# EMPIRICAL LIKELIHOOD FOR RANDOM SETS

KARUN ADUSUMILLI AND TAISUKE OTSU

Abstract. We extend the method of empirical likelihood to cover hypotheses involving the Aumann expectation of random sets. By exploiting the properties of random sets, we convert the testing problem into one involving a continuum of moment restrictions for which we propose two inferential procedures. The first, which we term marked empirical likelihood, corresponds to constructing a non-parametric likelihood for each moment restriction and assessing the resulting process. The second, termed sieve empirical likelihood, corresponds to constructing a likelihood for a vector of moments with growing dimension. We derive the asymptotic distributions under the null and sequence of local alternatives for both types of tests and prove their consistency. The applicability of these inferential procedures is demonstrated in the context of two examples on the mean of interval observations and best linear predictors for interval outcomes.

## 1. Introduction

Since the seminal paper by Owen (1988), a number of papers have sought to extend the method of empirical likelihood to cover various hypothesis testing problems. This is because empirical likelihood is a non-parametric likelihood method, thus inheriting the latter's good power properties while at the same being flexible and able to incorporate side information (Owen, 2001). A detailed overview of the many empirical likelihood based methods may be found in Owen (2001) and Chen and van Keilegom (2009).

The aim of this paper is to extend the method of empirical likelihood to cover hypotheses involving random sets. Informally, an object is said to be a random set if it has set-valued realizations. We refer to Molchanov (2005) for a comprehensive survey on the theory and applications of random sets. Recently, applications of random set methods have been discussed in the context of partial identification and inference in econometrics. Partial identification concerns the situation wherein a parameter of interest in not point identified but only as a set. This could be because of limitations in the data, e.g. interval or categorical data, or because the theoretical models do not provide enough restrictions to identify a unique value for the parameter, e.g. game theoretic models with multiple equilibria. Beresteanu and Molinari (2008) were the first to employ random set methods to obtain estimation and inference for partially identified models. Other applications of random set theory in the context of partial inference include Bontemps, Magnac and Maurin (2012), Chandrasekhar *et al.* (2012), Chernozukhov, Kocatulum and Menzel (2012), Kaido (2012), Kaido and Santos (2014). Molchanov and Molinari (2014) provide a review of the recent developments in the applications and use of random set theory in econometrics.

In this paper we consider hypothesis tests involving the Aumann expectation (which is a generalization of the expectation operator to random sets) of random convex and compact sets.

The authors would like to thank Hiroaki Kaido for helpful comments. Financial support from the ERC Consolidator Grant (SNP 615882) is gratefully acknowledged (Otsu).

By relying on the isomorphism between a convex set and its support function, we convert the testing problem to one involving inference on the support function that implies a continuum of moment restrictions for each direction of the support function. More generally, our procedures can be extended to random compact sets under the additional assumption that the underlying probability distribution is non-atomic. We propose two inferential procedures based on empirical likelihood for testing the continuum of moment restrictions, which we term marked empirical likelihood and sieve empirical likelihood.

The idea behind the marked empirical likelihood is to construct a non-parametric likelihood for each direction of the support function and assess the resulting process over the domain of all possible directions. Consistent tests are then obtained using Kolmogorov-Smirnov or Cramér-von Mises type test statistics. Similar methods have previously been employed by Einmahl and McKeague (2003) in the context of omnibus tests for non-parametric hypotheses and van Keilegom, Sellero and Manteiga (2008) in the context of testing for regression. The test statistic is shown to converge to a Gaussian process under the null and possess non-trivial power against local alternatives converging to the null at the $n^{-1/2}$ rate. Since the asymptotic distribution of the test statistic contains unknown parameters to be estimated, we provide a bootstrap calibration to approximate the critical values.

An alternative way to construct a likelihood function to test the continuum of moment restrictions is to employ a vector of moments with growing dimension. This is called the sieve empirical likelihood method. Empirical likelihood methods under a growing number of parameters have previously been considered in Hjort, McKeague and van Keilegom (2008) and Chen, Peng and Qin (2009) for example. In this context, we generalize the results of Hjort, McKeague and van Keilegom (2008) by allowing for arbitrary growth rates on the eigenvalues of the covariance matrix. In particular, the best condition on the growth rate (when the random sets are almost surely bounded) is shown to be $k^5\phi_k^{-6}/n \to 0$, where $k$ is the dimension of the vector of moments and $\phi_k$ is the smallest eigenvalue of the variance matrix. Under this assumption, and in line with the findings of most of the literature on growing number of moments, we find that the test statistic converges to the standard normal distribution. We also show that the test statistic has non-trivial power against local alternatives converging to the null at the rate $k^{1/4}n^{-1/2}$ in all directions and faster than $n^{-1/2}$ rate in some directions.

We further generalize both inferential procedures to allow for the presence of finite dimensional nuisance parameters. In particular, we find that in the presence of nuisance parameters, the sieve empirical likelihood is no longer first-order efficient in the sense that it is not internally studentized. We thus propose to add a penalty term in the dual representation of the empirical likelihood objective function to restore the efficiency. The penalty is akin to a one-step Newton-Raphson approximation that affects the variance term.

The applicability of our testing procedures is demonstrated using two examples on the mean of interval observations (reviewed in Manski (2003) and treated via random set methods in Beresteanu and Molinari (2008)) and best linear predictors for interval outcomes (Beresteanu and Molinari (2008), Bontemps, Magnac and Maurin (2012), Chandrasekhar *et al.* (2012)).

The structure of the paper is as follows. In Section 2 we introduce the notation and basic setup and propose two inferential procedures for testing hypotheses involving the mean (using the Aumann expectation) of random sets. We then illustrate these procedures with an example of the mean of interval observations. In Section 3 we extend these methods by including finite dimensional nuisance parameters which have to be estimated. As an example, we discuss the problem of inference on the set of best linear predictors with interval outcomes. All proofs are contained in the Appendix.

## 2. Prototype

We first introduce the notation and basic setup. The basic concepts discussed here follow from Molchanov (2005). Let $(\Omega, \mathcal{B}, \mu)$ be a probability space and $\mathbb{K}^d$ be the collection of all non-empty closed subsets of the $d$-dimensional Euclidean space $\mathbb{R}^d$. Let $\|\cdot\|$ and $\langle \cdot, \cdot \rangle$ be the Euclidean norm and inner product, respectively. The mapping $X : \Omega \mapsto \mathbb{K}^d$ is said to be a set-valued random variable (SVRV) if it holds $X^{-1}(C) = \{\omega \in \Omega : X(\omega) \cap C \neq \emptyset\} \in \mathcal{B}$ for each closed subset $C$ of $\mathbb{R}^d$. As a concept for the mathematical expectation of the SVRV $X$, we introduce the Aumann expectation $\mathbb{E}[X]$, that is

$$\mathbb{E}[X] = \left\{ \int_\Omega x d\mu : x \in \{x(\omega) \in X(\omega) \text{ a.s. and } \int_\Omega \|x\| \, d\mu < \infty\} \right\}.$$

Note that in this definition, $x(\omega)$ or $x$ is a random variable and $X(\omega)$ or $X$ is a SVRV. Hereafter, $\mathbb{E}[\cdot]$ denotes the Aumann expectation of a SVRV and $E[\cdot]$ denotes the expectation of a random variable. Let $\|A\|_H = \sup\{\|a\| : a \in A\}$ denote the Hausdorff norm of a set $A$. The SVRV $X$ is said to be integrably bounded if $E[\|X\|_H] < \infty$.

In this paper we consider SVRVs whose realizations are compact and convex sets. More generally, under the additional assumption of non-atomic probability measure $\mu$, the results in this paper can be extended to random compact sets by taking the convex hull operation, denoted by $\mathrm{co}(X)$. This follows from the result that $\mathbb{E}[X] = \mathbb{E}[\mathrm{co}(X)]$ for a compact SVRV $X$ if $\mu$ is non-atomic (Molchanov, 2005, Theorem 1.17 on p. 154). In particular, we note also that $\mathrm{co}(X)$ is a compact and convex valued SVRV if $X$ is a compact SVRV (by Molchanov (2005, Theorem 2.2.5 on p. 37) and the fact that the convex hull of a compact set is also compact). Hence for the remainder of this paper we restrict attention to compact and convex valued SVRVs with the implicit assumption that similar results also hold for compact sets when $\mu$ is non-atomic.

Let $\mathbb{K}^d_{kc}$ be the collection of all non-empty compact and convex sets of $\mathbb{R}^d$. The collection $\mathbb{K}^d_{kc}$ is endowed with the Hausdorff metric

$$d_H(A, B) = \max\{\sup_{a \in A} \inf_{b \in B} \|a - b\|, \sup_{b \in B} \inf_{a \in A} \|b - a\|\},$$

for compact and convex sets $A$ and $B$. It can be shown that the mapping $X : \Omega \mapsto \mathbb{K}^d_{kc}$ is a SVRV if and only if $X$ is $\mathcal{B}(\mathbb{K}^d_{kc})$ measurable, where $\mathcal{B}(\mathbb{K}^d_{kc})$ is the Borel $\sigma$-algebra generated by the Hausdorff metric on $\mathbb{K}^d_{kc}$ .[1] Let $\mathbb{S}^{d-1}$ be the unit sphere in $\mathbb{R}^d$ and $s(A, p) = \sup_{x \in A} \langle p, x \rangle$

[1]To observe this, define the Effros $\sigma$-algebra as the smallest $\sigma$-algebra generated by $\{A \in \mathbb{K}^d : A \cap C \neq \emptyset\}$ for $C$ running through the family of sets $\mathbb{K}^d$. It is known (e.g., Molchanov, 2005, p. 2) that the definition of SVRV is equivalent to $X : \Omega \mapsto \mathbb{K}^d$ being measurable with respect to the Effros $\sigma$-algebra. The claim then follows from

for $p \in \mathbb{S}^{d-1}$ be the support function of a compact convex set $A$. Note that by Beresteanu and Molinari (2008, equation (A.1)), the Hausdorff metric can be expressed in terms of support functions as

$$d_H(A,B) = \sup_{p \in \mathbb{S}^{d-1}} |s(A,p) - s(B,p)|,$$

for $A, B \in \mathbb{K}_{kc}^d$. Thus by the definition of $\mathcal{B}(\mathbb{K}_{kc}^d)$, $s(.,p) : \mathbb{K}_{kc}^d \mapsto \mathbb{R}$ is a measurable function for each $p \in \mathbb{S}^{d-1}$. This provides an alternative way to characterize the measurability of a compact and convex valued multifunction, i.e., $X : \Omega \mapsto \mathbb{K}_{kc}^d$ is a SVRV if and only if $\{s(X,p), p \in \mathbb{S}^{d-1}\}$ is a collection of random variables.

In this section, we consider empirical likelihood inference for testing the hypothesis

$$H_0 : \mathbb{E}[X] = \Theta_0 \text{ against } H_1 : \mathbb{E}[X] \neq \Theta_0,$$

based on an independent and identically distributed (i.i.d.) sequence of compact and convex SVRVs $\{X_1, \ldots, X_n\}$. We do not specify any parametric distribution form on $\mu$. By Molchanov (2005, Theorem 1.22 on p. 157) and linearity of the support function, under the assumption that $X$ is integrably bounded, the above testing problem is equivalent to testing

$$H_0 : E[s(X,p)] = s(\Theta_0, p) \text{ for all } p \in \mathbb{S}^{d-1} \text{ against } H_1 : E[s(X,p)] \neq s(\Theta_0,p) \text{ for some } p \in \mathbb{S}^{d-1},$$

where $E[\cdot]$ is the ordinary mathematical expectation with respect to the measure $\mu$. Therefore, inference on the Aumann expectation is equivalent to inference on the support function (or continuum of moment restrictions over $p \in \mathbb{S}^{d-1}$). Since this is a testing problem for infinite dimensional parameters without any parametric distributional assumption on the population $\mu$, it is of interest to develop certain non-parametric likelihood methods for this purpose. In the following subsections we propose two likelihood concepts and test statistics.

2.1. **Inference via marked empirical likelihood.** One way to construct a non-parametric likelihood function to test $H_0$ is to fix the direction $p \in \mathbb{S}^{d-1}$ in the support function and employ the empirical likelihood approach. For given $p$, the marked empirical likelihood function under the restriction $E[s(X,p)] = s(\Theta_0, p)$ is written as

$$\ell(p) = \max \left\{ \prod_{i=1}^n n w_i \middle| \sum_{i=1}^n w_i s(X_i, p) = s(\Theta_0, p), \ w_i \geq 0, \ \sum_{i=1}^n w_i = 1 \right\}.$$

For each $p$, a non-parametric version of Wilks' theorem (Owen, 1988) provides the null limiting distribution $-2 \log \ell(p) \xrightarrow{d} \chi_1^2$ under $H_0$. Note that the empirical likelihood function $\ell(p)$ marked by the direction $p$ imposes only a single restriction implied from the null $H_0$. In order to guarantee consistency against any departure from $H_0$, we need to assess the whole process $\{\ell(p) : p \in \mathbb{S}^{d-1}\}$ over the range of $\mathbb{S}^{d-1}$. Taking the supremum leads to the Kolmogorov-Smirnov type test

Molchanov (2005, Theorem 2.7 (iii) on p. 29) (which states that the Effros $\sigma$-algebra induced on the family of compact sets coincides with the Borel $\sigma$-algebra generated by the Hausdorff metric) and the fact that $\mathbb{K}_c^d$ and $\mathbb{K}_{kc}^d$ are both Effros measurable (Molchanov 2005, pages 20 and 64).

statistic[2]

$$K_n = \sup_{p \in \mathbb{S}^{d-1}} \{-2 \log \ell(p)\}.$$

In practice, the empirical likelihood function $\ell(p)$ can be computed by its dual form based on the Lagrange multiplier method, that is

$$\ell(p) = \prod_{i=1}^{n} \frac{1}{1 + \lambda\{s(X_i, p) - s(\Theta_0, p)\}},$$

where $\lambda$ solves the first-order condition $\sum_{i=1}^{n} \frac{s(X_i,p)-s(\Theta_0,p)}{1+\lambda\{s(X_i,p)-s(\Theta_0,p)\}} = 0$. The asymptotic property of the test statistic $K_n$ is obtained as follows.

**Theorem 1.** *Suppose $\{X_1, \ldots, X_n\}$ is an i.i.d. sequence of compact and convex SVRSs satisfying $E[\|X_i\|_H^{\xi}] < \infty$ for some $\xi > 2$ and $\inf_{p \in \mathbb{S}^{d-1}} \mathrm{Var}(s(X_i, p)) > 0$. Then the followings hold true.*

- **(i):** *Under the null hypothesis $H_0$,*

$$K_n \xrightarrow{d} \sup_{p \in \mathbb{S}^{d-1}} \frac{Z(p)^2}{E[Z(p)^2]},$$

  *where $Z$ is the Gaussian process with zero mean and covariance kernel $E[s(X,p)s(X,q)] - E[s(X,p)]E[s(X,q)]$.*
- **(ii):** *Under the alternative hypothesis $H_1$, $K_n$ diverges to infinity.*
- **(iii):** *Under the local alternative hypothesis $H_{1n} : E[s(X_i, p)] = s(\Theta_0, p) + n^{-1/2}\eta(p)$ for all $p \in \mathbb{S}^{d-1}$ with a continuous function $\eta$ on $\mathbb{S}^{d-1}$,*

$$K_n \xrightarrow{d} \sup_{p \in \mathbb{S}^{d-1}} \frac{\{Z(p) + \eta(p)\}^2}{E[Z(p)^2]}.$$

This theorem says that the null distribution of the Kolmogorov-Smirnov type statistic $K_n$ based on the marked empirical likelihood is characterized by the Gaussian process $Z$ and that the test based on $K_n$ can possess non-trivial power against local alternatives at distance proportional to $n^{-1/2}$ in terms of some continuous function from the null hypothesis. Beresteanu and Molinari (2008) consider local alternative hypotheses in the form of $H_{1n}^* : \mathbb{E}[X_i] = \Theta_{An}$ for some sequence $\{\Theta_{An}\}$ of non-empty compact convex sets given by $\Theta_{An} \oplus n^{-1/2}\Delta_1 = \Theta_0 \oplus n^{-1/2}\Delta_2$, where $\Delta_1$ and $\Delta_2$ are compact convex sets for which there exists a third convex set $\Delta_3$ such that $\Theta_0 = \Delta_1 \oplus \Delta_3$ (here $\oplus$ denotes the Minkowski summation). It is known that $s(\Psi_1 \oplus \Psi_2, \cdot) = s(\Psi_1, \cdot) + s(\Psi_2, \cdot)$ for any convex sets $\Psi_1$ and $\Psi_2$. Thus in terms of the support function, the local alternative hypothesis $H_{1n}^*$ is equivalently written as $s(\Theta_{An}, p) = s(\Theta_0, p) + n^{-1/2}\{s(\Delta_1, p) - s(\Delta_2, p)\}$ for all $p \in \mathbb{S}^{d-1}$. Consequently, Beresteanu and Molinari's (2008) local alternative $H_{1n}^*$ is equivalent to our $H_{1n}$ with $\eta(p) = s(\Delta_1, p) - s(\Delta_2, p)$. Therefore, the test statistic $K_n$ not only has non-trivial power against the choice of local alternatives in Beresteanu and Molinari (2008) but also against a much wider class of deviations from the null hypothesis.

[2] Although we focus on the Kolmogorov-Smirnov type statistic, it is also possible to employ other functionals of the marked empirical likelihood process, such as the Cramér-von Mises type statistic $\int_{p \in \mathbb{S}^{d-1}} -2 \log \ell(p) dp$.

Note that the limiting null distribution of the statistic $K_n$ contains parameters to be estimated. Thus we suggest estimating the critical values of the test statistic by the following bootstrap procedure. Let $\bar{s}(p) = n^{-1}\sum_{i=1}^n s(X_i, p)$.

**Algorithm.**

(1) Generate a bootstrap analog of support functions $\{s(X_i^*, p)\}_{i=1}^n$ of size $n$ by drawing from a random sample of the empirical distribution of $\{s(X_i, p)\}_{i=1}^n$ with replacement.

(2) Compute the bootstrap counterpart $K_n^* = \sup_{p\in\mathbb{S}^{d-1}}\{-2\log \ell^*(p)\}$, where

$$\ell^*(p) = \max\left\{\prod_{i=1}^n nw_i \middle| \sum_{i=1}^n w_i s(X_i^*, p) = \bar{s}(p),\ w_i \geq 0,\ \sum_{i=1}^n w_i = 1\right\}.$$

(3) Repeat Steps (1) and (2) to compute the empirical distribution $F_n^*$ of $K_n^*$.

(4) Estimate the $\alpha$-th quantile of the limiting distribution of $K_n$ by $\hat{c}_{n,\alpha} = \inf\{t : F_n^*(t) \geq 1-\alpha\}$.

The validity of this bootstrap procedure can be seen as follows. The bootstrap counterpart $K_n^*$ weakly converges to the limiting distribution of $K_n$, $\mu$-almost surely (a.s.) by an argument similar to the proof of Theorem 1 (i) along with the facts that (a) the empirical process $n^{1/2}\bar{m}^*(\cdot)$ weakly converges to $Z(\cdot)$, $\mu$-a.s. (Giné and Zinn, 1990, Theorem 2.4); and (b) $\sup_{p\in\mathbb{S}^{d-1}}|\hat{V}^*(p) - E[Z(p)^2]| \xrightarrow{p^*} 0$, $\mu$-a.s. (where $\bar{m}^*(\cdot) = n^{-1}\sum_{i=1}^n\{s(X_i^*, p) - \bar{s}(p)\}$, $\hat{V}^*(p) = n^{-1}\sum_{i=1}^n\{s(X_i^*, p) - \bar{s}(p)\}^2$, and $P^*$ denotes the bootstrap probability conditional on the data). The latter is a consequence of $\sup_{p\in\mathbb{S}^{d-1}}|\hat{V}^*(p) - \mathrm{Var}^*(s(X_i^*, p))| \xrightarrow{p^*} 0$, $\mu$-a.s. (Giné and Zinn, 1990, Theorem 2.6) and $\sup_{p\in\mathbb{S}^{d-1}}|\mathrm{Var}^*(s(X_i^*, p)) - E[Z(p)^2]| \to 0$ $\mu$-a.s. (by a uniform law of large numbers), where $\mathrm{Var}^*(s(X_i^*, p)) = n^{-1}\sum_{i=1}^n\{s(X_i, p) - \bar{s}(p)\}^2$ is the variance under the bootstrap distribution. Furthermore, by the Corollary of Lifshits (1982), the limiting distribution of $K_n$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$ since $\mathrm{Var}(Z(p)/E[Z(p)^2]^{1/2}) = 1 > 0$ for all $p \in \mathbb{S}^{d-1}$. It thus follows that the bootstrap critical value $\hat{c}_{n,\alpha}$ converges in probability to the $\alpha$-th quantile of the limiting distribution of $K_n$.

We now compare our empirical likelihood statistic $K_n$ with the existing one. Beresteanu and Molinari (2008) proposed a Wald type test statistic for $H_0$ based on the Hausdorff distance between the null hypothetical set $\Theta_0$ and the average $\frac{1}{n}\oplus_{i=1}^n X_i$ based on the Minkowski summation, that is

$$W_n = \sqrt{n} d_H\left(\frac{1}{n}\oplus_{i=1}^n X_i, \Theta_0\right).$$

It should be noted that for convex sets, the Wald type statistic $W_n$ is alternatively written by using the support functions as (see Beresteanu and Molinari, 2008, equation (A.1))

$$W_n = \sqrt{n}\sup_{p\in\mathbb{S}^{d-1}}\left|\frac{1}{n}\sum_{i=1}^n s(X_i, p) - s(\Theta_0, p)\right|.$$

Based on the proof of Theorem 1, we can see that

$$K_n^{1/2} = \sqrt{n}\sup_{p\in\mathbb{S}^{d-1}} E[Z(p)^2]^{-1/2}\left|\frac{1}{n}\sum_{i=1}^n s(X_i, p) - s(\Theta_0, p)\right| + o_p(1).$$

Therefore, while the Wald type statistic $W_n$ by Beresteanu and Molinari (2008) evaluates the contrast $\frac{1}{n}\sum_{i=1}^n s(X_i,p)-s(\Theta_0,p)$ over $p\in\mathbb{S}^{d-1}$, the empirical likelihood statistic $K_n$ evaluates the contrast normalized by its standard deviation.

2.2. **Inference via sieve empirical likelihood.** Another way to construct a non-parametric likelihood function to test $H_0$ is to incorporate the continuum of moment conditions $E[s(X,p)] = s(\Theta_0,p)$ for all $p\in\mathbb{S}^{d-1}$ by a vector of moments with growing dimension. Let $k=k_n\le n/2$ be a sequence of positive integers satisfying $k\to\infty$ as $n\to\infty$, and choose points (or sieve) $\{p_1,\ldots,p_k\}$ from $\mathbb{S}^{d-1}$ so that in the limit they form a dense subset of $\mathbb{S}^{d-1}$. The sieve empirical likelihood function under the restrictions $E[s(X,p_j)] = s(\Theta_0,p_j)$ for $j=1,\ldots,k$ is written as

$$l(p_1,\ldots,p_k)=\max\left\{\prod_{i=1}^n nw_i\left|\sum_{i=1}^n w_i s(X_i,p_j)=s(\Theta_0,p_j)\text{ for } j=1,\ldots,k,\ w_i\ge 0,\ \sum_{i=1}^n w_i=1\right.\right\}.$$

Let $m_k(X_i)=(s(X_i,p_1)-s(\Theta_0,p_1),\ldots,s(X_i,p_k)-s(\Theta_0,p_k))'$. In practice, the sieve empirical likelihood function $l(p_1,\ldots,p_k)$ can be computed by its dual form based on the Lagrange multiplier method, that is

$$l(p_1,\ldots,p_k)=\prod_{i=1}^n\frac{1}{1+\gamma' m_k(X_i)},$$

where $\gamma$ solves the first-order condition $\sum_{i=1}^n\frac{m_k(X_i)}{1+\gamma' m_k(X_i)}=0$. The test statistic for $H_0$ is

$$L_n=-2\log l(p_1,\ldots,p_k). \tag{1}$$

The asymptotic property of the sieve empirical likelihood statistic $L_n$ is obtained as follows. Let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ be the minimum and maximum eigenvalues of a matrix $A$, respectively. Also define $\phi_k=\lambda_{\min}(\mathrm{Var}(m_k(X_i)))$ and $\hat\phi_k=\lambda_{\min}(n^{-1}\sum_{i=1}^n m_k(X_i)m_k(X_i)')$.

**Theorem 2.** *Suppose $\{X_1,\ldots,X_n\}$ is an i.i.d. sequence of compact and convex SVRVs satisfying $E[\|X_i\|_H^\xi]<\infty$ for some $\xi\ge 4$. Also assume $k\to\infty$ and $(k^5\phi_k^{-6})^{\frac{\xi}{\xi-2}}/n\to 0$ as $n\to\infty$. Then the followings hold true.*

**(i):** *Under the null hypothesis $H_0$,*

$$\frac{L_n-k}{\sqrt{2k}}\xrightarrow{d}N(0,1).$$

**(ii):** *Under the alternative hypothesis $H_1$, $(L_n-k)/\sqrt{2k}$ diverges to infinity.*

**(iii):** *Under the sequence of local alternative hypotheses $H_{1n}: E[s(X_i,p)]=s(\Theta_0,p)+a_n\eta(p)$ for all $p\in\mathbb{S}^{d-1}$ with a function $\eta$ on $\mathbb{S}^{d-1}$, where $a_n=k^{1/4}/\sqrt{n\eta_k'\mathrm{Var}(m_k(X_i))^{-1}\eta_k}$ and $\eta_k=(\eta(p_1),\ldots,\eta(p_k))'$,*

$$\frac{L_n-k}{\sqrt{2k}}\xrightarrow{d}N(2^{-1/2},1).$$

The assumptions of Theorem 2 differ from those of Hjort, McKeague and van Keilegom (2009, Theorem 2.1) in that the lowest eigenvalue of $\mathrm{Var}(m_k(X_i))$ explicitly enters the rate condition for $k$. This is because the assumption of bounded eigenvalues imposed in Hjort, McKeague and van Keilegom (2009) is typically violated in our examples. Indeed for the special case when the eigenvalues of $\mathrm{Var}(m_k(X_i))$ are bounded from both above and below, inspection of the proof of Theorem 2 shows that the rate condition may be relaxed to $k^3/n \to 0$ which is the one obtained in Hjort, McKeague and van Keilegom (2009). It may be also noted from the proof that $|\hat{\phi}_k - \phi_k| = O_p(k/\sqrt{n})$, hence the rate condition may be equivalently written as $(k^5\hat{\phi}_k^{-6})^{\frac{\xi}{\xi-2}}/n \to 0$ where $\hat{\phi}_k$ is observable. Thus to apply Theorem 2, the number of points in the sieve should be chosen in such a way that $\hat{\phi}_k$ is not too low. One way to achieve this in practice is to use a hard threshold for $\hat{\phi}_k$ such as $\hat{\phi}_k \geq C(\ln n)^{-1}$ for some positive constant $C$. Then the rate condition will take the form of $k^5(\ln n)^6/n \to 0$.

We argue in the proof of Theorem 2 that $\lambda_{\max}(\mathrm{Var}(m_k(X_i))) = O(k)$. Thus Part (iii) of Theorem 2 assures that the test based on $L_n$ has non-trivial power against local alternatives at distance proportional to at least $k^{1/4}n^{-1/2}$ in terms of some continuous function from the null hypothesis. However this is not a strict bound as for some directions the statistic $L_n$ is able to distinguish alternatives that converge to the null faster than $n^{-1/2}$, for e.g. when $\eta_k$ is taken to be the eigenvector corresponding to the smallest eigenvalue of $\mathrm{Var}(m_k(X_i))$.

2.3. **Example: Mean of interval observations.** As an illustration, let us consider inference for the mean of interval observations. Suppose we observe an i.i.d. sequence of intervals $\{X_1, \ldots, X_n\}$, where $X_i = [x_{Li}, x_{Ui}] \subset \mathbb{R}$ with $x_{Li} \leq x_{Ui}$ almost surely. By Beresteanu and Molinari (2008, Lemma A.2), $\{X_1, \ldots, X_n\}$ is a sequence of i.i.d compact and convex SVRVs. In this case, $\mathbb{S}^{d-1} = \{-1, 1\}$ is finite, and the support function is written as

$$s(X_i, p) = \begin{cases} -x_{Li} & \text{for } p = -1, \\ x_{Ui} & \text{for } p = 1. \end{cases}$$

Therefore, testing $H_0 : \mathbb{E}[X_i] = \Theta_0 = [\theta_L, \theta_U]$ is equivalent to testing two moment restrictions $E[x_{Li}] = \theta_L$ and $E[x_{Ui}] = \theta_U$.

In this example, the Kolmogorov-Smirnov type test statistic based on the marked empirical likelihood is written as

$$K_n = 2\max\left\{\sum_{i=1}^n \log(1 + \lambda_L(-x_{Li} + \theta_L)), \sum_{i=1}^n \log(1 + \lambda_U(x_{Ui} - \theta_U))\right\},$$

where $\lambda_L$ and $\lambda_U$ solve $\sum_{i=1}^n \frac{-x_{Li}+\theta_L}{1+\lambda_L(-x_{Li}+\theta_L)} = 0$ and $\sum_{i=1}^n \frac{x_{Ui}-\theta_U}{1+\lambda_U(x_{Ui}-\theta_U)} = 0$, respectively. By applying Theorem 1,

$$K_n \xrightarrow{d} \max\{z_L^2/\sigma_L^2, z_U^2/\sigma_U^2\},$$

under $H_0$, where

$$\begin{pmatrix} z_L \\ z_U \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_L^2 & \rho\sigma_L\sigma_U \\ \rho\sigma_L\sigma_U & \sigma_U^2 \end{pmatrix}\right),$$

with $\sigma_L^2 = \mathrm{Var}(x_{Li})$, $\sigma_U^2 = \mathrm{Var}(x_{Ui})$, and $\rho = \mathrm{Corr}(x_{Li}, x_{Ui})$. Note that the (square of) Wald type statistic based on the Hausdorff distance by Beresteanu and Molinari (2008) is

$$W_n^2 = nd_H\left(\frac{1}{n}\oplus_{i=1}^n X_i, \Theta_0\right)^2 = n\max\left\{\left(-\frac{1}{n}\sum_{i=1}^n x_{Li} + \theta_L\right)^2, \left(\frac{1}{n}\sum_{i=1}^n x_{Li} - \theta_U\right)^2\right\},$$

which converges in distribution to $\max\{z_L^2, z_U^2\}$ under $H_0$.

By incorporating the restrictions $E[x_{Li}] = \theta_L$ and $E[x_{Ui}] = \theta_U$, the (sieve) empirical likelihood statistic in (1) reduces to the conventional one, that is

$$L_n = 2\sum_{i=1}^{n} \log(1 + \gamma' m_2(X_i)),$$

where $m_2(X_i) = (x_{Li} - \theta_L, x_{Ui} - \theta_U)'$ and $\gamma$ solves the first-order condition $\sum_{i=1}^{n} \frac{m_2(X_i)}{1+\gamma' m_2(X_i)} = 0$. By applying the standard arguments (Owen, 1988), it holds that $L_n \xrightarrow{d} \chi^2_2$ under $H_0$.

## 3. Generalization: Nuisance parameters

In the last section, we proposed two inference methods to test the simple hypothesis $H_0 : \mathbb{E}[X] = \Theta$ for a SVRV. However, for some cases, the null hypothesis is written in the form of

$$H_0 : \mathbb{E}[X] = \Theta_0(\nu),$$

where $\{\Theta(\nu) : \nu \in \mathbb{R}^r\}$ is a collection of convex sets in $\mathbb{R}^d$ indexed by $r$-dimensional nuisance parameters $\nu$. We consider the situation where $\nu = h(E[z])$ is a smooth function of means of a random vector $z \in \mathbb{R}^{r_1}$. We estimate $\nu$ by the method of moments estimator $\hat{\nu} = h(\bar{z})$ where $\bar{z} = n^{-1}\sum_{i=1}^{n} z_i$. This setup is general enough to accommodate the existing examples such as best linear prediction with interval outcomes (Beresteanu and Molinari (2008), Chandrasekhar *et al.* (2012)).

By using the support function, the testing problem can be written as

$H_0 : E[s(X,p)] = s(\Theta_0(\nu), p)$ for all $p \in \mathbb{S}^{d-1}$ against $H_1 : E[s(X,p)] \neq s(\Theta_0(\nu), p)$ for some $p \in \mathbb{S}^{d-1}$.

If we plug-in the method of moments estimator $\hat{\nu}$, the marked empirical likelihood function is obtained as

$$\ell(p, \hat{\nu}) = \max\left\{ \prod_{i=1}^{n} nw_i \middle| \sum_{i=1}^{n} w_i s(X_i, p) = s(\Theta_0(\hat{\nu}), p),\ w_i \geq 0,\ \sum_{i=1}^{n} w_i = 1 \right\},$$

and the Kolmogorov-Smirnov type test statistic is written as

$$K_n(\hat{\nu}) = \sup_{p \in \mathbb{S}^{d-1}} \{-2 \log \ell(p, \hat{\nu})\}.$$

To study the asymptotic properties of the test statistic $K_n(\hat{\nu})$, we impose the following assumptions.

**Assumption N.** *Suppose that $\{(X_1, z_1), \ldots, (X_n, z_n)\}$ is an i.i.d. sequence of pairs of compact and convex SVRSs and random vectors, and that $\nabla h(E[z])$ has full row rank, $\|\Theta_0(\hat{\nu})\|_H = O_p(1)$, and for some neighborhood $\mathcal{N}$ of $\nu$ there exists a function $G(\cdot;.) : \mathbb{S}^{d-1} \times \mathcal{N} \mapsto \mathbb{R}^r$ such that $G(p;\nu)$ is continuous with respect to $p$ at the true value $\nu$, $\sup_{p \in \mathbb{S}^{d-1}} \|G(p;\hat{\nu}) - G(p;\nu)\| \xrightarrow{p} 0$, and*

$$\sup_{p \in \mathbb{S}^{d-1}} |s(\Theta_0(\hat{\nu}), p) - s(\Theta_0(v), p) - G(p;\nu)'(\hat{\nu} - \nu)| = o_p(n^{-1/2}). \quad (2)$$

**Theorem 3.** *Suppose that Assumption N holds and that $E[\|X_i\|_H^{\xi}] < \infty$ for some $\xi > 2$, $\inf_{p \in \mathbb{S}^{d-1}} \mathrm{Var}(s(X_i, p)) > 0$, and $E[\|z_i\|^2] < \infty$. Then the followings hold true.*

**(i):** *Under the null hypothesis* $H_0$,

$$K_n(\hat{\nu}) \xrightarrow{d} \sup_{p\in\mathbb{S}^{d-1}} \frac{\tilde{Z}(p)^2}{\mathrm{Var}(s(X,p))},$$

*where* $\tilde{Z}(\cdot) = Z(\cdot) - G(\cdot;\nu)'\Gamma$ *is the Gaussian process implied from* $(Z(p),\Gamma)' \sim N(0,\tilde{V}(p))$ *and* $\tilde{V}(p)$ *is the covariance matrix of the vector* $(s(X,p),(\nabla h(E[z])'\{z-E[z]\})')$.

**(ii):** *Under the alternative hypothesis* $H_1$, $K_n(\hat{\nu})$ *diverges to infinity.*

**(iii):** *Under the local alternative hypothesis* $H_{1n} : E[s(X_i,p)] = s(\Theta_0(\nu),p) + n^{-1/2}\eta(p)$ *for all* $p \in \mathbb{S}^{d-1}$ *with a continuous function* $\eta$ *on* $\mathbb{S}^{d-1}$,

$$K_n(\hat{\nu}) \xrightarrow{d} \sup_{p\in\mathbb{S}^{d-1}} \frac{\{\tilde{Z}(p)+\eta(p)\}^2}{\mathrm{Var}(s(X,p))}.$$

In comparison to Theorem 1, the above theorem says that for the marked empirical likelihood with plug-in, the null distribution of the Kolmogorov-Smirnov type statistic $K_n(\hat{\nu})$ is characterized by the Gaussian process $\tilde{Z}(p)/\sqrt{\mathrm{Var}(s(X,p))}$, which reflects the contribution from the variance of nuisance parameter estimation. Since the null distribution contains parameters to be estimated, we propose the following bootstrap calibration. Let $\bar{s}(p) = n^{-1}\sum_{i=1}^n s(X_i,p)$ and $\hat{V}(p) = n^{-1}\sum_{i=1}^n \{s(X_i,p) - \bar{s}(p)\}^2$.

**Algorithm.**

(1) Generate a bootstrap sample $\{s(X_i^*,p), z_i^*\}_{i=1}^n$ of size $n$ by drawing from a random sample of the empirical distribution of $\{s(X_i,p), z_i\}_{i=1}^n$ with replacement.

(2) Denote $\hat{\nu}^* = h(\bar{z}^*)$ where $\bar{z}^* = n^{-1}\sum_{i=1}^n z_i^*$. Compute the bootstrap counterpart $K_n^*(\hat{\nu})$ by

$$K_n^*(\hat{\nu}) = \left\{\sqrt{n}\Big(n^{-1}\sum_{i=1}^n \{s(X_i^*,p) - \bar{s}(X,p)\} - G(p;\hat{\nu})'(\hat{\nu}^* - \hat{\nu})\Big)\hat{V}(p)^{-1/2}\right\}^2.$$

(3) Repeat Steps (1) and (2) to compute the empirical distribution $F_{n,\hat{\nu}}^*$ of $K_n^*(\hat{\nu})$.

(4) Estimate the $\alpha$-th quantile of the limiting distribution of $K_n(\hat{\nu})$ by $\hat{c}_{n,\alpha}(\hat{\nu}) = \inf\{t : F_{n,\hat{\nu}}^*(t) \geq 1-\alpha\}$.

The validity of this bootstrap procedure can be seen as follows. The bootstrap counterpart $K_n^*(\hat{\nu})$ weakly converges to the limiting distribution of $K_n(\hat{\nu})$ ($\mu$-a.s.) by the continuous mapping theorem and the facts that (a) $\{n^{-1/2}\sum_{i=1}^n \{s(X_i^*,p) - \bar{s}(p)\}, \sqrt{n}(\hat{\nu}^* - \hat{\nu}) : p \in \mathbb{S}^{d-1}\}$ weakly converges to $\{Z(p),\Gamma : p \in \mathbb{S}^{d-1}\}$, $\mu$-a.s. by Giné and Zinn (1990, Theorem 2.4) and $\sqrt{n}(\hat{\nu}^* - \hat{\nu}) \xrightarrow{d^*} \Gamma$, $\mu$-a.s. (since $\nabla h(E[z])$ exists); (b) $\sup_{p\in\mathbb{S}^{d-1}} \|G(p;\hat{\nu}) - G(p;\nu)\| \xrightarrow{p} 0$ by Assumption N; and (c) $\sup_{p\in\mathbb{S}^{d-1}} |\hat{V}(p) - E[Z(p)^2]| \xrightarrow{p} 0$ (by a uniform law of large numbers). Furthermore, under the additional assumption that $\inf_{p\in\mathbb{S}^{d-1}} \mathrm{Var}(\tilde{Z}(p)) > 0$, the Corollary of Lifshits (1982) assures that the limiting distribution of $K_n(\hat{\nu})$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$. Thus it follows that the bootstrap critical value $\hat{c}_{n,\alpha}(\hat{\nu})$ converges in probability to the $\alpha$-th quantile of the limiting distribution of $K_n(\hat{\nu})$.

We next consider a sieve empirical likelihood statistic with plug-in estimates for the nuisance parameters. The presence of estimated nuisance parameters implies that empirical likelihood no longer has the 'self-normalizing' property as the asymptotic variance of the nuisance parameter estimator enters into the limiting distribution. In order to restore self-normalization, we modify the empirical likelihood function by adding a penalty term. To describe the test statistic, we first define following notation. Let $m_k(X_i)$, $\tilde{m}_k(X_i)$, $\dot{m}_k(X_i)$, and $\hat{m}_k(X_i)$ be $k$-dimensional vectors whose $j$-th elements are given by

$$\begin{aligned}
m_{k,j}(X_i) &= s(X_i, p_j) - s(\Theta_0(\hat{\nu}), p_j), \\
\tilde{m}_{k,j}(X_i) &= s(X_i, p_j) - s(\Theta_0(\nu), p_j), \\
\dot{m}_{k,j}(X_i) &= s(X_i, p_j) - s(\Theta_0(\nu), p_j) - G(p_j; \nu)' \nabla h(E[z_i])'(z_i - E[z_i]), \\
\hat{m}_{k,j}(X_i) &= s(X_i, p_j) - s(\Theta_0(\hat{\nu}), p_j) - G(p_j; \hat{\nu})' \nabla h(\bar{z})'(z_i - \bar{z}),
\end{aligned}$$

respectively. Define $V_k = \mathrm{Var}(\tilde{m}_k(X_i))$, $\dot{V}_k = \mathrm{Var}(\dot{m}_k(X_i))$, $\hat{V}_k = n^{-1}\sum_{i=1}^n m_k(X_i)m_k(X_i)'$, $\bar{V}_k = n^{-1}\sum_{i=1}^n \hat{m}_k(X_i)\hat{m}_k(X_i)'$, $\dot{\phi}_k = \lambda_{\min}(\dot{V}_k)$, and $\bar{\phi}_k = \lambda_{\min}(\bar{V}_k)$. Then the penalized (dual) empirical likelihood test statistic is defined as

$$L_n(\hat{\nu}) = \sup_{\gamma \in \Gamma_n} G_n(\gamma),$$

where

$$G_n(\gamma) = 2\sum_{i=1}^n \log(1 + \gamma' m_k(X_i)) - n\gamma'(\bar{V}_k - \hat{V}_k)\gamma$$

and $\Gamma_n = \{\gamma \in \mathbb{R}^k : \|\gamma\| \leq C\bar{\phi}_k^{-3/2}\sqrt{k/n}\}$ for some positive constant $C$. In particular $C$ is chosen to satisfy $C > \max\{2C'\bar{\phi}_k^{-1/2}, 1\}$ where $C'$ is the positive constant obtained from $\|\bar{m}\| \leq C'\sqrt{k/n}$ w.p.a.1. The condition on $C$ ensures that the local maximum $\hat{\gamma}$ of $G_n(\gamma)$ lies in the interior of $\Gamma_n$ w.p.a.1 even in the case when $\dot{\phi}_k^{-1}$ is bounded. If $\dot{\phi}_k^{-1}$ diverges to infinity, this additional condition on $C$ may be dispensed with. Note that $G_n(\gamma)$ is well defined only in the region $S_n = \{\gamma \in \mathbb{R}^k : \gamma' m_k(X_i) > -1 \text{ for all } i = 1, \ldots, n\}$. However, since our assumptions guarantee $\max_{1\leq i\leq n} \sup_{\gamma\in\Gamma_n} |\gamma' m_k(X_i)| = o_p(1)$, it holds that $\Gamma_n \subseteq S_n$ w.p.a.1. It must be emphasized moreover that while the restrictions on $\Gamma_n$ are needed to derive the theoretical properties of the estimator, in practice it may be possible to avoid an explicit choice of $\Gamma_n$ since as we show later, $L_n(\hat{\nu})$ is equivalently characterized under the null as the local maximum of $G_n(\gamma)$ that is closest to 0.

The asymptotic properties of the test statistic $L_n(\hat{\nu})$ are obtained as follows.

**Theorem 4.** *Suppose that Assumption N holds and that $E[\|X_i\|_H^\xi] < \infty$ for some $\xi \geq 4$ and $E[\|z_i\|^4] < \infty$. For all $w$ in a neighborhood of $E[z]$ and all $\tilde{\nu}$ in a neighborhood of $\nu$, the derivatives $\nabla h$ and $G$ satisfy*

$$|\nabla h(w) - \nabla h(E[z])| \leq M\|w - E[z]\|^\alpha \text{ and } \sup_{p\in\mathbb{S}^{d-1}} |G(p;\tilde{\nu}) - G(p;\nu)| \leq M\|\tilde{\nu} - \nu\|^\alpha, \qquad (3)$$

*for some $\alpha \geq 2/3$ and $M > 0$ independent of $w$ and $\tilde{\nu}$. Furthermore, assume $k \to \infty$ and $(k^5\dot{\phi}_k^{-6})^{\frac{\xi}{\xi-2}}/n \to 0$ as $n \to \infty$. Then the followings hold true.*

**(i):** *Under the null hypothesis* $H_0$,

$$\frac{L_n(\hat{\nu}) - k}{\sqrt{2k}} \xrightarrow{d} N(0,1).$$

**(ii):** *Under the alternative hypothesis* $H_1$, $(L_n(\hat{\nu}) - k)/\sqrt{2k}$ *diverges to infinity.*

**(iii):** *Under the sequence of local alternative hypotheses* $H_{1n} : E[s(X_i, p)] = s(\Theta_0(\nu), p) + a_n \eta(p)$ *for all* $p \in \mathbb{S}^{d-1}$ *with a function* $\eta$ *on* $\mathbb{S}^{d-1}$, *where* $a_n = k^{1/4}/\sqrt{n\eta_k' \dot{V}_k \eta_k}$ *and* $\eta_k = (\eta(p_1), \ldots, \eta(p_k))'$,

$$\frac{L_n(\hat{\nu}) - k}{\sqrt{2k}} \xrightarrow{d} N(2^{-1/2}, 1).$$

The rate condition $(k^5 \dot{\phi}_k^{-6})^{\frac{\xi}{\xi-2}}/n \to 0$ is very similar to that in Theorem 2. Thus similar remarks on the choice of $k$ apply. In particular using the hard threshold rule $\bar{\phi}_k \geq C(\ln n)^{-1}$ gives the same rate condition $k^5(\ln n)^6/n \to 0$ as for Theorem 2. In comparison to Theorem 3, Theorem 4 requires further smoothness assumptions on $s(\Theta_o(\nu), \cdot)$ as seen by the Hölder continuity condition of equation (3). By taking $G(\cdot; \cdot)$ to be the derivative with respect to $\nu$ of $s(\Theta_0(\nu), \cdot)$, the condition (3) is satisfied with $\alpha = 1$ if at each $p \in \mathbb{S}^{d-1}$, $s(\Theta_0(\nu'), p)$ is twice differentiable with respect to $\nu'$ in some neighborhood of $\nu$.

Note that $L_n(\hat{\nu})$ is obtained as the maximum value of the criterion function $G_n(\gamma)$ over the set $\Gamma_n$. This is a convex optimization problem as shown by the following lemma.

**Lemma 1.** *Suppose that the assumptions of Theorem 4 are satisfied and* $(k^5 \bar{\phi}_k^{-6})^{\frac{\xi}{\xi-2}}/n \to 0$. *Then w.p.a.1,* $G_n(\gamma)$ *is a strictly concave function over the domain* $\Gamma_n$.

This lemma holds regardless of whether the null or the alternative holds true (recall from the previous discussion that the practitioner should choose $k$ and the sieve to satisfy $(k^5 \bar{\phi}_k^{-6})^{\frac{\xi}{\xi-2}}/n \to 0$). Using this lemma, we can obtain an alternative characterization of $L_n(\hat{\nu})$ under the null as the local maximum of $G_n(\gamma)$ that is closest to 0. Indeed, inspection of the proof of Theorem 4 shows that $G_n(\gamma)$ is maximized in the interior of $\Gamma_n$. This result is combined with the (strict) concavity of $G_n(\gamma)$ over $\Gamma_n$ to assure that both definitions of $L_n(\hat{\nu})$ are in fact equivalent. Defining $L_n(\hat{\nu})$ as a local maximum also makes it quite convenient for computation. In particular we could obtain $L_n(\hat{\nu})$ by setting 0 as the initial point of an optimization algorithm for finding local maxima. The convergence of this algorithm should generally be quite fast as $G_n(\cdot)$ is concave in this region. Although $G_n(\cdot)$ is not concave outside $\Gamma_n$ in general, the remark below provides some exceptions.

**Remark 1.** If $\dot{V}_k - V_k$ is positive definite, then $\Gamma_n$ can be taken to be $S_n$ that is the domain over which $G_n(\cdot)$ is well defined. To see this, we note from the proof of Theorem 4 that $\bar{V}_k$ and $\hat{V}_k$ are both consistent estimators of $\dot{V}_k$ and $V_k$. Thus since $\bar{V}_k - \hat{V}_k$ is positive definite, the criterion function $G_n(\cdot)$ is concave over its domain. Since $G_n(\gamma)$ is maximized in the interior of $\Gamma_n$, it follows that $L_n(\hat{\nu})$ is also the global maximum of $G_n(\gamma)$ over $S_n$.

One instance where the assumption of positive definiteness of $\dot{V}_k - V_k$ holds is when $\{X_i\}$ and $\{z_i\}$ are independent sequences and $\mathrm{Var}(z_i)$ is positive definite. Alternatively, it is possible to ensure this assumption in general by the following sample splitting method. First, we split the sample randomly into two (possibly unequal) parts. One part of the sample is used solely to obtain an estimate for $\nu$ denoted by $\hat{\nu}$. Such an estimate is then used as plug-in to form the penalized EL test statistic, $L_n(\hat{\nu})$ using observations only from the second sample. It is easy to verify that with this method, if $\mathrm{Var}(z_i)$ is positive definite, so is $\dot{V}_k - V_k$ in the second sample.

**Remark 2.** We now compare the statistic $L_n(\hat{\nu})$ with the standard (non-penalized) sieve empirical likelihood statistic with plug-in estimates for the nuisance parameters. The non-penalized version is given by

$$\tilde{L}_n(\hat{\nu}) = -2 \log l(p_1, \ldots, p_k, \hat{\nu}),$$

where

$$l(p_1, \ldots, p_k, \hat{\nu}) = \max \left\{ \prod_{i=1}^{n} n w_i \;\middle|\; \begin{array}{c} \sum_{i=1}^{n} w_i s(X_i, p_j) = s(\Theta_0(\hat{\nu}), p_j) \text{ for } j = 1, \ldots, k, \\ w_i \geq 0, \;\; \sum_{i=1}^{n} w_i = 1 \end{array} \right\}.$$

As noted earlier, $\tilde{L}_n(\hat{\nu})$ is not internally studentized. The intuition for this is as follows. The statistic $\tilde{L}_n(\hat{\nu})$ has the quadratic approximation $\tilde{L}_n(\hat{\nu}) \approx n\bar{m}_k' \hat{V}_k^{-1} \bar{m}_k$, where $\bar{m}_k = n^{-1} \sum_{i=1}^{n} m_k(X_i)$. While the $m_k(X_i)$'s are not i.i.d. as they contain the estimator $\hat{\nu}$, by using the Hölder continuity conditions, it can be seen that $\bar{m}_k$ is well approximated by the average of i.i.d. variables $\bar{\dot{m}}_k = n^{-1} \sum_{i=1}^{n} \dot{m}_k(X_i)$. We thus approximate $n\bar{m}_k' \hat{V}_k^{-1} \bar{m}_k$ by $n\bar{\dot{m}}_k' V_k^{-1} \bar{\dot{m}}_k$. But note that $E[\dot{m}_k(X_i) \dot{m}_k(X_i)'] = \dot{V}_k \neq V_k$ because $m_k(X_i)$ and $\dot{m}_k(X_i)$ do not have the same variance. This difference arises due to the variance of the estimator $\hat{\nu}$.

In contrast, for the case of $L_n(\hat{\nu})$ the penalty is chosen such that studentization is obtained. The intuition for this is as follows. Within the set $\Gamma_n$, the criterion function $G_n(\gamma)$ is uniformly close to $G_n^*(\gamma) = 2\sqrt{n}\gamma' \bar{m}_k - \gamma' \bar{V}_k \gamma$, which is maximized at $\gamma^* = \bar{V}_k^{-1} \sqrt{n} \bar{m}_k$. Thus we obtain the quadratic approximation $L_n(\hat{\nu}) \approx G_n^*(\gamma^*) = n\bar{m}_k' \bar{V}_k^{-1} \bar{m}_k$. But the latter is in turn approximated by $n\bar{\dot{m}}_k' \dot{V}_k^{-1} \bar{\dot{m}}_k$, which is now studentized.

3.1. **Example: Best linear predictor.** Suppose we wish to make inference on the best linear predictor of the interval outcome $Y_i = [y_{Li}, y_{Ui}] \subset \mathbb{R}$ with $y_{Li} \leq y_{Ui}$ almost surely, based on the explanatory variables $x_i \in \mathbb{R}^q$. Note that although $Y_i$ is an interval, $x_i$ is a vector. The set of best linear predictors of $Y_i$ given $x_i$ is identified as

$$\Theta(\Sigma) = \Sigma^{-1} \mathbb{E}[W_i],$$

where

$$\Sigma = \begin{pmatrix} 1 & Ex_i' \\ Ex_i & Ex_i x_i' \end{pmatrix}, \qquad W_i = \begin{pmatrix} [y_{Li}, y_{Ui}] \\ [x_i y_{Li}, x_i y_{Ui}] \end{pmatrix} \subset \mathbb{R}^{q+1}.$$

By Beresteanu and Molinari (2008, Lemmas A.4 and A.5), $\{W_1, \ldots, W_n\}$ is a sequence of i.i.d compact and convex SVRVs. Thus by properties of the Aumann expectation, $\Theta_0$ is a convex set. We consider the case where $x_i$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^q$. Beresteanu and Molinari (2008, Lemma A.8) showed that in this case $\Theta_0$ is strictly convex and subsequently $s(\Theta_0, p)$ is differentiable for all $p \in \mathbb{S}^q$. We therefore wish to test the null hypothesis $H_0 : \Sigma^{-1} \mathbb{E}[W_i] = \Theta_0$ for a strictly convex set $\Theta_0$. Using the support function, the null hypothesis is written as

$$H_0 : E[s(W, p)] = s(\Sigma\Theta_0, p) \text{ for all } p \in \mathbb{S}^q \text{ against } H_1 : E[s(W, p)] \neq s(\Sigma\Theta_0, p) \text{ for some } p \in \mathbb{S}^q.$$

Note that in this context the support function is given by $s(W_i,p) = [y_{Li} + (y_{Ui} - y_{Li})\mathbb{I}\{f(x_i,p) \geq 0\}]f(x_i,p)$ with $f(x_i,p) = (1, x_i')p$. Further, since $s(\Sigma\Theta_0, p) = s(\Theta_0, \Sigma p)$, the support function of the set $\Sigma\Theta_0$ can be computed from that of $\Theta_0$.

In this example, we treat $\Sigma$ as nuisance parameters. If $\Sigma$ is known, the empirical likelihood methods proposed in Section 2 apply. Here we consider the case where $\Sigma$ is unknown and estimated by $\hat{\Sigma} = \begin{pmatrix} 1 & n^{-1}\sum_{i=1}^n x_i' \\ n^{-1}\sum_{i=1}^n x_i & n^{-1}\sum_{i=1}^n x_i x_i' \end{pmatrix}$. Based on the notation of Theorems 3 and 4, we set as $X_i = W_i$, $\nu = \text{vec}(\Sigma)$, $\hat{\nu} = \text{vec}(\hat{\Sigma})$, $z_i = \text{vec}\big((1,x_i')'(1,x_i')\big)$, $h(z) = z$, and $\Theta_0(\hat{\nu}) = \hat{\Sigma}\Theta_0$. Also, let $\nabla s(\Theta_0,p)'$ be the Fréchet derivative of $s(\Theta_0,p)$ with respect to $p$ for $p \in \mathbb{R}^{q+1} \setminus \{0\}$, and set $G(p;\nu) = p \otimes \nabla s(\Theta_0, \Sigma p)$ where $\otimes$ represents the Kronecker product. In particular note that $G(p;\nu)'$ is the pointwise derivative of $s(\Sigma\Theta_0,p)$ with respect to $\nu = \text{vec}(\Sigma)$.

The null distributions of the test statistics $K_n(\text{vec}(\hat{\Sigma}))$ and $L_n(\text{vec}(\hat{\Sigma}))$ are obtained as follows.

**Proposition 1.** *Consider the setup of this subsection. Assume that $\{y_{Li}, y_{Ui}, x_i\}_{i=1}^n$ is i.i.d., where the distribution of $x_i$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^q$, and $\Sigma$ is full rank.*

**(i):** *Suppose $E\|y_{Li}\|^\xi < \infty$, $E\|y_{Ui}\|^\xi < \infty$, $E\|x_i y_{Li}\|^\xi < \infty$, $E\|x_i y_{Ui}\|^\xi < \infty$ for some $\xi > 2$, $E\|x_i\|^4 < \infty$, and $\text{Var}(y_{Li}|x_i), \text{Var}(y_{Ui}|x_i) \geq \sigma^2 > 0$ a.s. Then under $H_0$,*

$$K_n(vec(\hat{\Sigma})) \xrightarrow{d} \sup_{p \in \mathbb{S}^{d-1}} \frac{\tilde{Z}(p)^2}{\text{Var}(s(W_i,p))},$$

*where $\tilde{Z}(\cdot) = Z(\cdot) - G(\cdot;\nu)'\Gamma$ is the Gaussian process implied from $(Z(p),\Gamma)' \sim N(0,\tilde{V}(p))$ and $\tilde{V}(p)$ is the covariance matrix of the vector $(s(W_i,p), \{z_i - vec(\Sigma)\}')$.*

**(ii):** *Suppose $E\|y_{Li}\|^\xi < \infty$, $E\|y_{Ui}\|^\xi < \infty$, $E\|x_i y_{Li}\|^\xi < \infty$, $E\|x_i y_{Ui}\|^\xi < \infty$ for some $\xi \geq 4$, $E\|x_i\|^4 < \infty$, and $\nabla s(\Theta_0,p)$ is locally Hölder continuous of order $\alpha \geq 2/3$ over the domain $\mathbb{S}^q$. Also assume $k \to \infty$ and $(k^5 \dot{\phi}_k^{-6})^{\frac{\xi}{\xi-2}}/n \to 0$. Then under $H_0$,*

$$\frac{L_n(vec(\hat{\Sigma})) - k}{\sqrt{2k}} \xrightarrow{d} N(0,1).$$

Part (i) of this proposition is closely related to the result in Beresteanu and Molinari (2008, Theorem 4.3) who propose a Wald type test statistic. We employ the same assumptions as theirs except for the requirement $\xi > 2$ as opposed to $\xi \geq 2$. For Part (ii), we impose additional smoothness assumptions over the boundary $\partial\Theta_0$ of $\Theta_0$ in the form of a Hölder continuity assumption on $\nabla s(\Theta_0,p)$. In particular this is satisfied with $\alpha = 1$ if $s(\Theta_0,p)$ is twice differentiable at each point $p \in \mathbb{S}^q$. Also note that for Part (ii), the points $\{p_1,\ldots,p_k\}$ in the sieve have to be chosen to satisfy the rate condition $(k^5\dot{\phi}_k^{-6})^{\frac{\xi}{\xi-2}}/n \to 0$. This necessitates that the choice of the sieve should respect any constraints imposed by symmetry so as to make $\dot{\phi}_k \neq 0$. In practice this could be achieved by setting a hard threshold rule e.g. $\dot{\phi}_k \geq C(\ln n)^{-1}$ and/or using a suitable moment selection procedure that deselects redundant constraints.

Computing the test statistic $L_n(\text{vec}(\hat{\Sigma}))$ requires knowledge of $\nabla s(\Theta_0,p)$, the derivative of the support function $s(\Theta_0,p)$ at each $p \in \mathbb{S}^q$. Note that for strictly convex sets, $\nabla s(\Theta_0,p)$ is just a unique point $\theta_p \in \partial\Theta_0$ obtained from $s(\Theta_0,p) = p'\theta_p$. Hence $\nabla s(\Theta_0,p)$ is easily computed as a by-product of the maximization problem to compute $s(\Theta_0,p)$.

**Remark 3.** We now briefly discuss the case of discrete regressors. In such a case, $\Theta_0$ is a polytope with the set of all directions orthogonal to the exposed faces of the polytope given by $\mathcal{Q} = \{p \in \mathbb{S}^q : \mu\big(p'\Sigma(1, x_i')' = 0\big) \neq 0\}$ (Bontemps, Magnac and Maurin, 2012, Lemma 3). In particular with discrete regressors, the number of elements in $\mathcal{Q}$ is finite and may be obtained if $\Sigma$ is known by solving $p'\Sigma(1, x')' = 0$ for each possible realization $x$ of $x_i$. Thus with known $\Sigma$ it is easy to construct valid tests by simply choosing the moment conditions associated with directions in $\mathcal{Q}$. Such a technique is however clearly infeasible with unknown $\Sigma$. Chandrasekhar *et al.* (2012) address this issue by a data jittering technique to add an extra noise term so that $\Theta_0$ is now strictly convex. While we expect our test statistics to remain valid, a detailed analysis of their properties under data jittering is beyond the scope of this paper.

**Remark 4.** Chandrasekhar *et al.* (2012) generalized the notion of the best linear predictor to accommodate instrumental variables. In particular, if there exist $m$-dimensional instrumental variables $w_i$, they characterize a set of parameter values $\Theta$ of the 'best' linear approximators (as a function of $x_i$) for the SVRV $Y_i$ as

$$\Theta = \{\theta : \theta = \arg\min_{\tilde{\theta}} E[(y_i^* - x_i'\tilde{\theta})w_i'Ww_i(y_i^* - x_i'\tilde{\theta})],\ y_{Li} \leq y_i^* \leq y_{Ui}\}$$

where $W$ is some known positive definite $m \times m$ weight matrix. Using the Aumann expectation, the above set $\Theta$ is characterized as

$$\Gamma\Theta = \mathbb{E}[\tilde{W}_i],$$

where $\tilde{W}_i = x_i w_i' W w_i Y_i$ is a compact and convex random set, and $\Gamma = E[x_i w_i' W w_i x_i']$ is assumed to be full rank (which is guaranteed if $E[x_i x_i']$ is full rank). Therefore, a straightforward extension of Proposition 1 also covers this generalization.

**Remark 5.** We close this section with a few remarks on the incomplete linear instrumental variable model of Bontemps, Magnac and Maurin (2012). Consider the latent linear model

$$y_i^* = x_i'\theta + \epsilon_i,$$

where $y_i^*$ is an unobservable outcome variable, $x_i$ is a $p$-dimensional vector of observable explanatory variables, and $\epsilon_i$ is the error term. Suppose we observe the interval outcome $Y_i = [y_{Li}, y_{Ui}] \subset \mathbb{R}$ with $y_{Li} \leq y_i^* \leq y_{Ui}$ almost surely. Further suppose $x_i$ is endogenous in the sense that $E[x_i\epsilon_i] \neq 0$ but $m$-dimensional instrumental variables $w_i$ satisfying $E[w_i\epsilon_i] = 0$ are available. We wish to conduct inference on $\theta$ based on the observables $\{Y_i, x_i, w_i\}_{i=1}^n$. Assume $E[w_i w_i']$ is full rank and $E[w_i x_i']$ is full column rank. Then following Bontemps, Magnac and Maurin (2012), the set of values of $\theta$ that are consistent with the observed data is

$$\Theta = \{\theta : E[w_i x_i']\theta = E[w_i y_i^*],\ y_{Li} \leq y_i^* \leq y_{Ui}\}.$$

In terms of the Aumann expectation, the above set is characterized by

$$E[w_i x_i']\Theta = \mathbb{E}[w_i Y_i],$$

where $w_i Y_i = [w_i y_{Li}, w_i y_{Ui}]$. If $m = p$ (called just-identification), $E[w_i x_i']$ is full rank and the same argument to Proposition 1 applies. For the case of $m > p$ (called over-identification), in analogy with the generalized estimating equations, we may multiply $E[x_i w_i']W$ on both sides of the above equation to obtain

$$E[x_i w_i']WE[w_i x_i']\Theta = E[x_i w_i']W\mathbb{E}[w_i Y_i],$$

where $W$ is any positive definite weighing matrix (which ensures that $E[x_i w_i']WE[w_i x_i']$ is invertible). This inference problem however contains nuisance parameters on both sides of the null hypothesis which lies outside the scope of the testing framework considered so far. It remains only to conjecture that the marked EL test statistic can be extended to cover this case by using random function methods (Section 9.4 of van der Vaart, 1998) to take care of the nuisance parameters $E[x_i w_i']$ on the right hand side of the above equation.

## Appendix A. Mathematical Appendix

### A.1. **Proof of Theorem 1.**

*Proof of (i).* Observe that $|s(X,p) - s(X,q)| \leq \|X\|_H \|p-q\|$ a.s. for any $p, q \in \mathbb{S}^{d-1}$ (i.e., the support function is Lipschitz) and that $\mathbb{S}^{d-1}$ is compact. Thus by a standard empirical process argument (e.g., van der Vaart, 1998, Example 19.7), the process $\{s(X,p) - s(\Theta_0, p) : p \in \mathbb{S}^{d-1}\}$ is $\mu$-Donsker and consequently the empirical process $n^{-1/2}\sum_{i=1}^n \{s(X_i, \cdot) - s(\Theta_0, \cdot)\}$ weakly converges to $Z(\cdot)$. Next, the uniform law of large numbers guarantees $\sup_{p\in\mathbb{S}^{d-1}} |n^{-1}\sum_{i=1}^n \{s(X_i,p) - s(\Theta_0,p)\}^2 - E[Z(p)^2]| \overset{p}{\to} 0$. Finally, since the assumption $E[\|X_i\|_H^\xi] < \infty$ for $\xi > 2$ implies $E[\sup_{p\in\mathbb{S}^{d-1}} |s(X_i,p)|^\xi] < \infty$, a Borel-Cantelli lemma argument as in Owen (1988) ensures that $\max_{1\leq i\leq n} \sup_{p\in\mathbb{S}^{d-1}} |s(X_i,p) - s(\Theta_0,p)| = o(n^{1/2})$ with probability 1. Then the conclusion follows by a similar argument as in the proof of Hjort, McKeague and van Keilegom (2009, Theorem 2.1).

*Proof of (ii).* Let $g_i(p) = s(X_i,p) - s(\Theta_0,p)$. Under $H_1$, there exists $p^* \in \mathbb{S}^{d-1}$ such that $E[g_i(p^*)] \neq 0$. We prove the case of $E[g_i(p^*)] > 0$. The case of $E[g_i(p^*)] < 0$ is shown in the same manner. Pick any $\delta \in (0, 1/2)$. Observe that

$$\begin{aligned} -\log \ell(p^*) &= \sup_{\lambda\in\mathbb{R}} \sum_{i=1}^n \log(1 + \lambda g_i(p^*)) \geq \sum_{i=1}^n \log(1 + n^{-(1/2+\delta)} g_i(p^*)) \\ &= n^{1/2-\delta}\left\{\frac{1}{n}\sum_{i=1}^n g_i(p^*)\right\} + n^{-2\delta}\left\{\frac{1}{2n}\sum_{i=1}^n g_i(p^*)^2\right\} + O_p(n^{-2\delta}), \end{aligned}$$

where the first equality follows from the convex duality and the second equality follows from a Taylor expansion. Since the first term diverges to infinity and the other terms are negligible, the conclusion is obtained.

*Proof of (iii).* The proof is similar to that of Part (i) with the difference that the empirical process $n^{-1/2}\sum_{i=1}^n \{s(X_i,\cdot) - s(\Theta_0,\cdot)\} = n^{-1/2}\sum_{i=1}^n \{s(X_i,\cdot) - E[s(X_i,\cdot)]\} + \eta(\cdot)$ weakly converges to $Z(\cdot) + \eta(\cdot)$.

### A.2. **Proof of Theorem 2.**

*Proof of (i).* Define $g_i(p) = s(X_i,p) - s(\Theta_0,p)$, $V_k = \text{Var}(m_k(X_i))$, $\bar{m}_k = n^{-1}\sum_{i=1}^n m_k(X_i)$, and $\hat{V}_k = n^{-1}\sum_{i=1}^n m_k(X_i)m_k(X_i)'$. Since the process $\{g_i(p) : p \in \mathbb{S}^{d-1}\}$ is $\mu$-Donsker, we have

$$\|\bar{m}_k\| \leq \sqrt{k} \sup_{p\in\mathbb{S}^{d-1}} \left|\frac{1}{n}\sum_{i=1}^n g_i(p)\right| = O_p(\sqrt{k/n}). \tag{4}$$

Also, using the property of $\lambda_{\max}(\cdot)$,

$$\begin{aligned} |\lambda_{\max}(\hat{V}_k) - \lambda_{\max}(V_k)| &\leq \left\|\hat{V}_k - V_k\right\| \\ \leq k \sup_{p,q\in\mathbb{S}^{d-1}} &\left|\frac{1}{n}\sum_{i=1}^n \{g_i(p)g_i(q) - E[g_i(p)g_i(q)]\}\right| = O_p(k/\sqrt{n}), \end{aligned} \tag{5}$$

where the equality follows from the fact that the process $\{g_i(p)g_i(q) : p, q \in \mathbb{S}^{d-1}\}$ is $\mu$-Donsker.

For the conclusion of the theorem, it is sufficient to show the followings:

$$\frac{L_n - n\bar{m}_k'\hat{V}_k^{-1}\bar{m}_k}{\sqrt{2k}} \xrightarrow{p} 0, \tag{6}$$

$$\frac{n\bar{m}_k'\hat{V}_k^{-1}\bar{m}_k - k}{\sqrt{2k}} \xrightarrow{d} N(0,1). \tag{7}$$

We first show (6). By convex duality, $L_n = \sup_{\gamma \in S_n} F_n(\gamma)$, where $F_n(\gamma) = 2\sum_{i=1}^n \ln(1 + \gamma' m_k(X_i))$ and $S_n = \{\gamma \in \mathbb{R}^k : \gamma' m_k(X_i) > -1 \text{ for } i = 1, \ldots, n\}$. Define $D_n = \max_{1 \le i \le n} \|m_k(X_i)\|$, $\hat{\gamma} = \arg\max_{\gamma \in S_n} F_n(\gamma)$, and $\gamma^* = \hat{V}_k^{-1}\bar{m}_k$. Based on the proof of Hjort, McKeague and van Keilegom (2009, Proposition 4.1), the result in (6) is obtained if $(k^5\phi_k^{-6})^{\frac{\xi}{\xi-2}}/n \to 0$ and the followings hold:

$$(n^{-1/2}k^{3/2}\phi_k^{-3})D_n = o_p(1), \tag{8}$$

$$\|\gamma^*\| = O_p(\phi_k^{-1}\sqrt{k/n}), \tag{9}$$

$$\hat{\gamma} \text{ exists w.p.a.1 and } \|\hat{\gamma}\| = O_p(\phi_k^{-1}\sqrt{k/n}), \tag{10}$$

$$\lambda_{\max}(\hat{V}_k) = O_p(k). \tag{11}$$

For (8), observe that the assumption $E[\|X_i\|_H^\xi] < \infty$ implies $E[\sup_{p \in \mathbb{S}^{d-1}} |s(X_i, p)|^\xi] < \infty$ and thus $\sup_{k \in \mathbb{N}} E[\|k^{-1/2}m_k(X_i)\|^\xi] < \infty$. Then a similar argument as in the proof of Hjort, McKeague and van Keilegom (2009, Lemma 4.1) guarantees (8) under $(k^5\phi_k^{-6})^{\frac{\xi}{\xi-2}}/n \to 0$. For (9), note that $\|\gamma^*\| \le \hat{\phi}_k^{-1}O_p(\sqrt{k/n})$ by (4). Also $|\hat{\phi}_k - \phi_k| \le \left\|\hat{V}_k - V_k\right\| = O_p(k/\sqrt{n})$ by the property of $\lambda_{\min}(\cdot)$ and (5), which implies $|\hat{\phi}_k^{-1} - \phi_k^{-1}| = O_p(\phi_k^{-2}k/\sqrt{n})$. Consequently (9) is verified under $(k^5\phi_k^{-6})^{\frac{\xi}{\xi-2}}/n \to 0$. For (10), define $\Psi_n = \{\gamma \in \mathbb{R}^k : \|\gamma\| \le \phi_k^{-2}kn^{-1/2}\}$. Note that by (8) and the Cauchy-Schwarz inequality, $\sup_{\gamma \in \Psi_n, 1 \le i \le n} |\gamma' m_k(X_i)| = o_p(1)$. Thus $\Psi_n \subseteq S_n$ w.p.a.1 and it is possible to define $\tilde{\gamma} = \arg\max_{\gamma \in \Psi_n} F_n(\gamma)$ (note: $\tilde{\gamma}$ exists since $\Psi_n$ is a compact set). Hence w.p.a.1,

$$\begin{aligned} 0 &\le F_n(\tilde{\gamma})/n \le 2\tilde{\gamma}'\bar{m}_k + \left\{ \sup_{\gamma \in \Psi_n, 1 \le i \le n} \frac{-1}{\left(1 + \gamma' m_k(X_i)\right)^2} \right\} \tilde{\gamma}'\hat{V}_k\tilde{\gamma} \\ &\le 2\tilde{\gamma}'\bar{m}_k - \frac{1}{2}\tilde{\gamma}'\hat{V}_k\tilde{\gamma}, \end{aligned}$$

where the second inequality follows from an expansion around $\gamma = 0$. Also note that $\tilde{\gamma}'\hat{V}_k\tilde{\gamma} \ge \|\tilde{\gamma}\|^2 \phi_k(1 + o_p(1))$ by (5) while $|\tilde{\gamma}'\bar{m}_k| \le \|\tilde{\gamma}\| O_p(\sqrt{k/n})$ by (4). Therefore it must be the case that $\|\tilde{\gamma}\| = O_p(\phi_k^{-1}\sqrt{k/n})$. This in turn implies that $\tilde{\gamma}$ is an interior solution w.p.a.1 and is the global maximizer of $F_n(\gamma)$ (since $F_n(\gamma)$ is a concave function over its domain $S_n$). Therefore we obtain $\hat{\gamma} = \tilde{\gamma}$ w.p.a.1, which proves (10). Finally, the result in (11) is obtained by noting that $\sup_{k \in \mathbb{N}} E[\|m_k(X_i)/k^{1/2}\|^2] < \infty$ guarantees $\lambda_{\max}(\hat{V}_k) \le \left\|\hat{V}_k\right\| = O_p(k)$. Combining these results, the claim in (6) follows.

We now show (7). Decompose

$$
\begin{aligned}
\frac{n\bar{m}_k'\hat{V}_k^{-1}\bar{m}_k - k}{\sqrt{2k}} &= \frac{n^{-1}\sum_{i=1}^n m_k(X_i)'V_k^{-1}m_k(X_i) - k}{\sqrt{2k}} + \frac{\sum_{i\neq j} m_k(X_j)'V_k^{-1}m_k(X_i)}{n\sqrt{2k}} \\
&\quad + \frac{n\bar{m}_k'(\hat{V}_k^{-1} - V_k^{-1})\bar{m}_k}{\sqrt{2k}}. \qquad (12)
\end{aligned}
$$

First, by the Markov inequality the first term of (12) is $o_p(1)$ if $\phi_k^{-2}k/n \to 0$. Second, we show that the second term of (12) converges in distribution to $N(0,1)$. Since this term is a U-statistic, we show that the conditions for the central limit theorem of de Jong and Bierens (1994, Lemma 2) are satisfied. In particular, it is sufficient to verify that $E[H_n(m_k(X_1), m_k(X_2))^2] < \infty$ for each $n$ and that

$$
\lim_{n\to\infty} \{E[G_n(m_k(X_1), m_k(X_2))^2] + n^{-1}E[H_n(m_k(X_1), m_k(X_2))^4]\}/E[H_n(m_k(X_1), m_k(X_2))^2]^2 \to 0,
$$

where $H_n(u,v) = u'V_k^{-1}v/\sqrt{k}$ and $G_n(u,v) = E[H_n(m_k(X_1),u)H_n(m_k(X_1),v)]$. Arguing as in the proof of de Jong and Bierens (1994, Theorem 1), we obtain $E[H_n(m_k(X_1), m_k(X_2))^2] = 1$ and $E[G_n(m_k(X_1), m_k(X_2))^2] = 1/k$. Furthermore,

$$
\begin{aligned}
E[H_n(m_k(X_1), m_k(X_2))^4] &\leq k^{-2}E\big[\text{trace}^2[m_k(X_1)m_k(X_1)'V_k^{-1}]\text{trace}^2[m_k(X_2)m_k(X_2)'V_k^{-1}]\big] \\
&\leq k^{-2}E[\{m_k(X_1)'V_k^{-1}m_k(X_1)\}^2]E[\{m_k(X_2)'V_k^{-1}m_k(X_2)\}^2] \\
&\leq k^{-2}\phi_k^{-4}E[\|m_k(X_1)\|^4]^2 = O(\phi_k^{-4}k^2),
\end{aligned}
$$

where the first inequality follows from the Cauchy-Schwarz inequality and the equality follows from $\sup_{k\in\mathbb{N}} E[\left\|m_k(X_i)/k^{1/2}\right\|^{\xi}] < \infty$ with $\xi \geq 4$. Since $\phi_k^{-4}k^2/n \to 0$, the conditions for the central limit theorem of de Jong and Bierens (1994, Lemma 2) are seen to be satisfied. Finally, we show that the third term of (12) is negligible. By de Jong and Bierens (1994, Lemma 4a), it is bounded by $nk^{-1/2}\|\bar{m}_k\|^2\hat{\phi}_k^{-1}\phi_k^{-1}\left\|\hat{V}_k - V_k\right\|$. Thus using (4) and (5) the third term is $O_p(\sqrt{\phi_k^{-4}k^3/n})$ and consequently negligible. Therefore, the result in (7) follows.

*Proof of (ii).* Since in the limit the points $\{p_1, p_2, \ldots, p_k\}$ form a dense subset of $\mathbb{S}^{d-1}$ and the support function is continuous, under $H_1$ there exists an integer $N$ such that for all $n \geq N$ the set of points includes a direction $p^*$ for which $E[s(X_i,p^*) - s(\Theta_0,p^*)] \neq 0$. We prove the case of $E[s(X_i,p^*) - s(\Theta_0,p^*)] > 0$. The case of $E[s(X_i,p^*) - s(\Theta_0,p^*)] < 0$ is shown in the same manner. Pick any $\delta \in (0, 0.3)$. Recalling the definition of $S_n$ from Part (i), observe that

$$
\begin{aligned}
L_n &= \sup_{\gamma\in S_n} 2\sum_{i=1}^n \log(1 + \gamma' m_k(X_i)) \\
&\geq 2\sum_{i=1}^n \log(1 + n^{-(1/2+\delta)}\{s(X_i,p^*) - s(\Theta_0,p^*)\}),
\end{aligned}
$$

for all $n \geq N$, where the equality follows from the convex duality and the inequality follows by setting $\gamma = n^{-(1/2+\delta)}e^* \in S_n$ w.p.a.1, where $e^*$ is the unit vector that selects the component of $m_k(X_i)$ containing $p^*$. Then by an argument as in the proof of Theorem 1 (ii), $L_n$ diverges to infinity at the rate $n^{1/2-\delta}$. But $n^{1/2-\delta}/\sqrt{k} \to \infty$ and $k/n^{1/2-\delta} \to 0$ for $\delta \leq 0.3$, thus $(L_n - k)/\sqrt{2k} = (L_n/\sqrt{2k})(1 - k/L_n)$ is seen to diverge to infinity.

*Proof of (iii).* We use the same notation as Part (i). Define $\tilde{m}_k(X_i) = m_k(X_i) - E[m_k(X_i)]$ and $\bar{\tilde{m}}_k = n^{-1}\sum_{i=1}^n \tilde{m}_k(X_i)$. Note that $m_k(X_i) = \tilde{m}_k(X_i) + a_n\eta_k$. By similar arguments as used to show (4), it follows that $\bar{\tilde{m}}_k = O_p(\sqrt{k/n})$. Furthermore, expanding $\hat{V}_k = n^{-1}\sum_{i=1}^n \tilde{m}_k(X_i)\tilde{m}_k(X_i)' + a_n\eta_k\bar{\tilde{m}}_k' + a_n\bar{\tilde{m}}_k\eta_k' + a_n^2\eta_k\eta_k'$, straightforward algebra and analogous weak convergence arguments as in (5) assure that $\left\|\hat{V}_k - V_k\right\| = O_p(k/\sqrt{n})$.

We first show that $(L_n - n\bar{m}_k'\hat{V}_k^{-1}\bar{m}_k)/\sqrt{2k} \xrightarrow{p} 0$. By a similar argument as used to show (6) this follows if equations (8)-(11) hold. Indeed (8) and (11) follow by the same arguments as in the proof of part (i). To show (9), expand

$$\gamma^* = \hat{V}_k^{-1}\bar{\tilde{m}}_k + (\hat{V}_k^{-1} - V_k^{-1})a_n\eta_k + a_nV_k^{-1}\eta_k. \tag{13}$$

By similar arguments as in the proof of part (i), the first term of (13) is $O_p(\phi_k^{-1}\sqrt{k/n})$. Next, observe by $\lambda_{\max}(V_k) = O(k)$ (which implies $\inf_{k\in\mathbb{N}}\eta_kV_k^{-1}\eta_k > 0$) that $a_n\left\|\eta_k\right\| = O_p(k^{3/4}/\sqrt{n})$. This, along with $\left\|\hat{V}_k^{-1} - V_k^{-1}\right\| = O_p(\phi_k^{-2}k/\sqrt{n})$ ensures the second term of (13) is $O_p(\phi_k^{-2}k^{7/4}/n)$. Furthermore the third term of (13) is bounded by $a_n\phi_k^{-1/2}\|V_k^{-1/2}\eta_k\| = O_p(\sqrt{\phi_k^{-1}k^{1/2}/n})$ (since $\eta_kV_k^{-1}\eta_k = \|V_k^{-1/2}\eta_k\|^2$). Hence under $(k^5\phi_k^{-6})^{\frac{\xi}{\xi-2}}/n \to 0$, the first term of (13) dominates the other two ensuring $\gamma^* = O_p(\phi_k^{-1}\sqrt{k/n})$. This proves (9). Finally, to show (10), note that by the same arguments as in part (i), it still holds that $\tilde{\gamma}'\hat{V}_k\tilde{\gamma} \leq 4\tilde{\gamma}'\bar{m}_k$ which implies $\tilde{\gamma}'V_k\tilde{\gamma} \leq 4\tilde{\gamma}'\bar{m}_k + O_p(\phi_k^{-4}k^3/n^{3/2})$ since $\left\|\hat{V}_k - V_k\right\| = O_p(k/\sqrt{n})$ and $\|\tilde{\gamma}\| \leq \phi_k^{-2}kn^{-1/2}$ by definition of $\Psi_n$. Defining $\breve{\gamma} = V_k^{-1/2}\tilde{\gamma}$, we thus obtain under $(k^5\phi_k^{-6})^{\frac{\xi}{\xi-2}}/n \to 0$,

$$\|\breve{\gamma}\|^2 \leq 4\|\breve{\gamma}\|\Big(\|V_k^{-1/2}\bar{\tilde{m}}_k\| + a_n\|V_k^{-1/2}\eta_k\|\Big) + o_p\Big(\frac{\phi_k^{-1}k}{n}\Big).$$

Now $\|V_k^{-1/2}\bar{\tilde{m}}_k\| = O_p(\sqrt{\phi_k^{-1}k/n})$ since $\bar{\tilde{m}}_k = O_p(\sqrt{k/n})$ while $a_n\|V_k^{-1/2}\eta_k\| = k^{1/4}/\sqrt{n}$. Hence it must be the case that $\|\breve{\gamma}\| = O_p(\sqrt{\phi_k^{-1}k/n})$ which implies $\|\tilde{\gamma}\| \leq \phi_k^{-1/2}\|\breve{\gamma}\| = O_p(\phi_k^{-1}\sqrt{k/n})$ and consequently that $\tilde{\gamma}$ is an interior solution w.p.a.1 and the global maximizer of $F_n(\gamma)$. Thus (10) follows. Combining the above results proves $(L_n - n\bar{m}_k'\hat{V}_k^{-1}\bar{m}_k)/\sqrt{2k} \xrightarrow{p} 0$.

We now show that $\{n\bar{m}_k'\hat{V}_k^{-1}\bar{m}_k - k\}/\sqrt{2k} \xrightarrow{d} N(2^{-1/2}, 1)$. First, noting that $\|\bar{m}_k\| = O_p(k^{3/4}/\sqrt{n})$, by a similar argument as used to show the negligibility of the third term in (12), we obtain

$$\frac{n\bar{m}_k'\hat{V}_k^{-1}\bar{m}_k - k}{\sqrt{2k}} = \frac{n\bar{m}_k'V_k^{-1}\bar{m}_k - k}{\sqrt{2k}} + O_p(\sqrt{\phi_k^{-4}k^4/n}).$$

We can further decompose

$$\frac{n\bar{m}_k'V_k^{-1}\bar{m}_k - k}{\sqrt{2k}} = \frac{n\bar{\tilde{m}}_k'V_k^{-1}\bar{\tilde{m}}_k - k}{\sqrt{2k}} + \frac{1}{\sqrt{2}} + 2\frac{na_n\eta_k'V_k^{-1}\bar{\tilde{m}}_k}{\sqrt{2k}}. \tag{14}$$

Since $E[\tilde{m}_k(X_i)] = 0$ and $E[\tilde{m}_k(X_i)\tilde{m}_k(X_i)'] = V_k$, an argument similar to that used to show (7) implies that the first term of (14) converges in distribution to $N(0, 1)$. The third term of (14) can be written as $n^{-1}\sum_{i=1}^n z_{ni}$ where $z_{ni} = n\sqrt{2k^{-1}}a_n\eta_k'V_k^{-1}\{m_k(X_i) - E[m_k(X_i)]\}$ is an i.i.d.

sequence of random variables satisfying $E[z_{ni}] = 0$ and $E[z_{ni}^2] = 2nk^{-1/2}$. Thus, the Markov inequality implies $n^{-1}\sum_{i=1}^n z_{ni} \xrightarrow{p} 0$ and the conclusion follows.

A.3. **Proof of Theorem 3.**

*Proof of (i).* First, note that under Assumption N, the process $\{n^{-1/2}\sum_{i=1}^n\{s(X_i,p)-s(\Theta_0(\nu),p)\}, \sqrt{n}(\hat{\nu}-\nu) : p \in \mathbb{S}^{d-1}\}$ weakly converges to $\{Z(p),\Gamma : p \in \mathbb{S}^{d-1}\}$. Thus, by (2) and continuity of $G(p;\nu)$ with respect to $p$, the process $\{n^{-1/2}\sum_{i=1}^n s(X_i,\cdot) - s(\Theta_0(\hat{\nu}),\cdot)\}$ weakly converges to $\{Z(\cdot) - G(\cdot;\nu)'\Gamma\}$.

Next, decompose

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n\{s(X_i,p)-s(\Theta_0(\hat{v}),p)\}^2 &= \frac{1}{n}\sum_{i=1}^n\{s(X_i,p)-s(\Theta_0(v),p)\}^2 + \{s(\Theta_0(\nu),p)-s(\Theta_0(\hat{\nu}),p)\}^2 \\ &\quad +2\{s(\Theta_0(\nu),p)-s(\Theta_0(\hat{\nu}),p)\}\frac{1}{n}\sum_{i=1}^n\{s(X_i,p)-s(\Theta_0(v),p)\}.\end{aligned}$$

Using a uniform law of large numbers along with the fact that $\sup_{p\in\mathbb{S}^{d-1}}|s(\Theta(\hat{\nu}),p)-s(\Theta(v),p)| = o_p(1)$ (by (2)), we obtain $\sup_{p\in\mathbb{S}^{d-1}}|n^{-1}\sum_{i=1}^n\{s(X_i,p)-s(\Theta_0(\hat{\nu}),p)\}^2 - \mathrm{Var}(s(X,p))| \xrightarrow{p} 0$.

Finally note that by similar arguments as in the proof of Theorem 1 (i), $n^{-1/2}\max_{1\le i\le n}\sup_{p\in\mathbb{S}^{d-1}}|s(X_i,p)| = o_p(1)$. Thus by the assumption $\|\Theta_0(\hat{\nu})\|_H = O_p(1)$, we have $n^{-1/2}\max_{1\le i\le n}\sup_{p\in\mathbb{S}^{d-1}}|s(X_i,p)-s(\Theta_0(\hat{\nu}),p)| = o_p(1)$. Combining the above results, the conclusion follows by a similar argument as in the proof of Hjort, McKeague and van Keilegom (2009, Theorem 2.1).

*Proof of (ii).* Let $g_i(p) = s(X_i,p)-s(\Theta_0(\hat{\nu}),p)$. Under $H_1$, there exists $p^* \in \mathbb{S}^{d-1}$ such that $E[s(X_i,p^*)] - s(\Theta_o(\nu),p^*) \neq 0$. By (2) and a suitable law of large numbers, $n^{-1}\sum_{i=1}^n g_i(p^*) \xrightarrow{p} E[s(X_i,p^*)] - s(\Theta_0(\nu),p^*)$. Furthermore by similar arguments as in Part (i), $n^{-1}\sum_{i=1}^n g_i(p^*)^2 \xrightarrow{p} E[\{s(X_i,p^*)-s(\Theta_0(\nu),p^*)\}^2] < \infty$. The claim then follows in the same manner as the proof of Theorem 1 (ii).

*Proof of (iii).* The proof is analogous to that of Theorem 1 (iii) and is therefore omitted.

A.4. **Proof of Theorem 4.**

*Proof of (i).* Define $\bar{m}_k = n^{-1}\sum_{i=1}^n m_k(X_i)$ and $\bar{\dot{m}}_k = n^{-1}\sum_{i=1}^n \dot{m}_k(X_i)$. Note that by the mean value theorem (which is applicable since we assume the derivatives exist in a neighborhood of $\nu$), for each $p \in \mathbb{S}^{d-1}$ there exists some $\tilde{\nu}_{(p)}$ satisfying $\|\tilde{\nu}_{(p)} - \nu\| \le \|\hat{\nu}-\nu\|$ and $s(\Theta_0(\hat{\nu}),p) - s(\Theta_0(\nu),p) = G(p;\tilde{\nu}_{(p)})'(\hat{\nu}-\nu)$. Thus by (3) and the asymptotic expansion $\hat{\nu}-\nu = \nabla h(E[z_i])'n^{-1}\sum_{i=1}^n(z_i - E[z_i]) + O_p(n^{-(1+\alpha)/2})$,

$$\begin{aligned}\|\bar{m}_k - \bar{\dot{m}}_k\| &\le \sqrt{k}\sup_{p\in\mathbb{S}^{d-1}}\left\|s(\Theta_0(\hat{\nu}),p)-s(\Theta_0(\nu),p)-G(p;\nu)'(\hat{\nu}-\nu)\right\| + O_p(\sqrt{k/n^{1+\alpha}}) \\ &\le \sqrt{k}\,\|\hat{\nu}-\nu\|\sup_{p\in\mathbb{S}^{d-1}}\left\|G(p;\tilde{\nu}_{(p)})-G(p;\nu)\right\| + O_p(\sqrt{k/n^{1+\alpha}}) \\ &= O_p(\sqrt{k/n^{1+\alpha}}). \qquad (15)\end{aligned}$$

Also note that

$$\bar{\dot{m}}_k = O_p(\sqrt{k/n}), \qquad \bar{m}_k = O_p(\sqrt{k/n}), \tag{16}$$

where the first statement follows from the fact that the process $\{s(X_i, p) - s(\Theta_0(\nu)) - G(p;\nu)'\nabla h(E[z_i])'(z_i - E[z_i]); p \in \mathbb{S}^{d-1}\}$ is $\mu$-Donsker, and the second statement follows by (15). Next, by similar weak convergence arguments as used to show (5) along with $\sup_{p\in\mathbb{S}^{d-1}} |s(\Theta(\hat{\nu}), p) - s(\Theta(\nu), p)| = O_p(n^{-1/2})$ ((by (2), $\hat{\nu} - \nu = O_p(n^{-1/2})$ and continuity of $G(p;\nu)$ in $p$), straightforward algebra assures that

$$\left\|\hat{V}_k - V_k\right\| = O_p(k/\sqrt{n}). \tag{17}$$

Observe that

$$\begin{aligned}
& k^{-1/2}\left\|\dot{m}_k(X_i) - \hat{m}_k(X_i)\right\| \\
\leq & \sup_{p\in\mathbb{S}^{d-1}} \|s(\Theta_0(\nu), p) - s(\Theta_0(\hat{\nu}), p)\| + \|\nabla h(E[z_i]) - \nabla h(\bar{z})\| \, \|z_i - E[z_i]\| \sup_{p\in\mathbb{S}^{d-1}} \|G(p;\nu)\| \\
& + \|\nabla h(\bar{z})\| \, \|z_i - E[z_i]\| \sup_{p\in\mathbb{S}^{d-1}} \|G(p;\nu) - G(p;\hat{\nu})\| + \|\nabla h(\bar{z})\| \, \|\bar{z} - E[z_i]\| \sup_{p\in\mathbb{S}^{d-1}} \|G(p;\hat{\nu})\| \\
= & O_p(n^{-\alpha/2}) \|z_i - E[z_i]\| + O_p(n^{-1/2}),
\end{aligned}$$

where the equality follows from the assumptions in (2) and (3) combined with $\|\bar{z} - E[z_i]\| = O_p(n^{-1/2})$ and $\hat{\nu} - \nu = O_p(n^{-1/2})$. Also we can see that $\bar{V}_k - n^{-1}\sum_{i=1}^n \dot{m}_k(X_i)\dot{m}_k(X_i)'$ is bounded by $2n^{-1}\sum_{i=1}^n \{k^{1/2}\dot{g}_i\delta_i + \delta_i^2\}$, where

$$\dot{g}_i = \sup_{p\in\mathbb{S}^{d-1}} |s(X_i, p) - s(\Theta_0(\nu), p) - G(p;\nu)'\nabla h(E[z_i])'(z_i - E[z_i])|$$

and $\delta_i = \|\dot{m}_k(X_i) - \hat{m}_k(X_i)\|$. Note that our assumptions guarantee $E[\dot{g}_i^2] < \infty$. Thus substituting the expression for $\|\dot{m}_k(X_i) - \hat{m}_k(X_i)\|$ from the previous equation and using the law of large numbers, we obtain $\left\|\bar{V}_k - n^{-1}\sum_{i=1}^n \dot{m}_k(X_i)\dot{m}_k(X_i)'\right\| = O_p(\sqrt{k^2/n^\alpha})$. Also, we have $\left\|n^{-1}\sum_{i=1}^n \dot{m}_k(X_i)\dot{m}_k(X_i)' - \dot{V}_k\right\| = O_p(k/\sqrt{n})$ by analogous weak convergence arguments as used to show (5). Combining these results enables us to show

$$\left\|\bar{V}_k - \dot{V}_k\right\| = O_p(\sqrt{k^2/n^\alpha}). \tag{18}$$

For the conclusion of this theorem, it is sufficient to show the followings:

$$\frac{L_n(\hat{\nu}) - n\bar{m}_k'\bar{V}_k^{-1}\bar{m}_k}{\sqrt{2k}} \xrightarrow{p} 0, \tag{19}$$

$$\frac{n\bar{m}_k'\bar{V}_k^{-1}\bar{m}_k - k}{\sqrt{2k}} \xrightarrow{d} N(0,1). \tag{20}$$

We first show (19). Let $\hat{\gamma} \in \arg\max_{\gamma\in\Gamma_n} G_n(\gamma)$ (in fact $\hat{\gamma}$ is unique as implied from Lemma 1 proved below, though this is not needed for the proof) and $D_n = \max_{1\leq i\leq n} \|m_k(X_i)\|$. Also define $G_n^*(\gamma) = n(2\gamma'\bar{m}_k - \gamma'\bar{V}_k\gamma)$, which is maximized at $\gamma^* = \bar{V}_k^{-1}\bar{m}_k$. For (19), it is sufficient to show that $\hat{\gamma}, \gamma^* = O_p(\dot{\phi}_k^{-1}\sqrt{k/n})$, and $\sup_{\gamma\in\Omega_n\subseteq\Gamma_n} k^{-1/2}|G_n(\gamma) - G_n^*(\gamma)| \xrightarrow{p} 0$ where $\Omega_n = \{\gamma \in \mathbb{R}^k : \|\gamma\| \leq c\dot{\phi}_k^{-1}\sqrt{k/n}\}$ with $c > 0$ chosen to ensure $\Omega_n$ contains both $\hat{\gamma}$ and $\gamma^*$ w.p.a.1 and $\Omega_n \subseteq \Gamma_n$ (such a $c$ exists by the definition of $\Gamma_n$). Indeed these are shown by an argument similar to the proof of Hjort, McKeague and van Keilegom (2009, Proposition 4.1) if the following requirements are satisfied under $(k^5\dot{\phi}_k^{-6})^{\frac{\xi}{\xi-2}}/n \to 0$:

$$(n^{-1/2}k^{3/2}\dot{\phi}_k^{-3})D_n = o_p(1), \tag{21}$$

$$\|\gamma^*\| = O_p(\dot{\phi}_k^{-1}\sqrt{k/n}), \tag{22}$$

$$\lambda_{\max}(\hat{V}_k) = O_p(k), \tag{23}$$

$$\hat{\gamma} \text{ exists w.p.a.1 and } \|\hat{\gamma}\| = O_p(\dot{\phi}_k^{-1}\sqrt{k/n}). \tag{24}$$

To show (21), define $Y_{1i}$ and $Y_{2i}$ to be $k$ dimensional vectors whose $j$-th elements for $1 \le j \le k$ are given by $s(X_i, p_j)$ and $s(\Theta_0(\hat{\nu}), p_j)$, respectively. Note that $m_k(X_i) = Y_{1i} - Y_{2i}$. Now, $(n^{-1/2}k^{3/2}\dot{\phi}_k^{-3})\max_{1\le i\le n}\|Y_{1i}\| = o_p(1)$ by an argument similar to that used to show (8) and $\max_{1\le i\le n}\|Y_{2i}\| = O_p(k^{1/2})$ by the assumption $\|\Theta_0(\hat{\nu})\|_H = O_p(1)$. Thus under $k^4\dot{\phi}_k^{-6}/n \to 0$, (21) follows. Next, (22) follows similarly as (9) by an application of (18) and $k\dot{\phi}_k^{-2}/n^{-\alpha/2} \to 0$ for $\alpha \ge 2/3$. To show (23), observe that $\left\|\hat{V}_k - n^{-1}\sum_{i=1}^n \tilde{m}_k(X_i)\tilde{m}_k(X_i)'\right\| = O_p(k/\sqrt{n})$ by $\sup_{p\in\mathbb{S}^{d-1}}|s(\Theta(\hat{\nu}), p) - s(\Theta(\nu), p)| = O_p(n^{-1/2})$ and $\left\|n^{-1}\sum_{i=1}^n \tilde{m}_k(X_i)\tilde{m}_k(X_i)'\right\| = O_p(k)$ by the assumption $E[\|X_i\|_H^2] < \infty$. Hence using $\lambda_{\max}(\hat{V}_k) \le \left\|\hat{V}_k\right\|$ and the triangle inequality, (23) is verified. Finally, for (24), we first note that $\hat{\gamma}$ exists w.p.a.1 since $\Gamma_n \subseteq S_n$ w.p.a.1 and $\Gamma_n$ is a compact set. Thus letting $b_n = \max_{1\le i\le n}\sup_{\gamma\in\Gamma_n}\left\{1 - (1+\gamma' m_k(X_i))^{-2}\right\}$, an expansion around $\gamma = 0$ yields

$$0 \le G_n(\hat{\gamma}) \le n\{2\hat{\gamma}'\bar{m}_k - \hat{\gamma}'(\bar{V}_k - b_n\hat{V}_k)\hat{\gamma}\}.$$

Note that

$$b_n = O_p\left(\max_{1\le i\le n}\sup_{\gamma\in\Gamma_n}|\gamma' m_k(X_i)|\right) = O_p\left(D_n \sup_{\gamma\in\Gamma_n}\|\gamma\|\right) = o_p(\dot{\phi}_k^{3/2}k^{-1}),$$

where the first equality follows from $1 - (1+x)^{-2} = O(x)$ as $x \to 0$, the second by the Cauchy-Schwarz inequality, and the last by (21) and $\sup_{\gamma\in\Gamma_n}\|\gamma\| = O_p(\dot{\phi}_k^{-3/2}\sqrt{k/n})$; the latter in turn following from the definition of $\Gamma_n$ and $|\dot{\phi}_k^{-3/2} - \bar{\phi}_k^{-3/2}| = O_p(\dot{\phi}_k^{-5/2}k/n^{\alpha/2}) = o_p(\dot{\phi}_k^{-3/2})$ for $\alpha \ge 2/5$ by (18). Subsequently by the above, $\lambda_{\min}(\bar{V}_k - b_n\hat{V}_k) \ge \bar{\phi}_k - |b_n|\lambda_{\max}(\hat{V}_k) = \dot{\phi}_k(1+o_p(1))$, where to obtain the equality we also used (23) along with $\bar{\phi}_k - \dot{\phi}_k = o_p(\dot{\phi}_k)$ which is verified for $\alpha \ge 2/5$ using (18). Thus $\hat{\gamma}'(\bar{V}_k + b_n\hat{V}_k)\hat{\gamma} \ge \|\hat{\gamma}\|^2\dot{\phi}_k(1+o_p(1))$, which implies $\|\hat{\gamma}\| \le 2\dot{\phi}_k^{-1}\|\bar{m}_k\|(1+o_p(1))$. Therefore, by (16) it must be the case that $\hat{\gamma}$ is an interior solution w.p.a.1. (by the choice of $C$ in the definition of $\Gamma_n$) and that $\|\hat{\gamma}\| = O_p(\dot{\phi}_k^{-1}\sqrt{k/n})$. This proves (24). Combining these results, the claim in (19) follows.

We now show (20). We can decompose

$$\begin{aligned}\frac{n\bar{m}_k'\bar{V}_k^{-1}\bar{m}_k - k}{\sqrt{2k}} &= \frac{n\bar{m}_k'(\bar{V}_k^{-1} - \dot{V}_k^{-1})\bar{m}_k}{\sqrt{2k}} + \frac{n(\bar{m}_k - \bar{\dot{m}}_k)'\dot{V}_k^{-1}\bar{m}_k}{\sqrt{2k}} \\ &\quad + \frac{n\bar{\dot{m}}_k'\dot{V}_k^{-1}(\bar{m}_k - \bar{\dot{m}}_k)}{\sqrt{2k}} + \frac{n\bar{\dot{m}}_k'\dot{V}_k^{-1}\bar{\dot{m}}_k - k}{\sqrt{2k}}.\end{aligned} \tag{25}$$

By similar arguments as that used to show the negligibility of third term of (12), the first term of (25) is bounded by $O_p(\sqrt{\dot{\phi}_k^{-4}k^3/n^\alpha})$ using (18). Since $\alpha \ge 2/3$, negligibility of the first term follows. Next, by (15) and (16), the second term of (25) is bounded by $n\dot{\phi}_k^{-1}\|\bar{m}_k - \bar{\dot{m}}_k\|\|\bar{m}_k\|/\sqrt{2k} = O_p(\dot{\phi}_k^{-1}\sqrt{k/n^\alpha})$ and is thus negligible for $\alpha \ge 1/3$. Negligibility of the third term of (25) follows by a similar argument. Finally note that $E[\dot{m}_k(X_i)] = 0$ and $\text{Var}(\dot{m}_k(X_i)) = \dot{V}_k$. Therefore, by $\dot{\phi}_k^{-4} k^2/n \to 0$ and analogous arguments to that used to show (7), the last term of (25) converges in distribution to $N(0,1)$. Thus the result in (20) follows.

*Proof of (ii).* Since in the limit the points $\{p_1, p_2, \ldots, p_k\}$ form a dense subset of $\mathbb{S}^{d-1}$ and the support function is continuous, under $H_1$ there exists an integer $N$ such that for all $n \geq N$ the set of points includes a direction $p^*$ for which $E[s(X_i, p^*) - s(\Theta_0, p^*)] \neq 0$. Without loss of generality we prove the case of $E[s(X_i, p^*) - s(\Theta_0, p^*)] > 0$. Define $g_i(p) = s(X_i, p) - s(\Theta_0(\hat{\nu}), p)$, $\dot{g}_i(p) = s(X_i, p) - s(\Theta_0(\nu), p) - G(p; \nu)'\nabla h(E[z_i])'(z_i - E[z_i])$ and $\bar{g}_i(p) = s(X_i, p) - s(\Theta_0(\hat{\nu}), p) - G(p; \hat{\nu})'\nabla h(\bar{z})'(z_i - \bar{z})$. Pick any $\delta \in (0, 0.3)$ and observe that

$$
\begin{aligned}
L_n(\hat{\nu}) &\geq 2\sum_{i=1}^{n} \log(1 + n^{-(1/2+\delta)} g_i(p^*)) + n^{-2\delta}\Big\{\frac{1}{n}\sum_{i=1}^{n} g_i(p^*)^2 - \frac{1}{n}\sum_{i=1}^{n} \bar{g}_i(p^*)^2\Big\} \\
&= 2n^{1/2-\delta}\left\{\frac{1}{n}\sum_{i=1}^{n} g_i(p^*)\right\} - n^{-2\delta}\left\{\frac{1}{n}\sum_{i=1}^{n} \bar{g}_i(p^*)^2\right\} + O_p(n^{-2\delta}),
\end{aligned}
$$

for all $n \geq N$, where the inequality follows by setting $\gamma = n^{-(1/2+\delta)} e^* \in \Gamma_n$ w.p.a.1 where $e^*$ is the unit vector that selects the component of $m_k(X_i)$ containing $p^*$, and the equality follows from a Taylor expansion. Now, $n^{-1}\sum_{i=1}^{n} g_i(p^*) \xrightarrow{p} E[s(X_i, p^*)] - s(\Theta_0(\nu), p^*) \neq 0$ by suitable law of large numbers and $n^{-1}\sum_{i=1}^{n} \bar{g}_i(p^*)^2 \xrightarrow{p} E[\dot{g}_i(p^*)^2] < \infty$ by a similar argument used to show (18). Thus $L_n(\hat{\nu})$ diverges to infinity at the rate $n^{1/2-\delta}$ which implies, by arguing as in the proof of Theorem 2(ii), that $(L_n(\hat{\nu}) - k)/\sqrt{2k}$ diverges to infinity.

*Proof of (iii).* Define $\breve{m}_k(X_i) = \dot{m}_k(X_i) - E[\dot{m}_k(X_i)]$ and $\bar{\breve{m}}_k = n^{-1}\sum_{i=1}^{n} \breve{m}_k(X_i)$. Note that $\dot{m}_k(X_i) = \breve{m}_k(X_i) + a_n \eta_k$, where $a_n = O(k^{1/4}/\sqrt{n})$ (since $\lambda_{\max}(\dot{V}_k) \leq \|\dot{V}_k\| = O(k)$ by assumption of finite second moments). Now by the same reasoning as that used to show (15), (16) it follows that $\|\bar{m}_k - \bar{\breve{m}}_k\| = O_p(\sqrt{k/n^{1+\alpha}})$ and $\|\bar{\breve{m}}\| = O_p(\sqrt{k/n})$. Furthermore by modifying the arguments of (17) and (18) along the lines of the proof of Theorem 2(iii), it is still shown to be the case that $\left\|\hat{V}_k - V_k\right\| = O_p(k/\sqrt{n})$ and $\left\|\bar{V}_k - \dot{V}_k\right\| = O_p(\sqrt{k^2/n^{\alpha}})$.

We first prove that (19) still holds under $H_{1n}$. As argued in the proof of part (i), this follows if equations (21)-(24) hold. Indeed (21) and (23) follow by the same arguments as in the proof of part (i). To show (22), expand $\gamma^* = \bar{V}_k^{-1}(\bar{m}_k - \bar{\breve{m}}_k) + \bar{V}_k^{-1}\bar{\breve{m}}_k$. Now by standard arguments using $\bar{\phi}_k^{-1} = \dot{\phi}_k^{-1}(1 + o_p(1))$, it follows that $|\bar{V}_k^{-1}(\bar{m}_k - \bar{\breve{m}}_k)| = O_p(\dot{\phi}_k^{-1}\sqrt{k/n^{1+\alpha}})$. Further, by analogous arguments as in the proof of Theorem 2(iii), $|\bar{V}_k^{-1}\bar{\breve{m}}_k| = O_p(\dot{\phi}_k^{-1}\sqrt{k/n})$. Combining the above assures $\gamma^* = O_p(\dot{\phi}_k^{-1}\sqrt{k/n})$ which proves (22). Finally, for (24) note that by the same reasoning as in Part (i), we still have $\hat{\gamma}'(\bar{V}_k - b_n\hat{V}_k)\hat{\gamma} \leq 2\hat{\gamma}'\bar{m}_k$, which implies

$$
\hat{\gamma}'(\dot{V}_k - b_n V_k)\hat{\gamma} \leq 2\hat{\gamma}'\bar{m}_k + O_p(\dot{\phi}_k^{-3} k^2/n^{1+\alpha/2}),
$$

since $\left\|\hat{V}_k - V_k\right\| = O_p(k/\sqrt{n})$, $\left\|\bar{V}_k - \dot{V}_k\right\| = O_p(\sqrt{k^2/n^{\alpha}})$ and $\sup_{\gamma \in \Gamma_n}\|\gamma\| = O_p(\dot{\phi}_k^{-3/2}\sqrt{k/n})$; the latter in turn following from the definition of $\Gamma_n$ along with $\bar{\phi}_k^{-3/2} = \dot{\phi}_k^{-3/2}(1 + o_p(1))$. Defining $\breve{\gamma} = \dot{V}_k^{-1/2}\hat{\gamma}$ and substituting in the previous display equation, we obtain under $(k^5\dot{\phi}_k^{-6})^{\frac{\xi}{\xi-2}}/n \to$

0 and $\alpha \geq 2/3$,

$$\lambda_{\min}(I - b_n \dot{V}_k^{-1/2} V_k \dot{V}_k^{-1/2}) \|\breve{\gamma}\|^2 \leq 2\|\breve{\gamma}\| \left( \left\| \dot{V}_k^{-1/2}(\bar{m}_k - \bar{\bar{m}}_k) \right\| + \left\| \dot{V}_k^{-1/2} \breve{\bar{m}}_k \right\| + a_n \left\| \dot{V}_k^{-1/2} \eta_k \right\| \right) + o_p(\dot{\phi}_k^{-1} k/n).$$

Now $\lambda_{\min}(I - b_n \dot{V}_k^{-1/2} V_k \dot{V}_k^{-1/2}) \geq 1 - |b_n| \dot{\phi}_k^{-1} \lambda_{\max}(V_k) = 1 + o_p(1)$, where for the equality we used $\lambda_{\max}(V_k) = O(k)$ and the definition of $b_n$. Furthermore, note that $\left\| \dot{V}_k^{-1/2}(\bar{m}_k - \bar{\bar{m}}_k) \right\| = O_p(\sqrt{\dot{\phi}_k^{-1} k/n^{1+\alpha}})$, $\left\| \dot{V}_k^{-1/2} \breve{\bar{m}}_k \right\| = O_p(\sqrt{\dot{\phi}_k^{-1} k/n})$ and $a_n \left\| \dot{V}_k^{-1/2} \eta_k \right\| = O_p(k^{1/4}/\sqrt{n})$. Therefore, it must be the case that $\|\breve{\gamma}\| = O_p(\sqrt{\dot{\phi}_k^{-1} k/n})$, which implies $\|\hat{\gamma}\| \leq \dot{\phi}_k^{-1/2} \|\breve{\gamma}\| = O_p(\dot{\phi}_k^{-1} \sqrt{k/n})$. This proves (24). Combining the results, the claim in (19) follows.

We now show that $\{n \bar{m}_k' \bar{V}_k^{-1} \bar{m}_k - k\}/\sqrt{2k} \xrightarrow{d} N(2^{-1/2}, 1)$. Decompose,

$$\frac{n \bar{m}_k' \bar{V}_k^{-1} \bar{m}_k - k}{\sqrt{2k}} = \frac{n(\bar{m}_k - \bar{\bar{m}}_k)' \bar{V}_k^{-1} \bar{m}_k}{\sqrt{2k}} + \frac{n \bar{\bar{m}}_k' \bar{V}_k^{-1} (\bar{m}_k - \bar{\bar{m}}_k)}{\sqrt{2k}} + \frac{n \bar{\bar{m}}_k' (\bar{V}_k^{-1} - \dot{V}_k^{-1}) \bar{\bar{m}}_k}{\sqrt{2k}} + \frac{n \bar{\bar{m}}_k' \dot{V}_k^{-1} \bar{\bar{m}}_k - k}{\sqrt{2k}}. \quad (26)$$

Note that $a_n = O(k^{1/4}/\sqrt{n})$ implies $\bar{\bar{m}}_k = O_p(k^{3/4}/\sqrt{n})$ and subsequently, $\bar{m}_k = O_p(k^{3/4}/\sqrt{n})$. Thus by straightforward algebra and the fact that $|\bar{\phi}_k^{-1} - \dot{\phi}_k^{-1}| = o_p(\dot{\phi}_k^{-1})$, the first two terms of (26) are $O_p(\dot{\phi}_k^{-1} k^{3/4}/n^{\alpha/2})$ and therefore negligible under $(k^5 \dot{\phi}_k^{-6})^{\frac{\xi}{\xi-2}}/n \to 0$ and $\alpha \geq 1/3$. Next we show that the third term of (26) is negligible. To this end we further decompose the third term of (26) as

$$\frac{n \bar{\bar{m}}_k' (\bar{V}_k^{-1} - \dot{V}_k^{-1}) \bar{\bar{m}}_k}{\sqrt{2k}} = \frac{n \breve{\bar{m}}_k' (\bar{V}_k^{-1} - \dot{V}_k^{-1}) \breve{\bar{m}}_k}{\sqrt{2k}} + \frac{n a_n \breve{\bar{m}}_k' \bar{V}_k^{-1} (\dot{V}_k - \bar{V}_k) \dot{V}_k^{-1/2} (\dot{V}_k^{-1/2} \eta_k)}{\sqrt{2k}} + \frac{n a_n^2 \eta_k' \bar{V}_k^{-1} (\dot{V}_k - \bar{V}_k) \dot{V}_k^{-1/2} (\dot{V}_k^{-1/2} \eta_k)}{\sqrt{2k}}. \quad (27)$$

Now by similar arguments as that used to show the negligibility of third term of (12), the first term of (27) is $O_p(\sqrt{\dot{\phi}_k^{-4} k^3/n^{\alpha}})$. Next, the second term of (27) is bounded by

$$n k^{-1/2} a_n \|\breve{\bar{m}}\| \bar{\phi}_k^{-1} \dot{\phi}_k^{-1/2} \left\| \dot{V}_k - \bar{V}_k \right\| \left\| \dot{V}_k^{-1/2} \eta_k \right\| = O_p(\sqrt{\dot{\phi}_k^{-3} k^{5/2}/n^{\alpha}}),$$

where the equality follows from $\|\breve{\bar{m}}\| = O_p(\sqrt{k/n})$, $\left\| \bar{V}_k - \dot{V}_k \right\| = O_p(\sqrt{k^2/n^{\alpha}})$, and $|\bar{\phi}_k^{-1} - \dot{\phi}_k^{-1}| = o_p(\dot{\phi}_k^{-1})$. Furthermore, an analogous argument shows that the third term of (27) is $O_p(\sqrt{\dot{\phi}_k^{-3} k^3/n^{\alpha}})$. Combining these results proves that for $\alpha \geq 2/3$, the third term of (26) is negligible. Finally, by $\dot{\phi}_k^{-4} k^2/n \to 0$ and similar arguments to that used in the proof of Theorem 2 (iii), the last term of (26) converges in distribution to $N(2^{-1/2}, 1)$. Thus the claim follows.

A.5. **Proof of Lemma 1.** The claim follows if we show that uniformly over $\gamma \in \Gamma_n$, the second derivative of $G_n(\gamma)$ denoted by $D^2 G_n(\gamma)$ is negative definite w.p.a.1. Define the preference relation $\succeq$ over matrices as $A \succeq B$ if $A - B$ is positive semi-definite. Then note that for all $\gamma \in \Gamma_n$,

$$-\frac{1}{2n} D^2 G_n(\gamma) = \frac{1}{n} \sum_{i=1}^{n} \frac{m_k(X_i) m_k(X_i)'}{(1 + \gamma' m_k(X_i))^2} + \bar{V}_k - \hat{V}_k \succeq \bar{V}_k - |b_n| \hat{V}_k,$$

where $b_n = \sup_{\gamma\in\Gamma_n, 1\le i\le n}\left\{1-(1+\gamma' m_k(X_i))^{-2}\right\}$ is as defined in the proof of Theorem 4. Now by similar arguments as in the proof of Theorem 4 (i), we observe under $(k^5\bar{\phi}_k^{-6})^{\frac{\xi}{\xi-2}}/n \to 0$ that $(n^{-1/2}k^{3/2}\bar{\phi}_k^{-3})D_n = o_p(1)$, and consequently $|b_n| = o_p(\bar{\phi}_k^{3/2}k^{-1})$. Thus $\lambda_{\min}(\bar{V}_k - b_n\hat{V}_k) \ge \bar{\phi}_k - |b_n|\lambda_{\max}(\hat{V}_k) = \bar{\phi}_k(1+o_p(1))$, where to obtain the equality we also used (23) (note from the proof of Theorem 4 (i) that (23) holds regardless of whether under the null or the alternative). The above implies $\inf_{\gamma\in\Gamma_n}\{\lambda_{\min}(-D^2G_n(\gamma))\} \ge 2n\bar{\phi}_k(1+o_p(1)) > 0$ w.p.a.1, which proves the desired result.

A.6. **Proof of Proposition 1.**

*Proof of (i).* It is enough to show that the conditions of Theorem 3 are satisfied.

We first show that Assumption N is satisfied. The condition $\left\|\Theta_0(\text{vec}(\hat{\Sigma}))\right\|_H = O_p(1)$ is verified by $\left\|\Theta_0(\text{vec}(\hat{\Sigma}))\right\|_H = \left\|\hat{\Sigma}\Theta_0\right\|_H \le \left\|\hat{\Sigma}\right\| \|\Theta_0\|_H$ and $\hat{\Sigma} \xrightarrow{p} \Sigma$. Let $\nu = \text{vec}(\Sigma)$ and $G(p;\nu) = p \otimes \nabla s(\Theta_0, \Sigma p)$ for $p \in \mathbb{S}^q$. We now verify that $G(p;\nu)$ satisfies the properties set out in Assumption N. Since $x_i$ has no mass points, by Bontemps, Magnac and Maurin (2012, Lemma 3) $s(\Theta_0, p)$ is Fréchet differentiable with derivative $\nabla s(\Theta_0, p) = E\Big[(y_{Ui} - y_{Li})\mathbb{I}\{f(x_i,p)\ge 0\}(1, x_i')'\Big]$ at each $p \in \mathbb{R}^{q+1}\setminus\{0\}$. From the expression for $\nabla s(\Theta_0,p)$, it follows that $\nabla s(\Theta_0,p)$ is continuous at each $p \in \mathbb{R}^{q+1}\setminus\{0\}$ (since $x_i$ has no mass points). This proves continuity of $G(p;\nu)$ with respect to $p$. We now show $\sup_{p\in\mathbb{S}^q}\|G(p;\hat{\nu}) - G(p;\nu)\| \xrightarrow{p} 0$. Since $\nabla s(\Theta_0,p)$ is continuous at each $p\in\mathbb{S}^q$, it is also uniformly continuous on $\mathbb{S}^q$. So, we have

$$\sup_{p\in\mathbb{S}^{d-1}}\left\|\nabla s(\Theta_0,\hat{\Sigma}p) - \nabla s(\Theta_0,\Sigma p)\right\| = \sup_{p\in\mathbb{S}^{d-1}}\left\|\nabla s(\Theta_0, \hat{\Sigma}p/\left\|\hat{\Sigma}p\right\|) - \nabla s(\Theta_0, \Sigma p/\|\Sigma p\|)\right\| \xrightarrow{p} 0,$$

where the equality follows from $\nabla s(\Theta_0,p) = \nabla s(\Theta_0, p/\|p\|)$ for $p\in\mathbb{R}^{q+1}\setminus\{0\}$ (i.e., $\nabla s(\Theta_0,p)$ is homogenous of degree 0 in $p$), and the probability limit follows by uniform continuity and the fact that $\sup_{p\in\mathbb{S}^q}\left\|(\hat{\Sigma}p/\left\|\hat{\Sigma}p\right\|) - (\Sigma p/\|\Sigma p\|)\right\| \xrightarrow{p} 0$ if $\hat{\Sigma}\xrightarrow{p}\Sigma$ and $\Sigma$ is positive definite. Thus the claim $\sup_{p\in\mathbb{S}^q}\|G(p;\hat{\nu}) - G(p;\nu)\| \xrightarrow{p} 0$ follows. We now show that (2) holds. Indeed, this follows by an argument similar to the proof of Bontemps, Magnac and Maurin (2012, Lemma 13) which assures that $\sup_{p\in\mathbb{S}^q}|s(\Theta_0,\hat{\Sigma}p) - s(\Theta_0,\Sigma p) - \nabla s(\Theta_0,p)'(\hat{\Sigma}-\Sigma)p| = o_p(n^{-1/2})$. Combining these results, Assumption N is seen to be verified.

We next verify the other conditions of Theorem 3. The condition $E[\|W_i\|_H^{\xi}] < \infty$ for some $\xi > 2$ follows from $E[\sup_{p\in\mathbb{S}^q}|s(W_i,p)|^{\xi}] < \infty$, which is verified under the stated assumptions on the moments of $\{y_{Li}, y_{Ui}, x_i\}$. Furthermore the condition $E[\|z_i\|^2] < \infty$ follows from the assumption of finite fourth moments of $x_i$. It remains to show $\inf_{p\in\mathbb{S}^q}\text{Var}(s(W_i,p)) > 0$. To this end, we adapt the proof of Beresteanu and Molinari (2008, Theorem 4.3) to write

$$\begin{aligned}
\text{Var}(s(W_i,p)) &\ge E\left[\text{Var}\left(\left[y_{Li}\mathbb{I}\{f(x_i,p)<0\} + y_{Ui}\mathbb{I}\{f(x_i,p)\ge 0\}\right]f(x_i,p)\Big|x_i\right)\right]\\
&= E\left[f(x_i,p)^2\left\{\text{Var}(y_{Li}|x_i)\mathbb{I}\{f(x_i,p)<0\} + \text{Var}(y_{Ui}|x_i)\mathbb{I}\{f(x_i,p)\ge 0\}\right\}\right]\\
&\ge \sigma^2 p'\Sigma p > 0
\end{aligned}$$

for each $p\in\mathbb{S}^q$, where the first inequality follows by the law of iterated expectations, the second inequality follows from the assumption $\text{Var}(y_{Li}|x_i), \text{Var}(y_{Ui}|x_i) \ge \sigma^2 > 0$ a.s., and the last inequality follows by the assumption of full rank for $\Sigma$. We note that an extension of the above argument also suffices to show the stronger condition $\inf_{p\in\mathbb{S}^q} \mathrm{Var}(\tilde{Z}(p)) > 0$, which is needed for the validity of the bootstrap. Indeed, this follows from the fact that $\mathrm{Var}(\tilde{Z}(p)) = \mathrm{Var}(s(W_i, p) - G(p, \nu)'(z_i - \mathrm{vec}(\Sigma))) \geq E[\mathrm{Var}(s(W_i, p)|x_i)]$ since $z_i$ is a measurable function of $x_i$.

*Proof of (ii).* Note that the only additional assumption of Theorem 4 is the one corresponding to equation (3). Thus it is sufficient to prove that $G(p; \nu) = p \otimes \nabla s(\Theta_0, \Sigma p)$ satisfies equation (3). First note that as $\Sigma$ is assumed to be full rank, there exists a neighborhood $\mathcal{N}$ of $\Sigma$ such that $\inf_{p\in\mathbb{S}^q} \left\|\tilde{\Sigma}p\right\| \geq c > 0$ for all $\tilde{\Sigma} \in \mathcal{N}$ and some positive constant $c$. Then there exists a constant $M$ independent of $(\Sigma, \tilde{\Sigma}, p)$ such that for all $\Sigma, \tilde{\Sigma} \in \mathcal{N}$ and $p \in \mathbb{S}^q$,

$$\begin{aligned} \left\|\nabla s(\Theta_0, \Sigma p) - \nabla s(\Theta_0, \tilde{\Sigma} p)\right\| &= \left\|\nabla s(\Theta_0, \Sigma p / \|\Sigma p\|) - \nabla s(\Theta_0, \tilde{\Sigma} p / \left\|\tilde{\Sigma} p\right\|)\right\| \\ &\leq M \left\|(\Sigma p / \|\Sigma p\|) - (\tilde{\Sigma} p / \left\|\tilde{\Sigma} p\right\|)\right\|^{\alpha} \\ &\leq 2Mc^{-\alpha} \left\|\Sigma - \tilde{\Sigma}\right\|^{\alpha}, \end{aligned}$$

for some $\alpha \geq 2/3$, where the first inequality follows by the assumption of local Hölder continuity of $\nabla s(\Theta_0, p)$ of order $\alpha \geq 2/3$ on $\mathbb{S}^q$ (since on compact metric spaces, local Hölder continuity is equivalent to (global) Hölder continuity). Therefore, by $G(p; \nu) = p \otimes \nabla s(\Theta_0, \Sigma p)$, the claim follows.

## References


[1] Beresteanu, A. and F. Molinari (2008) Asymptotic properties for a class of partially identified models, *Econometrica*, 76, 763-814.

[2] Bontemps, C., Magnac, T. and E. Maurin (2012) Set identified linear models, *Econometrica*, 80, 1129-1155.

[3] Chandrasekhar, A., Chernozhukov, V., Molinari, F. and P. Schrimpf (2012) Inference for best linear approximations to set identified functions, *arXiv preprint arXiv:1212.5627.*

[4] Chen, S. X., Peng, L. and Y. L. Qin (2009) Effects of data dimension on empirical likelihood, *Biometrika*, 96, 711-722.

[5] Chen, S. X., and I. van Keilegom (2009) A review on empirical likelihood methods for regression, *Test*, 18, 415-447.

[6] Chernozhukov, V., Kocatulum, E. and K. Menzel (2012) Inference on sets in finance, *arXiv preprint arXiv:1211.4282*

[7] de Jong, R. M. and H. J. Bierens (1994) On the limit behavior of a chi-square type test if the number of conditional moments tested approaches infinity, *Econometric Theory*, 10, 70-90.

[8] Einmahl, J. H. J. and I. W. McKeague (2003) Empirical likelihood based hypothesis testing, *Bernoulli*, 9, 267-290.

[9] Giné, E. and J. Zinn (1990) Bootstrapping general empirical measures, *Annals of Probability*, 18, 851-869.

[10] Hjort, N. L., McKeague, I. W. and I. van Keilegom (2009) Extending the scope of empirical likelihood, *Annals of Statistics*, 37, 1079-1111.

[11] Kaido, H. (2010) A dual approach to inference for partially identified econometric models, Working Paper. Boston University.

[12] Kaido, H. and A. Santos (2014) Asymptotically efficient estimation of models defined by convex moment inequalities, *Econometrica,* 82, 387-413.

[13] Lifshits, M. A. (1982) On the absolute continuity of distributions of functionals of random processes, *Theory of Probability and Its Applications*, 27, 600-607.

[14] Manski, C. F. (2003) *Partial identification of probability distributions,* New York: Springer Verlag.

[15] Molchanov, I. (2005) *Theory of Random Sets*, Springer.

[16] Molchanov, I. and F. Molinari (2014) Applications of random set theory in econometrics, *Annual Review of Economics,* forthcoming.

[17] Owen, A. B. (1988) Empirical likelihood ratio confidence intervals for a single functional, *Biometrika*, 75, 237-249.

[18] Owen, A. B. (2001) *Empirical likelihood*, CRC press.

[19] van der Vaart, A. W. (1998) *Asymptotic Statistics*, Cambridge University Press.

[20] van Keilegom, I., Sellero, C. S. and W. G. Manteiga (2008) Empirical likelihood based testing for regression, *Electronic Journal of Statistics,* 2, 581-604.



Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK.

*E-mail address*: k.adusumilli@lse.ac.uk

Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK.

*E-mail address*: t.otsu@lse.ac.uk