

Hodges-Lehmann Optimality for Testing Moment Condition Models*

Ivan A. Canay [†]

Department of Economics
Northwestern University

Taisuke Otsu [‡]

Cowles Foundation & Department of Economics
Yale University

November 2009

Abstract

Economic models typically involve a set of moment conditions. The most widely used tests for moment conditions are the J -test associated with the generalized method of moments (GMM) and discrepancy tests associated with the family of generalized empirical likelihood (GEL). It is known that all of these tests have the same asymptotic properties under the null and local alternatives. This paper studies the Hodges and Lehmann (1956) optimality of these tests: evaluate these 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 of the tests bounded by some constant. We derive an optimal rate for this exponential rate of convergence and provide general conditions for a test to achieve this optimal rate. The results are applied to show that the GMM and GEL tests are Hodges-Lehmann optimal under mild conditions.

KEYWORDS: asymptotic optimality, large deviation, moment condition, generalized method of moments, generalized empirical likelihood.

JEL CLASSIFICATION: C12, C14.

*We would like to thank Andrew Barron and seminar participants at Indiana University and UCLA for helpful comments.

[†]Email: iacanay@northwestern.edu. Website: <http://faculty.wcas.northwestern.edu/~iac879>.

[‡]E-mail: taisuke.otsu@yale.edu. Website: <http://cowles.econ.yale.edu/faculty/otsu.htm>. Financial support from the National Science Foundation (SES-0720961) is gratefully acknowledged.

1 Introduction

Economic models typically involve a set of moment conditions. Probably the most widely used test to check the validity of such moment conditions is the J -test associated with the generalized method of moments (GMM) introduced by Hansen (1982). However, over the last two decades various one-step alternatives to the GMM and the associated moment condition tests have been proposed. These include empirical likelihood (EL) of Owen (1988), Qin and Lawless (1994), and Imbens (1997), continuous updating GMM of Hansen, Heaton, and Yaron (1996), exponential tilting of Kitamura and Stutzer (1997) and Imbens, Spady, and Johnson (1998), and Hellinger distance of Kitamura, Otsu, and Evdokimov (2009). Smith (1997) and Newey and Smith (2004) show that these approaches, as well as the Cressie and Read (1984) power divergence family of discrepancies, share a common structure typically known as generalized empirical likelihood (GEL). It is known that all of these tests have the same asymptotic properties under the null and local alternatives. This paper studies the Hodges and Lehmann (1956) optimality of these tests: evaluate these 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 of the tests bounded by some constant. We derive an optimal rate for this exponential rate of convergence and provide general conditions for a test to achieve this optimal rate. The results are applied to show that the GMM and GEL tests are Hodges-Lehmann optimal under mild conditions.

There are several existing papers that investigate statistical properties of the GMM and GEL methods beyond their first order asymptotic properties. Newey and Smith (2004) show that the EL estimator has lower higher order asymptotic bias than other GEL estimators and GMM, and the bias-adjusted EL estimator is higher order efficient relative to other bias corrected estimators. Imbens, Spady, and Johnson (1998) show that EL has an influence function that is unbounded, suggesting that exponential tilting might behave better under misspecification. This conjecture was later proved formally by Schennach (2007). These results, however, are concerned about point estimation problems rather than hypothesis testing problems for moment conditions. In terms of large deviations asymptotics, Kitamura (2001) and Kitamura, Santos, and Shaikh (2009) provide conditions under which EL is uniformly most powerful in a generalized Neyman-Pearson sense for testing moment restrictions. Additional large deviation optimality results include those in Kitamura and Otsu (2005), Canay (2008), and Otsu (2009). All these large deviation results share two characteristics. First, the type I error probability of the tests goes to zero asymptotically. Such approximations are often criticized as they do not reassemble the standard situation where a test statistic is compared to a fixed asymptotic critical value. Second, while the results prove that EL achieves some form of large deviation optimality, typically the possibility that other competing tests are also optimal remains an open question. The Hodges-Lehmann optimality analysis considered in this paper does not share any of these two aspects and contributes to the literature in three different ways. First, we derive an optimal rate of exponential growth of the power function to

one when the type I error probability is asymptotically bounded away from zero. This type of approximation represents the standard case where the researcher fixes a significance level and tests the null hypothesis by comparing a test statistic with a critical value such that the asymptotic size of the test is not above the fixed significance level. Second, we derive two different sets of sufficient conditions for a test to be Hodges-Lehmann optimal. These sufficient conditions apply for a wide range of testing problems (beyond testing moment restrictions) and therefore provide a useful starting point for the study of Hodges-Lehmann optimality in other semi-parametric contexts (as opposed to the existing literature on the Hodges-Lehmann optimality analysis, as in Serfling (1980, Ch. 10) and Nikitin (1995), which mostly focuses on parametric or some simple non-parametric models). Third, we apply our sufficient conditions for the problem of testing moment restrictions and show that the GMM and GEL tests are all Hodges-Lehmann optimal.

Our Hodges-Lehmann optimality analysis is split into four steps. The first step consists in deriving an optimal rate of exponential growth of the power function to one when the type I error probability is asymptotically bounded away from zero. Note that in such a setting the power of the test under fixed alternatives goes to one asymptotically as all the tests we consider are consistent. Second, since the derived optimal rate is related the Kullback-Leibler (KL) divergence between the alternative distribution and the set of the null distributions, EL arises as a likely candidate to achieve this optimal rate. We prove this conjecture formally. However, the proof of this result uses intensively the relationship between EL and the KL divergence and is not very informative about whether other tests might also achieve this optimal rate. Third, to analyze the Hodges-Lehmann optimality of the GMM tests, we provide a set of sufficient conditions for a test to achieve the optimal convergence rate based on lower semicontinuity in the weak topology of the mapping that defines the test statistic, and show that the GMM tests satisfy this sufficient condition. However, we give an example that shows that EL, while being Hodges-Lehmann optimal, does not satisfy such strong continuity condition. Finally, we use this example as a motivation for a second set of sufficient conditions and use this to show that the Cressie-Read family of test statistics, that include EL and exponential tilting as particular cases, are Hodges-Lehmann optimal.

The remainder of the paper is organized as follows. Section 2 introduces the basic notation, defines formally the notion of the Hodges-Lehmann optimality, and derives the optimal convergence rate of the type II error probability. Section 3 shows that the EL test is Hodges-Lehmann optimal. Section 4 provides two sets of sufficient conditions for a test to be Hodges-Lehmann optimal and shows that the GMM test and Cressie-Read family of tests are all optimal. Finally, Section 5 concludes.

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 Preliminaries

Suppose the econometrician observes an i.i.d. sample $\{x_i : i = 1, \dots, n\}$ generated from a probability measure P_0 with support \mathcal{X} and wishes to test the validity of moment conditions specified by some economic theory. The null hypothesis of the econometrician is written as

$$H_0 : E_{P_0}[m(x, \theta_0)] \equiv \int_{\mathcal{X}} m(x, \theta_0) dP_0 = 0, \quad \text{for some } \theta_0 \in \Theta \subseteq \mathbb{R}^k, \quad (2.1)$$

where $m : \mathcal{X} \times \Theta \rightarrow \mathbb{R}^q$ is a $q \times 1$ vector of known functions with $q > k$ (i.e., over-identified). The alternative hypothesis H_1 is that H_0 is false, i.e., $E_{P_0}[m(x, \theta)] \neq 0$ for any $\theta \in \Theta$. We do not make parametric assumptions on the distributional form of P_0 . Let \mathcal{M} be the space of probability measures on the Borel σ -field $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ and

$$\mathcal{P} \equiv \cup_{\theta \in \Theta} \{P \in \mathcal{M} : E_P[m(x, \theta)] = 0\} \quad (2.2)$$

be a set of probability measures satisfying the moment conditions at some point in Θ . Then the above testing problem can be alternatively written as

$$H_0 : P_0 \in \mathcal{P}, \quad H_1 : P_0 \in \mathcal{M} \setminus \mathcal{P}. \quad (2.3)$$

A test ϕ_n for H_0 against H_1 is a map from data into a binary decision, where $\phi_n = 0$ ($\phi_n = 1$) means acceptance (rejection) of the null hypothesis H_0 . We pay particular attention to tests that take the form of

$$\phi_n \equiv 1\{T(\hat{P}_n) > c_n\}, \quad (2.4)$$

where \hat{P}_n is the empirical measure, $T : \mathcal{M} \rightarrow \bar{\mathbb{R}}$ is a mapping to define the test statistic $T(\hat{P}_n)$, and $\{c_n : n \in \mathbb{N}\}$ is a sequence of positive real numbers monotonically decreasing to zero. Most existing tests fall into this class of tests as the following examples illustrate.

Example 1 [GMM] One common test for the null H_0 is based on the GMM of Hansen (1982). The (two-step) GMM J -statistic for the moment conditions (2.1) is defined as

$$J_n \equiv n \times \inf_{\theta \in \Theta} \bar{m}(\theta)' \Sigma(\hat{P}_n, \tilde{\theta}(\hat{P}_n))^{-1} \bar{m}(\theta),$$

where $\bar{m}(\theta) \equiv n^{-1} \sum_{i=1}^n m(x_i, \theta)$, $\Sigma(\hat{P}_n, \tilde{\theta}(\hat{P}_n)) \equiv n^{-1} \sum_{i=1}^n [m(x_i, \tilde{\theta}(\hat{P}_n)) - \bar{m}(\tilde{\theta}(\hat{P}_n))] [m(x_i, \tilde{\theta}(\hat{P}_n)) - \bar{m}(\tilde{\theta}(\hat{P}_n))]'$, and $\tilde{\theta}(\hat{P}_n) = \arg \min_{\theta \in \Theta} \bar{m}(\theta)' W \bar{m}(\theta)$ is a preliminary estimator of θ_0 using a $q \times q$ fixed weight matrix W . Note that this preliminary estimator can be regarded as a mapping of \hat{P}_n and

the J -statistic is written as $J_n = n \times T_{GMM}(\hat{P}_n)$, where

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

Since $J_n \Rightarrow \chi_{q-k}^2$ under H_0 with certain regularity conditions, the J -test is written as

$$\phi_{GMM,n} \equiv 1\{T_{GMM}(\hat{P}_n) > \chi_{q-k,1-\alpha}^2/n\},$$

where $\chi_{q-k,1-\alpha}^2$ is the $1 - \alpha$ quantile of the χ_{q-k}^2 distribution. ■

Example 2 [GEL] Alternatives to the GMM J -test include tests based on GEL, as in Smith (1997) and Newey and Smith (2004). Let $\rho : \mathcal{V} \rightarrow \mathbb{R}$ be a concave function on its domain \mathcal{V} , which is an open interval containing zero. Let $\Gamma_Q(\theta) \equiv \{\gamma \in \mathbb{R}^q : \Pr\{\gamma' m(x_i, \theta) \in \mathcal{V} : Q\} = 1\}$. The GEL over-identifying restriction test statistic is given by

$$H_n \equiv \inf_{\theta \in \Theta} \sup_{\gamma \in \Gamma_{\hat{P}_n}(\theta)} 2 \sum_{i=1}^n [\rho(\gamma' m(x_i, \theta)) - \rho(0)], \quad (2.5)$$

provided a solution for γ exists. The EL test statistic is a special case with $\rho(v) = \log(1 - v)$ and $\mathcal{V} = (-\infty, 1)$ (Imbens, 1997; Qin and Lawless, 1994; Smith, 1997). The exponential tilting test statistic is a special case with $\rho(v) = -\exp(v)$ and $\mathcal{V} = (-\infty, \infty)$ (Imbens, Spady, and Johnson, 1998; Kitamura and Stutzer, 1997; Smith, 1997). Other members include the continuous updating GMM (Hansen, Heaton, and Yaron, 1996) for $\rho(v) = -1/2(1 + v)^2$ and the Pearson's χ^2 for $\rho(v) = \sqrt{2v}$. Note that the GEL test statistic is written as $H_n = 2n \times T_{GEL}(\hat{P}_n)$, where

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

Since $H_n \Rightarrow \chi_{q-k}^2$ under H_0 with certain regularity conditions, the GEL test is written as

$$\phi_{GEL,n} \equiv 1\{T_{GEL}(\hat{P}_n) > \chi_{q-k,1-\alpha}^2/(2n)\}. \quad (2.6)$$

■

Our Hodges-Lehmann optimality analysis focuses on the convergence rate of the type II error probability, or power of the test ϕ_n , while fixing the limit of the type I error probability. Therefore, our analysis can directly investigate the asymptotic performance of the GMM test $\phi_{GMM,n}$ or GEL test $\phi_{GEL,n}$ above, which are employed in empirical research. On the other hand, the other approaches, such as the generalized Neyman-Pearson analysis of Kitamura (2001) and Kitamura, Santos, and Shaikh (2009) and the Bahadur analysis of Otsu (2009), consider the case where the type I error probabilities converge to zero. Therefore, intuitively, those approaches do not directly

analyze performance of the test in the form of $\phi_{GMM,n}$ or $\phi_{GEL,n}$ but analyze performance of the test in the form of $\tilde{\phi}_{GMM,n} \equiv 1\{J_n > n\chi_{q-k,1-\alpha}^2\} = 1\{T_{GMM}(\hat{P}_n) > \chi_{q-k,1-\alpha}^2\}$, for example, which is somewhat different from the testing procedure employed in practice. Also in general there is no clear logical relationships among those different optimality criteria.

2.1 Hodges-Lehmann Optimality

To assess the asymptotic performance of test procedures there exist several optimality criteria which usually consider problems that become harder as the sample size increases. A line of attack that is applicable to a wide range of cases is based on the theory of large deviations and has been used since the papers by Bahadur (1960), Chernoff (1952), and Hoeffding (1965), among others. Typically, test procedures are compared through their error probabilities and the various methods of comparison differ in the manner in which the type I and type II error probabilities vary with the sample size, and also in the manner in which the alternatives under consideration are required to behave. Letting α_n and β_n denote the type I and type II error probabilities of a test using a sample of size n , respectively, each performance criterion entails particular specifications regarding: (i) whether α_n goes to zero or not, (ii) whether β_n goes to zero or not, and (iii) whether the alternative hypotheses are fixed or get closer to the null with the sample size.¹

In this paper we focus on a particular notion of asymptotic optimality, where the type I error probability is asymptotically bounded above by some constant $\alpha \in (0, 1)$, the type II error probability decreases to zero as the sample size increases, and the alternative distribution is held fixed. This approach is known as the Hodges-Lehmann optimality analysis (see Hodges and Lehmann, 1956). Consider any test ϕ_n for the null H_0 that takes the form in (2.4). The type I error probability of the test ϕ_n is given by $\alpha_n = \sup_{P \in \mathcal{P}} E_P[\phi_n]$ so we say the test ϕ_n is uniformly asymptotically level $\alpha \in (0, 1)$ if

$$\limsup_{n \rightarrow \infty} \alpha_n = \limsup_{n \rightarrow \infty} \sup_{P \in \mathcal{P}} E_P[\phi_n] \leq \alpha. \quad (2.7)$$

On the other hand, we say that the test ϕ_n is pointwise asymptotically level $\alpha \in (0, 1)$ if

$$\limsup_{n \rightarrow \infty} E_P[\phi_n] \leq \alpha, \quad \text{for all } P \in \mathcal{P}. \quad (2.8)$$

Note that the uniform size control in (2.7) implies the pointwise size control (2.8) but not vice versa. On the other hand, the type II error probability evaluated at the alternative P_1 is given by $\beta_n = E_{P_1}[1 - \phi_n]$. The definition of the Hodges-Lehmann optimality is given below.

Definition 2.1 [Hodges-Lehmann] A test $\phi_{HL,n} : \mathcal{X}^n \rightarrow \{0, 1\}$ for the null $H_0 : P_0 \in \mathcal{P}$ against the alternative $H_1 : P_0 \in \mathcal{M} \setminus \mathcal{P}$ is called Hodges-Lehmann optimal at $P_1 \in \mathcal{M} \setminus \mathcal{P}$ if

1. $\phi_{HL,n}$ does not depend on P_1 ,

¹For a review on asymptotic comparisons of tests, see Serfling (1980, Ch. 10) and van der Vaart (1998, Ch. 14)

2. $\phi_{HL,n}$ satisfies (2.7) or (2.8),
3. for any alternative test ϕ_n that does not depend on P_1 and satisfies (2.7) or (2.8), it holds

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log E_{P_1}[1 - \phi_{HL,n}] \leq \liminf_{n \rightarrow \infty} \frac{1}{n} \log E_{P_1}[1 - \phi_n]. \quad (2.9)$$

Note that Definition 2.1 depends on the way that the tests asymptotically control the type I error probabilities. This is, given a restriction on the type I error probability (either (2.7) or (2.8)), a given 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 the alternative test. This is important since there are cases where it is not possible to find tests that are uniformly valid as in (2.7) (e.g., Bahadur and Savage (1956)), so such requirement might be too strong in some contexts. We will not discuss here conditions under which a given test is uniformly asymptotically level α because it does not affect our search of a Hodges-Lehmann optimal test.

Definition 2.1 defines an optimal test relative to *any* alternative test. The set of alternative tests for H_0 is potentially very large and therefore it might be infeasible to explore the inequality for every possible alternative test, given a candidate optimal test. The approach we take here divides the analysis in two parts. First, we show that in the class of tests taking the form of (2.4), there exists an optimal convergence rate for the type II error probability (or equivalently, a lower bound for $\liminf_{n \rightarrow \infty} n^{-1} \log E_{P_1}[1 - \phi_n]$). Then we investigate whether each candidate test achieves such optimal rate or not.

2.2 Optimal Convergence Rate of Type II Error Probability

We now 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 ,² and

$$K(Q, P) \equiv \begin{cases} \int_{\mathcal{X}} \log(dQ/dP) dQ & \text{if } Q \ll P \\ \infty & \text{otherwise} \end{cases} \quad (2.10)$$

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

Theorem 2.1 *For any test ϕ_n satisfying (2.7) or (2.8),*

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log E_{P_1}[1 - \phi_n] \geq -K(\mathcal{P}, P_1), \quad (2.11)$$

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

² Q is absolutely continuous with respect to P if $P(B) = 0$ implies $Q(B) = 0$ for every measurable set B .

Theorem 2.1 shows that the best exponential rate of growth of the power function depends on the KL divergence between the set of null probability distributions \mathcal{P} and the fixed alternative probability distribution P_1 . The bound in (2.11) is informative for alternatives such that $0 < K(\mathcal{P}, P_1) < \infty$. Under the assumptions employed in the next sections, the set of null measures \mathcal{P} is closed in the weak topology (see Lemma B.1), and this guarantees $K(\mathcal{P}, P_1) > 0$ for any $P_1 \in \mathcal{M} \setminus \mathcal{P}$.

Otsu (2009) studied the Bahadur optimality for testing the moment conditions (2.1), where the roles of the type I and type II error probabilities are interchanged: compare the convergence rates of the type I error probabilities toward zero with fixed values of the type II error probabilities. In Otsu (2009), the best possible rate of the type I error is obtained as $-K(P_1, \mathcal{P})$. Since the KL divergence is not symmetric (i.e., $K(P, Q) \neq K(Q, P)$) in general, this Bahadur bound is different from our Hodges-Lehmann bound.

There are two reasons to suspect that the EL test might achieve the lower bound in (2.11). On the one hand, the EL test for H_0 is based on the minimum KL divergence between the empirical measure \hat{P}_n and the set of null measures \mathcal{P} . On the other hand, Sanov's Theorem (see, e.g., Theorem 6.2.10 of Dembo and Zeitouni (1998)) says that the empirical measure \hat{P}_n satisfies the large deviation principle with the rate function $K(Q, P)$, and $K(Q, P)$ controls the large deviation limit behavior of the probability that \hat{P}_n falls into any subset of \mathcal{M} . Since a test induces a partition of \mathcal{M} into acceptance and rejection regions and EL uses a function tightly related to the large deviation behavior of the empirical measure, it is a likely candidate to achieve the bound in (2.11). In the next section we briefly describe EL and present conditions under which this conjecture holds true.

3 Optimality of Empirical Likelihood

EL is a data-driven nonparametric method of estimation and inference for moment condition models, which does not require weight matrix estimation like the GMM and is invariant to nonsingular linear transformations of the moment conditions. It was introduced by Owen(1988, 1990, 1991) and later studied in depth by Qin and Lawless (1994), Imbens, Spady, and Johnson (1998), Kitamura (2001), and Newey and Smith (2004), among others.

EL maximizes the nonparametric likelihood over distributions with an atom of probability on each x_i that impose the moment condition for a given $\theta \in \Theta$, and then picks the value of θ that maximizes the likelihood. The (restricted) empirical log-likelihood is

$$l_{EL}^r \equiv \sup_{\theta \in \Theta} \max_{p_1, \dots, p_n} \left\{ \sum_{i=1}^n \log(p_i) \left| p_i > 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i m(x_i, \theta) = 0 \right. \right\}$$

where p_i denotes the probability mass placed at x_i by a discrete distribution with support $\{x_1, \dots, x_n\}$. The unrestricted empirical log-likelihood, l_{EL}^u , is similar to l_{EL}^r except that the moment restriction

$\sum_{i=1}^n p_i m(x_i, \theta) = 0$ is not imposed. The solution in such a case is simply $\tilde{p}_i = 1/n$ and then $l_{EL}^u = -n \log(n)$. The EL ratio statistic arises by computing the difference between the restricted and unrestricted log-likelihoods,

$$\begin{aligned} ELR_n &\equiv 2\{l_{EL}^u - l_{EL}^r\} \\ &= \inf_{\theta \in \Theta} \min_{p_1, \dots, p_n} \left\{ 2 \sum_{i=1}^n \log \left(\frac{1/n}{p_i} \right) \middle| p_i > 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i m(x_i, \theta) = 0 \right\}. \end{aligned}$$

Therefore, large values of this statistic suggest the over-identifying restriction (2.1) is not supported by the data. The EL test statistic can be alternatively written as $ELR_n = 2n \times T_{EL}(\hat{P}_n)$, where

$$T_{EL}(Q) \equiv \inf_{P \in \mathcal{P}(Q)} K(Q, P), \quad (3.1)$$

and $\mathcal{P}(Q) \equiv \{J \in \mathcal{P} : J \ll Q, Q \ll J\}$. The infimum over the empty set is understood to be infinity. Intuitively, EL picks the measure in the set $\mathcal{P}(\hat{P}_n)$ that is closest to the empirical measure, where the closeness is defined in terms of the KL divergence. Under H_0 with certain regularity conditions, $ELR_n \Rightarrow \chi_{q-k}^2$ so that the EL test can be written as

$$\phi_{EL,n} \equiv 1\{ELR_n > \chi_{q-k, 1-\alpha}^2\} = 1\{T_{EL}(\hat{P}_n) > \chi_{q-k, 1-\alpha}^2 / (2n)\}.$$

Thus, the EL test takes the form in (2.4). It is known that the EL test statistic ELR_n is equivalent to H_n in (2.5) for $\rho(v) = \log(1-v)$. We impose the following conditions.

Condition 3.1 The support \mathcal{X} and the parameter space Θ are compact subsets of \mathbb{R}^d and \mathbb{R}^k , respectively.

Condition 3.2 The function $m : \mathcal{X} \times \Theta \rightarrow \mathbb{R}^q$ is continuous in both of its arguments.

Conditions 3.1 and 3.2 guarantee that the sets \mathcal{M} and \mathcal{P} of probability measures are compact under the weak topology (see, Theorem D.8 of Dembo and Zeitouni (1998) and Lemma B.1). The Hodges-Lehmann optimality of the EL test is presented as follows.

Theorem 3.1 *Under Conditions 3.1 and 3.2, it holds*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log E_{P_1}[1 - \phi_{EL,n}] = -K(\mathcal{P}, P_1),$$

for each $P_1 \in \mathcal{M} \setminus \mathcal{P}$, i.e., the EL test $\phi_{EL,n}$ is Hodges-Lehmann optimal at each $P_1 \in \mathcal{M} \setminus \mathcal{P}$.

The proof of Theorem 3.1 uses intensively the fact that EL is based on the KL divergence as a statistical criterion. As a consequence, it is not straightforward to extend the proof to show whether

other tests might also achieve this bound. The next section provides general sufficient conditions that allow us to study the Hodges-Lehmann optimality for a wide range of test procedures.

4 General Conditions for Optimality

4.1 Lower Semicontinuous Case: Optimality of GMM

There are several 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. For example, Tusnády (1977) shows that when the null hypothesis is simple (i.e., the null set of measures is a singleton), the majority of tests becomes Hodges-Lehmann optimal. Kallenberg and Kourouklis (1992) show that the Hodges-Lehmann optimality emerges in general when the acceptance region of a test converge to the null set of measures in a coarse way, *provided* the functional $T : \mathcal{M} \rightarrow \bar{\mathbb{R}}$ associated with the test $\phi_n \equiv 1\{T(\hat{P}_n) > c_n\}$ is continuous in the τ -topology. Kallenberg and Kourouklis (1992) apply their result to show that in parametric exponential family models several tests, including the likelihood ratio and t tests, are Hodges-Lehmann optimal. In this subsection we extend the result of Kallenberg and Kourouklis (1992) to show that the continuity assumption in the τ topology can be replaced with a lower semicontinuity assumption in the weak topology.³ In particular, we show that the lower semicontinuity of the mapping T in the weak topology combined with a condition to guarantee that the limit of the acceptance region of a test is not closer to the alternative distribution than the null set in terms of the KL divergence would be sufficient to ensure the Hodges-Lehmann optimality. Our conditions are specified as follows.

Condition 4.1 The mapping $T : \mathcal{M} \rightarrow \bar{\mathbb{R}}$ is lower semicontinuous in the weak topology on the set $\{Q \in \mathcal{M} : K(Q, P_1) < \infty\}$.

Condition 4.2 The mapping $T : \mathcal{M} \rightarrow \bar{\mathbb{R}}$ satisfies

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

These two conditions imply the Hodges-Lehmann optimality.

Theorem 4.1 *Assume that the mapping $T : \mathcal{M} \rightarrow \bar{\mathbb{R}}$ for the test $\phi_n \equiv 1\{T(\hat{P}_n) > c_n\}$ with $c_n \searrow 0$ satisfies Conditions 4.1 and 4.2. Then the test ϕ_n satisfies*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log E_{P_1} [1 - \phi_n] = -K(\mathcal{P}, P_1),$$

³Note that continuity in the weak topology implies continuity in the τ -topology, so requiring continuity in the weak topology would imply the condition in Kallenberg and Kourouklis (1992). Lower semicontinuity in the weak topology, on the other hand, does not imply τ -continuity. We believe that working with the weak topology is natural in our context since it is the topology compatible with weak convergence of probability measures.

for each $P_1 \in \mathcal{M} \setminus \mathcal{P}$, i.e., ϕ_n is Hodges-Lehmann optimal at each $P_1 \in \mathcal{M} \setminus \mathcal{P}$.

Theorem 4.1 is useful for analyzing the Hodges-Lehmann optimality for different tests in different contexts: given a test statistic based on a mapping $T : \mathcal{M} \rightarrow \bar{\mathbb{R}}$, we just need to verify whether Conditions 4.1 and 4.2 are satisfied or not. In our over-identifying restriction testing problem (2.1), this theorem is general enough to accommodate the (two-step) GMM J -test and the continuous updating GMM test.

Theorem 4.2 *Assume Conditions 3.1 and 3.2 hold. Define the two-step GMM J -test statistic as $J_n = n \times T_{GMM}(\hat{P}_n)$ and the continuous updating GMM test statistic as $n \times T_{CU}(\hat{P}_n)$, where*

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

for $\Sigma(Q, \theta) \equiv E_Q[(m(x, \theta) - E_Q m(x, \theta))(m(x, \theta) - E_Q m(x, \theta))']$ and $\tilde{\theta}(Q) : \mathcal{M} \rightarrow \Theta$ such that $\tilde{\theta}(Q_m) \rightarrow \tilde{\theta}(Q)$ whenever $Q_m \Rightarrow Q$, and

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

Then the mappings T_{GMM} and T_{CU} satisfy Conditions 4.1 and 4.2, and the two-step GMM J -test $\phi_{GMM, n} \equiv 1\{T_{GMM}(\hat{P}_n) > \chi_{q-k, 1-\alpha}^2/n\}$ and the continuous updating GMM test $\phi_{CU, n} \equiv 1\{T_{CU}(\hat{P}_n) > \chi_{q-k, 1-\alpha}^2/n\}$ are Hodges-Lehmann optimal for each $P_1 \in \mathcal{M} \setminus \mathcal{P}$ satisfying $0 < K(\mathcal{P}, P_1) < \infty$.

Conditions 4.1 and 4.2 are sufficient conditions meaning that even if a given test violates either Condition 4.1 or 4.2, we cannot conclude that such test is sub-optimal. In fact, we can use the Hodges-Lehmann optimality of the EL test in Theorem 3.1 to argue that these conditions turn out to be too strong in some cases. The examples below show that the mapping T_{EL} in (3.1) is neither lower nor upper semicontinuous in the weak topology, implying that Condition 4.1 is sufficient but not necessary.

Example 3 [T_{EL} is not lower semicontinuous] Suppose $m(x, \theta_0) = x$, $H_0 : E_{P_0}[X] = 0$, and $\mathcal{X} = [-x_L, x_H]$ for some $x_L > 0$ and $x_H > 0$. Note that Conditions 3.1 and 3.2 are satisfied. This example shows that $\Lambda_\eta \equiv \{Q \in \mathcal{M} : \inf_{P \in \mathcal{P}(Q)} K(Q, P) \leq \eta\}$ is not closed in the weak topology for any $\eta > 0$, meaning that T_{EL} is not lower semicontinuous in the weak topology. For a probability measure Q , let \mathcal{X}_Q denote the support of Q and $-x_{LQ}$ and x_{HQ} 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} . It is known that (see, e.g., Borwein and Lewis (1993))

$$\inf_{P \in \mathcal{P}(Q)} K(Q, P) = \max_{\gamma \in \Gamma_Q} \int_{\mathcal{X}} \log(1 + \gamma x) dQ,$$

where $\Gamma_Q = (-1/x_{H_Q}, 1/x_{L_Q})$ is the parameter space for the Lagrange multiplier γ . The corresponding ratios are equal to ∞ by definition if $x_{H_Q} \leq 0$ or $x_{L_Q} \leq 0$. 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 $\Gamma_{Q_m} = (-1/x^*, 1/x_L)$. This sequence weakly converges to the probability measure Q satisfying

$$Q(X = 0) = 1 - p, \quad Q(X = x^*) = p,$$

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 $\liminf_{m \rightarrow \infty} (x_{L_m} - x_{L_Q}) > 0$. 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 $Q \notin \Lambda_\eta$. Note that

$$\int_{\mathcal{X}} \log(1 + \gamma x) dQ_m = \log(1 - x_L \gamma)/m + \log(1 + \gamma x^*)p,$$

so that

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

and the EL test statistic increases to $\log(1 + x^*/x_L)p$ as $m \rightarrow \infty$. Thus, in order for Q_m to be in Λ_η for all $m \in \mathbb{N}$ (i.e., $T_{EL}(Q_m) \leq \eta$ for all $m \in \mathbb{N}$), x^* must satisfy

$$x^* \leq x_L(\exp(\eta/p) - 1). \tag{4.1}$$

However, in the case of the limit sequence Q we get

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

Note that $T_{EL}(Q) = \infty$ regardless of how small x^* or p might be, as long as both are positive. This happens because there is no $P \in \mathcal{P}$ that is equivalent to Q , since Q put mass on only one side of 0, and therefore $\mathcal{P}(Q) = \emptyset$. 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} for the EL test is not lower semicontinuous in the weak topology. ■

Example 4 [T_{EL} is not upper semicontinuous] Consider the sequence of measures satisfying $Q_m(X = 0) = 1 - 1/m$ and $Q_m(X = 1) = 1/m$ for $m \in \mathbb{N}$. If $T_{EL}(Q)$ were upper semicontinuous the set $\Lambda_\eta^r \equiv \{Q \in \mathcal{M} : T_{EL}(Q) > \eta\}$ would be closed for all $\eta > 0$. To see that this is not the case, note

that for the sequence Q_m ,

$$T_{EL}(Q_m) = \max_{\gamma \in (-1, \infty)} \log(1 + \gamma) \times \frac{1}{m} = \infty,$$

which implies $Q_m \in \Lambda_\eta^r$ for all $m \in \mathbb{N}$ and $\eta > 0$. However, $Q_m \Rightarrow Q$ where $Q\{X = 0\} = 1$ and $T_{EL}(Q) = 0$ since $Q \in \mathcal{P}$. Therefore, Λ_η^r is not closed and T_{EL} is not upper semicontinuous in the weak topology. ■

Example 3 is important for three reasons. First, it shows that the mapping T_{EL} does not satisfy Condition 4.1 and yet EL is Hodges-Lehmann optimal by Theorem 3.1. Second, it shows that lower semicontinuity fails due to the fact that weak convergence of Q_m to Q is not enough to say anything about the relationship between $\mathcal{P}(Q_m)$ and $\mathcal{P}(Q)$. In Example 3, the sequence Q_m is such that $\mathcal{P}(Q_m)$ is not empty along the sequence but $\mathcal{P}(Q) = \emptyset$. This makes the example simple but it is not necessary for the discontinuity to happen. It can be shown that lower semicontinuity fails even in cases where $\mathcal{P}(Q)$ is not empty. Third, this example suggests that it is important that $\eta > 0$ remains fixed to make the argument. If we consider a sequence of positive real numbers $\{\eta_m : m \in \mathbb{N}\}$ such that $\eta_m \searrow 0$, we would have by (4.1) that $x^* \rightarrow 0$ as $m \rightarrow \infty$ and in such case $T_{EL}(Q)$ would be equal to zero. This idea motivates an alternative sufficient condition for the Hodges-Lehmann optimality that does not require lower semicontinuity of the mapping T .

4.2 Discontinuous Case: Optimality of GEL

Based on the results of the previous subsection, this subsection derives the Hodges-Lehmann optimality under the following condition.

Condition 4.3 The mapping $T : \mathcal{M} \rightarrow \bar{\mathbb{R}}$ is such that whenever a sequence of measures $\{Q_m : m \in \mathbb{N}\}$ in \mathcal{M} and a positive sequence $\{\eta_m : m \in \mathbb{N}\}$ decreasing to zero satisfy $Q_m \Rightarrow Q \in \mathcal{M}$ and $T(Q_m) \leq \eta_m$ for all $m \in \mathbb{N}$, it follows that $T(Q) = 0$.

This condition is weaker than Condition 4.1 and allows discontinuities of the mappings to define tests. The Hodges-Lehmann optimality under Condition 4.3 is obtained as follows.

Theorem 4.3 *Assume that Conditions 3.1 and 3.2 hold and the mapping $T : \mathcal{M} \rightarrow \bar{\mathbb{R}}$ for the test $\phi_n \equiv 1\{T(\hat{P}_n) > c_n\}$ with $c_n \searrow 0$ satisfies Conditions 4.2 and 4.3. Then the test ϕ_n satisfies*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log E_{P_1}[1 - \phi_n] = -K(\mathcal{P}, P_1),$$

for each $P_1 \in \mathcal{M} \setminus \mathcal{P}$ satisfying $0 < K(\mathcal{P}, P_1) < \infty$, i.e., ϕ_n is Hodges-Lehmann optimal at each $P_1 \in \mathcal{M} \setminus \mathcal{P}$ satisfying $0 < K(\mathcal{P}, P_1) < \infty$.

Theorem 4.3 does not use Condition 4.1 but rather much weaker Condition 4.3. The theorem uses Conditions 3.1 and 3.2 but these assumptions were also used to prove optimality of EL and GMM in Theorems 3.1 and 4.2. Theorem 4.3 requires $K(\mathcal{P}, P_1) < \infty$, i.e., the alternative distribution P_1 should be absolutely continuous with respect to some distribution in the null set \mathcal{P} . This requirement means that the bound in Theorem 2.1 is informative.

To characterize the class of mappings that satisfy Condition 4.3, we consider the mappings $T_D : \mathcal{M} \rightarrow [0, \infty]$ taking the form of

$$T_D(Q) \equiv \inf_{P \in \mathcal{P}(Q)} D(Q, P),$$

for some distance or divergence $D : \mathcal{M} \times \mathcal{M} \rightarrow [0, \infty]$ between measures P and Q , where $\mathcal{P}(Q) \equiv \{P \in \mathcal{P} : P \ll Q, Q \ll P\}$. The test based on T_D is defined as $\phi_{D,n} \equiv 1\{T_D(\hat{P}_n) > c_n\}$ with some sequence of positive numbers $\{c_n : n \in \mathbb{N}\}$ decreasing to zero. The following theorem provides a sufficient condition that guarantees Condition 4.3 for the mapping T_D and the Hodges-Lehmann optimality of the test $\phi_{D,n}$.

Theorem 4.4 *Assume that Conditions 3.1 and 3.2 hold and the mapping T_D satisfies*

- (i) $D(Q, P)$ is lower semicontinuous under the weak topology for $P \in \mathcal{P}$ at given $Q \in \mathcal{M}$,
- (ii) $D(Q, P)$ is lower semicontinuous under the weak topology for $(Q, P) \in \mathcal{M} \times \mathcal{P}$,
- (iii) $D(Q, P) = 0$ if and only if $Q = P$.

Then the mapping T_D satisfies Conditions 4.2 and 4.3, and the test $\phi_{D,n}$ is Hodges-Lehmann optimal for each $P_1 \in \mathcal{M} \setminus \mathcal{P}$ satisfying $0 < K(\mathcal{P}, P_1) < \infty$.

This theorem is useful to establish the Hodges-Lehmann optimality of some members of the GEL test $\phi_{GEL,n}$ in (2.6). For example, consider the Cressie and Read (1984) family of criterion functions

$$\rho_\alpha(v) = -(1 + \alpha v)^{(\alpha+1)/\alpha} / (\alpha + 1),$$

for $\alpha \in \mathbb{R}$. Note that the GEL test defined by $\rho_\alpha(v)$ covers several existing tests, such as empirical likelihood ($\alpha = -1$), Hellinger distance ($\alpha = -1/2$), exponential tilting ($\alpha = 0$), and the continuous updating GMM ($\alpha = 1$). Theorem 2.2 of Newey and Smith (2004) shows that the dual form of the GEL test with $\rho_\alpha(v)$ is written as $\phi_{\alpha,n} = 1\{T_\alpha(\hat{P}_n) > \chi_{q-k, 1-\alpha}^2/n\}$, where

$$T_\alpha(Q) \equiv \inf_{P \in \mathcal{P}(Q)} D_\alpha(Q, P),$$

and

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

The mapping $D_\alpha : \mathcal{M} \times \mathcal{M} \rightarrow [0, \infty]$ is a special case of the so-called f -divergence (see, e.g., Liese and Vajda (1987)). Since the f -divergence (and also D_α) satisfies the semicontinuity requirements in Theorem 4.4 (Liese and Vajda, 1987, Theorem 1.47), it follows that the GEL test $\phi_{\alpha,n}$ with $\rho_\alpha(v)$ is Hodges-Lehmann optimal. This result is summarized in the following corollary.

Corollary 4.1 *Under Conditions 3.1 and 3.2, the GEL test $\phi_{\alpha,n} = 1\{T_\alpha(\hat{P}_n) > \chi_{q-k,1-\alpha}^2/n\}$, which includes the EL, exponential tilting, and continuous updating GMM tests, is Hodges-Lehmann optimal for each $P_1 \in \mathcal{M} \setminus \mathcal{P}$ satisfying $0 < K(\mathcal{P}, P_1) < \infty$.*

5 Concluding Remarks

This paper studies the Hodges-Lehmann optimality for testing moment condition models, which focuses on the convergence rates of the type II error probabilities under fixed alternatives with fixing the limits of the type I error probabilities at some level. We derive the optimal convergence rate of the type II error probabilities and propose two sets of sufficient conditions that imply the Hodges-Lehmann optimality of a test. In particular, we show that the existing GMM and GEL over-identifying restriction tests are Hodges-Lehmann optimal under mild regularity conditions. There are several directions of future research. It is interesting to extend our Hodges-Lehmann optimality analysis to conditional moment restrictions, dependent data, and other decision problems, such as model selection and inference for partially identified econometric models. Also our analysis may provide a basis for other large deviation optimality analysis, such as Chernoff (1952) optimality analysis.

Appendices

In the Appendices, we use 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 probability measures P and Q on \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 . The Lévy metric is compatible with the weak topology. Also, we use Sanov's Theorem (see, e.g., Theorem 6.2.10 of Dembo and Zeitouni (1998)) to analyze the large deviation behavior of the empirical measure \hat{P}_n , i.e.,

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

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

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

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

Appendix A Proof of Theorems

A.1 Proof of Theorem 2.1

Pick any $P_1 \in \mathcal{M} \setminus \mathcal{P}$. 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 $\bar{P}_0 \in \mathcal{P}$ such that $K(\bar{P}_0, P_1) < K(\mathcal{P}, P_1) + \epsilon < \infty$. Thus, $\bar{P}_0 \ll P_1$, and there exists the Radon-Nykodym derivative $r(x) \equiv \frac{d\bar{P}_0}{dP_1}$. Since $\{x_i : i = 1, \dots, n\}$ is an i.i.d sample, the strong law of large numbers implies

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \log r(x_i) = E_{\bar{P}_0}[\log r(x)] = K(\bar{P}_0, P_1) < \infty, \quad \bar{P}_0 - a.s. \quad (\text{A-1})$$

Now define the event

$$E_n = \left\{ \prod_{i=1}^n r(x_i) < \exp(n[K(\bar{P}_0, P_1) + \epsilon]) \right\}.$$

Let P^n be the n -fold product measure of a probability measure P . Observe that

$$\begin{aligned}
E_{P_1}[1 - \phi_n] &= \int 1\{T(\hat{P}_n) \leq c_n\} dP_1^n \\
&\geq \int_{E_n} 1\{T(\hat{P}_n) \leq c_n\} dP_1^n \\
&\geq \exp(-n[K(\bar{P}_0, P_1) + \epsilon]) \int_{E_n} 1\{T(\hat{P}_n) \leq c_n\} \prod_{i=1}^n r(X_i) dP_1^n \\
&= \exp(-n[K(\bar{P}_0, P_1) + \epsilon]) \int_{E_n} 1\{T(\hat{P}_n) \leq c_n\} d\bar{P}_0^n \\
&\geq \exp(-n[K(\bar{P}_0, P_1) + \epsilon])(1 - \Pr\{T(\hat{P}_n) > c_n : \bar{P}_0\} - \Pr\{E_n^c : \bar{P}_0\}),
\end{aligned}$$

where the first equality follows from the definition of ϕ_n in (2.4), the first inequality follows from the set inclusion relation, the second inequality follows from the definition of E_n , the second equality follows from the change of measures, and the last inequality follows from the set inclusion relation. By (2.7) or (2.8) we have

$$\limsup_{n \rightarrow \infty} \Pr\{T(\hat{P}_n) > c_n : \bar{P}_0\} = \limsup_{n \rightarrow \infty} E_{\bar{P}_0}[\phi_n] \leq \alpha < 1,$$

and by (A-1), $\lim_{n \rightarrow \infty} \Pr\{E_n^c : \bar{P}_0\} = 0$, so that

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log E_{P_1}[1 - \phi_n] \geq -K(\bar{P}_0, P_1) - \epsilon > -K(\mathcal{P}, P_1) - 2\epsilon,$$

where the second inequality follows from the definition of \bar{P}_0 . Since ϵ is arbitrary, the conclusion is obtained.

A.2 Proof of Theorem 3.1

Pick any $P_1 \in \mathcal{M} \setminus \mathcal{P}$. It is sufficient to show that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log E_{P_1}[1 - \phi_{EL,n}] \leq -K(\mathcal{P}, P_1). \quad (\text{A-2})$$

Let

$$\Lambda_\eta \equiv \left\{ Q \in \mathcal{M} : \inf_{P \in \mathcal{P}(Q)} K(Q, P) \leq \eta \right\}.$$

Since $\chi_{q-k, 1-\alpha}^2 / (2n) \searrow 0$,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log E_{P_1}[1 - \phi_{EL,n}] \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log E_{P_1}[1\{\hat{P}_n \in \Lambda_\eta\}],$$

for any constant $\eta > 0$. Since Λ_η is not closed in the weak topology, we create a closed set $\tilde{\Lambda}_\eta$ under the weak topology such that $\Lambda_\eta \subseteq \tilde{\Lambda}_\eta$. Define the closed ball by the Lévy metric with center Q and radius η as

$$\bar{B}_\eta(Q) \equiv \{Q' \in \mathcal{M} : d_L(Q', Q) \leq \eta\}.$$

Similarly, the open Lévy ball is defined as $B_\eta(Q) \equiv \{Q' \in \mathcal{M} : d_L(Q', Q) < \eta\}$. Let $\{\mathcal{A}\}^\eta \equiv \cup_{Q \in \mathcal{A}} B_\eta(Q)$ be the η -blowup for a subset $\mathcal{A} \subseteq \mathcal{M}$ by the open Lévy ball. Define $\tilde{\Lambda}_\eta$ as

$$\tilde{\Lambda}_\eta \equiv \left(\left\{ Q \in \mathcal{M} : \inf_{Q' \in \tilde{B}_{2\eta}(Q)} \inf_{P \in \mathcal{P}} K(Q', P) > \eta \right\}^\eta \right)^c,$$

for $\eta \in [0, \infty)$. Since $K(Q, P) \geq 0$ for any $P, Q \in \mathcal{M}$ and $K(Q, P) = 0$ if and only if $Q = P$,

$$\tilde{\Lambda}_0 = \left(\left\{ Q \in \mathcal{M} : \inf_{P \in \mathcal{P}} K(Q, P) > 0 \right\} \right)^c = \left\{ Q \in \mathcal{M} : \inf_{P \in \mathcal{P}} K(Q, P) = 0 \right\} = \mathcal{P}.$$

Since \mathcal{M} is compact under the Lévy metric or the weak topology (see, Theorem D.8 of Dembo and Zeitouni (1998)) and the blowup $\left\{ Q \in \mathcal{M} : \inf_{Q' \in \tilde{B}_{2\eta}(Q)} \inf_{P \in \mathcal{P}} K(Q', P) > \eta \right\}^\eta$ is open, the set $\tilde{\Lambda}_\eta$ is closed in the weak topology. Suppose that

$$\Lambda_\eta \subseteq \tilde{\Lambda}_\eta \quad \text{for each } \eta > 0, \tag{A-3}$$

and

$$\text{for any } \epsilon > 0, \text{ there exists some } \eta > 0 \text{ such that } K(\tilde{\Lambda}_\eta, P_1) + \epsilon \geq K(\tilde{\Lambda}_0, P_1). \tag{A-4}$$

Then by Sanov's theorem, for any $\epsilon > 0$ there exists some $\eta > 0$ such that

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log E_{P_1}[1\{\hat{P}_n \in \Lambda_\eta\}] &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log E_{P_1}[1\{\hat{P}_n \in \tilde{\Lambda}_\eta\}] \\ &\leq -K(\tilde{\Lambda}_\eta, P_1) \\ &\leq -K(\tilde{\Lambda}_0, P_1) + \epsilon \\ &= -K(\mathcal{P}, P_1) + \epsilon. \end{aligned}$$

Since ϵ is arbitrary, we obtain (A-2). Therefore, it is sufficient to show the statements in (A-3) and (A-4).

To show (A-3), pick any $\eta > 0$. Then pick any

$$Q^* \in \tilde{\Lambda}_\eta^c = \left\{ Q \in \mathcal{M} : \inf_{Q' \in \tilde{B}_{2\eta}(Q)} \inf_{P \in \mathcal{P}} K(Q', P) > \eta \right\}^\eta.$$

It is sufficient for (A-3) to show that $Q^* \in \Lambda_\eta^c$. If Q^* satisfies $\inf_{Q' \in \tilde{B}_{2\eta}(Q^*)} \inf_{P \in \mathcal{P}} K(Q', P) > \eta$, then we have

$$\inf_{P \in \mathcal{P}(Q^*)} K(Q^*, P) \geq \inf_{P \in \mathcal{P}} K(Q^*, P) > \eta,$$

and $Q^* \in \Lambda_\eta^c$. Even if not (i.e., Q^* lies in the blowup region), there exists $R \in \mathcal{M}$ such that $d_L(R, Q^*) < \eta$ and $K(R', \mathcal{P}) > \eta$ for all $R' \in \tilde{B}_{2\eta}(R)$. Since $Q^* \in \tilde{B}_{2\eta}(R)$, it holds $K(Q^*, \mathcal{P}) > \eta$ and we have $Q^* \in \Lambda_\eta^c$. Thus, (A-3) holds true.

Now we show (A-4). By Lemma B.3,

$$\tilde{\Lambda}_{\eta'} \subseteq \tilde{\Lambda}_\eta \quad \text{for all } \eta > 0 \text{ and } \eta' \in [0, \eta/2). \tag{A-5}$$

Then there exists a positive decreasing sequence $\{\eta_m : m \in \mathbb{N}\}$ such that $\eta_m \searrow 0$ and $\{K(\tilde{\Lambda}_{\eta_m}, P_1) : m \in \mathbb{N}\}$ is a positive non-increasing sequence having the limit $\lim_{m \rightarrow \infty} K(\tilde{\Lambda}_{\eta_m}, P_1) \leq K(\tilde{\Lambda}_0, P_1)$. For (A-4), it is

sufficient to show that

$$\text{there exists a subsequence } \{\eta_{m_j} : j \in \mathbb{N}\} \text{ of } \{\eta_m : m \in \mathbb{N}\} \text{ such that } \lim_{j \rightarrow \infty} K(\tilde{\Lambda}_{\eta_{m_j}}, P_1) \geq K(\tilde{\Lambda}_0, P_1). \quad (\text{A-6})$$

Since $\tilde{\Lambda}_\eta$ is compact in the weak topology (because \mathcal{M} is compact) for each $\eta \geq 0$ and $K(Q, P_1)$ is lower semicontinuous under the weak topology in $Q \in \mathcal{M}$, there exists a sequence $\{Q_m : m \in \mathbb{N}\}$ on \mathcal{M} such that $Q_m \in \tilde{\Lambda}_{\eta_m}$ and

$$K(\tilde{\Lambda}_{\eta_m}, P_1) = K(Q_m, P_1), \quad (\text{A-7})$$

for each $m \in \mathbb{N}$. From $Q_m \in \tilde{\Lambda}_{\eta_m}$, Q_m must satisfy

$$\eta_m \geq \inf_{(Q', P) \in \bar{B}_{2\eta_m}(Q_m) \times \mathcal{P}} K(Q', P),$$

for each $m \in \mathbb{N}$. Since $\bar{B}_{2\eta_m}(Q_m) \times \mathcal{P}$ is compact in the product topology based on the weak topologies of $\bar{B}_{2\eta_m}(Q_m)$ and \mathcal{P} (by Tychonoff's theorem (see, e.g., p. 143 of Kelley (1955)) with the compactness of $\bar{B}_{2\eta_m}(Q_m)$ and \mathcal{P}), and $K(Q', P)$ is lower semicontinuous (jointly) in $(Q', P) \in \mathcal{M} \times \mathcal{M}$, there exists a sequence of the pair of measures $\{(Q'_m, P_m) : m \in \mathbb{N}\}$ such that $Q'_m \in \bar{B}_{2\eta_m}(Q_m)$, $P_m \in \mathcal{P}$, and

$$\eta_m \geq K(Q'_m, P_m),$$

for each $m \in \mathbb{N}$. Since $\{(Q'_m, P_m) : m \in \mathbb{N}\}$ is a sequence on the compact product space $\mathcal{M} \times \mathcal{P}$, there exists a subsequence $\{(Q'_{m_j}, P_{m_j}) : j \in \mathbb{N}\}$ which converges under the product topology to some pair $(\bar{Q}', \bar{P}) \in \mathcal{M} \times \mathcal{P}$ and satisfies

$$K(\bar{Q}', \bar{P}) \leq \liminf_{j \rightarrow \infty} K(Q'_{m_j}, P_{m_j}) \leq \lim_{j \rightarrow \infty} \eta_{m_j} = 0,$$

where the first inequality follows from the lower semicontinuity of $K(Q', P)$ in $(Q', P) \in \mathcal{M} \times \mathcal{M}$. Therefore,

$$Q'_{m_j} \Rightarrow \bar{Q}' = \bar{P} \in \mathcal{P}. \quad (\text{A-8})$$

We now consider the weak convergence of $\{Q_{m_j} : j \in \mathbb{N}\}$. Since $d_L(Q'_{m_j}, Q_{m_j}) \leq 2\eta_{m_j}$ (because $Q'_{m_j} \in \bar{B}_{2\eta_{m_j}}(Q_{m_j})$), the triangle inequality implies

$$\begin{aligned} d_L(Q_{m_j}, \bar{Q}') &\leq d_L(Q_{m_j}, Q'_{m_j}) + d_L(Q'_{m_j}, \bar{Q}') \\ &\leq 2\eta_{m_j} + d_L(Q'_{m_j}, \bar{Q}'), \end{aligned}$$

and thus (A-8) implies

$$Q_{m_j} \Rightarrow \bar{Q}' = \bar{P} \in \mathcal{P}.$$

Therefore, the lower semicontinuity of $K(Q, P_1)$ in $Q \in \mathcal{M}$ implies

$$\begin{aligned} \lim_{j \rightarrow \infty} K(\tilde{\Lambda}_{\eta_{m_j}}, P_1) &= \liminf_{j \rightarrow \infty} K(Q_{m_j}, P_1) \\ &\geq K(\bar{Q}', P_1) \\ &= K(\bar{P}, P_1) \\ &\geq K(\tilde{\Lambda}_0, P_1), \end{aligned}$$

and the statement in (A-6) is obtained. Since this implies (A-4), the conclusion is obtained.

A.3 Proof of Theorem 4.1

By Lemma B.2 the function $\kappa(\eta) \equiv K(\Omega_\eta, P_1)$ with $\Omega_\eta \equiv \{Q \in \mathcal{M} : T(Q) \leq \eta\}$ is non-increasing and right continuous so that for any $\epsilon > 0$, there exists $\delta > 0$ such that $\kappa(\eta) - \kappa(\eta + \delta) < \epsilon$, meaning that $-\kappa(\eta + \delta) < -\kappa(\eta) + \epsilon$. The result then follows by similar steps to those in equation (A-13).

A.4 Proof of Theorem 4.2

Condition 4.2 is trivially satisfied for the two tests so we focus on Condition 4.1. Consider first the CU test,

$$T_{CU}(Q) \equiv \inf_{\theta \in \Theta} T_{CU}(Q, \theta), \quad T_{CU}(Q, \theta) \equiv \mu(Q, \theta)' \Sigma(Q, \theta)^{-1} \mu(Q, \theta), \quad (\text{A-9})$$

where $\Sigma(Q, \theta) = E_Q[(m(x, \theta) - \mu(Q, \theta))(m(x, \theta) - \mu(Q, \theta))']$ and the statistic is by definition equal to infinity if $\Sigma(Q, \theta)$ is singular and $\|\mu(Q, \theta)\| \neq 0$ and equal to zero if $\Sigma(Q, \theta)$ is singular and $\|\mu(Q, \theta)\| = 0$. By Conditions 3.1 and 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 functions of Q and θ in $\mathcal{M} \times \Theta$. Let $\det(A)$ denote the determinant of a matrix A and consider a sequence $\{(Q_m, \theta_m) : m \in \mathbb{N}\}$ such that $Q_m \Rightarrow Q^* \in \mathcal{M}$ and $\theta_m \rightarrow \theta^* \in \Theta$. If $\det(\Sigma(Q_m, \theta_m)) > 0$ for all m large enough, then $T_{CU}(Q_m, \theta_m)$ is a continuous transformation of $\mu(Q, \theta)$ and $\Sigma(Q, \theta)$ and therefore continuous in (Q, θ) , which implies $T_{CU}(Q^*, \theta^*) = \lim_{m \rightarrow \infty} T_{CU}(Q_m, \theta_m)$. If, on the contrary, $\det(\Sigma(Q_m, \theta_m)) = 0$ for all m large enough, then $T_{CU}(Q^*, \theta^*) = T_{CU}(Q_m, \theta_m) = \infty$ for all m large enough. It follows that $T_{CU}(Q^*, \theta^*) = \lim_{m \rightarrow \infty} T_{CU}(Q_m, \theta_m)$ for any sequence $\{(Q_m, \theta_m) : m \in \mathbb{N}\}$ such that $Q_m \Rightarrow Q^* \in \mathcal{M}$ and $\theta_m \rightarrow \theta^* \in \Theta$. By compactness of Θ , there exists $\bar{\theta} \in \Theta$ such that $T_{CU}(Q) = T_{CU}(Q, \bar{\theta})$. Now consider a sequence $Q_m \Rightarrow Q^*$. It follows that there exists a sequence $\{\bar{\theta}_m : m \in \mathbb{N}\}$ such that $T_{CU}(Q_m) = T_{CU}(Q_m, \bar{\theta}_m)$ for each $m \in \mathbb{N}$ and by compactness of Θ there exists a subsequence $\{\bar{\theta}_{m_j} : j \in \mathbb{N}\}$ such that $\bar{\theta}_{m_j} \rightarrow \theta^*$ for some $\theta^* \in \Theta$. Then

$$T_{CU}(Q^*) \leq T_{CU}(Q^*, \theta^*) = \lim_{j \rightarrow \infty} T_{CU}(Q_{m_j}, \bar{\theta}_{m_j}), \quad (\text{A-10})$$

and therefore $T_{CU}(Q)$ is lower semicontinuous in the weak topology. Now consider the case of the 2-step GMM test. In this case $\Sigma(Q, \tilde{\theta}(Q))$ is also continuous in Q by continuity of $\tilde{\theta}(Q)$. It follows that $T_{GMM}(Q, \theta) \equiv \mu(Q, \theta)' \Sigma(Q, \tilde{\theta}(Q))^{-1} \mu(Q, \theta)$ is continuous in Q and θ and therefore $T_{GMM}(Q)$ is lower semicontinuous by the same arguments used above.

A.5 Proof of Theorem 4.3

Let \bar{A} be the closure of a set $A \subseteq \mathcal{M}$ with respect to the weak topology. The first step involves proving that $\bar{\Omega}_\eta \subseteq \bar{\Omega}_{\eta'}$ for $\eta \leq \eta'$ (recall that $\Omega_\eta \equiv \{Q \in \mathcal{M} : T(Q) \leq \eta\}$). Note that $\bar{\Omega}_\eta = \Omega_\eta \cup \partial^* \Omega_\eta$, where

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

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 $\{Q_k : k \in \mathbb{N}\} \subseteq \Omega_{\eta'}$ for such sequence. Since $Q \in \bar{\Omega}_{\eta'}$, we obtain $\bar{\Omega}_\eta \subseteq \bar{\Omega}_{\eta'}$.

The second step is to prove that $\kappa(\eta) \equiv K(\bar{\Omega}_\eta, P_1)$ is continuous from the right at $\eta = 0$. Pick any sequence of positive numbers $\{\eta_m : m \in \mathbb{N}\}$ with $\eta_m \searrow 0$. Note that by Condition 4.2 and $0 < K(\mathcal{P}, P_1) < \infty$, we have $\kappa(0) < \infty$. Since $\bar{\Omega}_\eta \subseteq \bar{\Omega}_{\eta'}$ for $\eta \leq \eta'$, the function $\kappa(\cdot)$ is non-increasing in η , the limit $\lim_{n \rightarrow \infty} \kappa(\eta_m)$ exists and it holds $\lim_{n \rightarrow \infty} \kappa(\eta_m) \leq \kappa(0) < \infty$. 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, e.g., 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 $K(Q_m, P_1) = \kappa(\eta_m) < \infty$. Since the sequence $\{Q_m : m \in \mathbb{N}\}$ is on the compact set \mathcal{M} , there exists a subsequence $\{Q_{m_j} : j \in \mathbb{N}\}$ such that $Q_{m_j} \Rightarrow Q^*$ for some $Q^* \in \mathcal{M}$. Since $K(Q, P_1)$ is lower semicontinuous in Q ,

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

There are two possibilities. 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 in such case Condition 4.3 implies $T(Q^*) = 0$ meaning that $Q^* \in \Omega_0$. As a result,

$$\kappa(0) \geq \lim_{k \rightarrow \infty} \kappa(\eta_{m_k}) = \liminf_{k \rightarrow \infty} K(Q_{m_k}, P_1) \geq K(Q^*, P_1) \geq \kappa(0), \quad (\text{A-11})$$

and it follows that $\lim_{k \rightarrow \infty} \kappa(\eta_{m_k}) = \kappa(0)$. 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.4 that $T(Q^*) = 0$ and (A-11) follows.

Finally, since the function $\kappa(\cdot)$ is non-increasing and right continuous at $\eta = 0$, we have that for any $\epsilon > 0$ there exists $\delta > 0$ such that $\kappa(0) - \kappa(\delta) < \epsilon$, meaning that

$$-\kappa(\delta) < -\kappa(0) + \epsilon. \quad (\text{A-12})$$

Now, pick an arbitrary $\epsilon > 0$ and observe that

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log E_{P_1} [1\{T(\hat{P}_n) \leq c_n\}] &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log E_{P_1} [1\{\hat{P}_n \in \Omega_\delta : P_1\}] \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log E_{P_1} [1\{\hat{P}_n \in \bar{\Omega}_\delta : P_1\}] \\ &\leq -\kappa(\delta) \\ &< -\kappa(0) + \epsilon, \end{aligned} \quad (\text{A-13})$$

where the first inequality follows from $c_n \searrow 0$ and δ being positive, the second inequality follows from the set inclusion relation, the third inequality follows by Sanov's Theorem based on the fact that $\bar{\Omega}_\delta$ is closed in the weak topology, and the last inequality follows from (A-12). Since ϵ is arbitrary, we have $\limsup_{n \rightarrow \infty} \frac{1}{n} \log E_{P_1} [1\{T(\hat{P}_n) \leq c_n\}] \leq -K(\bar{\Omega}_0, P_1)$. The result follows by Condition 4.2, which implies $K(\bar{\Omega}_0, P_1) = K(\mathcal{P}, P_1)$.

A.6 Proof of Theorem 4.4

We first check Condition 4.2. If $Q \in \mathcal{P}$ then it follows that $Q \in \mathcal{P}(Q)$ and so $0 \leq T_D(Q) \leq D(Q, Q) = 0$, i.e., $Q \in \Omega_0^D \equiv \{Q \in \mathcal{M} : T_D(Q) = 0\}$. On the other hand, since $D(Q, P) = 0$ if and only if $Q = P$, if $Q \in \Omega_0^D$

it follows that $Q \in \mathcal{P}(Q) \subseteq \mathcal{P}$. Therefore, Condition 4.2 is verified. Now we check Condition 4.3. Pick any sequence $\{Q_m : m \in \mathbb{N}\}$ of measures in \mathcal{M} such that $Q_m \Rightarrow Q \in \mathcal{M}$ and $T_D(Q_m) \leq \eta_m$ for all $m \in \mathbb{N}$. Since the set \mathcal{P} is compact in the weak topology by Lemma B.1 and $D(Q, P)$ is lower semicontinuous in the weak topology for $P \in \mathcal{P}$, there exists a sequence $\bar{P}_m \in \mathcal{P}$ such that

$$D(Q_m, \bar{P}_m) = \inf_{P \in \mathcal{P}} D(Q_m, P),$$

for each $m \in \mathbb{N}$. Since $\{\bar{P}_m : m \in \mathbb{N}\}$ is a sequence on the compact set \mathcal{P} , there exists a subsequence $\{\bar{P}_{m_j} : j \in \mathbb{N}\}$ such that $\bar{P}_{m_j} \Rightarrow \bar{P} \in \mathcal{P}$. Since $D(Q, P)$ is lower semicontinuous for $(Q, P) \in \mathcal{M} \times \mathcal{P}$, it follows that

$$0 = \liminf_{j \rightarrow \infty} \eta_{m_j} \geq \liminf_{j \rightarrow \infty} T_D(Q_{m_j}) \geq \liminf_{j \rightarrow \infty} D(Q_{m_j}, \bar{P}_{m_j}) \geq D(Q, \bar{P}),$$

i.e., $D(Q, \bar{P}) = 0$. Since $D(Q, \bar{P}) = 0$ if and only if $Q = \bar{P}$, it follows that $\bar{P} \in \mathcal{P}(Q)$ and $T_D(Q) = \inf_{P \in \mathcal{P}(Q)} D(Q, P) = 0$, which completes the proof.

Appendix B Additional Lemmas

Lemma B.1 *Under Conditions 3.1 and 3.2, the set \mathcal{P} defined in (2.2) is closed in the weak topology.*

PROOF. Take a sequence $\{P_m : m \in \mathbb{N}\}$ in \mathcal{P} such that $P_m \Rightarrow P$ for some $P \in \mathcal{M}$. Since $P_m \in \mathcal{P}$ for all $m \in \mathbb{N}$, there exists a sequence $\{\theta_m : m \in \mathbb{N}\}$ such that

$$\int_{\mathcal{X}} m(x, \theta_m) dP_m = 0. \tag{B-1}$$

Since Θ is compact, there exists a subsequence $\{\theta_{m_j} : j \in \mathbb{N}\}$ such that $\theta_{m_j} \rightarrow \theta$ for some $\theta \in \Theta$. Therefore, it is sufficient to show that $E_P[m(x, \theta)] = 0$. To prove this, note that

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

where the first inequality follows from (B-1) and the triangle inequality, the second inequality follows by the Portmanteau Lemma (see van der Vaart, 1998, Lemma 2.2) as $m(\cdot, \theta)$ is bounded and continuous for all $\theta \in \Theta$, and the last equality follows by the uniform continuity of $m(x, \theta)$. ■

Lemma B.2 *Assume the mapping $T : \mathcal{M} \rightarrow \bar{\mathbb{R}}$ is lower semicontinuous in the weak topology on the set $\{Q \in \mathcal{M} : K(Q, P_1) < \infty\}$ for a given $P_1 \in \mathcal{M}$. Then the function $\kappa(\eta) \equiv K(\Omega_\eta, P_1)$ with $\Omega_\eta \equiv \{Q \in \mathcal{M} : T(Q) \leq \eta\}$ is continuous in η from the right.*

PROOF. First note that if $\eta_2 > \eta_1$ then $\Omega_{\eta_1} \subseteq \Omega_{\eta_2}$ 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} decreasing to some η such that

$\kappa(\eta) < \infty$. Since this function is non-increasing, $\kappa(\eta_m) \leq \kappa(\eta) < \infty$ for all $m \in \mathbb{N}$ and $\lim_{m \rightarrow \infty} \kappa(\eta_m)$ exists. By the lower semicontinuity of T , the set Ω_η is closed in the weak topology meaning that there exists $Q \in \Omega_\eta$ such that $K(Q, P) = K(\Omega_\eta, P)$. Therefore, for each $m \in \mathbb{N}$ there exists $Q_m \in \Omega_{\eta_m}$ such that $K(Q_m, P) = \kappa(\eta_m) \leq \kappa(\eta)$. Since $K(\cdot, P)$ has compact level sets for each $P \in \mathcal{M}$ (see Dupuis and Ellis, 1997, Lemma 1.4.3), there exists a subsequence $\{Q_{m_j}\}$ and $Q \in \mathcal{M}$ such that $Q_{m_j} \Rightarrow Q$ and $K(Q, P) \leq \liminf_{j \rightarrow \infty} K(Q_{m_j}, P) < \infty$ (by the lower semicontinuity of $K(Q, P)$ in $\mathcal{M} \times \mathcal{M}$). Since $T(Q_{m_j}) \leq \eta_{m_j}$ for each $j \in \mathbb{N}$ and T is lower semicontinuous, it follows that $T(Q) \leq \liminf_{j \rightarrow \infty} T(Q_{m_j}) \leq \liminf_{j \rightarrow \infty} \eta_{m_j} \leq \eta$. Therefore, $Q \in \Omega_\eta$ and $\kappa(\eta) \geq \lim_{j \rightarrow \infty} \kappa(\eta_{m_j}) = \liminf_{j \rightarrow \infty} K(Q_{m_j}, P) \geq K(Q, P) \geq \kappa(\eta)$. Thus, it follows that $\lim_{j \rightarrow \infty} \kappa(\eta_{m_j}) = \kappa(\eta)$. Note that the result also holds for $\eta \in \mathbb{R}$ such that $\kappa(\eta) = \infty$. To see this suppose by contradiction that $\kappa(\eta) = \infty$ but $\lim_{m \rightarrow \infty} \kappa(\eta_m)$ exists for a sequence $\{\eta_m\}$ with $\eta_m \searrow 0$. By the previous argument there exists $Q \in \Omega_\eta$ such that $K(Q, P) < \infty$, which violates $\kappa(\eta) = \infty$. ■

Lemma B.3 *The set*

$$\tilde{\Lambda}_\eta \equiv \left(\left\{ Q \in \mathcal{M} : \inf_{Q' \in \bar{B}_{2\eta}(Q)} \inf_{P \in \mathcal{P}} K(Q', P) > \eta \right\}^\eta \right)^c,$$

satisfies

$$\tilde{\Lambda}_{\eta'} \subseteq \tilde{\Lambda}_\eta \quad \text{for all } \eta > 0 \text{ and } \eta' \in [0, \eta/2).$$

PROOF. Pick any $\eta > 0$ and $\eta' \in [0, \eta/2)$. Then pick any $Q \in \tilde{\Lambda}_\eta^c$ where

$$\tilde{\Lambda}_\eta^c \equiv \left\{ Q \in \mathcal{M} : \inf_{Q' \in \bar{B}_{2\eta}(Q)} \inf_{P \in \mathcal{P}} K(Q', P) > \eta \right\}^\eta.$$

It is sufficient to show that

$$\inf_{P \in \mathcal{P}} K(Q', P) > \eta' \quad \text{for each } Q' \in \bar{B}_{2\eta'}(Q). \quad (\text{B-2})$$

Since $Q \in \tilde{\Lambda}_\eta^c$, there exists $R \in \mathcal{M}$ such that $d_L(Q, R) < \eta$ and $\inf_{P \in \mathcal{P}} K(R', P) > \eta > \eta'$ for all $R' \in \bar{B}_{2\eta}(R)$. Thus, it is sufficient for (B-2) to show that

$$Q' \in \bar{B}_{2\eta'}(R) \quad \text{for each } Q' \in \bar{B}_{2\eta'}(Q). \quad (\text{B-3})$$

Now, the triangle inequality implies that for each $Q' \in \bar{B}_{2\eta'}(Q)$,

$$\begin{aligned} d_L(Q', R) &\leq d_L(Q', Q) + d_L(Q, R) \\ &< 2\eta' + \eta \\ &< 2\eta. \end{aligned}$$

Therefore, (B-3) holds true and the conclusion is obtained. ■

Lemma B.4 *Under Condition 4.3, if $Q_m \in \partial^* \Omega_{\eta_m}$ for all $m \in \mathbb{N}$ with a sequence $\eta_m \searrow 0$, where*

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

and $Q_m \Rightarrow Q^* \in \mathcal{M}$, then $T(Q^*) = 0$.

PROOF. Pick any sequence of measures $\{Q_m : m \in \mathbb{N}\}$ 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 $Q^* \in \mathcal{M}$. Suppose that

$$\text{there exists } \{Q'_m : m \in \mathbb{N}\} \text{ such that } Q'_m \in \Omega_{\eta_m} \text{ for all } m \in \mathbb{N} \text{ and } Q'_m \Rightarrow Q^*. \quad (\text{B-4})$$

Then Condition 4.3 implies $T(Q^*) = 0$. So it is sufficient to show (B-4). To this end, 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

$$d_L(Q'_m, Q_m) \leq \epsilon/2 \quad (\text{B-5})$$

for all $m \in \mathbb{N}$. Also from $Q_m \Rightarrow Q^*$, there exists $M \in \mathbb{N}$ such that for all $m \geq M$,

$$d_L(Q_m, Q^*) \leq \epsilon/2. \quad (\text{B-6})$$

From the triangle inequality combined with (B-5) and (B-6), 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 (B-4) holds true completing the proof. ■

References

- BAHADUR, R. (1960): “On the Asymptotic Efficiency of Test and Estimators,” *Sankhya*, 22, 229–252.
- BAHADUR, R., AND L. J. SAVAGE (1956): “The Nonexistence of Certain Statistical Procedures in Nonparametric Problems,” *Annals of Mathematical Statistics*, 25, 1115–1122.
- BORWEIN, J. M., AND A. S. LEWIS (1993): “Partially-finite Programming in L_1 and the Existence of Maximum Entropy Estimates,” *SIAM Journal of Optimization*, 3, 248–267.
- CANAY, I. A. (2008): “EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity,” manuscript, Northwestern University.
- CHERNOFF, H. (1952): “A Measure of Asymptotic Efficiency for Test of a Hypothesis Based on the Sum of Observations,” *The Annals of Mathematical Statistics*, 23, 497–507.
- CRESSIE, N., AND R. C. READ (1984): “Multinomial goodness-of-fit tests,” *Journal of the Royal Statistical Society*.
- DEMBO, A., AND O. ZEITOUNI (1998): *Large Deviations Techniques and Applications*. Second Edition, Springer: New York.
- DUPUIS, P., AND R. S. ELLIS (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(4), 1029–1054.
- HANSEN, L. P., J. HEATON, AND A. YARON (1996): “Finite-Sample Properties of Some Alternative GMM Estimators,” *Journal of Business and Economic Statistics*, 14(3), 262–280.
- HODGES, J. L., AND E. L. LEHMANN (1956): “The Efficiency of Some Nonparametric Competitors of the t-test,” *Annals of Mathematical Statistics*, 27, 324–335.
- HOEFFDING, W. (1965): “Asymptotically Optimal Test for Multinomial Distributions (with Discussion),” *Annals of Mathematical Statistics*, 36, 369–408.
- IMBENS, G. (1997): “One-Step Estimators for Over-Identified Generalized Method of Moments Models,” *Econometrica*, 64(3), 359–383.
- IMBENS, G., R. H. SPADY, AND P. JOHNSON (1998): “Information Theoretic Approaches to Inference in Moment Condition Models,” *Econometrica*, 66(2), 333–357.

- KALLENBERG, W., AND S. KOUROUKLIS (1992): “Hodges-Lehmann Optimality of Tests,” *Statistics & Probability Letters*, 14.
- KELLEY, J. L. (1955): *General Topology*. Springer, New York.
- KITAMURA, Y. (2001): “Asymptotic Optimality of Empirical Likelihood for Testing Moment Restrictions,” *Econometrica*, 69(6), 1661–1672.
- KITAMURA, Y., AND T. OTSU (2005): “Minimax Estimation and Testing for Moment Condition Models via Large Deviations,” manuscript, Yale University.
- KITAMURA, Y., T. OTSU, AND K. EVDOKIMOV (2009): “Robustness, Infinitesimal Neighborhoods, and Moment Restrictions,” manuscript, Yale University.
- KITAMURA, Y., A. SANTOS, AND A. SHAIKH (2009): “On the Asymptotic Optimality of Empirical Likelihood for Testing Moment Restrictions,” manuscript, University of Chicago.
- KITAMURA, Y., AND M. STUTZER (1997): “An Information-Theoretic Alternative to Generalized Method of Moments Estimation,” *Econometrica*, 65(4), 861–874.
- LIESE, F., AND I. VAJDA (1987): *Convex Statistical Distances*. Teubner Leipzig.
- NEWBY, W. K., AND R. J. SMITH (2004): “Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators,” *Econometrica*, 72(1), 219–255.
- NIKITIN, Y. (1995): *Asymptotic Efficiency of Nonparametric Tests*. Cambridge University Press.
- OTSU, T. (2009): “On Bahadur Efficiency of Empirical Likelihood,” manuscript, Yale University.
- OWEN, A. (1988): “Empirical Likelihood Ratio Confidence Intervals for a Single Functional,” *Biometrika*, 75, 237–249.
- (1990): “Empirical Likelihood for Confidence Regions,” *Annals of Statistics*, 18(1), 90–120.
- (1991): “Empirical Likelihood for Linear Models,” *Annals of Statistics*, 19(4), 1725–1747.
- QIN, B. J., AND J. LAWLESS (1994): “Empirical Likelihood and General Estimating Equations,” *Annals of Statistics*, 22(1), 300–325.
- SCHENNACH, S. M. (2007): “Point Estimation with Exponentially Tilted Empirical Likelihood,” *Annals of Statistics*, 32(2), 634–672.
- SERFLING, R. J. (1980): *Approximation Theorems of Mathematical Statistics*. John Wiley, New York.

SMITH, R. J. (1997): “Alternative Semi-Parametric Likelihood Approaches to Generalized Method of Moments Estimation,” *Econometrica*, 50(4), 1029–1054.

TUSNÁDY, G. (1977): “On Asymptotic Optimality of Tests,” *The Annals of Statistics*, 5.

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