, 1997, Lemma 1.4.3), $\bar\Omega_\eta \cap \Omega'_\eta$ is also compact. Since $K(Q,P_1)$ is lower semicontinuous in $Q \in \mathcal{M}$ under the weak topology (Dupuis and Ellis, 1997, Lemma 1.4.3), the compactness of $\bar\Omega_\eta \cap \Omega'_\eta$ implies that there exists a measure $Q^* \in \bar\Omega_\eta \cap \Omega'_\eta$ such that $K(Q^*,P_1) = K(\bar\Omega_\eta \cap \Omega'_\eta, P_1)$. Since $K(\bar\Omega_\eta,P_1) = K(\bar\Omega_\eta \cap \Omega'_\eta, P_1)$ (otherwise there would exist $\tilde Q \in \bar\Omega_\eta \setminus \Omega'_\eta$ such that $K(\tilde Q,P_1) < K(Q^*,P_1)$. But using $\tilde Q \notin \Omega'_\eta$ and $Q^* \in \Omega'_\eta$, we would obtain a contradiction). Therefore, we have $K(Q^*,P_1) = K(\bar\Omega_\eta,P_1) \le \kappa(\eta) < \infty$.

Finally, we show $Q^* \in \Omega_\eta$, which implies $K(\bar\Omega_\eta,P_1) = \kappa(\eta)$. From $K(Q^*,P_1) = K(\bar\Omega_\eta,P_1)$, we can take a sequence $\{Q_m : m \in \mathbb{N}\}$ in Ω_η such that $Q_m \Rightarrow Q^*$. Since $Q^* \in \mathcal{Q}$ (from $Q^* \in \bar\Omega_\eta \cap \Omega'_\eta$) and $K(Q^*,P_1) < \infty$, Condition 2.2 guarantees $T(Q^*) \le \liminf_{m\to\infty} T(Q_m) \le \eta$ so $Q^* \in \Omega_\eta$. This completes the proof. □

Lemma B.2. *Under Conditions 2.1(a) and 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 of positive real numbers 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 Lemma B.1 it follows that 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 $K(Q, P_1) \leq \kappa(\eta) < \infty$. And by Condition 2.1(a) and the fact that $\{Q_m : M \in \mathbb{N}\} \subseteq \mathcal{Q}$, it follows that $Q \in \mathcal{Q}$.

Third, since $T(Q_{m_j}) \leq \eta_{m_j}$ for each $j \in \mathbb{N}$ and T is lower semicontinuous at Q such that $K(Q, P_1) < \infty$ 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, $Q \in \Omega_\eta$ and we can conclude that

$$\kappa(\eta) \geq \lim_{j\to\infty} \kappa(\eta_{m_j}) \geq \liminf_{j\to\infty} K(Q_{m_j}, P_1) \geq K(Q, P_1) \geq \kappa(\eta),$$

which means $\lim_{j\to\infty} \kappa(\eta_{m_j}) = \kappa(\eta)$.

Finally, note that the conclusion also holds for $\eta \in [0, \infty)$ such that $\kappa(\eta) = \infty$. To see this, suppose not, i.e., $\kappa(\eta) = \infty$ 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.3. *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

$$\exists\{Q'_m : m \in \mathbb{N}\} \text{ such that } Q'_m \in \Omega_{\eta_m} \text{ for all } m \in \mathbb{N} \text{ and } Q'_m \Rightarrow Q^*. \quad (8)$$

Then Condition 2.3 implies $T(Q^*) = 0$. So it is sufficient to show (8).

From $Q_m \in \partial^*\Omega_{\eta_m}$ and the definition of $\partial\Omega_{\eta_m}$, it follows that for all $m \in \mathbb{N}$ there exists a sequence $\{Q_{k(m)} : k(m) \in \mathbb{N}\}$ such that $Q_{k(m)} \in \Omega_{\eta_m}$ for all $k(m) \in \mathbb{N}$ and $d_L(Q_{k(m)}, Q_m) \to 0$ as $k(m) \to \infty$. Thus, there exist $k^*(m)$ such that for all $k(m) \geq k^*(m)$,

$$d_L(Q_{k(m)}, Q_m) \leq 1/m.$$

Now pick any $\epsilon > 0$. From the above display and the fact that $Q_m \Rightarrow Q^*$, it follows that there exists $M \in \mathbb{N}$ such that $d_L(Q_{k^*(m)}, Q_m) \leq \epsilon/2$ and $d_L(Q_m, Q^*) \leq \epsilon/2$ for all $m \geq M$. We can conclude that $d_L(Q_{k^*(m)}, Q^*) \leq d_L(Q_{k^*(m)}, Q_m) + d_L(Q_m, Q^*) \leq \epsilon$. Since ϵ is arbitrary, we obtain $Q_{k^*(m)} \in \Omega_{\eta_m}$ for all $m \in \mathbb{N}$ and $Q_{k^*(m)} \Rightarrow Q^*$ as $m \to \infty$, so that (8) holds true. Condition 2.3 implies $T(Q^*) = 0$ and this completes the proof. □

Lemma B.4. *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 $\epsilon \geq 0$.*

Proof. Pick any $\epsilon \geq 0$. 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. Take a sequence $\{Q_m : m \in \mathbb{N}\}$ in $\mathcal{Q}_\epsilon$ such that $Q_m \Rightarrow Q^* \in \mathcal{M}$. Note that for every $m \in \mathbb{N}$, there exits $\theta_m \in \Theta$ such that $\det(\Sigma(Q_m, \theta_m)) \geq \epsilon$. Also, by compactness of Θ there exists a subsequence $\{\theta_{m_k} : k \in \mathbb{N}\}$ of $\{\theta_m\}$ 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

$$\begin{aligned}
&\left\| \Sigma(Q_{m_k}, \theta_{m_k}) - \Sigma(Q^*, \theta^*) \right\| \\
&\leq \left\| \int_{\mathcal{X}} (g(x, \theta_{m_k}, Q_{m_k}) - g(x, \theta^*, Q^*)) dQ_{m_k} \right\| \\
&\quad + \left\| \int_{\mathcal{X}} g(x, \theta^*, Q^*)(dQ^* - dQ_{m_k}) \right\| \\
&\leq \sup_{x\in\mathcal{X}} \left\| g(x, \theta_{m_k}, Q_{m_k}) - g(x, \theta^*, Q^*) \right\| \\
&\quad + \left\| \int_{\mathcal{X}} g(x, \theta^*, Q^*)(dQ^* - dQ_{m_k}) \right\| \to 0,
\end{aligned} \quad (9)$$

as $k \to \infty$, 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 $\det(\Sigma(Q^*, \theta^*)) \geq \epsilon$ and so $\mathcal{Q}_\epsilon$ is closed.

We next show the closedness of $\mathcal{P}_\epsilon$. Take a sequence $\{P_m : m \in \mathbb{N}\}$ in $\mathcal{P}_\epsilon$ 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 $\{\theta_{m_k} : k \in \mathbb{N}\}$ such that $\theta_{m_k} \to \theta^*$ for some $\theta^* \in \Theta$. Therefore, it is sufficient to show that $E_{P^*}[m(x, \theta^*)] = 0$. To prove this, note that

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

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 $\mathcal{X} \times \Theta$. □

Lemma B.5. *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, $2n\ell_{CU}(\theta^*) \Rightarrow \chi^2_q$ under P^*, and therefore,

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

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. □

References

Bahadur, R.R., 1960. Stochastic comparison of test. Annals of Mathematical Statistics 31, 276–295.

Billingsley, P., 1995. Probability and Measure. Wiley-Interscience.

Billingsley, P., 1999. Convergence of Probability Measures. Wiley-Interscience.

Borwein, J.M., Lewis, A.S., 1993. Partially-finite programming 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.

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

Dembo, A., Zeitouni, O., 1998. Large Deviations Techniques and Applications, second ed. Springer, New York.

Dupuis, P., Ellis, R.S., 1997. A Weak Convergence Approach to the Theory of Large Deviations. In: 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., 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., Johnson, P., 1998. Information theoretic approaches to inference in moment condition models. Econometrica 66, 333–357.

Kallenberg, W., 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., Shaikh, A., 2012. On the asymptotic optimality of empirical likelihood for testing moment restrictions. Econometrica 80, 413–423.

Liese, F., Vajda, I., 1987. Convex Statistical Distances. Teubner, Leipzig.

Nelson, F.D., Savin, N.E., 1990. The danger of extrapolating asymptotic local power. Econometrica 58, 977–981.

Newey, W.K., 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., 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.