On the Asymptotic Optimality of Empirical Likelihood for Testing Moment Restrictions

Yuichi Kitamura
Department of Economics
Yale University
yuichi.kitamura@yale.edu

Andres Santos
Department of Economics
University of California - San Diego
a2santos@ucsd.edu

Azeem M. Shaikh
Department of Economics
University of Chicago
amshaikh@uchicago.edu

August 18, 2009

Abstract

In this paper we make two contributions. First, we show by example that empirical likelihood and other commonly used tests for parametric moment restrictions, including the GMM-based J-test of Hansen (1982), are unable to control the rate at which the probability of a Type I error tends to zero. From this it follows that, for the optimality claim for empirical likelihood in Kitamura (2001) to hold, additional assumptions and qualifications need to be introduced. The example also reveals that empirical and parametric likelihood may have non-negligible differences for the types of properties we consider, even in models in which they are first-order asymptotically equivalent. Second, under stronger assumptions than those in Kitamura (2001), we establish the following optimality result: (i) empirical likelihood controls the rate at which the probability of a Type I error tends to zero and (ii) among all procedures for which the probability of a Type I error tends to zero at least as fast, empirical likelihood maximizes the rate at which probability of a Type II error tends to zero for "most" alternatives. This result further implies that empirical likelihood maximizes the rate at which probability of a Type II error tends to zero for all alternatives among a class of tests that satisfy a weaker criterion for their Type I error probabilities.

KEYWORDS: Empirical Likelihood, Large Deviations, Hoeffding Optimality, Moment Restrictions

ACKNOWLEDGMENTS: The research of the first author has been supported by the National Science Foundation grants SES-0551271 and SES-0851759. The research of the third author has been supported by the National Science Foundation grant DMS-0820310.

1 Introduction

The purpose of this paper is two-fold. First, we show by example that empirical likelihood and other commonly used tests for parametric moment restrictions, including the GMM-based J-test proposed in Hansen (1982), are unable to control the rate at which the probability of a Type I error tends to zero. This fact has not been noted in previous research as this difficulty is not present in fully parametric models. The example shows in particular that, for the optimality claim for empirical likelihood in Kitamura (2001) to hold, additional assumptions and qualifications need to be introduced. It also reveals that empirical and parametric likelihood may have non-negligible differences for the types of properties we consider, even in models in which they are first-order asymptotically equivalent. This fact has also been unnoticed in previous research on empirical likelihood. Second, under stronger assumptions than those in Kitamura (2001), we establish a more qualified optimality result for empirical likelihood. This result further implies that empirical likelihood maximizes the rate at which probability of a Type II error tends to zero for all alternatives among a class of tests that satisfy a weaker criterion for their Type I error probabilities.

More concretely, let $P \in \mathbf{P}$ on $\mathcal{X} \subseteq \mathbf{R}^d$ and $g : \mathbf{R}^d \times \Theta \to \mathbf{R}^m$, where $\Theta \subseteq \mathbf{R}^r$, be given. Consider the null hypothesis

$$H_0 : P \in \mathbf{P}_0 \ , \tag{1}$$

where

$$\mathbf{P}_0 = \{P \in \mathbf{P} : E_P[g(X,\theta)] = 0 \text{ for some } \theta \in \Theta\} \ . \tag{2}$$

The alternative hypothesis is understood to be

$$H_1 : P \in \mathbf{P}_1 = \mathbf{P} \setminus \mathbf{P}_0 \ .$$

The problem is to test (1) based on $X_i, i = 1, \ldots n$, an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$. When $m > r$ this is typically referred to as a test of over-identifying restrictions. Hansen (1982) introduced a method based on the generalized method of moments as a means of testing (1). Subsequently, several alternatives to this test have also been proposed, including a continuously updated version of the generalized method of moments (Hansen et al. (1996)) and the empirical likelihood ratio test (Owen (1988) and Qin and Lawless (1994)) along with its variants (Kitamura and Stutzer (1997) and Imbens et al. (1998)). These alternatives are part of the generalized empirical likelihood framework studied in Newey and Smith (2004).

Following Kitamura (2001), we consider an asymptotic framework in which the probability of a Type I error tends to zero as the sample size, n, tends to infinity. This framework was first developed by Hoeffding (1965), who used it to study the asymptotic properties of tests of certain hypotheses about the parameters of a multinomial distribution. For this problem, he showed that among all tests for which the Type I error tends to zero at a suitable rate, the likelihood ratio test maximizes the rate at which the Type II error tends to zero for "most" alternatives. Such a property is sometimes called Hoeffding optimality. The generalization of the results of Hoeffding (1965) to tests of (1) is nontrivial because, in contrast to his setting, the null hypothesis is not required to be parametric and P is not assumed to have finite support.

As in Kitamura (2001), we restrict attention to nonrandomized tests of (1) based on the empirical distribution of the observations, $\hat{P}_n$. Any such test can be identified with a pair of sets of distributions, $(\Omega_{1,n}, \Omega_{2,n})$, such that the test accepts when $\hat{P}_n \in \Omega_{1,n}$ and rejects when $\hat{P}_n \in \Omega_{2,n}$. The empirical likelihood ratio test rejects when a certain (fixed) function of $\hat{P}_n$ exceeds a pre-specified value, η. For each $\eta > 0$, denote by $(\Lambda_1(\eta), \Lambda_2(\eta))$ the corresponding acceptance and rejection regions. Under weak assumptions on $\mathbf{P}$ and g, but stronger than the ones posited by Kitamura (2001), we prove that for all η sufficiently small

$$\sup_{P \in \mathbf{P}_0} \limsup_{n \to \infty} \frac{1}{n} \log P^n\{\hat{P}_n \in \Lambda_2(\eta)\} \leq -\eta \ . \tag{3}$$

where P^n is the n-fold product measure $\bigotimes_{i=1}^n P$. In this sense, empirical likelihood controls the rate at which the Type I error tends to zero. Moreover, we prove that any test $(\Omega_{1,n}, \Omega_{2,n})$ satisfying

$$\sup_{P \in \mathbf{P}_0} \limsup_{n \to \infty} \frac{1}{n} \log P^n\{\hat{P}_n \in \Omega_{2,n}^{\delta}\} \leq -\eta \tag{4}$$

for some $\delta > 0$, we have that

$$\limsup_{n \to \infty} \frac{1}{n} \log P^n\{\hat{P}_n \in \Omega_{1,n}\} \geq \limsup_{n \to \infty} \frac{1}{n} \log P^n\{\hat{P}_n \in \Lambda_1(\eta)\} \tag{5}$$

for "most" $P \in \mathbf{P}_1$. Here, the notation A^δ denotes the (open) δ-"blowup" of a set $A \subset \mathbf{M}$ with respect to Prokhorov-Lévy metric. More formally, $A^\delta = \cup_{P \in A} B(P, \delta)$, where $B(P, \delta)$ denotes an open ball with center P and radius δ with respect to the Prokhorov-Lévy metric. This is sometimes referred to as δ-"smoothing"; see Dembo and Zeitouni (1998) for further discussion of this technique. With this caveat in mind, this is the sense in which this result shows that empirical likelihood is more powerful at "most" alternatives than any other test that also controls the rate at which the Type I error tends to zero.

Part (a) of Theorem 2 in Kitamura (2001) claims, under very weak assumptions on $\mathbf{P}$ and g, that empirical likelihood controls the rate at which the Type I error tends to zero in the sense that (3) holds for *any* $\eta > 0$. We provide two examples that demonstrate that this claim is false without stronger assumptions and further qualifications. More specifically, we show that given the assumptions in Kitamura (2001), (3) is not satisfied for any $\eta > 0$. Importantly, our examples illustrate that if $\mathbf{P}$ is "too rich," then empirical likelihood, as well as the discussed alternative tests, will fail to satisfy (4) for any $\eta > 0$, which motivates the restrictions we ultimately place upon $\mathbf{P}$. Our examples also reveal that the the sort of asymptotic equivalence of empirical likelihood with parametric likelihood underlying many of their shared large-sample properties is insufficient for the types of properties we consider.

The remainder of the paper is organized as follows. In Section 2, we describe empirical likelihood more precisely and formulate it in terms of the empirical distribution, as is required for our analysis. In Section 3, we describe our examples that show empirical likelihood, the GMM-based J-test, and other commonly available tests fail to control their size in terms of large deviations. We then provide in Section 4 a precise statement of the optimality of empirical likelihood for testing (1) as described above. Proofs of all results are collected in the Appendix.

2 The Empirical Likelihood Ratio Test

Qin and Lawless (1994) propose testing (1) by rejecting for large values of the empirical likelihood ratio

$$\frac{L_n^{\text{constrained}}}{L_n^{\text{unconstrained}}} ,$$

where

$$L_n^{\text{constrained}} = \sup_{\theta \in \Theta} \sup \{ \prod_{1 \leq i \leq n} P\{X_i\} : P \in \mathbf{M}, P \ll \hat{P}_n, E_P[g(X_i, \theta)] = 0 \} \tag{6}$$

and $L_n^{\text{unconstrained}}$ is simply equal to n^{-n}. Here, $\mathbf{M}$ denotes the set of probability distributions on $\mathcal{X}$ (with the Borel $\sigma-$algebra) and the supremum over the empty set is understood to be zero. It is well known that such a test also has an information-theoretic interpretation. Let

$$\mathbf{P}(Q) = \bigcup_{\theta \in \Theta} \{ P \in \mathbf{M} : P \ll Q, Q \ll P, E_P[g(X_i, \theta)] = 0 \} .$$

The above test is equivalent to a test that rejects for large values of

$$\inf_{P \in \mathbf{P}(\hat{P}_n)} I(\hat{P}_n | P) , \tag{7}$$

where $I(Q|P)$ is the Kullback-Leibler divergence of P from Q defined as

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

Here, the infimum over the empty set is understood to be infinity. Note the importance of the restriction $P \ll \hat{P}_n$ for all $\mathbf{P}(\hat{P}_n)$. Indeed, without this requirement, it is possible to show that (7) is zero whenever $\{g(x, \theta) : x \in \mathcal{X}\} = \mathbf{R}^m$ for some $\theta \in \Theta$. In this case, one would never reject the null hypothesis.

In this language, empirical likelihood can be identified with a partition of $\mathbf{M}$ into the pair sets of distributions $(\Lambda_1(\eta), \Lambda_2(\eta))$, where

$$\Lambda_1(\eta) = \{ Q \in \mathbf{M} : \inf_{P \in \mathbf{P}(Q)} I(Q|P) < \eta \} \tag{8}$$

for some pre-specified $\eta > 0$ and

$$\Lambda_2(\eta) = \mathbf{M} \setminus \Lambda_1(\eta) . \tag{9}$$

Empirical likelihood rejects (1) whenever $\hat{P}_n \in \Lambda_2(\eta)$ and fails to reject (1) if $\hat{P}_n \in \Lambda_1(\eta)$.

3 Two Examples

Part (a) of Theorem 2 in Kitamura (2001) claims that the test based on (8) and (9) satisfies (3) for any $\eta > 0$ provided that

$$P\{\sup_{\theta \in \Theta} ||g(X,\theta)|| = \infty\} = 0 \text{ for all } P \in \mathbf{P} \tag{10}$$

$$g(x,\theta) \text{ is continuous at every } \theta \in \Theta \text{ for each } x \in \mathbf{R}^d \ . \tag{11}$$

The following example, however, shows that this claim does not hold without additional restrictions.

Example 3.1. Suppose $d = 1$, $m = 1$ and $g(x,\theta) = x$ for all $\theta \in \Theta$. Let $\mathbf{P}$ be any set of probability distributions satisfying (10) and (11) and containing

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

where P_c is the distribution that puts mass $1 - c$ on c and mass c on $-(1 - c)$. Then

$$\sup_{P \in \mathbf{P}_0} \limsup_{n \to \infty} \frac{1}{n} \log P^n\{\hat{P}_n \in \Lambda_2(\eta)\} \geq \sup_{P \in \mathbf{C}_0} \limsup_{n \to \infty} \frac{1}{n} \log P^n\{\hat{P}_n \in \Lambda_2(\eta)\} = 0 \tag{12}$$

for any $\eta > 0$. To see that (12) holds, let $\eta > 0$ be given and note that

$$P_c^n\{X_i = c \text{ for all } 1 \leq i \leq n\} = (1 - c)^n \ .$$

Moreover, when $X_i = c$ for all $1 \leq i \leq n$, we have that (7) is infinity, so $\hat{P}_n \in \Lambda_2(\eta)$. Thus,

$$(1 - c)^n \leq P_c^n\{\hat{P}_n \in \Lambda_2(\eta)\} \ ,$$

from which (12) follows. We conclude that (3) cannot be satisfied by a test based on (8) and (9) for any $\eta > 0$ without further assumptions on $\mathbf{P}$. ■

The above example suggests that if $\mathbf{P}$ is "too rich" then empirical likelihood can not satisfy (3) for any value of $\eta > 0$. It is important to note that this shortcoming is not unique to empirical likelihood and is shared by many commonly used tests. In particular, Example 3.1 applies to the test that rejects for large values of the absolute value of the t-statistic. Equivalently, it applies to the GMM-based J-test proposed in Hansen (1982). Hence, these tests are also unable to control size as in (3) if $\mathbf{P}$ is "too rich." The simplicity of Example 3.1 is illustrative but potentially misleading, as it suggests the problem is caused by measures that have "too little" mass on one side of zero. Example 3.2 shows this is actually not a necessary condition and also helps us uncover what drives the result in Example 3.1.

Example 3.2. As in the previous example, assume $d = 1$, $m = 1$ and $g(x,\theta) = x$ for all $\theta \in \Theta$. Let $\mathbf{P}$ be any set of probability distributions satisfying (10) and (11) and containing

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

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

$$R_{K,c}\{X_i = \frac{-2c}{(1-c)(K-1)}\} = \frac{1}{2} \qquad R_{K,c}\{X_i = \frac{2Kc}{(1-c)(K-1)}\} = \frac{1}{2} \ .$$

Then, empirical likelihood is unable to control size on $\mathbf{P}$, as (12) holds with $\mathbf{K}_0$ in place of $\mathbf{C}_0$. To see this, note that by direct calculation it is straightforward to obtain that

$$\inf_{P \in \mathbf{P}(R_{Kc})} I(R_{Kc}|P) = \frac{1}{2}\log(\frac{1+K}{2K}) + \frac{1}{2}\log(\frac{1+K}{2}) \tag{13}$$

which is greater than η for K sufficiently large. Denote such a choice by K_η. From (13), it is possible to show that

$$\liminf_{n\to\infty} \frac{1}{n} \log R^n_{K_\eta,c}\{\hat{P}_n \in \Lambda_2(\eta)\} = 0 \tag{14}$$

Define $A_n = \{X_i \neq -1 : \text{ for all } 1 \leq i \leq n\}$ and note that

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} \log & P^n_{K_\eta,c}\{\hat{P}_n \in \Lambda_2(\eta)\} \\
&\geq \liminf_{n\to\infty} \frac{1}{n} \log P^n_{K_\eta,c}\{\hat{P}_n \in \Lambda_2(\eta)|A_n\} + \liminf_{n\to\infty} \frac{1}{n} \log P^n_{K_\eta,c}\{A_n\} \\
&= \liminf_{n\to\infty} \frac{1}{n} \log R^n_{K_\eta,c}\{\hat{P}_n \in \Lambda_2(\eta)\} + \liminf_{n\to\infty} \frac{1}{n} \log P^n_{K_\eta,c}\{A_n\} = \log(1-c) \ . \qquad (15)
\end{aligned}$$

Letting c tend to zero, we see from (15) that (3) cannot be satisfied by a test based on (8) and (9) for any $\eta > 0$ without further assumptions on $\mathbf{P}$. ■

Note that the problem revealed in this example also applies to other commonly used tests such as the GMM-based J-test. Both Examples 3.1 and 3.2 rely on a sequence of distributions for which the rate at which the probability of a Type I error tends to zero itself tends to zero. These sequences are linked by

$$\lim_{c\to 0} P_c = \lim_{c\to 0} P_{K,c} = D_0 \ , \tag{16}$$

where D_0 is the degenerate distribution at 0 and the limit should be interpreted in the weak topology. The measure D_0 is unique in that it is the only measure satisfying the null hypothesis whose support has zero dimension. In more generality, the logic of these examples reveals that empirical likelihood fail to satisfy (3) for any $\eta > 0$ in the neighborhood of measures that satisfy the null hypothesis but whose support is contained in lower dimensional subspaces. We show in the next section that removing such neighborhood from the null space is sufficient to restore size control as in (3) for some $\eta > 0$.

Remark 3.1. Empirical and parametric likelihood often share desirable large-sample properties. To illustrate this phenomenon in a simple setting, fix $c_0 > 0$ and consider the binomial family

$$\mathbf{P}_{c_0} = \{P \in \mathbf{M} : P \ll P_{c_0}, \ P_{c_0} \ll P\} \ ,$$

where P_{c_0} is defined as in Example 3.1. Under the maintained assumption that $P \in \mathbf{P}_{c_0}$, the likelihood ratio test statistic for $H_0 : E_P[X] = 0$ versus $H_1 : E_P[X] \neq 0$ is then

$$\ell_{\text{par}} = I(\hat{P}_n | P_{c_0}) \ . \tag{17}$$

Similarly, once our sample includes both c_0 and $1 - c_0$ (and therefore $\mathbf{P}(\hat{P}_n) = \{P_{c_0}\}$), the empirical likelihood ratio statistic is simply

$$\ell_{\text{el}} = I(\hat{P}_n | P_{c_0}) \ . \tag{18}$$

Therefore from (17) and (18) it follows that $\ell_{\text{par}} = \ell_{\text{el}}$ with probability approaching one under any fixed $P \in \mathbf{P}_{c_0}$. This equivalence can be shown to hold in greater generality; see Newey and Smith (2004). For this reason, empirical likelihood inherits many of the desirable large-sample properties of parametric likelihood in this model. However, this sort of equivalence is too weak for the types of properties we consider. Specifically, while Example 3.1 reveals empirical likelihood can not satisfy (3) for $\eta > -\log(1 - c_0)$ on $\mathbf{P}_0 = \{P_{c_0}\}$, parametric likelihood is able to do so for any $\eta > 0$. See Theorem 3.5.4 in Dembo and Zeitouni (1998). ■

Remark 3.2. It is instructive to contrast empirical likelihood with parametric likelihood further. To this end, let $\mathbf{P}_{\text{par}} = \{P_\xi : \xi \in \Xi\}$, where $\Xi \subset \mathbf{R}^d$. Consider testing the null hypothesis that $P \in \mathbf{P}_{\text{par}}$ versus the alternative that $P \in \mathbf{M} \setminus \mathbf{P}_{\text{par}}$. It is possible to show that the likelihood ratio test rejects the null hypothesis for large values of

$$\inf_{P \in \mathbf{P}_{\text{par}}} I(\hat{P}_n | P) \ . \tag{19}$$

Heuristically, (19) is the distance between the empirical distribution $\hat{P}_n$ and the model $\mathbf{P}_{\text{par}}$. This representation of the likelihood ratio test is used by Zeitouni and Gutman (1991) to establish its Hoeffding optimality in this setting. On the other hand, in our analysis the model is given by $\mathbf{P}_0$ defined by (2), but the empirical likelihood ratio test does *not* reject for large values of

$$\inf_{P \in \mathbf{P}_0} I(\hat{P}_n | P), \tag{20}$$

which is the direct analogue of (19), but instead for large values of (7), contrary to equation (5) in Kitamura (2001). This modification is needed because $\mathbf{P}_0$ is "too large." In fact, the infimum in (20) may even be equal to zero. It is therefore reasonable to expect the need for additional conditions in establishing the Hoeffding optimality of empirical likelihood than those employed in the study of parametric likelihood. ■

4 The Main Result

The proofs of large deviation optimality results rely on large deviation principles for the empirical measure $\hat{P}_n$. These principles are often called Sanov's Theorem, of which several versions exist. We now state the result that will suffice for our purposes.

Theorem 4.1. *Let $\mathbf{M}(\Sigma)$ denote the space of probability measures on a Polish space Σ equipped with the weak topology. Suppose $P \in \mathbf{M}(\Sigma)$. Then,*

$$\limsup_{n\to\infty} \frac{1}{n} \log P^n\{\hat{P}_n \in G\} \leq - \inf_{Q\in G} I(Q|P)$$

for all closed sets $G \subset \mathbf{M}(\Sigma)$, and

$$\liminf_{n\to\infty} \frac{1}{n} \log P^n\{\hat{P}_n \in H\} \geq - \inf_{Q\in H} I(Q|P)$$

for all open sets $H \subset \mathbf{M}(\Sigma)$.

See Chapter 6.2 in Dembo and Zeitouni (1998) for different proofs of this result as well as refinements to stronger topologies.

Before stating the assumptions we require, we need to introduce some additional notation. Recall that $\mathbf{M}$ is the set of probability measures on $\mathcal{X}$ (with the Borel $\sigma-$algebra) and define for each $P \in \mathbf{M}$,

$$\Theta_0(P) = \{\theta \in \Theta : E_P[g(X_i, \theta)] = 0\} \tag{21}$$

We denote the set of distributions that agree with the hypothesized moment restriction by

$$\mathbf{M}_0 = \{P \in \mathbf{M} : \Theta_0(P) \neq \emptyset\} \tag{22}$$

As shown by our example, empirical likelihood is unable to satisfy (3) for $\mathbf{P}_0 = \mathbf{M}_0$. For this reason we impose the following assumptions on the model and $\mathbf{P}_0 \subset \mathbf{M}_0$,

Assumption 4.1. $\mathbf{P}_0 \subset \mathbf{M}_0$ *is closed in the weak topology.*

Assumption 4.2. *For each $P \in \mathbf{P}_0$, $\Theta_0(P)$ is a singleton denoted $\theta_0(P)$.*

Assumption 4.3. *For some $\epsilon > 0$,*

$$\sup_{P\in\mathbf{P}_0} \sup_{\gamma\neq 0} P\{\gamma' g(X_i, \theta_0(P)) \geq 0\} \leq 1 - \epsilon \ .$$

Assumption 4.4. *$\mathcal{X}$ and Θ are compact subsets of $\mathbf{R}^d$ and $\mathbf{R}^r$, respectively.*

Assumption 4.5. *$g : \mathcal{X} \times \Theta \to \mathbf{R}^m$ is continuous in both of its arguments.*

Assumption 4.1 is employed in showing $\mathbf{P}_0$ is "well separated" from the rejection region (see (31) below). It is left as a high level assumption, but we note closed sets in the weak topology are easily constructed by imposing moment restrictions on bounded continuous functions. Assumption 4.2 is employed to show $\theta_0(P)$ is continuous in $P \in \mathbf{P}_0$ under the weak topology. Continuity of $\theta_0(P)$ can in turn be employed to verify $\mathbf{P}_0$ and $\Lambda_2(\eta)$ are "well separated." Since we are typically interested in cases where $m > r$, we feel that Assumption 4.2 is not particularly restrictive. It may be possible to weaken it at the expense of a more complicated argument. Assumption 4.3 is made precisely to avoid Example 3.1. Assumption 4.4 implies that $\mathbf{M}$ is compact in the weak topology, a crucial point in showing that $\mathbf{P}_0$ and $\Lambda_2(\eta)$ are "well separated." Assumption 4.5 is straightforward.

Remark 4.1. Examples 3.1 and 3.2 illustrate that $\mathbf{P}_0$ must not contain neighborhoods of those $P \in \mathbf{M}_0$ whose supports are included in lower-dimensional subspaces. We denote these distributions

$$\mathbf{D}_0 = \{P \in \mathbf{M}_0 : \exists \theta \in \Theta_0(P) \text{ with } s(P, \theta) < m\} \tag{23}$$

where $s(P, \theta)$ denotes the dimension of the convex hull of the support of $g(X_i, \theta)$ under P, i.e.

$$s(P, \theta) = \dim(\text{co}(\text{supp}_P(g(X_i, \theta)))) \ . \tag{24}$$

The requirements imposed on $\mathbf{P}_0$ ensure that there exists a $\delta > 0$ such that $\mathbf{P}_0 \cap \mathbf{D}_0^\delta = \emptyset$. If, in addition, $\Theta_0(P)$ is a singleton for every $P \in \mathbf{M}_0 \setminus \mathbf{D}_0$, then for every $\delta > 0$ there exists a $\mathbf{P}_0$ with $\mathbf{M}_0 \setminus \mathbf{D}_0^\delta \subseteq \mathbf{P}_0$ and $\mathbf{P}_0$ satisfying Assumptions 4.1-4.3 (see Lemma 5.9 in the Appendix). Given the implications of Examples 3.1 and 3.2, the restrictions on $\mathbf{P}_0$ are therefore quite weak. ■

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

Theorem 4.2. *Let $X_i, i = 1, \ldots, n$ be an i.i.d. sequence of random variables with distribution $P \in \mathbf{M}$. Let $(\Lambda_1(\eta), \Lambda_2(\eta))$ be defined by (8) and (9). Suppose Assumptions 4.1 - 4.5 hold. Then, the following statements follow:*

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

$$\sup_{P \in \mathbf{P}_0} \limsup_{n \to \infty} \frac{1}{n} \log P^n\{\hat{P}_n \in \Lambda_2(\eta)\} \leq -\eta \ .$$

(b) If a test $(\Omega_{1,n}, \Omega_{2,n})$ satisfies

$$\sup_{P \in \mathbf{P}_0} \limsup_{n \to \infty} \frac{1}{n} \log P^n\{\hat{P}_n \in \Omega_{2,n}^\delta\} \leq -\eta \tag{25}$$

for some $\delta > 0$, then

$$\limsup_{n \to \infty} \frac{1}{n} \log Q^n\{\hat{P}_n \in \Omega_{1,n}\} \geq \limsup_{n \to \infty} \frac{1}{n} \log Q^n\{\hat{P}_n \in \Lambda_1(\eta)\}$$

for any Q satisfying

$$\inf_{P \in \mathbf{M}_0 \setminus \mathbf{P}_0} I(Q|P) \geq \eta \ . \tag{26}$$

PROOF: See Appendix. ■

Theorem 4.2 establishes the desired optimality property of empirical likelihood. When using empirical likelihood, the probability of a Type I error tends to zero at a rate that is uniform on $\mathbf{P}_0$.

Furthermore, for any distribution Q satisfying (26), the probability of a Type II error vanishes at a rate at least as fast as that of any procedure based on the empirical distribution $\hat{P}_n$ satisfying the requirement (25). We emphasize the non-local nature of the optimality property, in that it holds for all distributions Q satisfying (26). Condition (26) demands that Q be sufficiently "far" from the subset of $\mathbf{M}_0$ over which we do not demand control on the rate at which the probability of a Type I error tends to zero. While in (26) "far" is defined in terms of entropy, the following corollary shows that "far" may also be interpreted in terms of the Total Variation metric or the Prokhorov-Lévy metric.

Corollary 4.1. *Under Assumptions 4.1 - 4.5, Theorem 4.2 holds if condition* (26) *is replaced by*

$$\inf_{P \in \mathbf{M}_0 \setminus \mathbf{P}_0} d(Q, P) \geq \sqrt{\frac{\eta}{2}} \ ,$$

where $d(Q, P)$ may be either the Total Variation or Prokhorov-Lévy metric.

PROOF: See Appendix. ■

Remark 4.2. It is worthwhile to point out that any distribution Q satisfying

$$\inf_{P \in \mathbf{M}_0} I(Q|P) \geq \eta \tag{27}$$

also satisfies (26). Hence, part (b) of Theorem 4.2 applies to all distributions that are sufficiently "far" from the null hypothesis. The requirement (26) is weaker than (27) in the sense that it only requires the distribution Q to be "far" from distributions in $\mathbf{M}_0$ that are not in $\mathbf{P}_0$. In this sense, part (b) of Theorem 4.2 applies to alternatives that are "close" to the null hypothesis as well. See Figure 1 for a useful illustration of this feature of Theorem 4.2. ■

Remark 4.3. Note that we only smooth (in the sense of δ-"smoothing") alternative tests in Theorem 4.2. In contrast, much of the related literature smoothes both the original and alternative tests. See, for example, Dembo and Zeitouni (1998). As noted by Kitamura (2001), if one restricts attention to alternative tests that are "regular" in the sense that

$$\lim_{\delta \searrow 0} \sup_{P \in \mathbf{P}} \limsup_{n \to \infty} \frac{1}{n} P^n\{\hat{P}_n \in \Omega_{2,n}^{\delta}\} = \sup_{P \in \mathbf{P}} \limsup_{n \to \infty} \frac{1}{n} P^n\{\hat{P}_n \in \Omega_{2,n}\} \ ,$$

then one may avoid the use of δ-"smoothing" altogether. This condition has been used by Zeitouni and Gutman (1991), who also provide a sufficient condition for it. ■

The principal challenge in deriving our optimality result consists in showing that the empirical likelihood test satisfies (3) for η sufficiently small. Our strategy for establishing this result can be described in three steps:

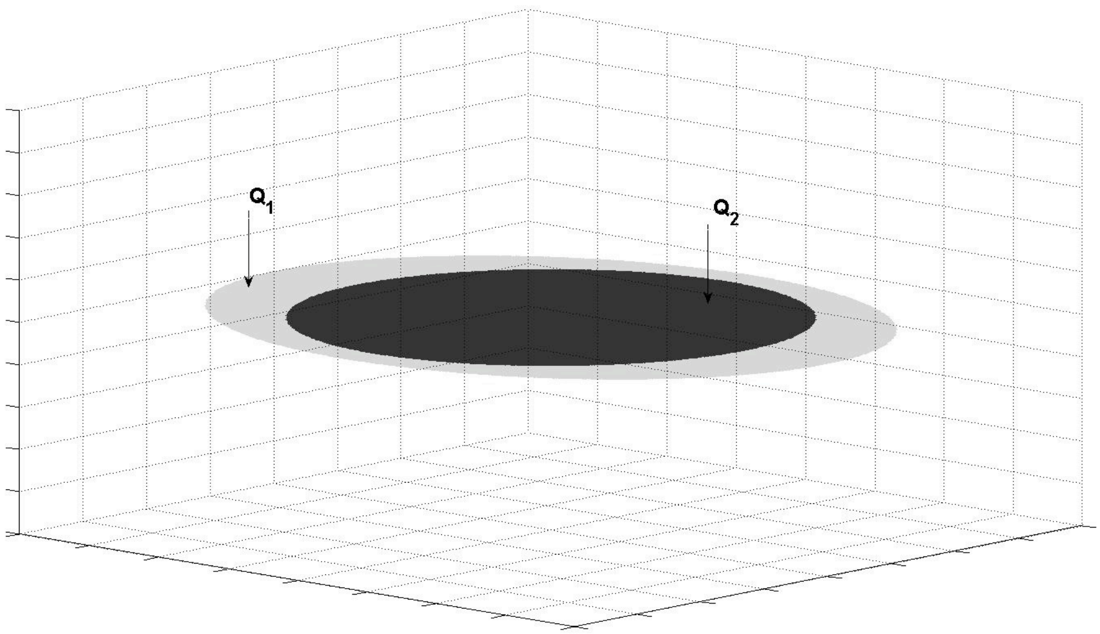

Figure 1: The larger and smaller (two-dimensional) ellipses represents $\mathbf{M}_0$ and $\mathbf{P}_0$, respectively (both are subsets of $\mathbf{M}$). Note that $\mathbf{M}_0$ has no interior relative to $\mathbf{M}$. Therefore, all points of $\mathbf{P}_0$ are on the boundary of $\mathbf{M}_0$. The distribution Q_1 is within $\eta > 0$ of $\mathbf{M}_0$ and $\mathbf{M}_0 \setminus \mathbf{P}_0$. In contrast, Q_2 is within $\eta > 0$ of $\mathbf{M}_0$, but not of $\mathbf{M}_0 \setminus \mathbf{P}_0$. Part (b) of Theorem 4.2 applies to Q_2, but not Q_1.

Step 1: Show $\Lambda_2(\eta) \subseteq \ddot{\Lambda}_2(\eta)$, where $\ddot{\Lambda}_2(\eta)$ is defined as follows. Let

$$\ddot{\mathbf{P}}(Q) = \bigcup_{\theta \in \Theta} \{P \in \mathbf{M} : P \ll Q, Q \ll P, s(Q,\theta) = m, E_P[g(X_i,\theta)] = 0\} \ , \tag{28}$$

The set $\ddot{\Lambda}_2(\eta)$ is then defined as

$$\ddot{\Lambda}_2(\eta) = \{Q \in \mathbf{M} : \inf_{P \in \ddot{\mathbf{P}}(Q)} I(Q|P) \geq \eta\} \tag{29}$$

Step 2: Show $\ddot{\Lambda}_2(\eta)$ is closed in the weak topology and employ Theorem 4.1 to establish

$$\limsup_{n\to\infty} \frac{1}{n} \log P^n\{\hat{P}_n \in \Lambda_2(\eta)\} \leq \limsup_{n\to\infty} \frac{1}{n} \log P^n\{\hat{P}_n \in \ddot{\Lambda}_2(\eta)\} \leq -\inf_{Q \in \ddot{\Lambda}_2(\eta)} I(Q|P) \tag{30}$$

Step 3: Establish that $\ddot{\Lambda}_2(\eta)$ and $\mathbf{P}_0$ are "well separated" in the sense that

$$\inf_{P \in \mathbf{P}_0} \inf_{Q \in \ddot{\Lambda}_2(\eta)} I(Q|P) \geq \eta \tag{31}$$

for $\eta > 0$ sufficiently small. Result (3) immediately follows from (30) and (31).

Remark 4.4. Note that $\ddot{\Lambda}_2(\eta)$ differs from $\Lambda_2(\eta)$ only through the difference between $\ddot{\mathbf{P}}(Q)$ and $\mathbf{P}(Q)$. In defining $\ddot{\mathbf{P}}(Q)$ in (28), we imposed the additional constraint $s(Q,\theta) = m$, which is not present in the definition of $\mathbf{P}(Q)$. This modification ensures that $\ddot{\Lambda}_2(\eta)$ is closed in the weak topology, as shown in the Appendix. A simple example establishes $\Lambda_2(\eta)$ is not closed. Let $\mathcal{X} = [-1, 1]$ and $g(x, \theta) = x$ for all θ. Further define D_0 to be the measure with $D_0\{X_i = 0\} = 1$ and D_n to be the measure satisfying

$$D_n\{X_i = 0\} = \frac{n-1}{n} \ , \ D_n\{X_i = 1\} = \frac{1}{n} \ .$$

Clearly, D_n converges to D_0 in the weak topology and $D_n \in \Lambda_2(\eta)$ for all n, but $D_0 \in \Lambda_1(\eta)$. See also Zeitouni and Gutman (1991). ■

As Examples 3.1 and 3.2 show, commonly used tests for (1) fail to control uniformly the rate at which the probability of a Type I error tends to zero in neighborhoods of $\mathbf{D}_0$. As a way of comparing among such procedures, it is interesting to examine optimality when we require

$$\sup_{P \in \mathbf{P}_0} \limsup_{n \to \infty} \frac{1}{n} \log P\{\hat{P}_n \in (\Omega_{2,n} \setminus \mathbf{D}_0^\epsilon)^\delta\} \leq -\eta \tag{32}$$

instead of (25). Requirement (32) should not be interpreted as "size" control, but rather as a benchmark for tests that fail to satisfy (25) for any $\eta > 0$ on $\mathbf{M}_0 \cap \mathbf{D}_0^\epsilon$. Clearly, given the weaker criterion for a Type I error probability in (32), any optimal test must have $\mathbf{D}_0^\epsilon \subseteq \Omega_{2,n}$. For this reason, we define:

$$\tilde{\Lambda}_1(\eta) = \Lambda_1(\eta) \setminus \mathbf{D}_0^\epsilon \qquad\qquad \tilde{\Lambda}_2(\eta) = \Lambda_2(\eta) \cup \mathbf{D}_0^\epsilon \tag{33}$$

where ϵ is an arbitrary positive constant and the dependence on ϵ is omitted in the notation. Note that the tests $(\Lambda_1(\eta), \Lambda_2(\eta))$ and $(\tilde{\Lambda}_1(\eta), \tilde{\Lambda}_2(\eta))$ may differ only on the event $\hat{P}_n \in \mathbf{D}_0^\epsilon$. We can use Theorem 4.2 to show the optimal test in this framework is given by the modified empirical likelihood test $(\tilde{\Lambda}_1(\eta), \tilde{\Lambda}_2(\eta))$.

Corollary 4.2. *Let $\mathbf{P} = \mathbf{M}$ and suppose Assumptions 4.4 and 4.5 hold. Suppose further that $\Theta_0(P)$ is a singleton for every $P \in \mathbf{M}_0 \setminus \mathbf{D}_0$, where $\mathbf{D}_0$ is defined in (23). Then, the following statements hold:*

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

$$\sup_{P \in \mathbf{P}_0} \limsup_{n \to \infty} \frac{1}{n} \log P^n\{\hat{P}_n \in \tilde{\Lambda}_2(\eta) \setminus \mathbf{D}_0^\epsilon\} \leq -\eta \ .$$

(b) If a test $(\Omega_{1,n}, \Omega_{2,n})$ satisfies

$$\sup_{P \in \mathbf{P}_0} \limsup_{n \to \infty} \frac{1}{n} \log P^n\{\hat{P}_n \in (\Omega_{2,n} \setminus \mathbf{D}_0^\epsilon)^\delta\} \leq -\eta \tag{34}$$

for some $\delta > 0$, then

$$\limsup_{n \to \infty} \frac{1}{n} \log Q^n\{\hat{P}_n \in \Omega_{1,n}\} \geq \limsup_{n \to \infty} \frac{1}{n} \log Q^n\{\hat{P}_n \in \tilde{\Lambda}_1(\eta)\} \tag{35}$$

for every probability measure Q.

Proof: See Appendix. ■

As mentioned earlier, (32) differs from (25) only in how the former treats distributions that are too close to distributions in $\mathbf{D}_0$. Remarkably, as a result of this rather simple modification, it is possible to remove the assumptions on $\mathbf{P}_0$ entirely. Moreover, in contrast to Theorem 4.2, (35) holds without qualifications on Q. This result may therefore provide some guidance when choosing among tests that have difficulty controlling the rate at which the Type I error tends to zero in neighborhoods of $\mathbf{D}_0$, such as tests based on (generalized) empirical likelihood or the GMM-based J-test.

5 Appendix

Lemma 5.1. *Let $\{B_1, \ldots, B_{d+1}\}$ be a collection of closed balls in $\mathbf{R}^d$ such that for any collection of points $\{g_1, \ldots, g_{d+1}\}$ with $g_i \in B_i$ for each $1 \le i \le d+1$,*

$$0 \in int(co(\{g_1, \ldots, g_{d+1}\})) \tag{36}$$

(relative to the topology on $\mathbf{R}^d$). Then, there exists $\epsilon > 0$ such that for all $0 \neq \gamma \in \mathbf{R}^d$ there exists $j = j(\gamma) \in \{1, \ldots, d+1\}$ such that $\gamma' g < 0$ and $|\gamma' g| \ge |g||\gamma|\epsilon$ for all $g \in B_j$.

PROOF: Let $\mathbf{B}(\gamma)$ be the maximal subset of $\{B_1, \ldots, B_{d+1}\}$ such that for all $B \in \mathbf{B}(\gamma)$ we have that $\gamma' g < 0$ for all $g \in B$. Note that the desired claim will follow if we can show (i) $\mathbf{B}(\gamma)$ is nonempty for any $0 \neq \gamma \in \mathbf{R}^d$ and (ii)

$$\inf_{0 \neq \gamma \in \mathbf{R}^d} \epsilon(\gamma) > 0 ,$$

where

$$\epsilon(\gamma) = \max_{B \in \mathbf{B}(\gamma)} \inf_{g \in B} \frac{|\gamma' g|}{|g||\gamma|} .$$

To establish (i), consider the hyperplane $H_\gamma = \{g \in \mathbf{R}^d : \gamma' g = 0\}$ and note that if $\gamma \neq 0$, then H_γ must strongly separate at least two balls $B_i, B_k \in \{B_1, \ldots, B_{d+1}\}$ with $i \neq k$. Otherwise, for either $\bar{\gamma} = \gamma$ or $\bar{\gamma} = -\gamma$, there exists a collection of points $\{g_1, \ldots, g_{d+1}\}$ with $g_i \in B_i$ and $\bar{\gamma}' g_i \ge 0$ for each $1 \le i \le d+1$, which contradicts (36). Therefore, since H_γ strongly separates at least two balls $B_i, B_k \in \{B_1, \ldots, B_{d+1}\}$ with $i \neq k$, it follows that there exists a $j = j(\gamma)$ such that $\gamma' g_j < 0$ for all $g \in B_j$.

To establish (ii), note that we may assume without loss of generality that $|\gamma| = 1$ and suppose by way of contradiction that there exists a sequence γ_n such that $\epsilon(\gamma_n) \to 0$. Since $|\gamma_n| = 1$, we have that there exists a subsequence γ_{n_k} such that $\gamma_{n_k} \to \gamma^*$ and $|\gamma^*| = 1$. Moreover, since $\mathbf{B}(\gamma^*) \subseteq \mathbf{B}(\gamma_{n_k})$ for all n_k sufficiently large, it follows that for such n_k

$$\epsilon(\gamma_{n_k}) \ge \max_{B \in \mathbf{B}(\gamma^*)} \inf_{g \in B} \frac{|\gamma_{n_k}' g|}{|g|} . \tag{37}$$

Next, note that

$$\inf_{g \in B_i} |g| > 0 \tag{38}$$

for $1 \le i \le d+1$. To see this, note that if (38) fails, there exists $1 \le i^* \le d+1$ such that $0 \in B_{i^*}$ since each B_i is closed. In this case, any collection of points $\{g_1, \ldots, g_{d+1}\}$ with $g_i \in B_i$ for $1 \le i \le d+1$ and $g_{i^*} = 0$ will not satisfy (36). It thus follows that $|\gamma_{n_k}' g|/|g| \to |\gamma^{*\prime} g|/|g|$ uniformly over $g \in B$ for each $B \in \mathbf{B}(\gamma^*)$. The righthand side of (37) therefore tends to $\epsilon(\gamma^*)$. But, since each $B \in \mathbf{B}(\gamma^*)$ is compact and there are only finitely many such B, $\epsilon(\gamma^*) > 0$. Hence, $\epsilon(\gamma_n) \not\to 0$, from which the desired claim follows. ■

Lemma 5.2. *If Assumptions 4.4 and 4.5 hold, then $\ddot{\Lambda}_2(\eta)$ is closed under the weak topology for any $\eta > 0$.*

PROOF: Let Q_n be a sequence such that $Q_n \to Q$ and $Q_n \in \ddot{\Lambda}_2(\eta)$ for all n. We wish to show that $Q \in \ddot{\Lambda}_2(\eta)$. Note that if $\ddot{\mathbf{P}}(Q) = \emptyset$, then

$$\inf_{P \in \ddot{\mathbf{P}}(Q)} I(Q|P) = +\infty ,$$

so $Q \in \ddot{\Lambda}_2(\eta)$. We may therefore assume further that $\ddot{\mathbf{P}}(Q) \neq \emptyset$.

Now suppose by way of contradiction that $Q \notin \ddot{\Lambda}_2(\eta)$. Define the set,

$$\ddot{\mathbf{P}}(Q,\theta) = \{P \in \mathbf{M} : P \ll Q, Q \ll P, s(Q,\theta) = m, E_P[g(X_i,\theta)] = 0\} \ , \tag{39}$$

and note that $\ddot{\mathbf{P}}(Q) = \bigcup_{\theta \in \Theta} \ddot{\mathbf{P}}(Q,\theta)$. Further define the set:

$$\Theta(Q) = \{\theta \in \Theta : \ddot{\mathbf{P}}(Q,\theta) \neq \emptyset\} \tag{40}$$

Since $\ddot{\mathbf{P}}(Q) \neq \emptyset$, it follows that $\Theta(Q) \neq \emptyset$ and therefore the Primal Constraint Qualification of Theorem 3.4 of Borwein and Lewis (1993) is satisfied for all $\theta \in \Theta(Q)$. Hence,

$$\inf_{P \in \ddot{\mathbf{P}}(Q)} I(Q|P) = \inf_{\theta \in \Theta(Q)} \max_{\gamma \in \mathbf{R}^m} \int \log(1 + \gamma' g(x,\theta)) dQ \ . \tag{41}$$

It follows that there exists $\theta^* \in \Theta(Q)$ such that

$$\max_{\gamma \in \mathbf{R}^m} \int \log(1 + \gamma' g(x,\theta^*)) dQ < \eta \ . \tag{42}$$

Further notice that since $\ddot{\mathbf{P}}(Q,\theta^*) \neq \emptyset$, by virtue of $\theta^* \in \Theta(Q)$, it follows that

$$s(Q,\theta^*) = m \ . \tag{43}$$

Next, we argue that

$$0 \in \text{int}(co(\text{supp}_Q(g(X_i,\theta^*)))) \tag{44}$$

(relative to the topology on $\mathbf{R}^m$). If this were not the case, then there exists a $0 \neq \gamma \in \mathbf{R}^m$ such that $\gamma' g(x,\theta^*) \geq 0$ for all $x \in \text{supp}(Q)$. Moreover, it must be the case that $\lambda' g(X_i,\theta^*) > 0$ with positive probability under Q, for otherwise $\text{supp}_Q(g(X_i,\theta^*))$ will be contained in a $m-1$ dimensional subspace of $\mathbf{R}^m$, which contradicts (43). For such γ we have for scalars α,

$$\lim_{\alpha \to \infty} \int \log(1 + \alpha\gamma' g(x,\theta^*)) dQ = \infty \ ,$$

which contradicts (42), so (44) is thus established.

We now show $\ddot{\mathbf{P}}(Q_n,\theta^*) \neq \emptyset$ for n sufficiently large. It follows from (44) that there exists a collection of points $\{g_1,\ldots,g_{s(Q,\theta^*)+1}\}$ in $\text{supp}_Q(g(X_i,\theta^*))$ such that

$$0 \in \text{int}(co(\{g_1,\ldots,g_{s(Q,\theta^*)+1}\}) \tag{45}$$

(relative to the topology on $\mathbf{R}^m$). For $1 \leq i \leq s(Q,\theta^*)+1$, let B_i be an open neighborhood of g_i so small that any collection of points $\{\tilde{g}_1,\ldots,\tilde{g}_{s(Q,\theta^*)+1}\}$ with $\tilde{g}_i \in \bar{B}_i$ for $1 \leq i \leq s(Q,\theta^*)+1$ will also satisfy (45) with $\tilde{g}_i$ in place of g_i. For $1 \leq i \leq s(Q,\theta^*)+1$, let

$$B_i^{-1} = \{x \in \mathcal{X} : g(x,\theta^*) \in B_i\} \ . \tag{46}$$

Since each B_i is open and $g(x,\theta^*)$ is continuous, each B_i^{-1} is also open. Moreover, since each B_i is an open neighborhood of a point in the support of $g(X_i,\theta^*)$ under Q,

$$Q\{X_i \in B_i^{-1}\} = Q\{g(X_i,\theta^*) \in B_i\} > 0 \ . \tag{47}$$

By the Pormanteau Lemma, we therefore have that for all n sufficiently large

$$Q_n\{X_i \in B_i^{-1}\} > 0 \tag{48}$$

for all $1 \leq i \leq s(Q,\theta^*) + 1$. Thus, for n sufficiently large (44) holds with Q_n in place of Q. It follows that $\ddot{\mathbf{P}}(Q_n,\theta^*) \neq \emptyset$ for n sufficiently large. Hence, the Primal Constraint Qualification of Theorem 3.4 of Borwein and Lewis (1993) is satisfied for such values of n, from which it follows that

$$\inf_{P \in \ddot{\mathbf{P}}(Q_n,\theta^*)} I(Q|P) = \max_{\gamma \in \mathbf{R}^m} \int \log(1 + \gamma' g(x,\theta^*)) dQ_n \ . \tag{49}$$

Let

$$\gamma_n^* \in \arg\max_{\gamma \in \mathbf{R}^m} \int \log(1 + \gamma' g(x,\theta^*)) dQ_n \ . \tag{50}$$

We now argue that the γ_n^* are uniformly bounded. If this were not the case, then for each $M > 0$ there would exist a subsequence $\gamma_{n_k}^*$ for which $|\gamma_{n_k}^*| > M$ for all k. By Lemma 5.1, there is an $\epsilon > 0$ and $j(\gamma_{n_k}^*) \in \{1, \ldots, s(Q,\theta^*)+1\}$ such that

$$\gamma_{n_k}^{*\prime} g < 0 \text{ and } |\gamma_{n_k}^{*\prime} g| \geq |g||\gamma_{n_k}^*|\epsilon \tag{51}$$

for all $g \in B_{j(\gamma_{n_k}^*)}$. There exists a further subsequence $\gamma_{n_{k_\ell}}^*$ along which $j(\gamma_{n_{k_\ell}}^*)$ is constant. Let $j^* = j(\gamma_{n_{k_\ell}}^*)$. For x such that $g(x,\theta^*) \in B_{j^*}$, we have from (51) that

$$\gamma_{n_{k_\ell}}^{*\prime} g(x,\theta^*) < 0 \text{ and } |\gamma_{n_{k_\ell}}^{*\prime} g(x,\theta^*)| \geq |g(x,\theta^*)||\gamma_{n_{k_\ell}}^*|\epsilon \ . \tag{52}$$

We have from (50) that

$$Q_{n_{k_\ell}}\{1 + \gamma_{n_{k_\ell}}^{*\prime} g(X_i,\theta^*) > 0\} = 1 \ ,$$

which, together with (52), implies that

$$Q_{n_{k_\ell}}\{g(X_i,\theta^*) \in B_{j^*}, |g(X_i,\theta^*)||\gamma_{n_{k_\ell}}^*|\epsilon > 1\} = 0 \ . \tag{53}$$

Hence,

$$Q_{n_{k_\ell}}\{g(X_i,\theta^*) \in B_{j^*}, |g(X_i,\theta^*)| > \frac{1}{\epsilon M}\} \leq Q_{n_{k_\ell}}\{g(X_i,\theta^*) \in B_{j^*}, |g(X_i,\theta^*)||\gamma_{n_{k_\ell}}^*|\epsilon > 1\} = 0 \ ,$$

where the inequality follows from the fact that $|\gamma_{n_{k_\ell}}^*| > M$ and the equality follows from (53). Thus, by the Pormanteau Lemma,

$$Q\{g(X_i,\theta^*) \in B_{j^*}, |g(X_i,\theta^*)| > \frac{1}{\epsilon M}\} = 0$$

which implies that

$$Q\{g(X_i,\theta^*) \in B_{j^*}\} = Q\{g(X_i,\theta^*) \in B_{j^*}, \ |g(X_i,\theta^*)| \leq \frac{1}{\epsilon M}\}$$

Letting $M \to \infty$, we conclude from (47) that $0 \in B_{j^*}$, which contradicts the requirement that any collection of points $\{\tilde{g}_1, \ldots, \tilde{g}_{s(Q,\theta^*)+1}\}$ with $\tilde{g}_i \in \bar{B}_i$ for $1 \leq i \leq s(Q,\theta^*)+1$ must satisfy (45) with $\tilde{g}_i$ in place of g_i. Hence, it must be the case that γ_n^* are uniformly bounded.

We therefore have that there exists a subsequence $\gamma_{n_k}^*$ such that $\gamma_{n_k}^* \to \gamma^*$ and $\gamma^* \in \mathbf{R}^m$. We will now argue that

$$Q\{1 + \gamma^{*\prime} g(X_i,\theta^*) > 0\} = 1 \ . \tag{54}$$

To this end, for $\delta > 0$ let

$$R_\delta^- = \{x \in \mathcal{X} : 1 + \gamma^{*\prime} g(x,\theta^*) < \delta\} \qquad R_\delta^+ = \{x \in \mathcal{X} : 1 + \gamma^{*\prime} g(x,\theta^*) \geq \delta\} \tag{55}$$

and note that

$$\begin{aligned}\int \log(1+\gamma_{n_k}^{*\prime} g(x,\theta^*))dQ_{n_k} &= \int_{R_\delta^-} \log(1+\gamma_{n_k}^{*\prime} g(x,\theta^*)dQ_{n_k} + \int_{R_\delta^+} \log(1+\gamma_{n_k}^{*\prime} g(x,\theta^*))dQ_{n_k} \\ &\leq \int_{R_\delta^-} \log(1+\gamma_{n_k}^{*\prime} g(x,\theta^*))dQ_{n_k} + \int_{R_\delta^+} \log(1+|\gamma_{n_k}^{*}||g(x,\theta^*)|)dQ_{n_k} \\ &\leq \int_{R_\delta^-} \log(1+\gamma_{n_k}^{*\prime} g(x,\theta^*))dQ_{n_k} + \log(1+M\max_{x\in\mathcal{X}}|g(x,\theta^*)|)\ , \end{aligned} \tag{56}$$

where the equality holds by inspection, the first inequalty holds by the Cauchy-Schwartz inequality, and the second inequality holds because $|\gamma_{n_k}^*| \leq M$. Since $1+\gamma_{n_k}^{*\prime} g(x,\theta^*) \to 1+\gamma^{*\prime} g(x,\theta^*)$ uniformly for $x \in \mathcal{X}$, we have that for k sufficiently large the integrand in the first term in (56) is bounded above by $\log(2\delta)$. Thus, for k sufficiently large (56) is bounded above by

$$Q_{n_k}\{X_i \in R_\delta^-\}\log(2\delta) + \log(1+M\max_{x\in\mathcal{X}}|g(x,\theta^*)|)\ . \tag{57}$$

But,

$$\liminf_{n_k\to\infty} Q_{n_k}\{X_i \in R_\delta^-\} \geq Q\{X_i \in R_\delta^-\} \geq Q\{X_i \in \overline{R_0^-}\}\ , \tag{58}$$

where the first inequality follows from the Pormanteau Lemma and the second inequality follows from the fact that $\overline{R_0^-} \subseteq R_\delta^-$ for all $\delta > 0$. If (54) fails, then from (58) we have that

$$\inf_{\delta>0}\liminf_{n_k\to\infty} Q_{n_k}\{X_i \in R_\delta^-\} > 0\ .$$

It now follows from (56) and (57) that

$$\int \log(1+\gamma_{n_k}^{*\prime} g(x,\theta^*))dQ_{n_k} < 0$$

for $\delta > 0$ sufficiently small, which contradicts (50). Hence, (54) is established.

To complete the proof, we argue that

$$\int \log(1+\gamma^{*\prime} g(x,\theta^*))dQ \geq \eta\ , \tag{59}$$

which will contradict (42), completing the proof. To this end, note that

$$\int \max\{\log(1+\gamma^{*\prime} g(x,\theta^*)),\log(\delta)\}dQ \tag{60}$$

$$= \int \max\{\log(1+\gamma^{*\prime} g(x,\theta^*)),\log(\delta)\}(dQ - dQ_{n_k}) \tag{61}$$

$$+ \int \max\{0,\log(\delta) - \log(1+\gamma_{n_k}^{*\prime} g(x,\theta^*))\}dQ_{n_k} \tag{62}$$

$$+ \int (\max\{\log(1+\gamma^{*\prime} g(x,\theta^*)),\log(\delta)\} - \max\{\log(1+\gamma_{n_k}^{*\prime} g(x,\theta^*)),\log(\delta)\})dQ_{n_k} \tag{63}$$

$$+ \int \log(1+\gamma_{n_k}^{*\prime} g(x,\theta^*))dQ_{n_k}\ . \tag{64}$$

By virtue of $Q_n \to Q$, (61) tends to zero, while (62) is nonnegative. Since

$$\max\{\log(1+\gamma^{*\prime}_{n_k} g(x,\theta^*)), \log(\delta)\} \to \max\{\log(1+\gamma^{*\prime} g(x,\theta^*)), \log(\delta)\}$$

uniformly on $x \in \mathcal{X}$, (63) tends to zero. Finally, because of (49), (50) and the fact that $Q_{n_k} \in \ddot{\Lambda}_2(\eta)$ for all k, (64) is weakly greater than η. Thus, (60) is weakly greater than η. By letting $\delta \searrow 0$, we see from the monotone convergence theorem that (59) holds, which contradicts (42). ■

Lemma 5.3. *Suppose $X_i, i = 1, \ldots, n$ is an i.i.d. sequence of random variables with distribution P on $\mathcal{X}$. Suppose further that Assumptions 4.2, 4.4 and 4.5 hold, $P \in \mathbf{P}_0$ and that there exists $\omega > 0$ such that all $Q \ll P$ with $P\{X_i \in supp(Q)\} > \exp(-\omega)$ satisfy $\ddot{\mathbf{P}}(Q, \theta_0(P)) \neq \emptyset$, where*

$$\ddot{\mathbf{P}}(Q,\theta) = \{P \in \mathbf{M} : P \ll Q, Q \ll P, s(Q,\theta) = m, E_P[g(X_i,\theta)] = 0\} \ . \tag{65}$$

Also let

$$\Gamma(\eta, P) = \{\gamma \in \mathbf{R}^m : e^{-\eta} \le P\{1 + \gamma' g(X, \theta_0(P)) \ge 0\} \le 1\} \ . \tag{66}$$

If $\eta < \omega$, and η satisfies

$$\inf_{\gamma \in \Gamma(\eta,P)} \sup_{\lambda_0,\lambda_1 \ge 0} \lambda_0 + \eta\lambda_1 - \exp(\lambda_0 - 1) \int \exp(\lambda_1 \log(1+\gamma' g(x,\theta_0(P)))) I\{1+\gamma' g(x,\theta_0(P)) > 0\} dP \ge \eta \tag{67}$$

then it follows that

$$\limsup_{n\to\infty} \frac{1}{n} \log P^n\{\hat{P}_n \in \Lambda_2(\eta)\} \le -\eta \ .$$

Proof: Let $\ddot{\Lambda}_2(\eta, P) = \{Q \in \mathbf{M} : \inf_{R \in \ddot{\mathbf{P}}(Q,\theta_0(P))} I(Q|R) \ge \eta\}$ and note that

$$\begin{aligned} \limsup_{n\to\infty} \frac{1}{n} \log P^n\{\hat{P}_n \in \ddot{\Lambda}_2(\eta)\} &\le -\inf_{Q \in \ddot{\Lambda}_2(\eta)} I(Q|P) \\ &\le -\inf_{Q \in \ddot{\Lambda}_2(\eta,P)} I(Q|P) \ , \end{aligned}$$

where the first inequality follows from Lemma 5.2 and Sanov's Thoerem, while the final inequality follows from $\ddot{\Lambda}_2(\eta) \subseteq \ddot{\Lambda}_2(\eta, P)$. To complete the proof, it therefore suffices to show that

$$\inf_{Q \in \ddot{\Lambda}_2(\eta,P)} I(Q|P) \ge \eta \tag{68}$$

for $\eta < \omega$ satisfying (67).

For $\mathcal{S} \subseteq \text{supp}(P)$, let $\mathbf{M}(\mathcal{S}) = \{Q \in \mathbf{M} : \text{supp}(Q) \subseteq \mathcal{S}\}$. Note that

$$\inf_{Q \in \mathbf{M}(\mathcal{S})} I(Q|P) = -\log(P\{X_i \in \mathcal{S}\}) \ . \tag{69}$$

To see this, observe that the lefthand side of (69) is greater or equal to the righthand side of (69) by Jensen's inequality and that $I(Q|P) = -\log(P\{X_i \in \mathcal{S}\})$ for Q given by the distribution P conditional on $\mathcal{S}$. Next, note that for any Q such that $P\{X_i \in \text{supp}(Q)\} \le \exp(-\eta)$, we have that

$$I(Q|P) \ge \inf_{\mu \in \mathbf{M}(\text{supp}(Q))} I(\mu|P) = -\log(P\{X_i \in \text{supp}(Q)\}) \ge \eta \ . \tag{70}$$

Note further that if Q is not dominated by P, then $I(Q|P) = +\infty$. Hence, for

$$\tilde{\Lambda}_2(\eta, P) = \{Q \in \ddot{\Lambda}_2(\eta, P) : Q \ll P, P\{X_i \in \text{supp}(Q)\} \ge \exp(-\eta)\}$$

we have that

$$\inf_{Q \in \ddot{\Lambda}_2(\eta,P)} I(Q|P) \geq \min \left\{ \inf_{Q \in \tilde{\Lambda}_2(\eta,P)} I(Q|P), \eta \right\} \ . \tag{71}$$

We may assume that $\tilde{\Lambda}_2(\eta, P) \neq \emptyset$, for otherwise the righthand side of (71) equals η, thus establishing (68). Furthermore, since $\eta < \omega$, we also have by assumption that for any $Q \in \tilde{\Lambda}_2(\eta, P)$, $\ddot{\mathbf{P}}(Q, \theta_0(P)) \neq \emptyset$. Hence, the Primal Constraint Qualification of Theorem 3.4 of Borwein and Lewis (1993) is satisfied, so for all $Q \in \tilde{\Lambda}_2(\eta, P)$ we have

$$\inf_{R \in \ddot{\mathbf{P}}(Q,\theta_0(P))} I(R|Q) = \max_{\gamma \in \mathbf{R}^m} \int \log(1 + \gamma' g(x, \theta_0(P)))dQ \geq \eta \ ,$$

where the inequality is implied by $Q \in \ddot{\Lambda}_2(\eta, P)$. Next, we define

$$\begin{aligned}
\Gamma &= \{\gamma \in \mathbf{R}^m : \exists Q \in \tilde{\Lambda}_2(\eta,P) \text{ s.t. } \gamma \in \arg\max_{\lambda \in \mathbf{R}^m} \int \log(1 + \lambda' g(x, \theta_0(P)))dQ\} \\
\mathbf{S}(\gamma) &= \{\mathcal{S} \subseteq \text{supp}(P) : \exists Q \in \tilde{\Lambda}_2(\eta,P) \text{ s.t. } \mathcal{S} = \text{supp}(Q), \gamma \in \arg\max_{\lambda \in \mathbf{R}^m} \int \log(1 + \lambda' g(x, \theta_0(P)))dQ\} \\
\mathbf{R}(\gamma, \mathcal{S}) &= \{Q \in \tilde{\Lambda}_2(\eta,P) : \gamma \in \arg\max_{\lambda \in \mathbf{R}^m} \int \log(1 + \lambda' g(x, \theta_0(P)))dQ, \mathcal{S} = \text{supp}(Q)\} \ .
\end{aligned}$$

With these definitions, we write

$$\tilde{\Lambda}_2(\eta, P) = \bigcup_{\gamma \in \Gamma} \bigcup_{\mathcal{S} \in \mathbf{S}(\gamma)} \mathbf{R}(\gamma, \mathcal{S}) \ .$$

Hence,

$$\inf_{Q \in \tilde{\Lambda}_2(\eta,P)} I(Q|P) = \inf_{\gamma \in \Gamma} \inf_{\mathcal{S} \in \mathbf{S}(\gamma)} \inf_{Q \in \mathbf{R}(\gamma,\mathcal{S})} I(Q|P) \ . \tag{72}$$

Note that if $Q \in \mathbf{R}(\gamma, \mathcal{S})$, then (i) $Q \ll P$, (ii) $\mathcal{S} = \text{supp}(Q)$ and (iii) $\int \log(1 + \gamma' g(x, \theta_0(P)))dQ \geq \eta$. We therefore have for $\delta > 0$ sufficiently small that

$$\begin{aligned}
\inf_{Q \in \mathbf{R}(\gamma,\mathcal{S})} I(Q|P) &\geq \inf\{\int_{\mathcal{S}} \log(\phi(x))\phi(x)dP : \phi(x) \in L^1(\mathcal{S}), \phi(x) > 0 \\
&\qquad \int_{\mathcal{S}} \log(1 + \gamma' g(x, \theta_0(P)))\phi(x)dP \geq \eta, \int_{\mathcal{S}} \phi(x)dP = 1\} \\
&\geq \inf\{\int_{\mathcal{S}} \log(\phi(x))\phi(x)dP : \phi(x) \in L^1(\mathcal{S}), \phi(x) > 0 \\
&\qquad \int_{\mathcal{S}} \log(1 + \gamma' g(x, \theta_0(P)))I\{x \in R_\delta^+\}\phi(x)dP \geq \eta, \int_{\mathcal{S}} \phi(x)dP = 1\} \qquad (73)
\end{aligned}$$

where the first inequality follows from the preceding statements (i), (ii) and (iii), and the second inequality follows from the definition of R_δ^+ in (55) but with $(\theta_0(P), \gamma)$ in place of (θ^*, γ^*).

We now use Corollary 4.8 of Borwein and Lewis (1992a) and part (vi) of Example 6.5 of Borwein and Lewis (1992b) to find the dual problem of (73). To this end, first note that since $\tilde{\Lambda}_2(\eta, P) \neq \emptyset$, we have that $\mathbf{R}(\gamma, \mathcal{S}) \neq \emptyset$ for at least one $\gamma \in \Gamma$ and $\mathcal{S} \in \mathbf{S}(\gamma)$. For any such γ and $\mathcal{S}$, we have as a result that there exists a $\phi(x)$ satisfying the constraints of (73). Next, note that the map $A : L^1(\mathcal{S}) \to \mathbf{R}$ defined by

$$A(\phi) = \int_{\mathcal{S}} \log(1 + \gamma' g(x, \theta_0(P)))I\{x \in R_\delta^+\}\phi(x)dP$$

is continuous because $\log(1 + \gamma' g(x, \theta_0(P)))I\{x \in R_\delta^+\}$ lies in $L^\infty(\mathcal{S})$ as a result of $\mathcal{S}$ being a subset of the compact set $\mathcal{X}$ and $g(x, \theta_0(P))$ being continuous on $\mathcal{X}$. Using Corollary 4.8 of Borwein and Lewis (1992a) and part (vi) of Example 6.5 of Borwein and Lewis (1992b) to find the dual problem of (73) implies

$$\inf_{Q \in \mathbf{R}(\gamma, \mathcal{S})} I(Q|P) \geq \liminf_{\delta \searrow 0} \sup_{\lambda_0, \lambda_1 \geq 0} \lambda_0 + \eta\lambda_1 - \exp(\lambda_0 - 1) \int_{\mathcal{S}} \exp(\lambda_1 \log(1 + \gamma' g(x, \theta_0(P))) I\{x \in R_\delta^+\}) dP \ . \quad (74)$$

By definition, for every $\mathcal{S} \in \mathbf{S}(\gamma)$ there exists Q such that $\mathcal{S} = \text{supp}(Q)$ and

$$\gamma \in \arg\max_{\lambda \in \mathbf{R}^m} \int \log(1 + \lambda' g(x, \theta_0(P))) dQ \ . \quad (75)$$

For any such Q, we must have that

$$Q\{1 + \gamma' g(X_i, \theta_0(P)) \leq 0\} = Q\{1 + \gamma' g(X_i, \theta_0(P)) \leq 0, X_i \in \mathcal{S}\} = 0 \ , \quad (76)$$

from which it follows that

$$P\{1 + \gamma' g(X_i, \theta_0(P)) \leq 0, X_i \in \mathcal{S}\} = 0 \quad (77)$$

as well. Hence, by letting $\delta \searrow 0$, we see by the monotone convergence theorem that

$$\int_{\mathcal{S}} \exp(\lambda_1 \log(1 + \gamma' g(x, \theta_0(P))) I\{x \in R_\delta^+\}) dP$$

is right-continuous at $\delta = 0$. Following arguments as in Lemma 17.29 in Aliprantis and Border (2006), it is possible to show the supremum in (74) is lower semicontinuous at $\delta = 0$ as well. Hence, the righthand side of (74) is greater than or equal to

$$\sup_{\lambda_0, \lambda_1 \geq 0} \lambda_0 + \eta\lambda_1 - \exp(\lambda_0 - 1) \int_{\mathcal{S}} \exp(\lambda_1 \log(1 + \gamma' g(x, \theta_0(P))) I\{1 + \gamma' g(x, \theta_0(P)) \geq 0\}) dP \ . \quad (78)$$

Since the integrand in (78) is nonnegative, we have from (74) and (77) and (78) that

$$\begin{aligned} &\inf_{\mathcal{S} \in \mathbf{S}(\gamma)} \inf_{Q \in \mathbf{R}(\gamma, \mathcal{S})} I(Q|P) \\ &\quad \geq \inf_{\mathcal{S} \in \mathbf{S}(\gamma)} \sup_{\lambda_0, \lambda_1 \geq 0} \lambda_0 + \eta\lambda_1 - \exp(\lambda_0 - 1) \int_{\mathcal{S}} \exp(\lambda_1 \log(1 + \gamma' g(x, \theta_0(P)))) I\{1 + \gamma' g(x, \theta_0(P)) > 0\} dP \\ &\quad \geq \sup_{\lambda_0, \lambda_1 \geq 0} \lambda_0 + \eta\lambda_1 - \exp(\lambda_0 - 1) \int \exp(\lambda_1 \log(1 + \gamma' g(x, \theta_0(P)))) I\{1 + \gamma' g(x, \theta_0(P)) > 0\} dP \quad (79) \end{aligned}$$

By definition, for every $\gamma \in \Gamma$ there exists a $Q \in \tilde{\Lambda}_2(\eta, P)$ such that γ satisfies (75). Thus, as before, (76) holds, from which it follows that

$$\text{supp}(Q) \subseteq \{x \in \mathbf{R}^d : 1 + \gamma' g(x, \theta_0(P)) > 0\} \ .$$

Therefore,

$$P(1 + \gamma' g(X_i, \theta_0(P)) \geq 0) \geq P(X_i \in \text{supp}(Q)) \geq \exp(-\eta)$$

by $Q \in \tilde{\Lambda}_2(\eta, P)$. Hence, $\gamma \in \Gamma(\eta, P)$, which implies $\Gamma \subseteq \Gamma(\eta, P)$. It therefore follows from (79) that

$$\begin{aligned} &\inf_{\gamma \in \Gamma} \inf_{\mathcal{S} \in \mathbf{S}(\gamma)} \inf_{Q \in \mathbf{R}(\gamma, \mathcal{S})} I(Q|P) \geq \\ &\quad \inf_{\gamma \in \Gamma(\eta, P)} \sup_{\lambda_0, \lambda_1 \geq 0} \lambda_0 + \eta\lambda_1 - \exp(\lambda_0 - 1) \int \exp(\lambda_1 \log(1 + \gamma' g(x, \theta_0(P)))) I\{1 + \gamma' g(x, \theta_0(P)) > 0\} dP \ . \end{aligned}$$

The desired claim (68) thus follows for $\eta < \omega$, satisfying (67). ■

Lemma 5.4. *If Assumptions 4.2, 4.4 and 4.5 hold, then $\theta_0(P)$ is continuous on $\mathbf{P}_0$ under the weak topology.*

Proof: Let $P_n \to P$ with $P_n \in \mathbf{P}_0$ for all n and denote

$$\theta_n = \theta_0(P_n)$$

where $\theta_0(P_n) \neq \emptyset$ by $P_n \in \mathbf{P}_0$ and $\theta_0(P_n)$ is a singleton by Assumption 4.2. Let θ^* be a limit point of $\{\theta_n\}$ and θ_{n_k} a subsequence such that $\theta_{n_k} \to \theta^*$. It then follows that,

$$\begin{aligned} |\int g(x,\theta^*)dP| &= \lim_{n_k\to\infty} |\int g(x,\theta^*)dP_{n_k}| \\ &= \lim_{n_k\to\infty} |\int (g(x,\theta^*) - g(x,\theta_{n_k}))dP_{n_k}| \\ &\leq \lim_{n_k\to\infty} \sup_{x\in\mathcal{X}} |g(x,\theta^*) - g(x,\theta_{n_k})| \\ &= 0 \end{aligned} \tag{80}$$

where the first equality follows by $P_n \to P$ and $g(x,\theta^*)$ continuous and bounded. The second equality is implied by $\theta_{n_k} = \theta_0(P_{n_k})$, the inequality follows by inspection and the final result is due to the uniform continuity of $g(x,\theta)$. Hence,

$$\theta^* = \theta_0(P) \ . \tag{81}$$

It follows that $\theta_0(P)$ is the unique limit point of $\{\theta_n\}$, which establishes the claim of the Lemma. ■

Lemma 5.5. *If Assumptions 4.2, 4.3, 4.4 and 4.5 hold, then for any δ such that $0 < \delta < \epsilon$,*

$$\Gamma(\eta, P) = \{\gamma \in \mathbf{R}^m : e^{-\eta} \leq P\{1 + \gamma' g(X, \theta_0(P)) \geq 0\} \leq 1\} \tag{82}$$

is nonempty, compact valued and upper hemicontinuous on $(\eta, P) \in [0, -\log(1-\epsilon+\delta)] \times \mathbf{P}_0$ under the product of the topology on $\mathbf{R}$ and the weak topology.

Proof: The correspondence $\Gamma(\eta, P)$ is clearly not empty since $0 \in \Gamma(\eta, P)$ for all $(\eta, P) \in [0, -\log(1 - \epsilon + \delta)] \times \mathbf{P}_0$. To establish upper hemicontinuity we wish to show that if $P_n \to P$ and $\eta_n \to \eta$ with $(\eta_n, P_n) \in [0, -\log(1-\epsilon+\delta)] \times \mathbf{P}_0$ for all n, then any sequence $\{\gamma_n\}_{n=1}^{\infty}$ with $\gamma_n \in \Gamma(\eta_n, P_n)$ for all n, has a limit point in $\Gamma(\eta, P)$. For this purpose we first show the sequence $\{\gamma_n\}_{n=1}^{\infty}$ is uniformly bounded. Suppose by way of contradiction,

$$\limsup_{n\to\infty} |\gamma_n| = +\infty \tag{83}$$

It follows that there exists a subsequence satisfying

$$|\gamma_{n_k}| \geq n_k \tag{84}$$

In addition, by compactness there exists an additional subsequence such that

$$\frac{\gamma_{n_{k_l}}}{|\gamma_{n_{k_l}}|} \to \gamma_1 \tag{85}$$

Along such a subsequence, however, we have

$$
\begin{aligned}
e^{-\eta} &= \lim_{n_{k_l}\to\infty} e^{-\eta_{n_{k_l}}} \\
&\leq \limsup_{n_{k_l}\to\infty} P_{n_{k_l}}\{1+\gamma_{n_{k_l}}' g(X,\theta_0(P_{n_{k_l}}))\geq 0\} \\
&= \limsup_{n_{k_l}\to\infty} P_{n_{k_l}}\{\frac{\gamma_{n_{k_l}}'}{|\gamma_{n_{k_l}}|} g(X,\theta_0(P_{n_{k_l}}))\geq -\frac{1}{|\gamma_{n_{k_l}}|}\} \\
&\leq \liminf_{\epsilon\searrow 0}\limsup_{n_{k_l}\to\infty} P_{n_{k_l}}\{\frac{\gamma_{n_{k_l}}'}{|\gamma_{n_{k_l}}'|} g(X,\theta_0(P_{n_{k_l}}))\geq -\epsilon\} \\
&\leq \liminf_{\epsilon\searrow 0}\limsup_{n_{k_l}\to\infty} P_{n_{k_l}}\{\gamma_1' g(X,\theta_0(P))\geq -2\epsilon\} \\
&\leq \liminf_{\epsilon\searrow 0} P\{\gamma_1' g(X,\theta_0(P))\geq -2\epsilon\} \\
&= P\{\gamma_1' g(X,\theta_0(P))\geq 0\} \qquad (86)
\end{aligned}
$$

where the first equality follows by assumption and the first inequality by $\gamma_{n_{k_l}} \in \Gamma(\eta_{n_{k_l}}, P_{n_{k_l}})$ for all l. The second equality follows by inspection. The second inequality is implied by (84) and the third inequality by $\theta_0(P_{n_{k_l}}) \to \theta_0(P)$ by Lemma 5.4, (85) and the unform continuity of $g(x,\theta)$. The final inequality and equality follow by the Portmanteau and Bounded Convergence theorems respectively. Hence,

$$1-\epsilon < e^{-\eta} \leq P\{\gamma_1' g(X,\theta_0(P))\geq 0\} \qquad (87)$$

by (86) and $\eta_{n_{k_l}} \in [0, -\log(1-\epsilon+\delta)]$ for all l. Result (87), however, contradicts $P\in\mathbf{P}_0$.

Because the sequence $\{\gamma_n\}_{n=1}^{\infty}$ is uniformly bounded, it follows that there exists a subsequence such that

$$\lim_{n_j\to\infty} \gamma_{n_j} = \gamma_2 \qquad (88)$$

To conclude establishing upper hemicontinuity we show $\gamma_2 \in \Gamma(\eta, P)$, which is implied by

$$
\begin{aligned}
e^{-\eta} &= \lim_{n_j\to\infty} e^{-\eta_{n_j}} \\
&\leq \limsup_{n_j\to\infty} P_{n_j}\{1+\gamma_{n_j}' g(X,\theta_0(P_{n_j}))\geq 0\} \\
&\leq \liminf_{\epsilon\searrow 0}\limsup_{n_j\to\infty} P_{n_j}\{1+\gamma_2' g(X,\theta_0(P))\geq -\epsilon\} \\
&\leq \liminf_{\epsilon\searrow 0} P\{1+\gamma_2' g(X,\theta_0(P))\geq -\epsilon\} \\
&= P\{1+\gamma_2' g(X,\theta_0(P))\geq 0\} \qquad (89)
\end{aligned}
$$

where the first equality follows by assumption and the first inequality by $\gamma_{n_j} \in \Gamma(\eta_{n_j}, P_{n_j})$ for all j. By Lemma 5.4, $\theta_0(P_{n_j}) \to \theta_0(P)$ and therefore the second inequality follows by the uniform continuity of $g(x,\theta)$. The final inequality and equality follow by the Portmanteau and Bounded Convergence theorems respectively.

The arguments in (83)-(86) but for $\{\gamma_n\}_{n=1}^{\infty}$ an unbounded sequence in $\Gamma(\eta, P)$ and $\eta_n = \eta$, $P_n = P$ for all n show $\Gamma(\eta, P)$ is bounded. Similarly, the arguments in (89) but with $\eta_n = \eta$ and $P_n = P$ for all n show $\Gamma(\eta, P)$ is closed. Hence, $\Gamma(\eta, P)$ is compact. ■

Lemma 5.6. *If Assumptions 4.2, 4.4 and 4.5 hold, then the function*

$$f(\lambda_1,\gamma,P) = \int (1+\gamma' g(x,\theta_0(P)))^{\lambda_1} I\{1+\gamma' g(x,\theta_0(P))>0\}dP$$

is lower semicontinuous on $(\lambda_1,\gamma,P)\in\mathbf{R}_+\times\mathbf{R}^m\times\mathbf{P}_0$ *where* $\mathbf{P}_0$ *is endowed with the weak toplogy.*

PROOF: Let $(\lambda_{1,n}, \gamma_n, P_n) \to (\lambda_1, \gamma, P)$. In order to establish the Lemma we aim to show that,

$$\liminf_{n\to\infty} f(\lambda_{1,n}, \gamma_n, P_n) \geq f(\lambda, \gamma, P) \tag{90}$$

For this purpose, we define the auxiliary variable,

$$\epsilon_{m_0} \equiv \sup_{x\in\mathcal{X}, n\geq m_0} |\gamma_n' g(x, \theta_0(P_n)) - \gamma' g(x, \theta_0(P))| \tag{91}$$

Notice that due to Lemma 5.4 and Assumption 4.4 we have $\lim_{m_0\to\infty} \epsilon_{m_0} = 0$. Also define,

$$\bar{\lambda}_{1,m_0} \equiv \sup_{n\geq m_0} \lambda_n \qquad\qquad \underline{\lambda}_{1,m_0} \equiv \inf_{n\geq m_0} \lambda_n \tag{92}$$

as well as the function:

$$L_{m_0}(u) = u^{\underline{\lambda}_{1,m_0}} I\{u > 1\} + u^{\bar{\lambda}_{1,m_0}} I\{0 < u \leq 1\} \tag{93}$$

The notice that pointwise in $x \in \mathcal{X}$ we have that:

$$\begin{aligned}
\inf_{m\geq m_0} & (1 + \gamma_m' g(x, \theta_0(P_m)))^{\lambda_m} I\{1 + \gamma_m' g(x, \theta_0(P_m)) > 0\} \\
&\geq \inf_{m\geq m_0} (1 + \gamma_m' g(x, \theta_0(P_m)))^{\lambda_m} I\{1 + \gamma' g(x, \theta_0(P)) > \epsilon_{m_0}\} \\
&\geq \inf_{m\geq m_0} (1 + \gamma' g(x, \theta_0(P)) - \epsilon_{m_0})^{\lambda_m} I\{1 + \gamma' g(x, \theta_0(P)) > \epsilon_{m_0}\} \\
&\geq L_{m_0}(1 + \gamma' g(x, \theta_0(P)) - \epsilon_{m_0})
\end{aligned} \tag{94}$$

where the first two inequalities are implied by (91) and the final one follows by (93) and direct calculation. Next, exploiting standard manipulations and (94) we are able to conclude

$$\begin{aligned}
\liminf_{n\to\infty} f(\lambda_{1,n}, \gamma_n, P_n) &= \lim_{n\to\infty} \inf_{n_0\geq n} \int (1 + \gamma_{n_0}' g(x, \theta_0(P_{n_0})))^{\lambda_{n_0}} I\{1 + \gamma_{n_0}' g(x, \theta_0(P_{n_0})) > 0\} dP_{n_0} \\
&\geq \lim_{n\to\infty} \inf_{n\geq n_0} \inf_{m\geq m_0} \int (1 + \gamma_{m_0}' g(x, \theta_0(P_{m_0})))^{\lambda_{m_0}} I\{1 + \gamma_{m_0}' g(x, \theta_0(P_{m_0})) > 0\} dP_{n_0} \\
&\geq \liminf_{m_0\to\infty} \liminf_{n\to\infty} \inf_{m\geq m_0} \int (1 + \gamma_{m_0}' g(x, \theta_0(P_{m_0})))^{\lambda_{m_0}} I\{1 + \gamma_{m_0}' g(x, \theta_0(P_{m_0})) > 0\} dP_n \\
&\geq \liminf_{m_0\to\infty} \liminf_{n\to\infty} \int \inf_{m\geq m_0} (1 + \gamma_{m_0}' g(x, \theta_0(P_{m_0})))^{\lambda_{m_0}} I\{1 + \gamma_{m_0}' g(x, \theta_0(P_{m_0})) > 0\} dP_n \\
&\geq \liminf_{m_0\to\infty} \liminf_{n\to\infty} \int L_{m_0}(1 + \gamma' g(x, \theta_0(P)) - \epsilon_{m_0}) dP_n
\end{aligned} \tag{95}$$

Further, observe from (93) that if $\bar{\lambda}_{1,m_0} > 0$, then $L_{m_0}(u)$ is continuous, while if $\bar{\lambda}_{1,m_0} = 0$ then we have $L_{m_0}(u) = I\{u > 0\}$. In both cases, since $g(x, \theta_0(P))$ is continuous and $\mathcal{X}$ is compact, we obtain by the Portmanteau Lemma and $P_n \to P$ in the weak topology,

$$\begin{aligned}
\liminf_{m_0\to\infty} \liminf_{n\to\infty} \int L_{m_0}(1 + \gamma' g(x, \theta_0(P)) - \epsilon_{m_0}) dP_n &\geq \liminf_{m_0\to\infty} \int L_{m_0}(1 + \gamma' g(x, \theta_0(P)) - \epsilon_{m_0}) dP \\
&\geq \int \liminf_{m_0\to\infty} L_{m_0}(1 + \gamma' g(x, \theta_0(P)) - \epsilon_{m_0}) dP
\end{aligned} \tag{96}$$

where the second inequality follows by Fatou's Lemma. Finally, by $\bar{\lambda}_{1,m_0} \to \lambda_1$, $\underline{\lambda}_{1,m_0} \to \lambda_1$ and $\epsilon_{m_0} \to 0$, direct calculation reveals that pointwise in $x \in \mathcal{X}$ we have,

$$\liminf_{m_0\to\infty} L_{m_0}(1 + \gamma' g(x, \theta_0(P)) - \epsilon_{m_0}) \geq (1 + \gamma' g(x, \theta_0(P)))^{\lambda_1} I\{1 + \gamma' g(x, \theta_0(P)) > 0\} \tag{97}$$

Combining (95), (96) and (97) establishes the claim of the Lemma. ■

Lemma 5.7. *Suppose Assumptions 4.2, 4.4 and 4.5 hold and for* $(\lambda_0, \lambda_1, \eta, \gamma, P) \in [0,2]^2 \times \mathbf{R}_+ \times \mathbf{R}^m \times \mathbf{P}_0$ *with* $\mathbf{P}_0$ *endowed with the weak topology, define the function*

$$F(\lambda_0, \lambda_1, \eta, \gamma, P) = \lambda_0 + \eta(\lambda_1 - 1) - e^{\lambda_0 - 1} \int (1 + \gamma' g(x, \theta_0(P))^{\lambda_1} I\{1 + \gamma' g(x, \theta_0(P)) > 0\} dP \ .$$

In addition, consider the following correspondences,

$$\begin{aligned} E(\eta, \gamma, P) &= \{(\lambda_0, \lambda_1, y) \in [0,2]^2 \times \mathbf{R} : y \leq F(\lambda_0, \lambda_1, \eta, \gamma, P)\} \\ \Pi(\eta, \gamma, P) &= \{y \in \mathbf{R} : (\lambda_0, \lambda_1, y) \in E(\eta, \gamma, P) \text{ for some } (\lambda_0, \lambda_1) \in [0,2]^2\} \end{aligned}$$

It then follows that $\Pi(\eta, \gamma, P)$ *is lower hemicontinuous on* $\mathbf{R}_+ \times \mathbf{R} \times \mathbf{P}_0$.

Proof: As in Lemma 5.6 we define the function,

$$f(\lambda_1, \gamma, P) = \int (1 + \gamma' g(x, \theta_0(P)))^{\lambda_1} I\{1 + \gamma' g(x, \theta_0(P)) > 0\} dP \tag{98}$$

We first show that $f(\lambda_1, \gamma, P)$ is continuous at all points on $[0,2] \times \mathbf{R}^m \times \mathbf{P}_0$ with $\lambda_1 \neq 0$. For this purpose, let $(\lambda_{1,n}, \gamma_n, P_n) \to (\lambda_1, \gamma, P)$ and note that by Lemma 5.4 and $\mathcal{X}$ being compact, we have:

$$\lim_{n \to \infty} \sup_{x \in \mathcal{X}} |\gamma_n' g(x, \theta_0(P_n)) - \gamma' g(x, \theta_0(P))| = 0 \tag{99}$$

Further, notice that since $\lambda_1 > 0$, then by $\lambda_{1,n} \to \lambda_1$ we have $\lambda_{1,n} > 0$ for n large enough, which implies,

$$\lim_{n \to \infty} \sup_{x \in \mathcal{X}} |(1 + \gamma_n' g(x, \theta_0(P_n)))^{\lambda_{1,n}} I\{1 + \gamma_n' g(x, \theta_0(P_n)) > 0,\ 1 + \gamma' g(x, \theta_0(P)) \leq 0\}| = 0 \tag{100}$$

as a result of (99). By direct calculations we then obtain from (99) and (100) that,

$$\lim_{n \to \infty} \sup_{x \in \mathcal{X}} |(1+\gamma_n' g(x, \theta_0(P_n)))^{\lambda_{1,n}} I\{1+\gamma_n' g(x, \theta_0(P_n)) > 0\} - (1+\gamma' g(x, \theta_0(P)))^{\lambda_1} I\{1+\gamma' g(x, \theta_0(P)) > 0\}| = 0 \tag{101}$$

By (101) and noting that the integrand is a continuous bounded function for $\lambda_1 > 0$, $P_n \to P$ establishes:

$$f(\lambda_{1,n}, \gamma_n, P_n) = \int (1 + \gamma' g(x, \theta_0(P)))^{\lambda_1} I\{1 + \gamma' g(x, \theta_0(P)) > 0\} dP_n + o(1) \to f(\lambda_1, \gamma, P) \tag{102}$$

hence proving the desired continuity of $f(\lambda_1, \gamma, P)$ at all points $(\lambda_1, \gamma, P) \in [0,2] \times \mathbf{R}_+ \times \mathbf{P}_0$ with $\lambda_1 > 0$.

We now establish lower hemicontinuity of $\Pi(\eta, \gamma, P)$. This requires showing that for any $y \in \Pi(\eta, \gamma, P)$ and $(\eta_n, \gamma_n, P_n) \to (\eta, \gamma, P)$ there exists a subsequence $(\eta_{n_k}, \gamma_{n_k}, P_{n_k})$ and $y_{n_k} \in \Pi(\eta_{n_k}, \gamma_{n_k}, P_{n_k})$ with $y_{n_k} \to y$. Since $y \in \Pi(\eta, \gamma, P)$, there exist a $(\lambda_0(y), \lambda_1(y)) \in [0,2]^2$ with:

$$y \leq F(\lambda_0(y), \lambda_1(y), \eta, \gamma, P) \ . \tag{103}$$

If $\lambda_1(y) > 0$, then we immediately have from (102) that,

$$F(\lambda_0(y), \lambda_1(y), \eta_n, \gamma_n, P_n) \to F(\lambda_0(y), \lambda_1(y), \eta, \gamma, P) \tag{104}$$

from which it follows that there exist $y_n \in \Pi(\eta_n, \gamma_n, P_n)$ with $y_n \to y$. To address the case $\lambda_1(y) = 0$, notice:

$$\begin{aligned} \limsup_{n \to \infty} F(\lambda_0(y), 0, \eta_n, \gamma_n, P_n) &= \lambda_0(y) + \eta - e^{\lambda_0(y) - 1} \times \liminf_{n \to \infty} P_n\{1 + \gamma_n' g(x, \theta_0(P_n)) > 0\} \\ &\geq \lambda_0(y) + \eta - e^{\lambda_0(y) - 1} \times \liminf_{n \to \infty} P\{1 + \gamma' g(x, \theta_0(P)) > 0\} \\ &= F(\lambda_0(y), 0, \eta, \gamma, P) \end{aligned} \tag{105}$$

where the inequality is implied by $P_n \to P$, (99), the extended continuous mapping theorem of Theorem 1.11.1 in van der Vaart and Wellner (1996) and the Portmanteau Lemma. The final equality in (105) is definitional. The existence of a subsequence $(\gamma_{n_k}, \eta_{n_k}, P_{n_k})$ with $y_{n_k} \in \Pi(\gamma_{n_k}, \eta_{n_k}, P_{n_k})$ and $y_{n_k} \to y$ then follows. ■

Lemma 5.8. *If Assumption 4.2, 4.3, 4.4 and 4.5 hold, then for every $Q \in \mathbf{P}_0$ there exists an open neighborhood $N(Q)$ in $\mathbf{P}_0$ with respect to the weak topology and a $\bar{\eta}(Q) > 0$ such that for all $\eta \in [0, \bar{\eta}(Q)]$,*

$$\inf_{P \in N(Q)} \inf_{\gamma \in \Gamma(\eta, P)} \sup_{\lambda_0, \lambda_1 \geq 0} \lambda_0 + \eta(\lambda_1 - 1) - e^{\lambda_0 - 1} \int (1 + \gamma' g(x, \theta_0(P)))^{\lambda_1} I\{1 + \gamma' g(x, \theta_0(P)) > 0\} dP \geq 0 \quad (106)$$

Proof: First notice that since by Lemma 5.5 the correspondence $\Gamma(0, Q)$ is compact valued, there exists a compact set A such that,

$$\Gamma(0, Q) \subset A$$

Furthermore, since by Lemma 5.5, $\Gamma(\eta, P)$ is also upper hemicontinuous at $(\eta, P) = (0, Q)$, there exists a $\alpha(Q) > 0$ and an open neighborhood $B(Q)$ in $\mathbf{P}_0$ such that for all $0 \leq \eta \leq \alpha(Q)$ and $P \in B(Q)$, we have

$$\Gamma(\eta, P) \subset A \quad (107)$$

Thus, since $[0, 2]^2 \subset \mathbf{R} \times \mathbf{R}_+$, it immediately follows that for all $0 \leq \eta \leq \alpha(Q)$ and $P \in B(Q)$,

$$\inf_{\gamma \in \Gamma(\eta, P)} \sup_{\lambda_0, \lambda_1 \geq 0} \lambda_0 + \eta(\lambda_1 - 1) - e^{\lambda_0 - 1} \int (1 + \gamma' g(x, \theta_0(P)))^{\lambda_1} I\{1 + \gamma' g(x, \theta_0(P)) > 0\} dP$$
$$\geq \inf_{\gamma \in A} \sup_{(\lambda_0, \lambda_1) \in [0,2]^2} \lambda_0 + \eta(\lambda_1 - 1) - e^{\lambda_0 - 1} \int (1 + \gamma' g(x, \theta_0(P)))^{\lambda_1} I\{1 + \gamma' g(x, \theta_0(P)) > 0\} dP \quad (108)$$

We will establish the Lemma by showing that for η sufficiently small, the right hand side of (108) is non-negative on an open neighborhood of Q. For this purpose, define the function

$$F(\lambda_0, \lambda_1, \eta, \gamma, Q) = \lambda_0 + \eta(\lambda_1 - 1) - e^{\lambda_0 - 1} \int (1 + \gamma' g(x, \theta_0(Q))^{\lambda_1} I\{1 + \gamma' g(x, \theta_0(Q)) > 0\} dQ \quad (109)$$

By Lemma 5.6, Lemma 5.7 and Theorem 2 in Ausubel and Deneckere (1993), it follows that

$$C(\gamma, \eta, Q) = \max_{(\lambda_0, \lambda_1) \in [0,2]^2} F(\lambda_0, \lambda_1, \eta, \gamma, Q) \quad (110)$$

is continuous on $(\gamma, \eta, Q) \in \mathbf{R}^m \times \mathbf{R}_+ \times \mathbf{P}_0$. Moreover, since A is compact, applying Berge's Theorem of the Maximum establishes that the correspondence

$$\Xi(\eta, P) = \arg\min_{\gamma \in A} C(\gamma, \eta, P) \quad (111)$$

is well defined and upper hemicontinuous on $\mathbf{R}_+ \times \mathbf{P}_0$.

We now show $\Xi(0, Q) = \{0\}$. If $\gamma \in A \backslash \Gamma(0, Q)$, then $Q\{1 + \gamma' g(X, \theta_0(Q)) \geq 0\} < 1$, and hence

$$F(1, 0, 0, \gamma, Q) = 1 - Q\{1 + \gamma' g(X, \theta_0(Q)) > 0\} > 0 \quad (112)$$

On the other hand, for any $0 \neq \gamma \in \Gamma(0, Q)$, we have $Q\{1 + \gamma' g(X, \theta_0(Q)) \geq 0\} = 1$. Therefore,

$$F(1, 1, 0, \gamma, Q) = 1 - \int (1 + \gamma' g(X, \theta_0(Q))) dQ = 0 \quad (113)$$

by virtue of $Q \in \mathbf{P}_0$. Further, since $Q \in \mathbf{P}_0$, Assumption 4.3 implies that for $\gamma \neq 0$,

$$0 < Q\{\gamma' g(X, \theta_0(Q)) \geq 0\} < 1 \tag{114}$$

Next, use the dominated convergence theorem to exchange the order of differentiation and integration in (113) and conclude that for $0 \neq \gamma \in \Gamma(0, Q)$:

$$\frac{\partial}{\partial \lambda_1} F(1, \lambda_1, 0, \gamma, Q)\Big|_{\lambda_1 = 1} = \int (1 + \gamma' g(x, \theta_0(Q))) \log(1 + \gamma' g(x, \theta_0(Q))) I\{1 + \gamma' g(x, \theta_0(Q)) > 0\} dQ > 0 \ , \tag{115}$$

where the inequality holds by (114) which implies $\gamma' g(x, \theta_0(Q))$ is not constant on $\text{supp}_Q(g(X_i, \theta_0(Q)))$ and therefore Jensen's inequality holds strictly. Hence, if $0 \neq \gamma \in \Gamma(0, Q)$, there exists $1 \leq \tilde{\lambda}_1 \leq 2$ such that

$$F(1, \tilde{\lambda}_1, 0, \gamma, Q) > 0 \tag{116}$$

Thus, so far we have established through (112) and (116) that if $0 \neq \gamma \in A$ then

$$C(\gamma, 0, Q) > 0$$

On the other hand, it follows from direct calculation that $C(0, 0, Q) = 0$, and hence we conclude,

$$\Xi(0, Q) = \{0\} \tag{117}$$

Next notice that continuity of $g(x, \theta)$ in (x, θ) and compactness of $\mathcal{X}$ and Θ implies that

$$\sup_{\theta \in \Theta} \sup_{x \in \mathcal{X}} |g(x, \theta)| < \infty \tag{118}$$

Furthermore, since as argued $\Xi(\eta, P)$ is upper hemicontinuous at $(\eta, P) = (0, Q)$, it follows from (117) and (118) that there exists a $\alpha(Q) \geq \bar{\eta}(Q) > 0$ and open neighborhood $N(Q) \subseteq B(Q)$ such that if $\eta \in [0, \bar{\eta}(Q)]$ and $P \in N(Q)$ then,

$$\sup_{\gamma \in \Xi(\eta, P)} |\gamma| < \frac{1}{\sup_{x \in \mathcal{X}} |g(x, \theta_0(P))|} \tag{119}$$

We therefore conclude that if $0 \leq \eta \leq \bar{\eta}(Q)$, $P \in N(Q)$ and $\gamma \in \Xi(\eta, P)$ then

$$P\{1 + \gamma' g(X, \theta_0(P)) \geq 0\} = 1 \ .$$

It follows that if $0 \leq \eta \leq \bar{\eta}(Q)$ and $P \in N(Q)$, then

$$\Xi(\eta, P) \subseteq \Gamma(0, P) \ .$$

Consequently, we obtain that for all $0 \leq \eta \leq \bar{\eta}(Q)$ and $P \in N(Q)$,

$$\begin{aligned} &\min_{\gamma \in A} \max_{(\lambda_0, \lambda_1) \in [0,2]^2} \lambda_0 + \eta(\lambda_1 - 1) - e^{\lambda_0 - 1} \int (1 + \gamma' g(x, \theta_0(P)))^{\lambda_1} 1\{1 + \gamma' g(x, \theta_0(P)) > 0\} dP \\ &= \min_{\gamma \in \Gamma(0, P)} \max_{(\lambda_0, \lambda_1) \in [0,2]^2} \lambda_0 + \eta(\lambda_1 - 1) - e^{\lambda_0 - 1} \int (1 + \gamma' g(x, \theta_0(P)))^{\lambda_1} 1\{1 + \gamma' g(x, \theta_0(P)) > 0\} dP \end{aligned} \tag{120}$$

Arguing as in (113) it then follows that $F(1, 1, 0, \gamma, P) = 0$ for all $\gamma \in \Gamma(0, P)$. To conclude note that since the minimum is attained, we establish using (120) that,

$$\min_{\gamma \in \Gamma(0, P)} \max_{(\lambda_0, \lambda_1) \in [0,2]^2} \lambda_0 + \eta(\lambda_1 - 1) - e^{\lambda_0 - 1} \int (1 + \gamma' g(x, \theta_0(P)))^{\lambda_1} 1\{1 + \gamma' g(x, \theta_0(P)) > 0\} dP \geq 0 \tag{121}$$

Therefore (108), (120) and (121) establish the claim of the Lemma. ■

PROOF OF PART (A) OF THEOREM 4.2: First observe that since $\mathbf{P}(Q) \subseteq \ddot{\mathbf{P}}(Q)$ it follows that $\Lambda_2(\eta) \subseteq \ddot{\Lambda}_2(\eta)$. Hence:

$$\limsup_{n\to\infty} \frac{1}{n} \log P^n\{\hat{P}_n \in \Lambda_2(\eta)\} \leq \limsup_{n\to\infty} \frac{1}{n} \log P^n\{\hat{P}_n \in \ddot{\Lambda}_2(\eta)\} \tag{122}$$

The proof then proceeds by showing the conditions of Lemma 5.3 hold for all $P \in \mathbf{P}_0$ if $\eta > 0$ is sufficiently small. Define

$$\omega_1 = -\log(1-\epsilon) \tag{123}$$

We first show that for all $P \in \mathbf{P}_0$, if $Q \ll P$ and $P\{X \in \text{supp}(Q)\} > \exp(-\omega_1)$, then $\ddot{\mathbf{P}}(Q, \theta_0(P)) \neq \emptyset$. For this purpose note that:

$$\begin{aligned} \sup_{\gamma\neq 0} P\{X \in \text{supp}(Q), \gamma' g(X, \theta_0(P)) \geq 0\} &\leq \sup_{\gamma \neq 0} P\{\gamma' g(X,\theta_0(P)) \geq 0\} \\ &\leq 1-\epsilon \\ &< P\{X \in \text{supp}(Q)\} \end{aligned} \tag{124}$$

where the first inequality follows by inspection, the second inequality by $P \in \mathbf{P}_0$ and the last inequality by hypothesis. It follows from (124) that for all $\gamma \in \mathbf{R}^m$

$$P\{X \in \text{supp}(Q), \gamma' g(X,\theta_0(P)) \geq 0\} > 0 \tag{125}$$

$$P\{X \in \text{supp}(Q), \gamma' g(X,\theta_0(P)) < 0\} > 0 \tag{126}$$

Hence, there exists no hyperplane separating $\text{supp}_Q(g(X_i,\theta_0(P)))$ and $\{0\}$, which implies

$$0 \in \text{int}(co(\text{supp}_Q(g(X_i,\theta_0(P)))))$$

(relative to the topology on $\mathbf{R}^m$). We therefore conclude $\ddot{\mathbf{P}}(Q,\theta_0(P)) \neq \emptyset$ as desired.

To complete the proof, we verify that (67) holds uniformly in $P \in \mathbf{P}_0$ for $\eta > 0$ sufficiently small. By Lemma 5.8, for every $P \in \mathbf{P}_0$, there exists an $\bar{\eta}(P) > 0$ and an open neighborhood in the weak topology $N(P)$ such that for all $0 \leq \eta \leq \bar{\eta}(P)$ we have,

$$\inf_{Q\in N(P)} \inf_{\gamma\in\Gamma(\eta,Q)} \sup_{\lambda_0,\lambda_1\geq 0} \lambda_0 + \eta(\lambda_1 - 1) - e^{\lambda_0 - 1}\int (1+\gamma' g(x,\theta_0(Q)))^{\lambda_1} I\{1+\gamma' g(x,\theta_0(Q)) > 0\} dQ \geq 0$$

By Theorem 15.11 in Aliprantis and Border (2006), $\mathbf{M}$ is compact under the weak topology, and hence since $\mathbf{P}_0 \subseteq \mathbf{M}$ is closed, it is compact as well. Consequently, as

$$\mathbf{P}_0 = \bigcup_{P\in\mathbf{P}_0} N(P)$$

and $N(P)$ are open for all $P \in \mathbf{P}_0$, compactness implies the existence of a finite subcover such that

$$\mathbf{P}_0 = \bigcup_{i=1}^{k} N(P_i) \tag{127}$$

To conclude, let

$$\omega_2 = \min\{\bar{\eta}(P_1), \ldots, \bar{\eta}(P_k)\} \tag{128}$$

and notice that by construction $\omega_2 > 0$ and in addition, for all $0 \le \eta \le \omega_2$

$$\inf_{P \in \mathbf{P}_0} \inf_{\gamma \in \Gamma(\eta, P)} \sup_{\lambda_0, \lambda_1 \ge 0} \lambda_0 + \eta(\lambda_1 - 1) - e^{\lambda_0 - 1} \int (1 + \gamma' g(x, \theta_0(P)))^{\lambda_1} I\{1 + \gamma' g(x, \theta_0) > 0\} dP \ge 0 \tag{129}$$

Letting $\bar{\eta} = \min\{\omega_1, \omega_2\}$ implies the conditions of Lemma 5.3 are satisfied for all $P \in \mathbf{P}_0$ and $0 \le \eta \le \bar{\eta}$, which establishes the claim (a) of the Theorem. ■

Proof of part (b) of Theorem 4.2: The proof closely follows arguments in Kitamura (2001) and Dembo and Zeitouni (1998). Define the set of probability measures,

$$\mathbf{R}(\eta) = \{Q \in \mathbf{M} : \inf_{P \in \mathbf{M}_0 \setminus \mathbf{P}_0} I(Q|P) \ge \eta\} \tag{130}$$

We first aim to show that the proposition,

$$\Lambda_1(\eta) \cap \mathbf{R}(\eta) \subseteq \Omega_{1,n} \cap \mathbf{R}(\eta) \tag{131}$$

holds for all $n > n_0$ and n_0 sufficiently large. Suppose by way of contradiction that there exists an infinite sequence of probability measures $\{\xi_n\}_{n=1}^{\infty}$ such that $\xi_n \in \Lambda_1(\eta) \cap \mathbf{R}(\eta)$ and $\xi_n \in \Omega_{2,n} \cap \mathbf{R}(\eta)$. Since $\mathbf{M}$ is compact under the weak topology by Theorem 15.11 in Aliprantis and Border (2006), there exists a subsequence ξ_{n_k} such that

$$\xi_{n_k} \to \xi \tag{132}$$

for some $\xi \in \mathbf{M}$. Hence, there exists a k_0 such that for all $k \ge k_0$ it follows that $\xi_{n_k} \in B(\xi, \delta/2)$ and therefore $B(\xi, \delta/2) \subset \Omega^{\delta}_{2,n_k}$. Hence, by Sanov's Theorem and various inclusions restrictions,

$$\begin{aligned} \sup_{P \in \mathbf{P}_0} \limsup_{n \to \infty} \frac{1}{n} \log P^n\{\hat{P}_n \in \Omega^{\delta}_{2,n}\} &\ge \sup_{P \in \mathbf{P}_0} \liminf_{n_k \to \infty} \frac{1}{n_k} \log P^{n_k}\{\hat{P}_{n_k} \in \Omega^{\delta}_{2,n_k}\} \\ &\ge \sup_{P \in \mathbf{P}_0} \liminf_{n \to \infty} \frac{1}{n} \log P^n\{\hat{P}_n \in B(\xi, \delta/2)\} \\ &\ge \sup_{P \in \mathbf{P}_0} - \inf_{Q \in B(\xi, \delta/2)} I(Q|P) \\ &\ge \sup_{P \in \mathbf{P}_0} -I(\xi_{n_{k_0}}|P) \end{aligned} \tag{133}$$

Since $\xi_{n_{k_0}} \in \Lambda_1(\eta) \cap \mathbf{R}(\eta)$, it must be that

$$\inf_{P \in \mathbf{M}_0} I(\xi_{n_{k_0}}|P) \le \inf_{P \in \mathbf{P}(\xi_{n_{k_0}})} I(\xi_{n_{k_0}}|P) < \eta \tag{134}$$

by virtue of $\xi_{n_{k_0}} \in \Lambda_1(\eta)$ and $\mathbf{P}(\xi_{n_{k_0}}) \subseteq \mathbf{M}_0$. Furthermore, since $\xi_{n_{k_0}} \in \mathbf{R}(\eta)$ we have,

$$\inf_{P \in \mathbf{M}_0 \setminus \mathbf{P}_0} I(\xi_{n_{k_0}}|P) \ge \eta \tag{135}$$

Hence, combining (134), (135) and $\mathbf{P}_0 \subset \mathbf{M}_0$ we conclude,

$$\inf_{P \in \mathbf{P}_0} I(\xi_{n_{k_0}}|P) < \eta \tag{136}$$

Therefore, it follows from results (133) and (138) that,

$$\sup_{P \in \mathbf{P}_0} \limsup_{n \to \infty} \frac{1}{n} \log P^n\{\hat{P}_n \in \Omega^{\delta}_{2,n}\} > -\eta \tag{137}$$

which contradicts the assumptions on $(\Omega_{1,n}, \Omega_{2,n})$ and hence (131) must be true. Therefore,

$$\limsup_{n\to\infty} \frac{1}{n} \log Q^n\{\hat{P}_n \in \Lambda_1(\eta)\} \leq \limsup_{n\to\infty} \frac{1}{n} Q^n\{\hat{P}_n \in \Omega_{1,n}\} \tag{138}$$

for all $Q \in \mathbf{R}(\eta)$, which establishes claim (b) of the Theorem. ■

Proof of Corollary 4.1: Let $d_{TV}(Q,P)$ and $d_{PL}(Q,P)$ denote the Total Variation and Prokhorov-Lévy metrics between measures Q and P respectively. The claim of the corollary then immediately follows from Theorem 4.2 and the inequalities $I(Q|P) \geq 2d^2_{TV}(Q,P) \geq 2d^2_{PL}(Q,P)$. ■

Lemma 5.9. *(i) If Assumptions 4.1-4.5 hold, then there exists a $\delta > 0$ such that $\mathbf{P}_0 \cap \mathbf{D}_0^\delta = \emptyset$; (ii) If Assumptions 4.4-4.5 hold and $\Theta_0(P)$ is a singleton for every $P \in \mathbf{M}_0 \setminus \mathbf{D}_0$, then for every $\delta > 0$ there exists a $\mathbf{P}_0$ satisfying Assumptions 4.1-4.3 and $\mathbf{M}_0 \setminus \mathbf{D}_0^\delta \subseteq \mathbf{P}_0$.*

Proof: To establish the first claim of the Lemma, suppose by way of contradiction that there exists a sequence $\{P_n\}$ with $P_n \in \mathbf{P}_0$ for all n such that

$$\lim_{n\to\infty} \inf_{Q\in\mathbf{D}_0} d(Q,P_n) = 0 \tag{139}$$

where $d(Q,P)$ is any metric compatible with the weak topology. By Theorem 15.11 in Aliprantis and Border (2006), $\mathbf{M}$ is compact in the weak topology, and hence $\mathbf{P}_0 \subset \mathbf{M}$ is as well by virtue of being closed. Therefore there exists a $P^* \in \mathbf{P}_0$ and subsequence P_{n_k} such that $P_{n_k} \to P^*$. Hence, we obtain from (139) that

$$\inf_{Q\in\mathbf{D}_0} d(Q,P^*) \leq \lim_{n_k\to\infty} \inf_{Q\in\mathbf{D}_0} d(Q,P_{n_k}) + \lim_{n\to\infty} d(P_{n_k},P^*) = 0 \tag{140}$$

Therefore, there exists a sequence $\{Q_n\}$ with $Q_n \in \mathbf{D}_0$ for all n and $Q_n \to P^*$. Hence, there is a sequence $\{\theta_n\}$ with $\theta_n \in \Theta_0(Q_n)$ and $s(Q_n,\theta_n) < m$ for all n, while by compactness of Θ there is a subsequence θ_{n_k} with $\theta_{n_k} \to \theta^*$. Further, it follows from (80) that

$$\int g(x,\theta^*)dP^* = 0 \quad . \tag{141}$$

Since $P^* \in \mathbf{P}_0$, it must be that $\Theta_0(P^*) = \{\theta^*\}$ and $s(P^*,\theta^*) = m$. However, arguing as in (45)-(48) in turn implies $s(Q_{n_k},\theta_{n_k}) = m$ for k sufficiently large, contradicting that $s(Q_n,\theta_n) < m$ for all n.

For the second claim, notice that the arguments in (80) imply $\mathbf{M}_0$ is closed with respect to the weak topology. Hence, by defining

$$\mathbf{P}_0 \equiv \overline{\mathbf{M}_0 \setminus \mathbf{D}_0^\delta} \tag{142}$$

it follows that $\mathbf{P}_0$ satisfies Assumptions 4.1-4.2 and $\mathbf{M}_0 \setminus \mathbf{D}_0^\delta \subseteq \mathbf{P}_0$. We verify $\mathbf{P}_0$ satisfies Assumption 4.3 by way of contradiction. Suppose instead that

$$\sup_{P\in\mathbf{P}_0} \sup_{\|\gamma\|=1} P\{\gamma' g(X_i,\theta_0(P)) \geq 0\} = 1 \tag{143}$$

Letting $\mathbf{S}^m$ denote the unit sphere in $\mathbf{R}^{d_m}$, (143) and compactness of $\mathbf{P}_0 \times \mathbf{S}^m$ implies there exists a sequence $(P_n,\gamma_n) \in \mathbf{P}_0 \times \mathbf{S}^m$ for all n satisfying $(P_n,\gamma_n) \to (P^*,\gamma^*) \in \mathbf{P}_0 \times \mathbf{S}^m$ and

$$\lim_{n\to\infty} P_n\{\gamma_n' g(X_i,\theta_0(P_n)) \geq 0\} = 1 \tag{144}$$

Defining the sets $A_n^+ = \{x \in \mathcal{X} : \gamma_n' g(x, \theta(P_n)) > 0\}$ and $A_n^- = \{x \in \mathcal{X} : \gamma_n' g(x, \theta(P_n)) < 0\}$ we then obtain from (144), $\int g(x, \theta_0(P_n)) dP_n = 0$ and $g(x, \theta)$ bounded on $\mathcal{X} \times \Theta$ that,

$$\begin{aligned}
\limsup_{n\to\infty} \int |\gamma_n' g(x, \theta_0(P_n))| dP_n &= \limsup_{n\to\infty} \{ \int_{A_n^+} \gamma_n' g(x, \theta_0(P_n)) dP_n - \int_{A_n^-} \gamma_n' g(x, \theta_0(P_n)) dP_n \} \\
&= \limsup_{n\to\infty} \int_{A_n^-} 2|\gamma_n' g(x, \theta_0(P_n))| dP_n \\
&\leq \limsup_{n\to\infty} \sup_{x\in\mathcal{X}, \theta\in\Theta} 2|g(x,\theta)| \times P_n\{A_n^-\} \\
&= 0 \qquad (145)
\end{aligned}$$

Since $(P_n, \gamma_n) \to (P^*, \gamma^*)$, Lemma 5.4 and compactness imply $\sup_{x\in\mathcal{X}} |\gamma_n' g(x, \theta_0(P_n)) - \gamma^{*\prime} g(x, \theta_0(P^*)| \to 0$. Hence, (145), $P_n \to P^*$ and $g(x, \theta_0(P^*))$ continuous and bounded yield

$$\begin{aligned}
\int |\gamma^{*\prime} g(x, \theta_0(P^*))| dP^* &\leq \limsup_{n\to\infty} \int |\gamma^{*\prime} g(x, \theta_0(P^*))| (dP^* - dP_n) \\
&+ \limsup_{n\to\infty} \int |\gamma^{*\prime} g(x, \theta_0(P^*)) - \gamma_n' g(x, \theta_0(P_n))| dP_n + \limsup_{n\to\infty} \int |\gamma_n' g(x, \theta_0(P_n))| dP_n = 0 \qquad (146)
\end{aligned}$$

It follows from (146) that $P^* \in \mathbf{D}_0$, which contradict $P^* \in \mathbf{P}_0$ by (142). ■

Proof of Corollary 4.2: By Lemma 5.9 there exists a $\mathbf{P}_0$ satisfying Assumptions 4.1-4.3 such that $\mathbf{M}_0 \setminus \mathbf{D}_0^{\frac{\epsilon}{2}} \subseteq \mathbf{P}_0$. Therefore, by Theorem 4.2 there exists an $\bar{\eta}_1(\epsilon) > 0$ such that for all $\bar{\eta}_1(\epsilon) \geq \eta > 0$ we have

$$\sup_{P\in\mathbf{M}_0\setminus\mathbf{D}_0^{\frac{\epsilon}{2}}} \limsup_{n\to\infty} \frac{1}{n} \log P^n\{\hat{P}_n \in \tilde{\Lambda}_2(\eta) \setminus \mathbf{D}_0^{\epsilon}\} \leq \sup_{P\in\mathbf{P}_0} \limsup_{n\to\infty} \frac{1}{n} \log P^n\{\hat{P}_n \in \Lambda_2(\eta)\} \leq -\eta \qquad (147)$$

Let $d_{PL}(Q, P)$ be the Prokhorov-Lévy metric between measures Q and P. The inclusion $\tilde{\Lambda}_2(\eta) \setminus \mathbf{D}_0^{\epsilon} \subseteq (\mathbf{D}_0^{\epsilon})^c$, Sanov's Theorem and the inequality $I(Q|P) \geq 2d_{PL}^2(Q, P)$ then imply:

$$\begin{aligned}
\sup_{P\in\mathbf{M}_0\cap\mathbf{D}_0^{\frac{\epsilon}{2}}} \limsup_{n\to\infty} \frac{1}{n} \log P^n\{\hat{P}_n \in \tilde{\Lambda}_2(\eta) \setminus \mathbf{D}_0^{\epsilon}\} &\leq \sup_{P\in\mathbf{M}_0\cap\mathbf{D}_0^{\frac{\epsilon}{2}}} \limsup_{n\to\infty} \frac{1}{n} \log P^n\{\hat{P}_n \in (\mathbf{D}_0^{\epsilon})^c\} \\
&\leq \sup_{P\in\mathbf{M}_0\cap\mathbf{D}_0^{\frac{\epsilon}{2}}} - \inf_{Q\in(\mathbf{D}_0^{\epsilon})^c} I(Q|P) \\
&\leq - \inf_{P\in\mathbf{M}_0\cap\mathbf{D}_0^{\frac{\epsilon}{2}}} \inf_{Q\in(\mathbf{D}_0^{\epsilon})^c} 2d_{PL}^2(Q, P) \qquad (148)
\end{aligned}$$

Therefore, results (147), (148) and setting $\bar{\eta}(\epsilon) \leq \min\{\bar{\eta}_1(\epsilon), \epsilon^2/2\}$ establishes part (a) of the Corollary. Furthermore, the same arguments as in (132)-(137) yield $\tilde{\Lambda}_1(\eta) \cup \mathbf{D}_0^{\epsilon} \subseteq \Omega_{1,n} \cup \mathbf{D}_0^{\epsilon}$, which implies $\tilde{\Lambda}_1(\eta) \subseteq \Omega_{1,n}$ thus yielding part (b). ■

References

Aliprantis, C. D. and Border, K. C. (2006). *Infinite Dimensional Analysis – A Hitchhiker's Guide.* Springer-Verlag, Berlin.

Ausubel, L. and Deneckere, R. (1993). A generalized theorem of the maximum. *Economic Theory*, **3** 99–107.

Borwein, J. M. and Lewis, A. S. (1992a). Partially finite convex programming part I : Quasi-relative interiors and duality theory. *Mathematical Programming*, **57** 15–48.

Borwein, J. M. and Lewis, A. S. (1992b). Partially finite convex programming part II : Explicit lattice models. *Mathematical Programming*, **57** 49–83.

Borwein, J. M. and Lewis, A. S. (1993). Partially finite programming in L_1 and the existence of maximum entropy estimates. *SIAM Journal of Optimization*, **3** 248–267.

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

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, **50** 891–916.

Hansen, L. P., Heaton, J. and Yaron, A. (1996). Finite sample properties of some alternative GMM estimators. *Journal of Business and Economics Statistics*, **14** 262–280.

Hoeffding, W. (1965). Asymptotically optimal tests for multinomial distributions. *Annals of Mathematical Statistics*, **36** 369–401.

Imbens, G., Spady, R. and Johnson, P. (1998). Information theoretic approaches to inference in moment condition models. *Econometrica*, **66** 333–357.

Kitamura, Y. (2001). Asymptotic optimality of empirical likelihood for testing moment restrictions. *Econometrica*, **69** 1661–1672.

Kitamura, Y. and Stutzer, M. (1997). An information-theoretic alternative to generalized method of moments esitmation. *Econometrica*, **65** 861–874.

Newey, W. K. and Smith, R. J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica*, **72** 219–255.

Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75** 237–249.

Qin, J. and Lawless, J. (1994). Empirical likelihood and generalized estimating equations. *Annals of Statistics*, **22** 300–325.

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics.* Springer-Verlag, New York.

Zeitouni, O. and Gutman, M. (1991). On universal hypothesis testing via large deviations. *IEEE Transactions on Information Theory*, **37** 285–290.