

# ASYMPTOTIC PROPERTIES FOR A CLASS OF PARTIALLY IDENTIFIED MODELS

---

*Arie Beresteanu*  
*Francesca Molinari*

THE INSTITUTE FOR FISCAL STUDIES  
DEPARTMENT OF ECONOMICS, UCL  
**cemmap** working paper CWP10/06

# Asymptotic Properties for a Class of Partially Identified Models\*

Arie Beresteanu  
Department of Economics  
Duke University  
arie@econ.duke.edu

Francesca Molinari  
Department of Economics  
Cornell University  
fm72@cornell.edu

First Version: August 15, 2005. This Version: June 5, 2006

## Abstract

We propose inference procedures for partially identified population features for which the population identification region can be written as a transformation of the Aumann expectation of a properly defined *set valued random variable* (SVRV). An SVRV is a mapping that associates a set (rather than a real number) with each element of the sample space. Examples of population features in this class include sample means and best linear predictors with interval outcome data, and parameters of semiparametric binary models with interval regressor data. We extend the analogy principle to SVRVs, and show that the sample analog estimator of the population identification region is given by a transformation of a Minkowski average of SVRVs. Using the results of the mathematics literature on SVRVs, we show that this estimator converges in probability to the identification region of the model with respect to the Hausdorff distance. We then show that the Hausdorff distance between the estimator and the population identification region, when properly normalized by  $\sqrt{n}$ , converges in distribution to the supremum of a Gaussian process whose covariance kernel depends on parameters of the population identification region. We provide consistent bootstrap procedures to approximate this limiting distribution. Using similar arguments as those applied for vector valued random variables, we develop a methodology to test assumptions about the true identification region and to calculate the power of the test. We show that these results can be used to construct a *confidence collection*, that is a collection of sets that, when specified as null hypothesis for the true value of the population identification region, cannot be rejected by our test.

**JEL Classifications:** C14.

**Keywords:** Partial Identification, Confidence Collections, Set-Valued Random Variables.

---

\*We are grateful to Larry Blume for introducing us to the theory of Set Valued Random Variables and for many useful conversations, and to Ilya Molchanov for extremely valuable suggestions on several aspects of this paper. We benefited from comments by Levon Barseghyan, Victor Chernozhukov, Ani Guerdjikova, Stefan Hoderlein, Bo Honoré, Joel Horowitz, Nick Kiefer, Enno Mammen, Chuck Manski, Massimo Marinacci, Karl Mosler, Ulrich Müller, Morten Nielsen, Whitney Newey, Vladimir Rotar, Marco Scarsini, Elie Tamer, Tymon Tatur, Tim Vogelsang, Martin Wells, seminar participants at Cornell, Duke, Harvard-MIT, Mannheim, Northwestern, Princeton, Torino University, UCL, and from the opportunity to present this work at Northwestern University's "Mini-Conference on Incomplete Econometric Models," at the "Triangle Econometrics Conference," and at the 2006 "New York Econometrics Camp." We thank Jenny Law and Liyan Yang for able research assistantship. Financial support from the Center for Analytic Economics at Cornell University and from the Micro Incentives Research Center at Duke University is gratefully acknowledged.

# 1 Introduction

This paper contributes to the growing literature on inference for partially identified population features. These features include vectors of parameters or of statistical functionals characterizing a population probability distribution of interest, for which the sampling process and the maintained assumptions reveal that they lie in a set – their identification region – which is not necessarily a singleton (Manski (2003)). Much of this literature focuses on cases in which the population identification region is an interval on the real line (e.g., Manski (1989)), or can be defined as the set of minimizers of a criterion function (e.g., Chernozhukov, Hong & Tamer (2004)). In these cases the analogy principle is applied to replace either the extreme points of the interval or the criterion function with their sample analogs, to obtain an estimator for the identification region. Limit theorems for sequences of scalar or vector valued random variables are then employed to establish consistency of these estimators, and to construct confidence sets that asymptotically cover each point in the identification region (Imbens & Manski (2004)) or the entire identification region (Chernozhukov et al. (2004)) with at least a prespecified probability.

In this paper, we introduce a novel approach for estimation and inference for a certain class of partially identified population features. The key insight that leads to our approach is the observation that, within this class, the identification region of the vector of parameters (or statistical functionals) of interest is given by the “expectation” of a measurable mapping that associates a set (rather than a real number or a real vector) with each element of the sample space. In the mathematics literature, this measurable mapping is called a *set valued random variable* (SVRV). Just as one can think of the identification region of a parameter vector as a set of parameter vectors, one can think of an SVRV as a set of random variables (Aumann (1965)). We extend the analogy principle to SVRVs and estimate the identification region, which is the “expectation” of an SVRV, by its sample analog, which is a “sample average” of SVRVs. The expressions “expectation” and “sample average” used above are in quotation marks because when working with sets, a particular expectation operator needs to be used, the *Aumann expectation*. Similarly, a particular summation operator needs to be used, the *Minkowski summation*, in order to get the set analog of the sample average, which is called *Minkowski sample average*.<sup>1</sup> Approaching the problem from this perspective is beneficial because it allows the researcher to perform, in the space of sets, operations which are analogs to

---

<sup>1</sup>These concepts are formalized in Section 2.

those widely used in the space of vectors. In particular, we show how to use limit theorems for SVRVs to conduct inference for the entire identification region of a population feature of interest, in a way that is completely analogous to how inference would be conducted if this feature were point identified.

**Overview.** Our methodology applies to the class of partially identified population features for which the identification region is equal to a transformation of the Aumann expectation of a properly defined SVRV. Examples of population features in this class include means and best linear predictors with interval outcome data, and parameter vectors characterizing semiparametric binary models with interval regressor data (under the assumptions of Magnac & Maurin (2005)). The SVRVs whose Aumann expectation is equal to the identification region of the population feature of interest can be constructed from observable random variables.

The mathematics literature on SVRVs provides Laws of Large Numbers and Central Limit Theorems for the Hausdorff distance between the Minkowski average of a random sample of SVRVs, and their Aumann expectation. Using these results, we show that our estimator of the identification region, which is given by a transformation of the Minkowski average of a random sample of SVRVs, is  $\sqrt{n}$ -consistent, in the sense that the Hausdorff distance between the estimated set and the population identification region converges to zero at the rate  $O_p(1/\sqrt{n})$ . This result does not depend on whether the random variables used to construct the SVRVs have a continuous or a discrete distribution. When the Hausdorff distance between our estimator and the population identification region is properly normalized by  $\sqrt{n}$ , it converges in distribution to the supremum of a Gaussian process whose covariance kernel depends on parameters of the population identification region which can be consistently estimated.

Having an asymptotic distribution for the Hausdorff distance between the estimated set and the population identification region allows us to perform hypotheses testing. In particular, we introduce a test which is consistent against any fixed alternative. We then extend the notion of local alternatives to partially identified models, derive the asymptotic distribution of our test against these local alternatives, and show that the test is locally asymptotically unbiased. Our test statistic is given by the Hausdorff distance between the estimated set and the population identification region, and the test rejects a particular null hypothesis for the population identification region if this statistic exceeds a critical value. We show that this critical value can be consistently estimated

using bootstrap methods.

The test can be inverted to construct what we call a *confidence collection* for the population identification region. The confidence collection is given by the collection of all sets that, when specified as null hypothesis for the true value of the population identification region, cannot be rejected by our test. Its main property is that (asymptotically) the population identification region is one of its elements with a prespecified confidence level  $(1 - \alpha)$ .

Hence for this class of problems there is a complete analogy, at the level of estimation and inference, between the approach usually adopted for point identified parameters and the approach that we propose for partially identified parameters. In particular, when point identified, the parameters of interest can be consistently estimated using a transformed sample average of the data. The resulting estimator has an asymptotically normal distribution. The confidence region for the parameter vector is given by a collection of vectors – that is, a collection of points in the relevant space – and can be obtained through the inversion of a properly specified test statistic. In the partially identified case, our results show that the identification region of each of these parameter vectors can be consistently estimated using a transformed Minkowski average of the data. The Hausdorff distance between this estimator and its population counterpart has an asymptotic distribution which is a function of a Gaussian process. The confidence region for the population identification region is given by a collection of sets (rather than points) and can be obtained through the inversion of the test statistic that we propose.

Our inferential approach targets the entire identification region of a partially identified population feature, and provides asymptotically exact size critical values with which to test hypotheses and construct confidence collections. However, there are applications in which the researcher might want to test hypotheses and construct confidence sets for the “true” value of the population feature of interest, following the insight of Imbens & Manski (2004). For this case, we show that our methodology can be modified to conduct conservative tests of hypotheses, and construct conservative confidence sets that asymptotically cover each point in the identification region with a prespecified probability (details in Section 6.2).

**Structure of the Paper.** Because the literature on SVRVs has never been used before in econometrics, we devote the first part of the paper (Section 2) to introducing the main ingredients of this theory. In Section 3, we propose our test statistic, establish its properties, provide a consis-

tent bootstrap procedure to estimate its limiting distribution, show how the test statistic can be inverted to obtain the confidence collection, and provide a simple characterization of the collection. In Section 4, we apply our results to the simple problem of inference for the mean of a random variable for which only interval data are available. In Section 5, we introduce a  $\sqrt{n}$ -consistent estimator of the population identification region for the best linear predictor parameters when only interval data are available for the outcome variable, but the covariates are perfectly observed. We then derive the asymptotic distribution of the Hausdorff distance between this estimator and the population identification region, and provide a consistent bootstrap procedure to obtain its critical values. Section 6 describes extensions of the results in Sections 3-5. In particular, it outlines how to conduct inference for the semiparametric binary model with interval regressor data of Magnac & Maurin (2005), and how to construct confidence sets that asymptotically cover each point in the identification region with at least a prespecified probability. Section 7 presents Monte Carlo results evaluating the finite sample performance of our estimators and test statistics, and Section 8 concludes. An Appendix contains all the proofs.

**Related Literature.** Consistent estimators for specific partially identified population features have been proposed, among others, by Manski (1989), Horowitz and Manski (1997, 1998, 2000), Manski & Tamer (2002), Chernozhukov et al. (2004), Honoré & Tamer (2004), Andrews, Berry & Jia (2004), and Chernozhukov, Hahn & Newey (2005). The development of methodologies that allow for the construction of confidence regions for partially identified population features is a topic of current research. Horowitz and Manski (1998, 2000) consider the case in which the identification region of the parameter of interest is an interval whose lower and upper bounds can be estimated from sample data, and propose confidence intervals that asymptotically cover the entire identification region with fixed probability. For the same class of problems, Imbens & Manski (2004) suggest shorter confidence intervals that (asymptotically) uniformly cover each point in the identification region, rather than the entire region, with at least a prespecified probability  $(1 - \alpha)$ . Chernozhukov et al. (2004) are the first to address the problem of construction of confidence sets for identification regions of parameters obtained as the solution of the minimization of a criterion function. They provide methods to construct confidence sets that cover the entire identification region with probability asymptotically equal to  $(1 - \alpha)$ , as well as confidence sets that asymptotically cover each point in the identification region with at least probability  $(1 - \alpha)$ . They also develop resampling methods

to implement these procedures. Shaikh (2005) analyzes the same problem, and proposes various subsampling procedures. Andrews et al. (2004) consider economic models of entry in which the equilibrium conditions place a set of inequality restrictions on the parameters of the model. These restrictions may only allow the researcher to identify a set of parameter values consistent with the observable data. They suggest a procedure to obtain confidence regions that asymptotically cover the identified set with at least probability  $(1 - \alpha)$  by looking directly at the distribution of the inequality constraints. Pakes, Porter, Ho & Ishii (2005) consider single agent and multiple agent structural models in which again equilibrium conditions impose moment inequality restrictions on the parameters of interest. They suggest a conservative specification test for the value of the estimated parameters. Rosen (2006) considers related models in which the parameter of interest is partially identified by a finite number of moment inequalities. He establishes a connection between these models and the literature on multivariate one-sided hypothesis tests, and shows that for this class of models, conservative confidence sets can be constructed by inverting a statistic that tests the hypothesis that a given element of the parameter space satisfies the restrictions of the model. Guggenberger, Hahn & Kim (2006) further explore specification tests for parameter vectors obtained as the solution of moment inequalities. Molinari (2005) considers misclassification models in which the identification region for the distribution of the misclassified variable can be calculated using a nonlinear programming estimator, the constraints of which depend on parameters that can be consistently estimated. She proposes conservative confidence sets given by the union of the identification regions obtained through the nonlinear programming estimator by replacing the estimated parameters with the elements of their Wald confidence ellipsoid. Galichon & Henry (2006) propose a specification test for partially identified structural models based on an extension of the Kolmogorov-Smirnov test for Choquet capacities.

## **2 Background: Asymptotic Theory for Set Valued Random Variables**

In this section we provide a succinct presentation of the theory of SVRVs, following closely the treatment in Molchanov (2005) (limit theorems for SVRVs are also discussed in Li, Ogura & Kreinovich (2002)). We first rigorously define an SVRV and describe the basic operations on sets that we use in the paper. Then we review how SVRVs can be represented as elements of a functional space.

Similarly, we review how the Hausdorff distance between SVRVs can be represented as a transformation of elements of a functional space. For the purposes of this paper, this is the key result of the literature on SVRVs: it enables us to derive the asymptotic properties of partially identified models using functional Laws of Large Numbers and Central Limit Theorems which are of familiar use in econometrics.

**Notation.** Throughout the paper (with a few exceptions), we reserve the use of capital Latin letters to sets and SVRVs; we use lower case Latin letters for random variables and boldface lower case Latin letters for random vectors. We denote sets of parameters by capital Greek letters, scalar valued parameters by lower case Greek letters and vector valued parameters by boldface lower case Greek letters.

### Set Valued Random Variables, Basic Operations, Norm of a Set and Distance between Sets

Let  $(\Omega, \mathcal{A})$  be a measurable space and let  $K(\mathfrak{R}^d)$  be the collection of all non-empty closed subsets of  $\mathfrak{R}^d$ . In what follows,  $\mathfrak{R}^d$  denotes the Euclidean space, equipped with the Euclidean norm (which is denoted by  $\|\cdot\|$ ). As the name suggests, an SVRV is a measurable mapping  $F : \Omega \rightarrow K(\mathfrak{R}^d)$  that associates a set to each point in the sample space. However, “measurable” in this context has a slightly different meaning than in the case of random variables.

**Definition 1** *A map  $F : \Omega \rightarrow K(\mathfrak{R}^d)$  is called a **set valued random variable (SVRV)** if for each closed subset  $C$  of  $\mathfrak{R}^d$ ,  $F^{-1}(C) = \{\omega \in \Omega : F(\omega) \cap C \neq \emptyset\} \in \mathcal{A}$ .*

The measurability concept used above is different from the more familiar one for vector valued random variables because it must be “restrictive enough to ensure that all functionals of interest [of the SVRV] become random variables” (Molchanov (2005), page 1). An example of a relevant functional of SVRVs which, given Definition 1, is a random variable, is the Hausdorff distance (formally defined below) between two SVRVs. A few examples can help clarify the concept of an SVRV:

#### Example 1

- a) (Trivial) Consider a random variable  $y$ . Then the singleton  $\{y\}$  is an SVRV.
- b) If  $y$  is a random variable, then  $F \equiv (-\infty, y]$  is an SVRV.



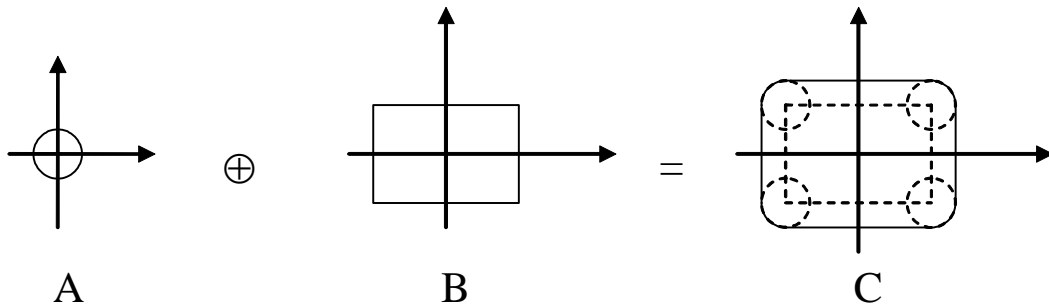


Figure 1: Minkowski summation of a ball ( $A$ ) and a square ( $B$ ).

- c) If  $y_1, y_2$  are random variables, then  $F \equiv [y_1, y_2]$  is an SVRV.  
d) If  $\mathbf{y}$  is a random vector, and  $x$  is a random variable, then the random ball  $B_x(\mathbf{y})$  is an SVRV.

The basic operations on sets that we use are as follows.

**Definition 2** Let  $A$  and  $B$  be two sets in  $\mathfrak{R}^d$ , let  $\mathbf{c}$  be a vector in  $\mathfrak{R}^d$  and let  $\lambda$  be a scalar in  $\mathfrak{R}$ .

Then,

- a)  $\mathbf{c} + A = \{\mathbf{r} \in \mathfrak{R}^d : \mathbf{r} = \mathbf{c} + \mathbf{a}, \mathbf{a} \in A\}$  (translation),  
b)  $\lambda A = \{\mathbf{r} \in \mathfrak{R}^d : \mathbf{r} = \lambda \mathbf{a}, \mathbf{a} \in A\}$  (dilation or homothety),  
c)  $A \oplus B = \{\mathbf{r} \in \mathfrak{R}^d : \mathbf{r} = \mathbf{a} + \mathbf{b}, \mathbf{a} \in A, \mathbf{b} \in B\}$  (Minkowski summation).

The Minkowski summation is a commutative and associative operation. If at least one of the summands is empty, the summation is empty. In general  $A \oplus A \supseteq 2A$ , with equality holding if  $A$  is a convex set. If  $A$  and  $B$  are convex sets, then  $A \oplus B$  is a convex set (for additional properties of the Minkowski summation, see Molchanov (2005), page 397). Notably, the Minkowski summation is not an invertible operation: given two sets  $A$  and  $B$ , it might be impossible to find a set  $C$  such that  $A \oplus C = B$ . For example, if  $A$  is a ball and  $B$  is a rectangle, there cannot exist a set  $C$  such that  $A \oplus C = B$ , because the Minkowski summation of  $A$  and  $C$  smooths out the vertices of  $C$  (see Figure 1). We return to this issue when discussing the limit theorems for SVRVs.

In what follows, we denote the Minkowski summation of a sequence of sets  $A_i$ ,  $i = 1, \dots, n$  by  $\bigoplus_{i=1}^n A_i$ , and reserve the notation  $\sum_{i=1}^n a_i$  ( or  $\sum_{i=1}^n \mathbf{a}_i$ ) to regular summations of scalars (or vectors). We refer to  $\frac{1}{n} \bigoplus_{i=1}^n A_i$  as the Minkowski average of the sets  $A_i$ ,  $i = 1, \dots, n$ .

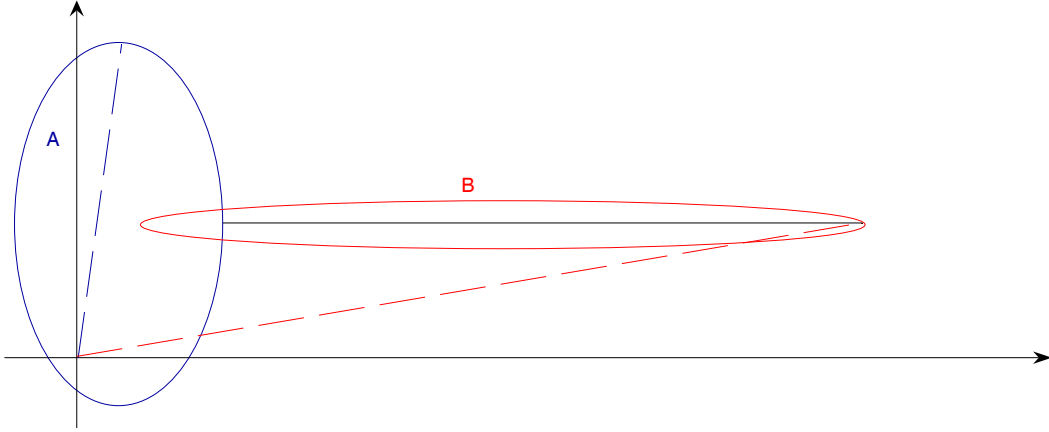


Figure 2: Hausdorff distance between the sets  $A$  and  $B$  (solid line), and Hausdorff norm of the set  $A$  and of the set  $B$  (dashed lines).

Other important concepts are the distance between a point and a set, the distance between two sets, and the norm of a set:

**Definition 3** Let  $A$  and  $B$  be two subsets of  $\mathfrak{R}^d$  and  $\mathbf{c}$  be a vector in  $\mathfrak{R}^d$ . Then we denote:

- a) The **distance** between a point and a set as  $d(\mathbf{c}, A) = \inf_{\mathbf{a} \in A} \|\mathbf{a} - \mathbf{c}\|$ ,
- b) The **Hausdorff distance** between two sets as

$$H(A, B) = \max \left\{ \sup_{\mathbf{a} \in A} \inf_{\mathbf{b} \in B} \|\mathbf{a} - \mathbf{b}\|, \sup_{\mathbf{b} \in B} \inf_{\mathbf{a} \in A} \|\mathbf{a} - \mathbf{b}\| \right\},$$

- c) The **Hausdorff norm** of a set as

$$\|A\|_H = H(A, \{0\}) = \sup \{\|\mathbf{a}\| : \mathbf{a} \in A\}.$$

Figure 2 illustrates the Hausdorff distance between two sets, and the Hausdorff norm of a set.

We notice that if  $A$  and  $B$  are unbounded subsets of  $\mathfrak{R}^d$ ,  $H(A, B)$  may be infinite. However, denoting by  $K_k(\mathfrak{R}^d)$  the set of non-empty compact subsets of  $\mathfrak{R}^d$ , the space  $(K_k(\mathfrak{R}^d), H(\cdot, \cdot))$  is a complete and separable metric space (Li et al. (2002), Theorems 1.1.2 and 1.1.3). Observing that

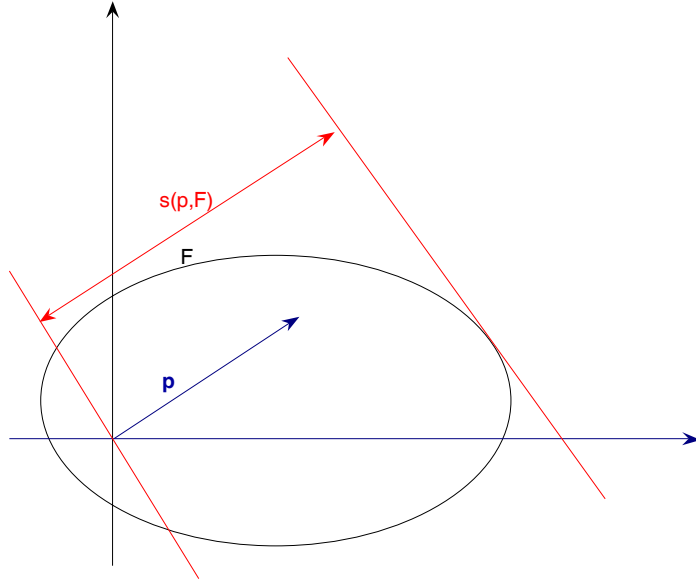


Figure 3: Support function of  $F$  at  $\mathbf{p}$ .

the set  $K_{kc}(\mathfrak{R}^d)$  of non-empty compact and convex subsets of  $\mathfrak{R}^d$  is a subset of  $K_k(\mathfrak{R}^d)$ , we obtain the same conclusion for  $(K_{kc}(\mathfrak{R}^d), H(\cdot, \cdot))$ .

### Support Function

Limit theorems for SVRVs are presented in Theorems 2.2 and 2.3 below. These theorems rely on representing sets as elements of a functional space. In this section we illustrate how this is accomplished by means of the support function of a set.

**Definition 4** Let  $F \in K(\mathfrak{R}^d)$ . Then the **support function** of  $F$  at  $\mathbf{p} \in \mathfrak{R}^d$ , denoted  $s(\mathbf{p}, F)$ , is given by

$$s(\mathbf{p}, F) = \sup_{\mathbf{f} \in F} \langle \mathbf{p}, \mathbf{f} \rangle.$$

In Definition 4,  $\langle \cdot, \cdot \rangle$  denotes the inner product in  $\mathfrak{R}^d$ . To gain insight on the support function, assume that  $\mathbf{p}$  is such that  $\|\mathbf{p}\| = 1$ . Then, Schneider (1993, page 37) provides the following useful intuition: “the support function  $s(\mathbf{p}, F)$  is the signed distance of the support plane to  $F$  with exterior normal vector  $\mathbf{p}$  from the origin; the distance is negative if and only if  $\mathbf{p}$  points into the open half space containing the origin” (see Figure 3).

The importance of the support function for our purposes can be more easily appreciated if we consider compact convex sets. A set in this class is given by the intersection of its supporting halfspaces: if  $F \in K_{kc}(\mathfrak{R}^d)$ , then (Rockafellar (1970), Chapter 13)

$$F = \bigcap_{\mathbf{p} \in \mathfrak{R}^d} \{\mathbf{f} : \langle \mathbf{p}, \mathbf{f} \rangle \leq s(\mathbf{p}, F)\}.$$

Hence such a set can be conveniently described by specifying the position of its support planes, given their exterior normal vectors. This is accomplished by the support function.

It is easy to show that the support function of a (not necessarily convex or bounded) set in  $K(\mathfrak{R}^d)$  is sublinear, and that for two SVRVs  $F_1, F_2 \in K(\mathfrak{R}^d)$  it satisfies<sup>2</sup>

$$(2.1) \quad s(\mathbf{p}, F_1 \oplus F_2) = s(\mathbf{p}, F_1) + s(\mathbf{p}, F_2),$$

$$(2.2) \quad s(\mathbf{p}, \lambda F) = \lambda s(\mathbf{p}, F), \quad \lambda \in \mathfrak{R}_+.$$

Equations (2.1)-(2.2) allow one to convert the Minkowski average of a sequence of sets  $F_i$ ,  $i = 1, \dots, n$  into the average of the corresponding support functions. In particular,

$$s\left(\mathbf{p}, \frac{1}{n} \bigoplus_{i=1}^n F_i\right) = \frac{1}{n} \sum_{i=1}^n s(\mathbf{p}, F_i).$$

The support function is instrumental in understanding how the solution to a problem of maximizing a linear objective function under convex constraints changes as the coefficients of the linear objective function change (see Rockafellar (1970), Chapter 13). As such, the support function is an important tool in economic analysis.<sup>3</sup>

For our purposes, the support function provides the first step towards establishing a connection between SVRVs and elements of a functional space. In particular, an SVRV  $F : \Omega \rightarrow K_{kc}(\mathfrak{R}^d)$  can be represented through the support function of its realizations; this (as discussed in the next section) results in a random variable whose realizations are continuous functions. Similarly, a Minkowski average of SVRVs  $F_i : \Omega \rightarrow K_{kc}(\mathfrak{R}^d)$ ,  $i = 1, \dots, n$ , can be represented through the sample average of the corresponding support functions  $s(\mathbf{p}, F_i)$ ,  $i = 1, \dots, n$ .

---

<sup>2</sup>Additionally, one can show that the support function of a bounded set  $F \in K(\mathfrak{R}^d)$  is Lipschitz with Lipschitz constant  $\|F\|_H$ , that is

$$|s(\mathbf{p}, F) - s(\mathbf{q}, F)| \leq \|F\|_H \|\mathbf{p} - \mathbf{q}\|, \quad \mathbf{p}, \mathbf{q} \in \mathfrak{R}^d$$

(Molchanov (2005), Theorem F.1).

<sup>3</sup>Indirect utility functions, expenditure functions, cost functions and profit functions are examples of support functions. See Mas-Colell, Whinston & Green (1995) (especially Chapter 3) for a discussion of support functions in the context of Duality Theory and Demand Analysis.

In the next section we discuss how to represent the Hausdorff distance between SVRVs as a transformation of the difference between elements of a functional space. This is accomplished through Hörmander's Embedding Theorem, which establishes the relation between the set of continuous functions on the unit sphere and the set of convex, compact non-empty sets in  $\mathfrak{R}^d$ .

### Hörmander's Embedding Theorem and Shapley-Folkman's Theorem

Define  $\mathbb{S} = \{\mathbf{p} \in \mathfrak{R}^d : \|\mathbf{p}\| = 1\}$  and  $\mathbb{B} = \{\mathbf{p} \in \mathfrak{R}^d : \|\mathbf{p}\| \leq 1\}$  to be, respectively, the unit sphere and the unit ball in  $\mathfrak{R}^d$ . Equation (2.2) implies that when considering the support function of a set, it suffices to restrict attention to vectors  $\mathbf{p} \in \mathbb{S}$ . Let  $\mathbb{C}(\mathbb{S})$  be the set of continuous functions from  $\mathbb{S}$  to  $\mathfrak{R}$ , and let  $\mathbb{C}(\mathbb{B})$  be the set of continuous functions from  $\mathbb{B}$  to  $\mathfrak{R}$ . For  $h \in \mathbb{C}(\mathbb{S})$ , let  $\|h\|_{\mathbb{C}(\mathbb{S})} = \sup_{\mathbf{p} \in \mathbb{S}} |h(\mathbf{p})|$  be the  $\mathbb{C}(\mathbb{S})$ -norm, and for  $g \in \mathbb{C}(\mathbb{B})$ , let  $\|g\|_{\mathbb{C}(\mathbb{B})} = \sup_{\mathbf{p} \in \mathbb{B}} |g(\mathbf{p})|$  be the  $\mathbb{C}(\mathbb{B})$ -norm. Hörmander's Embedding Theorem (Li et al. (2002), Theorem 1.1.12) ensures that  $(K_{kc}(\mathfrak{R}^d), H(\cdot, \cdot))$  can be isometrically embedded into a closed convex cone in  $\mathbb{C}(\mathbb{S})$  (or in  $\mathbb{C}(\mathbb{B})$ ). In particular, whenever  $F \in K_{kc}(\mathfrak{R}^d)$ , its support function  $s(\mathbf{p}, F)$  is a continuous function of  $\mathbf{p}$  and belongs to  $\mathbb{C}(\mathbb{S})$  (and to  $\mathbb{C}(\mathbb{B})$ ). For any  $F, F_1, F_2 \in K_{kc}(\mathfrak{R}^d)$  the map  $j : F \rightarrow s(\cdot, F)$  has the properties given in equations (2.1)-(2.2), and additionally

$$(2.3) \quad H(F_1, F_2) = \|s(\cdot, F_1) - s(\cdot, F_2)\|_{\mathbb{C}(\mathbb{S})} = \|s(\cdot, F_1) - s(\cdot, F_2)\|_{\mathbb{C}(\mathbb{B})}$$

A simple proof (based on the work of Hörmander (1954)) of the result in (2.3) is given in Giné, Hahn & Zinn (1983), Lemma 1.1. Hence the mapping  $j$  is an isometry and it preserves the operations of closed addition and multiplication by non-negative scalars. Therefore  $j(K_{kc}(\mathfrak{R}^d))$  is a convex cone in  $\mathbb{C}(\mathbb{S})$  (and in  $\mathbb{C}(\mathbb{B})$ ) and it is closed since  $(K_{kc}(\mathfrak{R}^d), H(\cdot, \cdot))$  is closed.

If  $F : \Omega \rightarrow K_{kc}(\mathfrak{R}^d)$  is an SVRV, then  $s(\cdot, F(\cdot))$  is a  $\mathbb{C}(\mathbb{S})$ -valued (as well as a  $\mathbb{C}(\mathbb{B})$ -valued) random variable (measurability follows from Lemma 5.1 in Artstein (1972); continuity follows from Hörmander's Embedding Theorem). Formally, a  $\mathbb{C}(\mathbb{S})$ -valued random system is defined as follows:

**Definition 5** A mapping  $T : \mathbb{S} \times \Omega \rightarrow \mathfrak{R}$  is called a  $\mathbb{C}(\mathbb{S})$ -valued random system if

- a)  $T_{\cdot, \omega} \equiv T(\cdot, \omega)$  is a function in  $\mathbb{C}(\mathbb{S})$  for every  $\omega \in \Omega$ ,
- b)  $T_{\mathbf{p}, \cdot} \equiv T(\mathbf{p}, \cdot)$  is a random variable for every  $\mathbf{p} \in \mathbb{S}$ .

The  $\mathbb{C}(\mathbb{S})$ -norm of a random system  $\|T\|_{\mathbb{C}(\mathbb{S})} = \left\{ \|T_{\cdot, \omega}\|_{\mathbb{C}(\mathbb{S})} \right\}_{\omega \in \Omega}$  is a mapping from  $\Omega$  to  $\mathfrak{R}$  defined by  $\omega \mapsto \|T_{\cdot, \omega}\|_{\mathbb{C}(\mathbb{S})}$ . The **expectation function** of a  $\mathbb{C}(\mathbb{S})$ -valued random system is  $\tau : \mathbb{S} \rightarrow \mathfrak{R}$  such

that  $\tau(\mathbf{p}) = E(T_{\mathbf{p},\cdot})$ . The **covariance kernel** of a  $\mathbb{C}(\mathbb{S})$ -valued random system is  $\gamma : \mathbb{S} \times \mathbb{S} \rightarrow \mathfrak{R}$  such that  $\gamma(\mathbf{p}, \mathbf{q}) = \text{Cov}(T_{\mathbf{p},\cdot}, T_{\mathbf{q},\cdot})$  for  $(\mathbf{p}, \mathbf{q}) \in \mathbb{S} \times \mathbb{S}$ .

A special type of  $\mathbb{C}(\mathbb{S})$ -valued random system is a  $\mathbb{C}(\mathbb{S})$ -valued Gaussian system where  $T_{\mathbf{p},\cdot}$  is a Gaussian random variable. The limit theorems for SVRVs presented below involve Gaussian processes on the unit sphere.

For the analysis in this paper, the importance of Hörmander's Embedding Theorem lies in the fact that Laws of Large Numbers and Central Limit Theorems are available for  $\mathbb{C}(\mathbb{S})$ -valued random variables. Hence these limit theorems can be translated to limit theorems for the Hausdorff distance between compact convex SVRVs. The requirement in Hörmander's Embedding Theorem that the SVRVs need to be convex valued is actually not restrictive, as Theorem 2.1 below shows. Let  $\text{co}A$  denote the convex hull of a set  $A$ .

**Theorem 2.1 (Shapley-Folkman)** *Let  $A_i \in K_k(\mathfrak{R}^d)$ ,  $i = 1, \dots, n$  such that  $\|A_i\|_H \leq c$ . Then*

$$H\left(\bigoplus_{i=1}^n A_i, \bigoplus_{i=1}^n \text{co}A_i\right) \leq c\sqrt{d}.$$

**Proof.** Starr (1969). ■

This result is crucial because it shows that Minkowski averaging SVRVs which are compact valued but not necessarily convex is asymptotically “convexifying,” as noted by Artstein & Vitale (1975). In particular, the Hausdorff distance between the Minkowski average of a sequence of SVRVs  $\{F_i\}_{i=1}^n$  and the Minkowski average of their convex hulls goes to zero at the rate  $\frac{1}{n}$ .

### Expectation of an SVRV and Independently and Identically Distributed (*i.i.d.*) SVRVs

In order to introduce a Strong Law of Large Numbers and a Central Limit Theorem for SVRVs, we need to define the notion of expectation of SVRVs. We start by defining a set of functions that are used in the definition of the expected value of an SVRV. Let  $(\Omega, \mathcal{A}, \mu)$  be a probability space.

**Definition 6** *For any SVRV  $F$ , a (measurable) **selection** of  $F$  is a random vector  $\mathbf{f}$  (taking values in  $\mathfrak{R}^d$ ) such that  $\mathbf{f}(\omega) \in F(\omega)$   $\mu$ -a.s. We denote by  $\mathcal{S}(F)$  the set of all selections from  $F$ .*

The concept of selection of an SVRV goes back to Aumann's (1965) idea of thinking of SVRVs (or random correspondences in Aumann's work) as bundles of random variables – the selections of the SVRVs. We observe that if  $F$  is a measurable closed-valued almost surely non-empty SVRV in

$K(\mathfrak{R}^d)$ ,  $\mathcal{S}(F)$  is non-empty (Aumann (1965); see also Li et al. (2002), Theorem 1.2.6). Denoting by  $\mathbf{L}^1 = \mathbf{L}^1(\Omega, \mathfrak{A})$  the space of  $\mathfrak{A}$ -measurable random variables with values in  $\mathfrak{R}^d$  such that the  $\mathbf{L}^1$ -norm

$$\|\boldsymbol{\xi}\|_1 = E(\|\boldsymbol{\xi}\|)$$

is finite, we obtain that for an SVRV  $F$  defined on  $(\Omega, \mathfrak{A})$  the family of all integrable selections of  $F$  is given by

$$\mathcal{S}^1(F) = \mathcal{S}(F) \cap \mathbf{L}^1.$$

We now turn to defining integrably bounded SVRVs:

**Definition 7** *A SVRV  $F : \Omega \rightarrow K(\mathfrak{R}^d)$  is called **integrably bounded** if  $\|F\|_H = \sup\{\|\mathbf{f}\| : \mathbf{f} \in F\}$  has a finite expectation.*

Observe that for any  $\mathbf{f} \in \mathcal{S}(F)$ ,  $\|\mathbf{f}\| \leq \|F\|_H$ . Hence, all selections of an integrably bounded SVRV are integrable (and  $\mathcal{S}^1(F) = \mathcal{S}(F)$ ). Moreover, by Hörmander's Embedding Theorem  $\|s(\cdot, F)\|_{\mathbb{C}(\mathbb{S})} = \|F\|_H$ . Hence, if  $F$  is integrably bounded, then  $E[|s(\mathbf{p}, F)|]$  is finite for all  $\mathbf{p} \in \mathbb{S}$ . To build intuition on the connection between the upcoming limit theorems for SVRVs and the limit theorems for random variables, notice that the concept of an integrably bounded SVRV corresponds to the concept of an absolutely integrable random variable.

We are now ready to introduce the expectation of an SVRV  $F$ , which is based on the Aumann integral and is referred to as Aumann expectation (Aumann (1965)). We denote this expectation by  $\mathbb{E}[F]$ , and reserve the notation  $E[\cdot]$  for the expectation of random variables and random vectors.

**Definition 8** *The Aumann expectation of an SVRV  $F$  is given by*

$$\mathbb{E}[F] = \left\{ \int_{\Omega} \mathbf{f} \, d\mu : \mathbf{f} \in \mathcal{S}^1(F) \right\}$$

where  $\int_{\Omega} \mathbf{f} \, d\mu$  is taken coordinate wise. If  $F$  is integrably bounded, then

$$\mathbb{E}[F] = \left\{ \int_{\Omega} \mathbf{f} \, d\mu : \mathbf{f} \in \mathcal{S}(F) \right\}$$

Clearly, since  $\mathcal{S}(F)$  is non-empty, the Aumann expectation of an integrably bounded SVRV is non-empty. Moreover, if  $F$  is an integrably bounded random compact set on a non-atomic probability space, or if  $F$  is an integrably bounded random convex compact set, then  $\mathbb{E}[F]$  is convex and coincides with  $\mathbb{E}[coF]$ , and  $E[s(\mathbf{p}, F)] = s(\mathbf{p}, \mathbb{E}[F])$  (Artstein (1974)).

Finally we need to introduce the concept of independently and identically distributed SVRVs. Let  $\mathcal{B}(K(\mathfrak{R}^d))$  be the Borel field of  $K(\mathfrak{R}^d)$  with respect to the Hausdorff metric  $H$ . Then it follows from Definition 1 that an SVRV  $F : \Omega \rightarrow K(\mathfrak{R}^d)$  is  $\mathcal{B}(K(\mathfrak{R}^d))$ -measurable (this result is stated in Molchanov (2005), Theorem 1.2.3; for a proof see Castaing & Valadier (1977) or Li et al. (2002), Theorem 1.2.3). Define

$$\mathcal{A}_F = \sigma \left\{ F^{-1}(A) : A \in \mathcal{B}(K(\mathfrak{R}^d)) \right\}$$

to be the  $\sigma$ -algebra generated by inverse images of sets in  $\mathcal{B}(K(\mathfrak{R}^d))$ . Also let  $\mu_F(A) = \mu(F^{-1}(A))$  for  $A \in \mathcal{B}(K(\mathfrak{R}^d))$  denote the distribution of  $F$ .

**Definition 9** Let  $F_1$  and  $F_2$  be two  $\mathcal{B}(K(\mathfrak{R}^d))$ -measurable SVRVs defined on the same measurable space  $\Omega$ .  $F_1$  and  $F_2$  are independent if  $\mathcal{A}_{F_1}$  and  $\mathcal{A}_{F_2}$  are independent,  $F_1$  and  $F_2$  are identically distributed if  $\mu_{F_1}$  and  $\mu_{F_2}$  are identical.

## Limit Theorems

We are now ready to state the Strong Law of Large Numbers (SLLN) and the Central Limit Theorem (CLT) for SVRVs. Similarly to the familiar results for vector valued random variables, these limit theorems relate the Minkowski average of an *i.i.d.* sample of SVRVs to their Aumann expectation.<sup>4</sup> Observe that having a sequence of independent random elements ensures that the relevant (product) probability space is nonatomic, so that the Aumann expectation is a convex set.

**Theorem 2.2 (SLLN)** Let  $\{F, F_i : i \in \mathbb{N}\}$  be *i.i.d.* non-empty, integrably bounded, compact valued SVRVs in  $K_k(\mathfrak{R}^d)$ . Then

$$\lim_{n \rightarrow \infty} H \left( \frac{1}{n} \bigoplus_{i=1}^n F_i, \mathbb{E}[F] \right) = 0 \quad \mu - a.s.$$

**Proof.** Artstein & Vitale (1975). ■

The proof in Artstein & Vitale (1975) uses the triangle inequality to show that

$$(2.4) \quad H \left( \frac{1}{n} \bigoplus_{i=1}^n F_i, \mathbb{E}[F] \right) \leq H \left( \frac{1}{n} \bigoplus_{i=1}^n F_i, \frac{1}{n} \bigoplus_{i=1}^n coF_i \right) + H \left( \frac{1}{n} \bigoplus_{i=1}^n coF_i, \mathbb{E}[F] \right)$$

---

<sup>4</sup>There are also Laws of Large Numbers for independent but not identically distributed SVRVs, and for set valued martingales. See Molchanov (2005).



An application of the Shapley-Folkman's Theorem then implies that the first term on the right hand side of (2.4) converges to zero *a.s.* as  $n \rightarrow \infty$ . The second term converges to zero *a.s.* by Hörmander's Embedding Theorem and because

$$s \left( \mathbf{p}, \frac{1}{n} \bigoplus_{i=1}^n coF_i \right) = \frac{1}{n} \sum_{i=1}^n s(\mathbf{p}, coF_i) \xrightarrow{a.s.} E[s(\mathbf{p}, coF)] = s(\mathbf{p}, \mathbb{E}[coF]) = s(\mathbf{p}, \mathbb{E}[F])$$

by the Strong Law of Large Numbers in Banach spaces (e.g. Mourier (1955)).

Theorem 2.2 states that the Minkowski average of an *i.i.d.* sample of SVRVs converges in the Hausdorff distance to their Aumann expectation as  $n \rightarrow \infty$ . We are now left to discuss the rate at which this convergence occurs, as well as to find the limiting distribution. When working with random variables this is achieved by taking the difference between the sample average of the variables and their expectation, and normalizing it with a growing sequence. But as discussed in Molchanov (2005, pages 213-214), the Minkowski addition is not an invertible operation. One cannot get around this problem by considering SVRVs with Aumann expectation equal to the singleton zero, because all integrable random compact sets with Aumann expectation equal to zero are singletons (Molchanov (2005), Proposition 2.1.30). However, one can avoid having to deal with subtraction by considering the normalized Hausdorff distance  $\sqrt{n}H \left( \frac{1}{n} \bigoplus_{i=1}^n F_i, \mathbb{E}[F] \right)$ . The limit of this sequence is a random variable equal to the supremum of a Gaussian random function on  $\mathbb{S}$ . Theorem 2.3 presents this result.

**Theorem 2.3 (CLT)** *Let  $\{F, F_i : i \in \mathbb{N}\}$  be i.i.d. non-empty, compact valued SVRVs such that  $E \left[ \|F\|_H^2 \right] < \infty$ . Then,*

$$\sqrt{n}H \left( \frac{1}{n} \bigoplus_{i=1}^n F_i, \mathbb{E}[F] \right) \xrightarrow{d} \|z\|_{\mathbb{C}(\mathbb{S})}$$

where  $z$  is a Gaussian random system with (a)  $\tau(\mathbf{p}) = 0$  for all  $\mathbf{p} \in \mathbb{S}$ , and (b)

$$\gamma(\mathbf{p}, \mathbf{q}) = E[s(\mathbf{p}, F) s(\mathbf{q}, F)] - E[s(\mathbf{p}, F)] E[s(\mathbf{q}, F)].$$

**Proof.** Weil (1982) and Giné et al. (1983). ■

The (independently developed) proofs in Weil (1982) and in Giné et al. (1983) use the triangle inequality and Shapley-Folkman's Theorem to show that

$$\sqrt{n} \left| H \left( \frac{1}{n} \bigoplus_{i=1}^n F_i, \mathbb{E}[F] \right) - H \left( \frac{1}{n} \bigoplus_{i=1}^n coF_i, \mathbb{E}[F] \right) \right| \leq \sqrt{n}H \left( \frac{1}{n} \bigoplus_{i=1}^n F_i, \frac{1}{n} \bigoplus_{i=1}^n coF_i \right) \xrightarrow{p} 0.$$

Hörmander's Embedding Theorem implies that

$$\sqrt{n}H\left(\frac{1}{n}\bigoplus_{i=1}^n coF_i, \mathbb{E}[F]\right) = \left\| \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n s(\cdot, coF_i) - E[s(\cdot, F)]\right) \right\|_{\mathbb{C}(\mathbb{S})}$$

and the Central Limit Theorem for  $\mathbb{C}(\mathbb{S})$ -valued random variables of Jain & Marcus (1975) implies that

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n s(\mathbf{p}, coF_i) - E[s(\mathbf{p}, F)]\right) \xrightarrow{d} z(\mathbf{p}),$$

where  $z(\mathbf{p})$  satisfies (a) and (b) in Theorem 2.3. Hence, the result follows by the continuous mapping theorem and because

$$E[s(\mathbf{p}, F)s(\mathbf{q}, F)] - E[s(\mathbf{p}, F)]E[s(\mathbf{q}, F)] = E[s(\mathbf{p}, coF)s(\mathbf{q}, coF)] - E[s(\mathbf{p}, coF)]E[s(\mathbf{q}, coF)].$$

### 3 Testing and Confidence Collections

#### Hypothesis Testing

Given a non-empty, compact and convex (non-random) set  $\Psi_0$ , in order to test the null hypothesis

$$\mathfrak{H}_0 : \mathbb{E}[F] = \Psi_0$$

against the alternative

$$\mathfrak{H}_A : \mathbb{E}[F] \neq \Psi_0$$

at a prespecified significance level  $\alpha \in (0, 1)$ , we propose the following criterion:

$$\text{Reject } \mathfrak{H}_0 \text{ if } \sqrt{n}H\left(\frac{1}{n}\bigoplus_{i=1}^n coF_i, \Psi_0\right) > c_\alpha,$$

$$\text{Do not Reject } \mathfrak{H}_0 \text{ if } \sqrt{n}H\left(\frac{1}{n}\bigoplus_{i=1}^n coF_i, \Psi_0\right) \leq c_\alpha,$$

where  $c_\alpha$  is chosen so that

$$(3.1) \quad \Pr\left\{\|z\|_{\mathbb{C}(\mathbb{S})} > c_\alpha\right\} = \alpha.$$

Observe that the test statistic uses the Hausdorff distance between  $\Psi_0$  and  $\frac{1}{n}\bigoplus_{i=1}^n coF_i$ , rather than the Hausdorff distance between  $\Psi_0$  and  $\frac{1}{n}\bigoplus_{i=1}^n F_i$ . Using  $\bigoplus_{i=1}^n coF_i$  rather than  $\bigoplus_{i=1}^n F_i$  to construct the test statistic improves the small sample properties of the test without changing its asymptotic behavior. By Theorem 2.2,  $\frac{1}{n}\bigoplus_{i=1}^n coF_i$  is a consistent estimator of  $\mathbb{E}[F]$ , and by

Theorem 2.3,  $\sqrt{n}H\left(\frac{1}{n}\bigoplus_{i=1}^n \text{co}F_i, \mathbb{E}[F]\right)$  and  $\sqrt{n}H\left(\frac{1}{n}\bigoplus_{i=1}^n F_i, \mathbb{E}[F]\right)$  have the same asymptotic distribution. In a finite sample, when the sets  $F_i$ ,  $i = 1, \dots, n$ , are not convex, their Minkowski average needs not be a convex set, although of course it converges to a convex set as the sample size increases. In a finite sample this can lead to rejecting a null hypothesis which is true, only because a non-convex Minkowski average is compared to a convex set  $\Psi_0$ . This can result in an oversized test. Our choice for the test statistic avoids this problem.

Since the limiting distribution of our test statistics depends on parameters to be estimated, we obtain the critical values using the following bootstrap procedure:

**Algorithm 3.1**

- 1) Generate a bootstrap sample of size  $n$ ,  $\{F_i^* : i = 1, \dots, n\}$ , such that  $F_i^*$ ,  $i = 1, \dots, n$ , are i.i.d. and  $\mu_*(F_1^* = F_j) = \frac{1}{n}$  for each  $j = 1, \dots, n$ , where  $\mu_*$  denotes the bootstrap probability law.
- 2) Compute

$$(3.2) \quad r_n^*(\mathbf{p}) \equiv \sqrt{n}H\left(\frac{1}{n}\bigoplus_{i=1}^n \text{co}F_i^*, \frac{1}{n}\bigoplus_{i=1}^n \text{co}F_i\right).$$

- 3) Use the results of  $b$  repetitions of Steps 1-2 to compute the empirical distribution of  $r_n^*$  at a point  $t$ , denoted by  $J_n(t)$ .
- 4) Estimate the critical value  $c_\alpha$  defined in equation (3.1) by

$$(3.3) \quad \hat{c}_{\alpha n} = \inf \{t : J_n(t) \geq 1 - \alpha\}.$$

The results of Giné & Zinn (1990), along with an application of the continuous mapping theorem, guarantee the validity of this bootstrap procedure. In particular, the following result holds:

**Proposition 3.1** *Let the assumptions of Theorem 2.3 hold. Then  $r_n^* \xrightarrow{d} \|z\|_{\mathbb{C}(\mathbb{S})}$ , where  $r_n^*$  is defined in equation (3.2) and  $z$  is a Gaussian random system with  $\tau(\mathbf{p}) = 0$  for all  $\mathbf{p} \in \mathbb{S}$ , and with covariance kernel  $\gamma(\mathbf{p}, \mathbf{q}) = E[s(\mathbf{p}, F)s(\mathbf{q}, F)] - E[s(\mathbf{p}, F)]E[s(\mathbf{q}, F)]$ . If in addition  $\text{Var}(z(\mathbf{p})) > 0$  for each  $\mathbf{p} \in \mathbb{S}$ , then*

$$(3.4) \quad \hat{c}_{\alpha n} = c_\alpha + o_p(1),$$

where  $\hat{c}_{\alpha n}$  is defined in (3.3).

**Proof.** See Appendix. ■

Alternatively we could obtain a consistent estimator of the critical value  $c_\alpha$  using the following procedure. Simulate the distribution of the supremum of a Gaussian random system with mean function equal to zero for each  $\mathbf{p} \in \mathbb{S}$  and covariance kernel equal to

$$\hat{\gamma}(\mathbf{p}, \mathbf{q}) = \frac{1}{n} \sum_{i=1}^n s(\mathbf{p}, F_i) s(\mathbf{q}, F_i) - \frac{1}{n} \sum_{i=1}^n s(\mathbf{p}, F_i) \frac{1}{n} \sum_{i=1}^n s(\mathbf{q}, F_i).$$

Observe that the supremum is to be taken over  $\mathbf{p} \in \mathbb{S}$ , hence over a known set (the unit sphere) which does not need to be estimated. It then follows from the Strong Law of Large Numbers in Banach spaces of Mourier (1955) that

$$\hat{\gamma}(\mathbf{p}, \mathbf{q}) \xrightarrow{a.s.} E[s(\mathbf{p}, F) s(\mathbf{q}, F)] - E[s(\mathbf{p}, F)] E[s(\mathbf{q}, F)].$$

Therefore, by the same argument as in the proof of Proposition 3.1, if  $Var(z(\mathbf{p})) > 0$  for each  $\mathbf{p} \in \mathbb{S}$ , the critical values of the simulated distribution consistently estimate the critical values of  $\|z\|_{\mathcal{C}(\mathbb{S})}$ .

We now show that our test is consistent against any fixed alternative hypothesis in  $\mathfrak{H}_A$ .

**Theorem 3.2** *Let  $\{F, F_i : i \in \mathbb{N}\}$  be i.i.d. non-empty, compact valued SVRVs such that  $E[\|F\|_H^2] < \infty$ , and let  $\Psi_A$  be a non-empty, compact and convex valued set such that  $\mathbb{E}[F] = \Psi_A \neq \Psi_0$ . Let  $Var(z(\mathbf{p})) > 0$  for each  $\mathbf{p} \in \mathbb{S}$ , where  $z(\mathbf{p})$  is defined in Theorem 2.3. Then*

$$\lim_{n \rightarrow \infty} \Pr \left\{ \sqrt{n}H \left( \frac{1}{n} \bigoplus_{i=1}^n coF_i, \Psi_0 \right) > \hat{c}_{\alpha n} \right\} = 1 \quad \mu - a.s.$$

**Proof.** See Appendix. ■

We conclude the discussion of our test by determining its power against local alternatives at distance (proportional to)  $1/\sqrt{n}$  from the null hypothesis. Suppose we are interested in the power of our test of  $\mathfrak{H}_0$  against a sequence of non-empty, compact and convex alternative sets  $\{\Psi_{An}\}$  such that

$$(3.5) \quad \Psi_{An} \in K_{kc}(\mathfrak{R}^d) : \Psi_{An} \oplus \frac{1}{\sqrt{n}}\Delta_1 = \Psi_0 \oplus \frac{1}{\sqrt{n}}\Delta_2,$$

where  $\Delta_1$  and  $\Delta_2$  are (non-random) non-empty, compact and convex sets and

$$(3.6) \quad \begin{aligned} \kappa &= H(\Delta_2, \Delta_1) = \sqrt{n}H \left( \Psi_0 \oplus \frac{1}{\sqrt{n}}\Delta_2, \Psi_0 \oplus \frac{1}{\sqrt{n}}\Delta_1 \right) \\ &= \sqrt{n}H \left( \Psi_{An} \oplus \frac{1}{\sqrt{n}}\Delta_1, \Psi_0 \oplus \frac{1}{\sqrt{n}}\Delta_1 \right) = \sqrt{n}H(\Psi_{An}, \Psi_0) < \infty, \end{aligned}$$

with the second equality following from the properties of the Hausdorff distance between two sets (see, e.g., DeBlasi & Iervolino (1969)). As  $\kappa$  increases, the local alternatives get farther away from the null, with a resulting increase in the power of the test.

This choice of local alternatives encompasses the case in which  $\Delta_1 = \{0\}$ ,  $\Delta_2 \neq \{0\}$  and the sets  $\Psi_{A_n}$  shrink and/or shift to become equal to  $\Psi_0$ , and the case in which  $\Delta_1 \neq \{0\}$ ,  $\Delta_2 = \{0\}$  and the sets  $\Psi_{A_n}$  enlarge and/or shift to become equal to  $\Psi_0$ . This allows us to consider a large class of deviations from the null hypothesis. In particular, our results apply both to the case in which the null is a subset of the true identification region, and to the case that it is a superset of it. The asymptotic distribution of our test under these local alternatives is given by the following theorem.

**Theorem 3.3** *Let  $\{F, F_i : i \in \mathbb{N}\}$  be i.i.d. non-empty, compact valued SVRVs such that  $E[\|F\|_H^2] < \infty$ , let  $\Psi_0$  be a non-empty, compact and convex valued set, and let  $\{\Psi_{A_n}\}$  be the sequence of sets defined in (3.5). Then*

$$\sqrt{n}H\left(\frac{1}{n}\bigoplus_{i=1}^n coF_i, \Psi_0\right) \xrightarrow{d} \|w\|_{\mathbb{C}(\mathbb{S})}$$

under  $\Psi_{A_n}$ , where  $w$  is a Gaussian random system with

- a)  $\tau(\mathbf{p}) = s(\mathbf{p}, \Delta_2) - s(\mathbf{p}, \Delta_1)$  for all  $\mathbf{p} \in \mathbb{S}$ , and
- b)  $\gamma(\mathbf{p}, \mathbf{q}) = E[s(\mathbf{p}, F)s(\mathbf{q}, F)] - E[s(\mathbf{p}, F)]E[s(\mathbf{q}, F)]$ .

**Proof.** See Appendix. ■

It then follows from Anderson’s Lemma (Anderson (1955)) and the discussion in Section 6 of Andrews (1997) that our test is asymptotically locally unbiased, i.e.

$$\lim_{n \rightarrow \infty} \Pr\left\{\sqrt{n}H\left(\frac{1}{n}\bigoplus_{i=1}^n coF_i, \Psi_0\right) > \hat{c}_{\alpha n}\right\} = \Pr\left\{\|w\|_{\mathbb{C}(\mathbb{S})} > c_\alpha\right\} \geq \Pr\left\{\|z\|_{\mathbb{C}(\mathbb{S})} > c_\alpha\right\} = \alpha,$$

where we have assumed that the law of  $\|z\|_{\mathbb{C}(\mathbb{S})}$  is continuous and strictly increasing at  $c_\alpha$ .

### Confidence Collections

We now propose a procedure to construct what we call a “confidence collection,” that is the collection of all sets that, when specified as null hypothesis for the true value of the population identification region, cannot be rejected by our test. We denote this collection by  $CC^{\mathbb{E}[F]}$ . The confidence collection is based on exploiting the duality between confidence regions and hypothesis

tests that is of familiar use for point identified models, and it has the property that asymptotically the set  $\mathbb{E}[F]$  is a member of such collection with a prespecified confidence level  $1 - \alpha$ .

Inverting the test statistic described above, we construct  $CC^{\mathbb{E}[F]}$  as

$$CC^{\mathbb{E}[F]} = \left\{ \tilde{\Psi} \in K_{kc}(\mathfrak{R}^d) : \sqrt{n}H \left( \frac{1}{n} \bigoplus_{i=1}^n coF_i, \tilde{\Psi} \right) \leq \hat{c}_{\alpha n} \right\},$$

where  $\hat{c}_{\alpha n}$  is defined in equation (3.3). Hence if the law of  $\|z\|_{\mathbb{C}(\mathfrak{S})}$  is continuous

$$\lim_{n \rightarrow \infty} \Pr \left\{ \mathbb{E}[F] \in CC^{\mathbb{E}[F]} \right\} = 1 - \alpha.$$

In practice, it can be computationally challenging to calculate all the sets that belong to the confidence collection. However, the union of all sets in  $CC^{\mathbb{E}[F]}$  can be calculated in a particularly simple way. The following Theorem shows how.

**Theorem 3.4** *Let  $\mathcal{U}_n = \bigcup \left\{ \tilde{\Psi} : \tilde{\Psi} \in CC^{\mathbb{E}[F]} \right\}$ ,  $B_{\hat{c}_{\alpha n}} = \left\{ \mathbf{b} \in \mathfrak{R}^d : \|\mathbf{b}\| \leq \frac{\hat{c}_{\alpha n}}{\sqrt{n}} \right\}$ ,  $\mathfrak{S}_n = \frac{1}{n} \bigoplus_{i=1}^n coF_i \oplus B_{\hat{c}_{\alpha n}}$ . Then  $\mathcal{U}_n = \mathfrak{S}_n$ .*

**Proof.** See Appendix. ■

This result implies that one can check whether a specific vector, say  $\psi$  is a plausible value for the parameter of interest by checking whether it is included in  $\mathfrak{S}_n$ . If  $\psi \in \mathfrak{S}_n$ , it means that there exists a set  $\tilde{\Psi} \in CC^{\mathbb{E}[F]}$  such that  $\psi \in \tilde{\Psi}$ . An interesting consequence of Theorem 3.4 is that the union of all sets in  $CC^{\mathbb{E}[F]}$  is also included in  $CC^{\mathbb{E}[F]}$ , and thus represents the biggest set that cannot be rejected as a null hypothesis.

## 4 Population Mean with Interval Data

Suppose that one is interested in the population mean of a random variable  $y$ ,  $E(y)$ . Suppose further that one does not observe the realizations of  $y$ , but rather the realizations of two real valued random variables  $y_L, y_U$  such that  $\Pr \{y_L \leq y \leq y_U\} = 1$ . Let

$$(4.1) \quad Y = [y_L, y_U].$$

Assume that  $\{(y_{iL}, y_{iU})\}_{i=1}^n$  are *i.i.d* random vectors. Let  $Y_i$  be the mapping defined as in (4.1) using  $(y_{iL}, y_{iU})$ . Then we have the following result:

**Theorem 4.1** Let  $\{(y_{iL}, y_{iU})\}_{i=1}^n$  be i.i.d real valued random vectors such that  $\Pr\{y_{iL} \leq y_i \leq y_{iU}\} = 1$ . Then  $\{Y, Y_i : i \in \mathbb{N}\}$  are i.i.d. SVRVs. If in addition  $E(|y_L|) < \infty$ ,  $E(|y_U|) < \infty$ , then

$$\lim_{n \rightarrow \infty} H \left( \frac{1}{n} \bigoplus_{i=1}^n Y_i, \mathbb{E}[Y] \right) = 0 \quad \mu - a.s.$$

**Proof.** See Appendix. ■

Manski (1989) and Manski & Tamer (2002) show that  $[E(y_L), E(y_U)]$  is the sharp bound for  $E[y]$ . The following theorem shows that the Aumann expectation of the SVRV  $Y$  defined in equation (4.1) is the sharp bound.

**Theorem 4.2**  $\mathbb{E}[Y] = [E(y_L), E(y_U)]$ .

**Proof.** See Appendix. ■

We are now ready to state the CLT result for this problem. Since in this application we operate in  $\mathfrak{R}^1$ ,  $\mathfrak{S} = \{-1, 1\}$ .

**Theorem 4.3** Let  $\{(y_{iL}, y_{iU})\}_{i=1}^n$  be i.i.d real valued random vectors such that  $\Pr\{y_{iL} \leq y_i \leq y_{iU}\} = 1$ . Assume that  $E(y_L^2) < \infty$ ,  $E(y_U^2) < \infty$ . Then,

$$\sqrt{n}H \left( \frac{1}{n} \bigoplus_{i=1}^n Y_i, \mathbb{E}(Y) \right) \xrightarrow{d} \max\{|z_{-1}|, |z_1|\}$$

where

$$(4.2) \quad \begin{pmatrix} z_{-1} \\ z_1 \end{pmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \text{Var}(y_L) & -\text{Cov}(y_L, y_U) \\ -\text{Cov}(y_L, y_U) & \text{Var}(y_U) \end{bmatrix} \right).$$

**Proof.** See Appendix. ■

The random variable appearing on the right-hand side of the limit in Theorem 4.3 is the maximum of the absolute value of two random variables with the joint distribution given in (4.2). Replacing the covariance matrix in (4.2) with a consistent estimator, both the cumulative distribution function and the probability density function of  $\max\{|z_{-1}|, |z_1|\}$  can be easily evaluated (they both have closed form expressions that are easy to derive), and critical values for  $\alpha \in (0, 1)$  can be obtained from the implied quantile function.

**Remark 4.4** The result of Theorem 4.3 can be extended to the entire class of models that give interval bounds (in  $\mathfrak{R}$ ) for the parameter of interest, provided that one can derive the joint asymptotic

distribution of the endpoints of the interval. This observation was already made by Chernozhukov et al. (2004). In our context, denoting by  $[\alpha, \beta]$  the population bound, by  $[\hat{\alpha}, \hat{\beta}]$  the estimated bound, and by  $a_n$  a growing sequence, if

$$a_n \begin{pmatrix} \hat{\alpha} - \alpha \\ \hat{\beta} - \beta \end{pmatrix} \xrightarrow{d} \begin{pmatrix} u_{-1} \\ u_1 \end{pmatrix}$$

then  $a_n H([\hat{\alpha}, \hat{\beta}], [\alpha, \beta]) = a_n \max\{|\hat{\alpha} - \alpha|, |\hat{\beta} - \beta|\} \xrightarrow{d} \max\{|u_{-1}|, |u_1|\}$ .

## 5 Best Linear Prediction with Interval Outcome Data

Suppose that one is interested in the parameters of the Best Linear Predictor (BLP) of a random variable  $y$  conditional on a random vector  $\mathbf{x}$ . Suppose further that one does not observe the realizations of  $y$ , but rather the realizations of two real valued random variables  $y_L, y_U$  such that  $\Pr\{y_L \leq y \leq y_U\} = 1$ . Let  $Y$  be defined as in equation (4.1). Throughout this Section, we maintain the following Assumption:

**Assumption 5.1** *Let  $(y, y_L, y_U, \mathbf{x})$  be a random vector in  $\mathfrak{R} \times \mathfrak{R} \times \mathfrak{R} \times \mathfrak{R}^d$  such that  $\Pr\{y_L \leq y \leq y_U\} = 1$ . The researcher observes a random sample  $\{(y_{iL}, y_{iU}, \mathbf{x}_i) : i = 1, \dots, n\}$  from the joint distribution of  $(y_L, y_U, \mathbf{x})$ .*

The proofs of the Propositions and Theorems for this section, given in the Appendix, consider the general case of  $\mathbf{x} \in \mathfrak{R}^d$ . To simplify the notation, in this section we introduce ideas restricting attention to the case that  $x \in \mathfrak{R}$  (though our assumptions are written for the general case  $d \geq 1$ ). To simplify the notation in what follows, let

$$\Sigma \equiv \begin{bmatrix} 1 & E(x) \\ E(x) & E(x^2) \end{bmatrix}, \quad \hat{\Sigma}_n \equiv \begin{bmatrix} 1 & \bar{x} \\ \bar{x} & \bar{x}^2 \end{bmatrix}.$$

Assume the following:

**Assumption 5.2**  $E(|y_L|) < \infty, E(|y_U|) < \infty, E(|y_L x_k|) < \infty, E(|y_U x_k|) < \infty, k = 1, \dots, d$ .

**Assumption 5.3**  $\Sigma$  is of full rank.



Then in the point identified case the population best linear predictor  $[\theta_1, \theta_2]$  solves the equations

$$\begin{aligned} E(y) &= \theta_1 + \theta_2 E(x), \\ E(xy) &= \theta_1 E(x) + \theta_2 E(x^2), \end{aligned}$$

and its sample analog  $[\hat{\theta}_1, \hat{\theta}_2]$  solves the equations

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n y_i &= \hat{\theta}_1 + \hat{\theta}_2 \frac{1}{n} \sum_{i=1}^n x_i, \\ \frac{1}{n} \sum_{i=1}^n x_i y_i &= \hat{\theta}_1 \frac{1}{n} \sum_{i=1}^n x_i + \hat{\theta}_2 \frac{1}{n} \sum_{i=1}^n x_i^2. \end{aligned}$$

In the interval outcomes case  $Y$  is an SVRV. We first introduce some additional notation to accommodate set valued variables. Let

$$\begin{aligned} G(\omega) &= \left\{ \left( \begin{bmatrix} y(\omega) \\ x(\omega) \cdot y(\omega) \end{bmatrix} : y(\omega) \in Y(\omega) \right) \right\}, \\ G_i(\omega) &= \left\{ \left( \begin{bmatrix} y_i(\omega) \\ x_i(\omega) \cdot y_i(\omega) \end{bmatrix} : y_i(\omega) \in Y_i(\omega) \right) \right\}. \end{aligned}$$

Lemma A.3 in the Appendix shows that  $G$  and  $G_i$ ,  $i \in \mathbb{N}$ , are SVRVs.

We define the population set-valued best linear predictor as

$$\begin{aligned} (5.1) \quad \Theta &= \left[ \begin{array}{cc} 1 & E(x) \\ E(x) & E(x^2) \end{array} \right]^{-1} \mathbb{E}[G] \\ &= \left\{ \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} : \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} = \left[ \begin{array}{cc} 1 & E(x) \\ E(x) & E(x^2) \end{array} \right]^{-1} \begin{bmatrix} E(y) \\ E(xy) \end{bmatrix}, \begin{bmatrix} y \\ xy \end{bmatrix} \in \mathcal{S}^1(G) \right\}. \end{aligned}$$

Before proceeding to deriving the estimator of  $\Theta$  and its asymptotic properties, we show that the identification region  $\Theta$  defined in (5.1) is identical to the identification region for the BLP obtained following the approach in Manski (2003) and denoted by  $\Theta^M$ :

$$(5.2) \quad \Theta^M = \left\{ \begin{bmatrix} \theta_1^M \\ \theta_2^M \end{bmatrix} : \begin{bmatrix} \theta_1^M \\ \theta_2^M \end{bmatrix} = \arg \min \int (y - \theta_1 - \theta_2 x)^2 d\eta, \eta \in \mathbf{P}_{yx} \right\},$$

where

$$(5.3) \quad \mathbf{P}_{yx} = \left\{ \eta : \Pr(y_L \leq t, x \leq x_0) \geq \eta((-\infty, t], (-\infty, x_0]) \geq \Pr(y_U \leq t, x \leq x_0) \forall t \in \mathfrak{R}, \forall x_0 \in \mathfrak{R}, \right. \\ \left. \eta((-\infty, +\infty), (-\infty, x_0]) = \Pr(x \leq x_0) \forall x_0 \in \mathfrak{R} \right\}.$$

**Proposition 5.1** *Suppose that one observes real valued random variables  $y_L$ ,  $y_U$ ,  $\mathbf{x}$  such that  $\Pr\{y_L \leq y \leq y_U\} = 1$  and  $\mathcal{S}^1(G) \neq \emptyset$ . Let  $\Theta$  and  $\Theta^M$  be defined respectively as in (5.1) and (5.2)-(5.3). Then  $\Theta = \Theta^M$ .*

**Proof.** See Appendix. ■

Given these preliminaries, we apply the analogy principle and define the sample analog of  $\Theta$  as

$$(5.4) \quad \hat{\Theta}_n = \begin{bmatrix} 1 & \bar{x} \\ \bar{x} & \bar{x}^2 \end{bmatrix}^{-1} \frac{1}{n} \bigoplus_{i=1}^n G_i,$$

where  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ ,  $\bar{x}^2 = \frac{1}{n} \sum_{i=1}^n x_i^2$ .

Observe that by construction  $\Theta$ , as defined in equation (5.1), is a convex set. Lemma A.7 in the Appendix describes further its geometry, showing that when  $x$  has an absolutely continuous distribution and  $G$  is integrably bounded,  $\Theta$  is a strictly convex set, that is, it doesn't have a flat face. The estimated set  $\hat{\Theta}_n$  is a convex polytope, because it is given by a finite Minkowski sum of segments in  $\mathfrak{R}^d$  (also called a *zonotope*, see Molchanov (2005) page 204).

Theorem 5.2 below shows that under mild regularity conditions on the moments of  $(y_L, y_U, \mathbf{x})$  as in Assumption 5.2,  $\hat{\Theta}_n$  is a consistent estimator of  $\Theta$ . Under the additional Assumption 5.4 below, this convergence occurs at the rate  $O_p\left(\frac{1}{\sqrt{n}}\right)$ .

**Assumption 5.4**  $E\left(|y_L|^2\right) < \infty$ ,  $E\left(|y_U|^2\right) < \infty$ ,  $E\left(|y_L x_k|^2\right) < \infty$ ,  $E\left(|y_U x_k|^2\right) < \infty$ ,  $E\left(|x_k|^4\right) < \infty$ ,  $k = 1, \dots, d$ .

**Theorem 5.2** *Let Assumptions 5.1, 5.2, and 5.3 hold. Define  $\Theta$  and  $\hat{\Theta}_n$  as in (5.1) and (5.4) respectively. Then  $H\left(\hat{\Theta}_n, \Theta\right) \xrightarrow{a.s.} 0$ . If in addition Assumption 5.4 holds, then  $H\left(\hat{\Theta}_n, \Theta\right) = O_p\left(\frac{1}{\sqrt{n}}\right)$ .*

**Proof.** See Appendix. ■

The proof of Theorem 5.2 is based on an extension of Slutsky's Theorem to SVRVs, which we provide in Lemma A.5 in the Appendix. The rate of convergence that we obtain is  $\frac{1}{\sqrt{n}}$  irrespective of whether  $x$  has a continuous or a discrete distribution. To derive the asymptotic distribution of  $H\left(\hat{\Theta}_n, \Theta\right)$  we need to impose an additional condition on the distribution of  $x$  :

**Assumption 5.5** *The distribution of  $\mathbf{x}$  is absolutely continuous with respect to Lebesgue measure on  $\mathfrak{R}^{d+1}$ .*

Theorem 5.3 gives the asymptotic distribution of  $\sqrt{n}H\left(\hat{\Theta}_n, \Theta\right)$ .

**Theorem 5.3** *Let Assumptions 5.1, 5.3, 5.4, and 5.5 hold.*

(i) *Then  $\sqrt{n}H\left(\hat{\Theta}_n, \Theta\right) \xrightarrow{d} \|v\|_{\mathbb{C}(\mathbb{S})}$  where  $v$  is a linear function of a vector Gaussian random system with  $E[v(\mathbf{p})] = 0$  for all  $\mathbf{p} \in \mathbb{S}$ , and*

$$(5.5) \quad E[v(\mathbf{p})v(\mathbf{q})] = E[s(\mathbf{p}, \Sigma^{-1}G)s(\mathbf{q}, \Sigma^{-1}G)] - E[s(\mathbf{p}, \Sigma^{-1}G)]E[s(\mathbf{q}, \Sigma^{-1}G)] \\ - \langle \boldsymbol{\xi}_{\mathbf{p}}, \boldsymbol{\kappa}_{\mathbf{q}, \mathbf{p}} \rangle - \langle \boldsymbol{\kappa}_{\mathbf{p}, \mathbf{q}}, \boldsymbol{\xi}_{\mathbf{q}} \rangle + \langle \boldsymbol{\xi}_{\mathbf{p}}, V_{\mathbf{p}, \mathbf{q}} \boldsymbol{\xi}_{\mathbf{q}} \rangle,$$

where for  $\mathbf{p}, \mathbf{q} \in \mathbb{S}$  the singleton  $\boldsymbol{\xi}_{\mathbf{p}} = \Theta \cap \{\boldsymbol{\vartheta} \in \mathfrak{R}^{d+1} : \langle \boldsymbol{\vartheta}, \mathbf{p} \rangle = s(\mathbf{p}, \Theta)\}$ , and the matrix  $V_{\mathbf{p}, \mathbf{q}}$  and the vector  $\boldsymbol{\kappa}_{\mathbf{p}, \mathbf{q}}$  are given, respectively, in equations (A.8) and (A.9) in the Appendix.

(ii) *Assume in addition that  $\text{Var}(y_L | \mathbf{x}), \text{Var}(y_U | \mathbf{x}) \geq \sigma^2 > 0$   $P(\mathbf{x})$  – a.s. Then  $\text{Var}(v(\mathbf{p})) > 0$  for each  $\mathbf{p} \in \mathbb{S}$ , and therefore the law of  $\|v\|_{\mathbb{C}(\mathbb{S})}$  is absolutely continuous with respect to Lebesgue measure on  $\mathfrak{R}$ .*

**Proof.** See Appendix. ■

The difficulty in obtaining the asymptotic distribution of  $\sqrt{n}H\left(\hat{\Theta}_n, \Theta\right)$  lies in the fact that there are two sources of randomness in  $\hat{\Theta}_n = \hat{\Sigma}_n^{-1} \frac{1}{n} \bigoplus_{i=1}^n G_i$ , which we need to disentangle. The first source of randomness is the SVRV  $\frac{1}{n} \bigoplus_{i=1}^n G_i$ , while the second one is the random matrix  $\hat{\Sigma}_n^{-1}$ . If instead of the random matrix  $\hat{\Sigma}_n^{-1}$  we had the non-random matrix  $\Sigma^{-1}$  premultiplying  $\frac{1}{n} \bigoplus_{i=1}^n G_i$ , then a straightforward extension of Theorem 2.3 could be applied. Because there is no analog of Slutsky’s Theorems in the SVRVs literature (except, to the best of our knowledge, for the result that we establish in Lemma A.5, which however only provides an upper bound on  $H\left(\hat{\Theta}_n, \Theta\right)$ ), to prove Theorem 5.3 we proceed in steps. First, we use Hörmander’s embedding theorem to express  $\sqrt{n}H\left(\hat{\Theta}_n, \Theta\right)$  in terms of the uniform distance between the support functions of (a transformation of)  $\hat{\Theta}_n$  and  $\Theta$ . Then we focus on the support functions of (a transformation of) these sets, and rewrite them as the sum of two (not necessarily independent) elements. We derive their joint asymptotic distribution using an application of the Delta Method to  $\mathbb{C}(\mathbb{S})$ –valued random variables. However, the use of this Functional Delta Method (as it is explained in the Appendix) requires differentiability of the support function of  $\Theta$ ; Lemma A.7 in the Appendix shows that this condition is satisfied when  $x$  has an absolutely continuous distribution.

Alternatively (for example if  $x$  has a discrete distribution<sup>5</sup>) an extremely simple approximation allows the researcher to circumvent the randomness in  $\hat{\Sigma}_n$  at the cost of a slower rate of

---

<sup>5</sup>When  $x$  has a discrete distribution, even the population identification region  $\Theta$  is a polytope, given by a Minkowski sum of segments. As a result, when  $x$  is discrete the support function of  $\Theta$  is not differentiable and the Functional

convergence. The approximation is as follows: let  $m_n$  be a sequence of integers which goes to infinity slower than  $n$  :  $m_n \rightarrow \infty$  and  $\frac{m_n}{n} \rightarrow 0$  as  $n \rightarrow \infty$  (for example,  $m_n = \lfloor \frac{n}{\log n} \rfloor$ ). Let  $\bar{G}_{m_n} = \frac{1}{m_n} (G_1 \oplus G_2 \oplus \dots \oplus G_{m_n})$  be a Minkowski average of a random subsample of size  $m_n$  from the  $n$  observations of  $G$  (that is, the  $m_n$  observations are independently and identically distributed, each with selection expectation  $\mathbb{E}[G]$ ). Let

$$(5.6) \quad \hat{\Theta}_{n,m_n} = \hat{\Sigma}_n^{-1} \bar{G}_{m_n}.$$

**Corollary 5.4** *Let Assumptions 5.1, 5.3, and 5.4 hold. Pick a random subsample of size  $m_n$  such that  $m_n \rightarrow \infty$  and  $\frac{m_n}{n} \rightarrow 0$  as  $n \rightarrow \infty$ . Define  $\Theta$  and  $\hat{\Theta}_{n,m_n}$  as in (5.1) and (5.6) respectively. Then  $\sqrt{m_n} H(\hat{\Theta}_{n,m_n}, \Theta) \xrightarrow{d} \|z\|_{\mathbb{C}(\mathbb{S})}$  where  $z$  is a Gaussian random system with  $E[z(\mathbf{p})] = 0$  for all  $\mathbf{p} \in \mathbb{S}$ , and*

$$(5.7) \quad E[z(\mathbf{p})z(\mathbf{q})] = E[s(\mathbf{p}, \Sigma^{-1}G) s(\mathbf{q}, \Sigma^{-1}G)] - E[s(\mathbf{p}, \Sigma^{-1}G)] E[s(\mathbf{q}, \Sigma^{-1}G)].$$

*Assume in addition that  $\text{Var}(y_L | \mathbf{x}), \text{Var}(y_U | \mathbf{x}) \geq \sigma^2 > 0$   $P(\mathbf{x}) - a.s.$  Then  $\text{Var}(z(\mathbf{p})) > 0$  for each  $\mathbf{p} \in \mathbb{S}$ , and therefore the law of  $\|z\|_{\mathbb{C}(\mathbb{S})}$  is absolutely continuous with respect to Lebesgue measure on  $\mathfrak{R}$ .*

**Proof.** See Appendix. ■

Similarly to the discussion in Section 3, the asymptotic distribution of our test statistics depends on parameters to be estimated. Hence we again obtain the critical values using bootstrap procedures. Here we detail a bootstrap procedure for the approximation of the critical values of the limiting distribution in Theorem 5.3 (those for the limiting distribution in Corollary 5.4 can be obtained similarly):

### Algorithm 5.1

1) Generate a bootstrap sample of size  $n$ ,  $\{(y_{iL}^*, y_{iU}^*, x_i^*) : i = 1, \dots, n\}$ , by drawing a random sample from the joint empirical distribution of the vector  $\{(y_{iL}, y_{iU}, x_i) : i = 1, \dots, n\}$ , denoted  $\hat{P}_n$ , with replacement. Use this sample to construct bootstrap versions of  $G_i$ , denoted  $G_i^*$ , and  $\hat{\Sigma}_n^{-1}$ , denoted  $\hat{\Sigma}_n^{-1*}$ .

2) Compute  $r_n^* \equiv \sqrt{n} H(\hat{\Sigma}_n^{-1*} \cdot \frac{1}{n} \bigoplus_{i=1}^n G_i^*, \hat{\Theta}_n)$ .

---

Delta Method cannot be applied. In this case, tedious calculations allow one to express the extreme points of  $\Theta$  as functions of moments of  $(y_L, y_U, x)$ . Hence the exact asymptotic distribution of  $H(\hat{\Theta}_n, \Theta)$  can be derived.

3) Use the results of  $b$  repetitions of Steps 1-2 to compute the empirical distribution of  $r_n^*$  at a point  $t$ , denoted by  $J_n(t, \hat{P}_n)$ .

4) Estimate the critical value  $c_\alpha^{BLP}$  such that  $\Pr\{\|v\|_{\mathbb{C}(\mathbb{S})} > c_\alpha^{BLP}\} = \alpha$  by

$$(5.8) \quad \hat{c}_{\alpha n}^{BLP} = \inf\left\{t : J_n(t, \hat{P}_n) \geq 1 - \alpha\right\}.$$

The asymptotic validity of this procedure follows by an application of the Delta Method for the bootstrap (in the case of the bootstrap algorithm that one would use to approximate the critical values of the limiting distribution in Corollary 5.4, its validity follows easily from Theorem 2.4 of Giné & Zinn (1990)). In particular, the following result holds:

**Proposition 5.5** *Let the assumptions of Theorem 5.3-(i) hold. Then  $r_n^* \xrightarrow{d} \|v\|_{\mathbb{C}(\mathbb{S})}$ , where  $v$  is a linear function of a vector Gaussian random system with  $E[v(\mathbf{p})] = 0$  for all  $\mathbf{p} \in \mathbb{S}$ , and with the covariance kernel given in equation (5.5). If in addition the assumptions of Theorem 5.3-(ii) hold, then  $\hat{c}_{\alpha n}^{BLP} = c_\alpha^{BLP} + o_p(1)$ , where  $\hat{c}_{\alpha n}^{BLP}$  is defined in (5.8).*

**Proof.** See Appendix. ■

Alternatively we could obtain consistent estimates of  $c_\alpha^{BLP}$  by simulating the distribution of the supremum of a linear function of a vector Gaussian random system with mean function equal to zero for each  $\mathbf{p} \in \mathbb{S}$  and with a covariance kernel which consistently estimates the covariance kernel in (5.5). The formal result is as follows:

**Proposition 5.6** *Let the assumptions of Theorem 5.3 hold. Let  $\hat{v}$  be a linear function of a vector Gaussian random system with  $E[\hat{v}(\mathbf{p})] = 0$  for all  $\mathbf{p} \in \mathbb{S}$ , and with covariance kernel given by*

$$\begin{aligned} E[\hat{v}(\mathbf{p})\hat{v}(\mathbf{q})] &= \frac{1}{n} \sum_{i=1}^n s(\mathbf{p}, \hat{\Sigma}_n^{-1}G_i) s(\mathbf{q}, \hat{\Sigma}_n^{-1}G_i) - \frac{1}{n} \sum_{i=1}^n s(\mathbf{p}, \hat{\Sigma}_n^{-1}G_i) \frac{1}{n} \sum_{i=1}^n s(\mathbf{q}, \hat{\Sigma}_n^{-1}G_i) \\ &\quad - \langle \hat{\xi}_{\mathbf{p}}, \hat{\kappa}_{\mathbf{q}, \mathbf{p}} \rangle - \langle \hat{\kappa}_{\mathbf{p}, \mathbf{q}}, \hat{\xi}_{\mathbf{q}} \rangle + \langle \hat{\xi}_{\mathbf{p}}, \hat{V}_{\mathbf{p}, \mathbf{q}} \hat{\xi}_{\mathbf{q}} \rangle, \end{aligned}$$

where  $\hat{V}_{\mathbf{p}, \mathbf{q}}$  and  $\hat{\kappa}_{\mathbf{p}, \mathbf{q}}$  are the sample analogs of  $V_{\mathbf{p}, \mathbf{q}}$  and  $\kappa_{\mathbf{p}, \mathbf{q}}$ , and

$$\hat{\xi}_{\mathbf{p}} \in \hat{\Xi}_{n, \mathbf{p}} = \hat{\Theta}_n \cap \left\{ \boldsymbol{\vartheta} \in \mathbb{R}^{d+1} : \langle \boldsymbol{\vartheta}, \mathbf{p} \rangle = s(\mathbf{p}, \hat{\Theta}_n) \right\}, \mathbf{p} \in \mathbb{S}.$$

Then if  $\text{Var}(v(\mathbf{p})) > 0$  for each  $\mathbf{p} \in \mathbb{S}$ ,  $\inf\left\{t : \Pr\{\|\hat{v}\|_{\mathbb{C}(\mathbb{S})} > t\} \geq 1 - \alpha\right\} = c_\alpha^{BLP} + o_p(1)$ , where  $c_\alpha^{BLP}$  is defined in step 4 of Algorithm 5.1.

**Proof.** See Appendix. ■

## 6 Extensions

In this Section we briefly outline how the general methodology proposed in Section 3 can be (i) applied to problems (other than inference for the population mean and the BLP with interval outcome data) in which the identification region is defined through moment conditions that are linear in the Aumann expectation of an SVRV, and (ii) modified to test hypotheses about the “true” value of a partially identified population feature. The first extension is exemplified looking at the semiparametric binary model with interval regressor data analyzed by Magnac & Maurin (2005). For the sake of brevity, the discussion below omits rigorous proofs of the results.<sup>6</sup>

### 6.1 Semiparametric Binary Models with Interval Regressor Data

Magnac & Maurin (2005) study identification in semi-parametric binary regression models of the form  $y = \mathbb{I}(\mathbf{x}'\boldsymbol{\beta} + v + \varepsilon > 0)$ , where  $\mathbb{I}(\cdot)$  is the indicator function of the event in brackets,  $\mathbf{x}$  is a vector of potentially endogenous regressors, and only interval data  $(v_L, v_U)$  such that  $\Pr(v_L \leq v \leq v_U) = 1$  are available on  $v$ . Under a suitable set of assumptions (in the spirit of Lewbel (2000)), the identification region of  $\boldsymbol{\beta}$  is defined through moment conditions which are linear in the Aumann expectation of a properly specified SVRV.

Magnac & Maurin (2005) assume that the support of  $(v_L, v_U)$  is finite, so that  $v$  belongs to one of  $k - 1$  intervals. They denote by  $v^*$  the number of the interval in which  $v$  lies. The support of  $v$  conditional on  $(\mathbf{x}, \mathbf{z}, v^* = j)$  is assumed to be  $[v_j, v_{j+1}[$ . The overall support is  $[v_1, v_k[$ , where  $v_1$  and  $v_k$  are finite. The error term  $\varepsilon$  is assumed to have the following properties: it is uncorrelated with a set of instruments  $\mathbf{z}$ ; it is independent of  $v$  and  $v^*$  conditionally on  $\mathbf{x}$  and  $\mathbf{z}$ ; and it has a conditional support sufficiently smaller than the support of  $v$  (see Magnac & Maurin (2005), Assumption L in Section 2, and Assumption ID in Section 4.1). Under these assumptions, Magnac & Maurin (2005, Theorem 6) show that the identification region for  $\boldsymbol{\beta}$  is given by a set with a structure very similar to that in equation (5.1), although in their case the analog of  $Y$  is a random segment whose extreme points are a function of  $(\mathbf{x}, \mathbf{z})$ . In particular, denote by

$$\begin{aligned} \bar{\phi}(\mathbf{x}, \mathbf{z}) &= \sum_{j=1, \dots, k-1} [\Pr(y = 1 | \mathbf{x}, \mathbf{z}, v^* = j + 1) - \Pr(y = 1 | \mathbf{x}, \mathbf{z}, v^* = j)] (v_{j+1} - v_j), \\ \underline{\phi}(\mathbf{x}, \mathbf{z}) &= \sum_{j=1, \dots, k-1} [\Pr(y = 1 | \mathbf{x}, \mathbf{z}, v^* = j) - \Pr(y = 1 | \mathbf{x}, \mathbf{z}, v^* = j - 1)] (v_{j+1} - v_j), \end{aligned}$$

---

<sup>6</sup>Which can be obtained from the authors upon request.

where by convention  $\Pr(y = 1 | \mathbf{x}, \mathbf{z}, v^* = j) = 0$  for  $j = 0$ , and  $\Pr(y = 1 | \mathbf{x}, \mathbf{z}, v^* = j) = 1$  for  $j = k$ .

Magnac and Maurin show that the identification region for  $\beta$ , denoted  $B$ , is

$$B = \left\{ \beta \in \mathbb{R}^{d+1} : E \left[ \mathbf{z} \left( \mathbf{x}'\beta - \left( \frac{(v_{v^*+1} - v_{v^*})}{P(v^* | \mathbf{x}, \mathbf{z})} y - v_k \right) \right) \right] = E[\mathbf{z}u(\mathbf{x}, \mathbf{z})], \quad u(\mathbf{x}, \mathbf{z}) \in ]\underline{\phi}(\mathbf{x}, \mathbf{z}), \bar{\phi}(\mathbf{x}, \mathbf{z})[ \right\}.$$

This implies that

$$(6.1) \quad \begin{aligned} B &\subset [E(\mathbf{z}\mathbf{x}')^{-1} E \left[ \mathbf{z} \left( \frac{(v_{v^*+1} - v_{v^*})}{P(v^* | \mathbf{x}, \mathbf{z})} y - v_k \right) \right] \oplus [E(\mathbf{z}\mathbf{x}')^{-1} \mathbb{E}[M], \\ M &= \{ (\mathbf{z} \cdot u(\mathbf{x}, \mathbf{z}) : u(\mathbf{x}, \mathbf{z}) \in ]\underline{\phi}(\mathbf{x}, \mathbf{z}), \bar{\phi}(\mathbf{x}, \mathbf{z})[) \}, \end{aligned}$$

where  $M$  is the closure of the SVRV  $\{[\mathbf{z}u(\mathbf{x}, \mathbf{z})], u(\mathbf{x}, \mathbf{z}) \in ]\underline{\phi}(\mathbf{x}, \mathbf{z}), \bar{\phi}(\mathbf{x}, \mathbf{z})[ \}$  which appears in Magnac & Maurin.

As a subject of work in progress we are exploring the possibility of estimating the set  $B$  by replacing the expectations in equation (6.1) with the corresponding sample averages, and the Aumann expectation in equation (6.1) with a Minkowski sample average of random segments. The distribution functions  $P(v^* | \mathbf{x}, \mathbf{z})$  and  $\Pr(y = 1 | \mathbf{x}, \mathbf{z}, v^*)$  are estimated by kernel methods. Denote the resulting estimator by  $\hat{B}_n$ . Our work in progress studies suitable regularity conditions under which consistency of  $\hat{B}_n$  (with respect to the Hausdorff distance) can be established by combining standard methods for semiparametric two-step estimators, with the result in Theorem 5.2 above. Similarly, we are exploring suitable regularity conditions under which one can obtain the asymptotic distribution of  $H(\hat{B}_n, B)$  combining the result in Theorem 5.3 above with the argument of Theorem 8.11 in Newey & McFadden (1994).

## 6.2 Testing Procedure and Confidence Sets for Points in the Identification Region

Our inferential approach as introduced in Section 3 targets the entire identification region of a partially identified population feature, and provides asymptotically exact size critical values with which to test hypotheses and construct confidence collections. However, there are applications in which the researcher is interested in testing hypotheses and construct confidence sets for the “true” value of the population feature, following the insight of Imbens & Manski (2004). For this case, our methodology can be modified to conduct conservative tests of hypotheses, and construct conservative confidence sets that asymptotically cover each point in the identification region with a prespecified probability.

## Hypothesis Testing

Given a real valued vector  $\boldsymbol{\psi}_0$ , suppose that one wants to test the null hypothesis

$$\mathfrak{H}_0 : \boldsymbol{\psi}_0 \in \mathbb{E}[F]$$

against the alternative

$$\mathfrak{H}_A : \boldsymbol{\psi}_0 \notin \mathbb{E}[F]$$

at a prespecified significance level  $\alpha \in (0, 1)$ . Under the null hypothesis,

$$(6.2) \quad d(\boldsymbol{\psi}_0, \mathbb{E}[F]) = \inf_{\mathbf{f} \in \mathbb{E}[F]} \|\mathbf{f} - \boldsymbol{\psi}_0\| = 0.$$

Hence, by the triangle inequality and Theorem 1.1.9 in Li et al. (2002),

$$(6.3) \quad \begin{aligned} d\left(\boldsymbol{\psi}_0, \frac{1}{n} \bigoplus_{i=1}^n F_i\right) &\leq d(\boldsymbol{\psi}_0, \mathbb{E}[F]) + \sup_{\mathbf{f} \in \mathbb{E}[F]} d\left(\mathbf{f}, \frac{1}{n} \bigoplus_{i=1}^n F_i\right) \\ &= \sup_{\mathbf{f} \in \mathbb{E}[F]} d\left(\mathbf{f}, \frac{1}{n} \bigoplus_{i=1}^n F_i\right) \\ &= \sup_{\mathbf{p} \in \mathbb{B}} \left\{ s(\mathbf{p}, \mathbb{E}[F]) - s\left(\mathbf{p}, \frac{1}{n} \bigoplus_{i=1}^n F_i\right) \right\} \\ &= \sup_{\mathbf{p} \in \mathbb{B}} \left\{ - \left( \frac{1}{n} \sum_{i=1}^n s(\mathbf{p}, F_i) - E[s(\mathbf{p}, F_i)] \right) \right\}. \end{aligned}$$

where the first equality follows from equation (6.2). Based on the last equality in the expression above, under the same assumptions of Theorem 2.3 one can show that

$$\sqrt{n} \sup_{\mathbf{f} \in \mathbb{E}[F]} d\left(\mathbf{f}, \frac{1}{n} \bigoplus_{i=1}^n F_i\right) \xrightarrow{d} \sup_{\mathbf{p} \in \mathbb{B}} \{-z(\mathbf{p})\}$$

where  $z$  is a Gaussian random system with  $E[z(\mathbf{p})] = 0$  for all  $\mathbf{p} \in \mathbb{B}$ , and  $E[z(\mathbf{p})z(\mathbf{q})] = E[s(\mathbf{p}, F)s(\mathbf{q}, F)] - E[s(\mathbf{p}, F)]E[s(\mathbf{q}, F)]$ .

Hence, letting  $\bar{F}_n \equiv \frac{1}{n} \bigoplus_{i=1}^n c o F_i$ , one can test  $\mathfrak{H}_0$  against  $\mathfrak{H}_A$  using the following criterion:<sup>7</sup>

$$\text{Reject } \mathfrak{H}_0 \text{ if } \quad \sqrt{n}d(\boldsymbol{\psi}_0, \bar{F}_n) > \tilde{c}_\alpha,$$

$$\text{Do not Reject } \mathfrak{H}_0 \text{ if } \quad \sqrt{n}d(\boldsymbol{\psi}_0, \bar{F}_n) \leq \tilde{c}_\alpha,$$

where  $\tilde{c}_\alpha$  is chosen so that

$$(6.4) \quad \Pr \left\{ \sup_{\mathbf{p} \in \mathbb{B}} \{-z(\mathbf{p})\} > \tilde{c}_\alpha \right\} = \alpha.$$

---

<sup>7</sup>As in Section 3, using  $\bigoplus_{i=1}^n c o F_i$  rather than  $\bigoplus_{i=1}^n F_i$  to construct the test statistic improves the small sample properties of the test without changing its asymptotic behavior.



By construction, given the result in equation (6.3),

$$\lim_{n \rightarrow \infty} \Pr \{ \sqrt{nd} (\boldsymbol{\psi}_0, \bar{F}_n) > \tilde{c}_\alpha \} \leq \lim_{n \rightarrow \infty} \Pr \left\{ \sqrt{n} \sup_{\mathbf{f} \in \mathbb{E}[F]} d(\mathbf{f}, \bar{F}_n) > \tilde{c}_\alpha \right\} = \alpha.$$

This testing procedure is therefore conservative, but one can show that it preserves the property of rejecting a false null hypothesis with probability approaching 1 as the sample size increases. The critical value  $\tilde{c}_\alpha$  can be consistently estimated through a bootstrap procedure similar to the one in Algorithm 3.1 (we denote this estimator by  $\hat{c}_{\alpha n}$ ).

### Confidence Sets

Exploiting the duality between confidence sets and hypothesis tests that is of familiar use for point identified models, we can invert the test statistic described above to construct a confidence set with the property that it covers each point in the identification region with at least a prespecified probability. In particular, we construct the set  $CI^\psi$  as

$$CI^\psi = \left\{ \tilde{\boldsymbol{\psi}} \in \mathfrak{R}^d : \sqrt{nd} (\tilde{\boldsymbol{\psi}}, \bar{F}_n) \leq \hat{c}_{\alpha n} \right\}.$$

Hence if the law of  $\sup_{\mathbf{p} \in \mathbb{B}} \{-z(\mathbf{p})\}$  is continuous, for each  $\tilde{\boldsymbol{\psi}} \in \mathbb{E}[F]$ ,

$$\lim_{n \rightarrow \infty} \Pr \left\{ \tilde{\boldsymbol{\psi}} \in CI^\psi \right\} \geq 1 - \alpha.$$

For the case of partially identified scalar valued parameters for which the identification region is an interval, Imbens & Manski (2004) propose confidence sets which asymptotically cover the true value of the parameter uniformly across different values for the width of the bound. The derivation of conditions under which the coverage of the confidence sets described in this Section is uniform is the subject of work in progress.

Similarly to the case of the union of all the sets in the confidence collection  $CC^{\mathbb{E}[F]}$  introduced in Section 3, the (exact) set  $CI^\psi$  can be calculated in a particularly simple way. One can show that

$$CI^\psi = \bar{F}_n \oplus B_{\hat{c}_{\alpha n}} \subset \mathcal{U}_n,$$

where  $B_{\hat{c}_{\alpha n}} = \left\{ \mathbf{b} \in \mathfrak{R}^d : \|\mathbf{b}\| \leq \frac{\hat{c}_{\alpha n}}{\sqrt{n}} \right\}$  and  $\mathcal{U}_n$  is defined in Theorem 3.4. This result confirms what one might have intuitively expected: when the researcher is interested in a confidence set for the “true” value of the population feature of interest, the set  $CI^\psi$  gives a smaller (and therefore less conservative) confidence region than the one obtained as the union of all sets in  $CC^{\mathbb{E}[F]}$ .

Table 1: Size and rejection rates of the test against local alternatives (nominal level = 0.05) –  $\mathbb{E}(y)$

sample size	$\delta$						
	0	0.5	1	2	4	8	16
$n = 100$	0.0464	0.0504	0.0710	0.1681	0.5328	0.9764	1.0000
$n = 200$	0.0462	0.0520	0.0721	0.1671	0.5276	0.9808	1.0000
$n = 500$	0.0509	0.0553	0.0773	0.1665	0.5268	0.9820	1.0000
$n = 1000$	0.0473	0.0542	0.0743	0.1654	0.5186	0.9823	1.0000
$n = 2000$	0.0497	0.0555	0.0767	0.1706	0.5249	0.9833	1.0000

## 7 Monte Carlo Results

In this section we conduct a series of Monte Carlo experiments to evaluate the properties of the test proposed in Section 3, as applied to the problem of inference for the mean and for the parameters of the BLP with interval outcome data. We use the same data set as in Chernozhukov et al. (2004) to conduct inference for (i) the mean (logarithm of) wage, and (ii) the returns to education, for men between the ages of 20 and 50. The data are taken from the March 2000 wave of the Current Population Survey (CPS), and contain 13,290 observations on income and education. The wage variable is artificially bracketed to create interval-valued outcomes. A detailed description of the data construction procedure appears in Chernozhukov et al. (2004, Section 4). Denote the logarithm of the lower bound of the wage, the logarithm of the upper bound of the wage, and the level of education in years, by  $y_L$ ,  $y_U$  and  $x$  respectively. We treat the empirical joint distribution of  $(y_L, y_U, x)$  in the CPS sample as the population distribution and draw small samples from it.

The first Monte Carlo experiment looks at the asymptotic properties of the estimator for the population mean with interval data (see Section 4). Denote by  $Y = [y_L, y_U]$  the interval-valued log of wages. The population identification region is

$$\Psi_0 = \mathbb{E}(Y) = [4.4347, 4.9674].$$

We draw 25,000 small samples of sizes 100, 200, 500, 1000, and 2000 from the CPS “population.” For each sample, we estimate the critical value  $c_\alpha$ ,  $\alpha = 0.05$ , implementing Algorithm 3.1 with 2000 bootstrap repetitions.<sup>8</sup> We build the local alternatives using equation (3.5) with  $\Delta_1 = \{0\}$

<sup>8</sup>Matlab and Fortran 90 codes for computing the Minkowski sample average of intervals in  $\mathfrak{R}$  and  $\mathfrak{R}^2$ , the Hausdorff distance between polytopes in  $\mathfrak{R}^2$ , and for implementing Algorithm 3.1 and Algorithm 5.1 are available upon request from the authors.

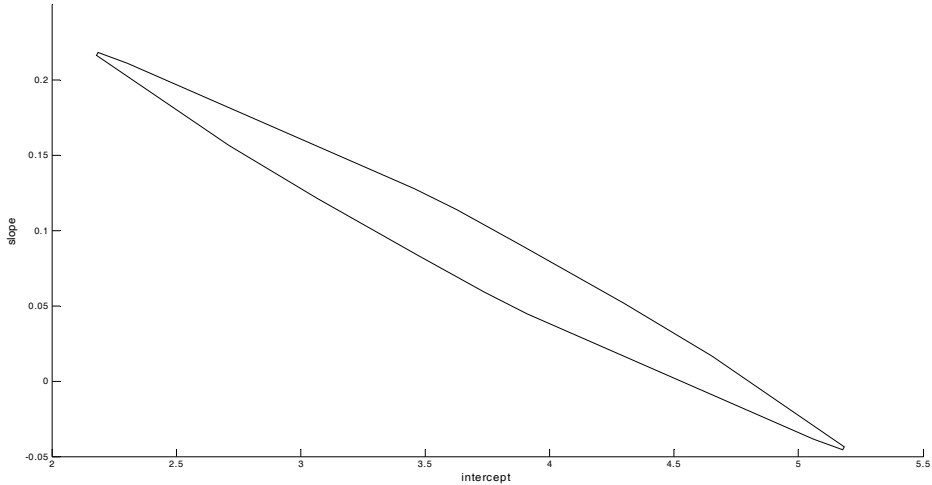


Figure 4: The population identification region of the parameters of the BLP

and  $\Delta_2 = \delta \frac{\{0.5\}}{\sqrt{n}}$ . More specifically, the local alternatives are defined as  $\Psi_{A_n}(\delta) = \Psi_0 \oplus \delta \frac{\{0.5\}}{\sqrt{n}}$  for  $\delta \in \{0, \frac{1}{2}, 1, 2, 4, 8, 16\}$ , where 0.5 is the width (approximated to the first decimal point) of  $\mathbb{E}(Y)$ . In terms of equation (3.6), the Hausdorff distance between the local alternative and the null is  $\kappa = \frac{1}{2}\delta$ . Results for this Monte Carlo experiment appear in Table 1.<sup>9</sup> As the first column of this Table shows, the empirical size of our test is quite close to the nominal level of 0.05. As  $\delta$  (and therefore  $\kappa$ ) increases, the rejection rates increase for each given sample size, and for a given  $\delta$ , the rejection rates are stable across sample sizes. These results are invariant with the width of  $\mathbb{E}(Y)$ .

The second Monte Carlo experiment uses both the interval valued information on the logarithm of wage and the information about education. Here we implement the test at the 0.05 level for the best linear predictor of the logarithm of wage given education. Figure 4 depicts  $\Theta$ , the population identification region of the parameters of the BLP given in equation (5.1).

To trace the increase in power as we get farther away from the null, we again use a sequence of local alternatives. The local alternatives are defined as  $\Psi_{A_n}(\delta) = \Psi_0 \oplus \frac{\{(3,0.3)\}}{\sqrt{n}}$  for  $\delta \in \{0, \frac{1}{2}, 1, 2, 4, 8, 16\}$ , where  $\Psi_0 = \Theta$  is the polytope in Figure 5, and the vector (3, 0.3) gives

---

<sup>9</sup>We also conducted Monte Carlo experiments in which the critical value  $c_\alpha$  is obtained by simulations. We considered both the case in which the covariance kernel was known (i.e., it was the covariance kernel obtained from the entire CPS “population”), and the case in which it was replaced by a consistent estimator. The results are comparable to those reported in Table 1. However, the procedure in Algorithm 3.1 is computationally less burdensome.

Table 2: Size and rejection rates of the test against local alternatives (nominal level = 0.05) – BLP

sample size	$\delta$						
	0	0.5	1	2	4	8	16
$n = 100$	0.0663	0.0759	0.0930	0.1518	0.3516	0.8644	1.0000
$n = 200$	0.0660	0.0757	0.0927	0.1502	0.3488	0.8807	1.0000
$n = 500$	0.0613	0.0686	0.0830	0.1344	0.3361	0.8901	1.0000
$n = 1000$	0.0569	0.0632	0.0772	0.1262	0.3283	0.8934	1.0000
$n = 2000$	0.0539	0.0589	0.0726	0.1201	0.3245	0.9006	1.0000

the width (approximated to the first decimal point) of the projection of  $\Theta$  on each axis. In terms of equation (3.6) the distance between the local alternative and the null is  $\kappa = 3.015 \cdot \delta$ . We draw 25,000 small samples of sizes 100, 200, 500, 1000, and 2000 from the CPS “population.” For each sample, we estimate the critical value  $c_\alpha$ ,  $\alpha = 0.05$ , implementing Algorithm 5.1 with 2000 bootstrap repetitions. The rejection rates of the null and local alternatives are shown in Table 2. As the first column of this Table shows, for small samples (e.g.,  $n = 100$ ) the test is a little oversized; however, as  $n$  increases this size distortion disappears. As  $\delta$  (and therefore  $\kappa$ ) increases, the rejection rates increase for each given sample size.

In summary, our Monte Carlo experiments show that the test described in section 3 performs well even with samples of size as small as 100. The rejection rates of the null are very close to 0.05. The power against the local alternatives grows rapidly as the alternatives get far away from the null.

## 8 Conclusions

This paper introduces a methodology to conduct estimation and inference for partially identified population features in a completely analogous way to how estimation and inference are conducted when the population features is point identified. In particular, we show that for a certain class of partially identified population features, which include means and best linear predictors with interval outcome data, the identification region is given by a transformation of the Aumann expectation of an SVRV. Extending the analogy principle to SVRVs, we prove that this expectation can be  $\sqrt{n}$ -consistently estimated (with respect to the Hausdorff distance) by a transformation of the Minkowski average of a sample of SVRVs which can be constructed from observable random

variables. When the Hausdorff distance between the proposed estimator and the population identification region is normalized by  $\sqrt{n}$ , this statistic converges in distribution to the supremum of a Gaussian process whose covariance kernel depends on parameters of the population identification region. We introduce consistent bootstrap procedures to estimate the quantiles of this distribution.

The asymptotic distribution result allows us to introduce a procedure to test whether the population identification region is equal to a particular set. This test procedure is consistent against any fixed alternative, and is locally asymptotically unbiased. A Monte Carlo exercise shows that our test performs well even with samples of size as small as 100. The rejection rates of the null are very close to the nominal level, and the power of the test against local alternatives grows rapidly as the alternatives get far away from the null. The test statistic can be inverted to construct a confidence collection for the population identification region. The confidence collection is given by the collection of all sets that, when specified as null hypothesis for the true value of the population identification region, cannot be rejected by our test. Its main property is that (asymptotically) the population identification region is one of its elements with a prespecified confidence level  $(1 - \alpha)$ .

## A Appendix: Lemmas and Proofs

We first define the notation that we use throughout the appendix. Let  $\langle \cdot, \cdot \rangle$  denote the inner product in  $\mathfrak{R}^d$ . In this appendix, we use the capital Greek letters  $\Pi$  and  $\Sigma$  to denote matrices, and we use other capital Greek letters to denote sets of parameters. It will be obvious from the context whether a capital Greek letter refers to a matrix or to a set of parameters. For a finite  $d \times d$  matrix  $\Pi$  and a set  $A \subset \mathfrak{R}^d$  let  $\Pi A = \{\mathbf{r} \in \mathfrak{R}^d : \mathbf{r} = \Pi \mathbf{a}, \mathbf{a} \in A\}$ , and observe that  $s(\mathbf{p}, \Pi A) = s(\Pi' \mathbf{p}, A)$ , where  $\Pi'$  denotes the transposed matrix. We denote the matrix norm induced by the euclidean vector norm (i.e., the 2-norm) as  $\|\Pi\| = \max_{\|\mathbf{p}\|=1} \|\Pi \mathbf{p}\| = \sqrt{\lambda_{\max}}$ , where  $\lambda_{\max}$  is the largest eigenvalue of  $\Pi' \Pi$ ; observe that the eigenvalues of a given matrix are a continuous function of the elements of that matrix (Meyer (2000), Chapter 5). This matrix norm is compatible with its underlying vector norm (i.e. the euclidean norm), so that  $\|\Pi \mathbf{a}\| \leq \|\Pi\| \|\mathbf{a}\|$ .

### A.1 Proofs for Section 3

#### A.1.1 Proof of Proposition 3.1

**Proof.** To establish the asymptotic validity of this procedure, observe that by Theorem 2.4 in Giné & Zinn (1990) for each  $\mathbf{p} \in \mathbb{S}$ ,

$$\sqrt{n} \left[ \frac{1}{n} \sum_{i=1}^n s(\mathbf{p}, \text{co}F_i^*) - \frac{1}{n} \sum_{i=1}^n s(\mathbf{p}, \text{co}F_i) \right] \xrightarrow{d} z(\mathbf{p}).$$

The result then follows by the continuous mapping theorem using standard arguments (e.g., Politis, Romano & Wolf (1999) Chapter 1).

For the processes considered in this paper, it follows from Theorem 1 in Tsirel'son (1975) and from the Corollary in Lifshits (1982) that  $\text{Var}(z(\mathbf{p})) > 0$  for each  $\mathbf{p} \in \mathbb{S}$  is a sufficient condition for the law of  $\|z\|_{\mathbb{C}(\mathbb{S})}$  to be absolutely continuous with respect to Lebesgue measure on  $\mathfrak{R}$ . The result then follows by standard arguments. ■

#### A.1.2 Proof of Theorem 3.2

**Proof.** By triangle inequality,

$$H(\Psi_A, \Psi_0) \leq H\left(\frac{1}{n} \bigoplus_{i=1}^n \text{co}F_i, \Psi_0\right) + H\left(\frac{1}{n} \bigoplus_{i=1}^n \text{co}F_i, \Psi_A\right)$$

Hence,

$$\begin{aligned} \lim_{n \rightarrow \infty} \Pr \left\{ \sqrt{n} H\left(\frac{1}{n} \bigoplus_{i=1}^n \text{co}F_i, \Psi_0\right) > \hat{c}_{\alpha n} \right\} &\geq \lim_{n \rightarrow \infty} \Pr \left\{ \sqrt{n} H(\Psi_A, \Psi_0) - \sqrt{n} H\left(\frac{1}{n} \bigoplus_{i=1}^n \text{co}F_i, \Psi_A\right) > \hat{c}_{\alpha n} \right\} \\ &= \lim_{n \rightarrow \infty} \Pr \left\{ H(\Psi_A, \Psi_0) > H\left(\frac{1}{n} \bigoplus_{i=1}^n \text{co}F_i, \Psi_A\right) + \frac{c_\alpha + o_p(1)}{\sqrt{n}} \right\} \\ &= \Pr \{ H(\Psi_A, \Psi_0) > 0 \} = 1 \end{aligned}$$

where the second to last equality follows from Theorem 2.2 and because  $\hat{c}_{\alpha n} \xrightarrow{P} c_{\alpha} < \infty$  (see equation (3.4)), and the last equality follows because  $H(\Psi_A, \Psi_0)$  is a positive constant. ■

### A.1.3 Proof of Theorem 3.3

**Proof.** To simplify the notation, let  $\bar{F}_n = \frac{1}{n} \bigoplus_{i=1}^n \text{co}F_i$ . By Hörmander's embedding theorem,

$$\sqrt{n}H(\bar{F}_n, \Psi_0) = \sqrt{n} \sup_{\mathbf{p} \in \mathbb{S}} |s(\mathbf{p}, \bar{F}_n) - s(\mathbf{p}, \Psi_0)|.$$

Let

$$\begin{aligned} h(\cdot) &= s(\cdot, F) - s(\cdot, \Psi_0) = s(\cdot, F) - s\left(\cdot, \Psi_{A_n} \oplus \frac{1}{\sqrt{n}} \Delta_1\right) + s\left(\cdot, \Psi_{A_n} \oplus \frac{1}{\sqrt{n}} \Delta_1\right) - s(\cdot, \Psi_0) \\ &= s(\cdot, F) - s(\cdot, \Psi_{A_n}) - s\left(\cdot, \frac{1}{\sqrt{n}} \Delta_1\right) + s\left(\cdot, \Psi_0 \oplus \frac{1}{\sqrt{n}} \Delta_2\right) - s(\cdot, \Psi_0) \\ &= s(\cdot, F) - s(\cdot, \Psi_{A_n}) + s\left(\cdot, \frac{1}{\sqrt{n}} \Delta_2\right) - s\left(\cdot, \frac{1}{\sqrt{n}} \Delta_1\right). \end{aligned}$$

Similarly,

$$h_k(\cdot) = s(\cdot, F_k) - s(\cdot, \Psi_0) = s(\cdot, F_k) - s(\cdot, \Psi_{A_n}) + s\left(\cdot, \frac{1}{\sqrt{n}} \Delta_2\right) - s\left(\cdot, \frac{1}{\sqrt{n}} \Delta_1\right).$$

Hence under the local alternative  $h, h_1, h_2, \dots$  are  $\mathbb{C}(\mathbb{S})$ -valued i.i.d. random variables with

$$E[h(\mathbf{p})] = \frac{1}{\sqrt{n}} s(\mathbf{p}, \Delta_2) - \frac{1}{\sqrt{n}} s(\mathbf{p}, \Delta_1).$$

The result then follows from Proposition 3.1.9 in Li et al. (2002). ■

### A.1.4 Proof of Theorem 3.4

**Proof.** We first show that  $\mathfrak{S}_n \subset \mathcal{U}_n$ . This follows because  $\mathfrak{S}_n$  is a convex compact set (the Minkowski sum of convex sets is convex), and because

$$\begin{aligned} \sqrt{n}H\left(\frac{1}{n} \bigoplus_{i=1}^n \text{co}F_i, \mathfrak{S}_n\right) &= \sqrt{n}H\left(\frac{1}{n} \bigoplus_{i=1}^n \text{co}F_i, \frac{1}{n} \bigoplus_{i=1}^n \text{co}F_i \oplus B_{\hat{c}_{\alpha n}}\right) \\ &= \sqrt{n}H(\{0\}, B_{\hat{c}_{\alpha n}}) = \hat{c}_{\alpha n} \end{aligned}$$

where the last equality follows from the definition of  $B_{\hat{c}_{\alpha n}}$ . Hence  $\mathfrak{S}_n \subset CC^{\mathbb{E}[F]}$ . We now show that  $\mathcal{U}_n \subset \mathfrak{S}_n$ . Take any  $\tilde{\mathbf{f}} \in \mathcal{U}_n$ . Then by definition  $\exists \tilde{\Psi} \in CC^{\mathbb{E}[F]}$  such that  $\tilde{\mathbf{f}} \in \tilde{\Psi}$  and  $\sqrt{n}H\left(\frac{1}{n} \bigoplus_{i=1}^n \text{co}F_i, \tilde{\Psi}\right) \leq \hat{c}_{\alpha n}$ . Choose  $\bar{\mathbf{f}} \in \bar{F}_n = \frac{1}{n} \bigoplus_{i=1}^n \text{co}F_i$  such that

$$\bar{\mathbf{f}} = \arg \inf_{\mathbf{f} \in \bar{F}_n} \|\tilde{\mathbf{f}} - \mathbf{f}\|$$

Let  $\tilde{\mathbf{b}} = \tilde{\mathbf{f}} - \bar{\mathbf{f}}$ ; then by construction  $\tilde{\mathbf{f}} = \bar{\mathbf{f}} + \tilde{\mathbf{b}}$ , and  $\tilde{\mathbf{b}} \in B_{\hat{c}_{\alpha n}}$  because

$$\|\tilde{\mathbf{b}}\| = \|\tilde{\mathbf{f}} - \bar{\mathbf{f}}\| = \inf_{\mathbf{f} \in \bar{F}_n} \|\tilde{\mathbf{f}} - \mathbf{f}\| \leq \sup_{\mathbf{g} \in \tilde{\Psi}} \inf_{\mathbf{f} \in \bar{F}_n} \|\mathbf{g} - \mathbf{f}\| \leq \frac{\hat{c}_{\alpha n}}{\sqrt{n}},$$

where the last inequality follows because  $\sqrt{n}H\left(\frac{1}{n} \bigoplus_{i=1}^n \text{co}F_i, \tilde{\Psi}\right) \leq \hat{c}_{\alpha n}$ . Hence  $\tilde{\mathbf{f}} \in \mathfrak{S}_n$ . ■

## A.2 Proofs for Section 4

### A.2.1 Proof of Theorem 4.1

This result can be easily proved by using the following Lemmas.

**Lemma A.1** *Let  $Y = [y_L, y_U]$ , where  $y_L, y_U$  are real valued random variables such that  $\Pr(y_L \leq y_U) = 1$ . Then  $Y$  is a compact valued non-empty SVRV.*

**Proof.** Theorem 1.2.5 in Molchanov (2005) implies that  $Y$  is an SVRV iff  $s(\mathbf{p}, Y)$  is a random variable for each  $\mathbf{p} \in \{-1, 1\}$ . Since  $y_L(\omega) \leq y_U(\omega) \mu - a.s.$ ,

$$s(p, Y(\omega)) = \max\{py_L(\omega), py_U(\omega)\} = \begin{cases} y_U(\omega) & \text{if } p = 1, \\ -y_L(\omega) & \text{if } p = -1. \end{cases}$$

From the fact that  $y_L$  and  $y_U$  are random variables, also  $s(\mathbf{p}, Y)$  is a random variable and the claim follows. The fact that  $Y$  takes almost surely compact values follows because  $y_L$  and  $y_U$  are real valued random variables. The fact that  $Y$  is non-empty follows because  $\Pr\{y_L \leq y_U\} = 1$ . ■

**Lemma A.2** *Given an i.i.d. sequence  $\{y_{iL}, y_{iU}\}_{i=1}^n$ , where  $y_{iL}, y_{iU}$  are real valued random variables such that  $\Pr(y_{iL} \leq y_{iU}) = 1$  for each  $i$ , let  $Y_i = [y_{iL}, y_{iU}]$ . Then  $\{Y_i : i \in \mathbb{N}\}$  are i.i.d SVRVs.*

**Proof.**

(i)  $\{Y_i\}_{i=1}^n$  are identically distributed. For  $A \in \mathcal{B}(K(\mathfrak{R}))$ ,  $\mu_{Y_i}(A) = \mu(Y_i^{-1}(A)) = \Pr\{[y_{iL}, y_{iU}] \cap A \neq \emptyset\}$ .

Since  $\{(y_{iL}, y_{iU})\}_{i=1}^n$  are i.i.d.,  $\mu_{Y_i} = \mu_{Y_j}$  for all  $i$  and  $j$ .

(ii)  $\{Y_i\}_{i=1}^n$  are independent. For all  $C_1, \dots, C_n$  in  $K_k(\mathfrak{R})$ ,

$$\begin{aligned} \mu(Y_1^{-1}(C_1), \dots, Y_n^{-1}(C_n)) &= \mu(Y_1 \cap C_1 \neq \emptyset, \dots, Y_n \cap C_n \neq \emptyset) \\ &= \Pr\{[y_{1L}, y_{1U}] \cap C_1 \neq \emptyset, \dots, [y_{nL}, y_{nU}] \cap C_n \neq \emptyset\} \\ &= \prod_{i=1}^n \Pr\{[y_{iL}, y_{iU}] \cap C_i \neq \emptyset\} \end{aligned}$$

where the last equality comes from the fact that  $\{(y_{iL}, y_{iU})\}_{i=1}^n$  are independent. The result then follows by Proposition 1.1.19 in Molchanov (2005). ■

**Proof. Of Theorem 4.1.**

Lemmas A.1 and A.2 show that  $\{Y, Y_i : i \in \mathbb{N}\}$  are i.i.d. compact valued non-empty SVRVs. To verify that the SVRVs are integrably bounded, observe that

$$\int \|Y\|_H d\mu \leq E(|y_L|) + E(|y_U|) < \infty,$$

where the last inequality follows from the assumptions in Theorem 4.1. All the conditions of Theorem 2.2 are therefore satisfied. ■



### A.2.2 Proof of Theorem 4.2

**Proof.** Let  $f(\omega) \in Y(\omega) = [y_L(\omega), y_U(\omega)]$   $\mu - a.s.$  Then  $\int_{\Omega} f d\mu \in [E(y_L), E(y_U)]$ . This implies,  $\mathbb{E}[Y] \subset [E(y_L), E(y_U)]$ . Conversely, since  $[E(y_L), E(y_U)]$  is a convex set, any  $b \in [E(y_L), E(y_U)]$  can be written as  $b = \alpha E(y_L) + (1 - \alpha) E(y_U)$  for some  $\alpha \in [0, 1]$ . Define

$$f_b(\omega) = \alpha y_L(\omega) + (1 - \alpha) y_U(\omega).$$

Then  $\int_{\Omega} f_b(\omega) d\mu = b$  and  $f_b \in \mathcal{S}(Y)$ , and therefore  $[E(y_L), E(y_U)] \subset \mathbb{E}[Y]$ . ■

### A.2.3 Proof of Theorem 4.3

**Proof.** Our random sets are *i.i.d.*, compact valued (i.e. in  $K_k(\mathfrak{R})$ ) and non-empty by Lemmas A.1 and A.2.

We know that  $\int_{\Omega} \|Y(\omega)\|_H^2 d\mu \leq \int_{\Omega} y_L(\omega)^2 d\mu + \int_{\Omega} y_U(\omega)^2 d\mu$ , which is finite by assumption. Therefore, all the conditions in Theorem 2.3 are satisfied. Then, by Theorem 2.3,

(i)  $E(z_{-1}) = E(z_1) = 0$ ,

(ii)  $E(z_p \cdot z_q) = E[s(p, Y) s(q, Y)] - E[s(p, Y)] \cdot E[s(q, Y)]$  where  $p, q \in \{-1, 1\}$ .

By the definition of the support function,

$$s(p, Y)(\omega) = \begin{cases} y_U(\omega) & \text{if } p = 1, \\ -y_L(\omega) & \text{if } p = -1. \end{cases}$$

The covariance kernel in the theorem follows by simple algebra. ■

## A.3 Proofs for Section 5

In the text we defined  $\Theta$ ,  $\Theta^M$ , and the relevant set valued random variables for the simple case  $x \in \mathfrak{R}$ . Here we extend these definitions for the case of a column vector  $\mathbf{x} \in \mathfrak{R}^d$ , and then prove all the results for this more general case:

$$(A.1) \quad G(\omega) = \left\{ \left( \begin{bmatrix} y(\omega) \\ \mathbf{x}(\omega) \cdot y(\omega) \end{bmatrix} : y(\omega) \in Y(\omega) \right) \right\},$$

$$(A.2) \quad G_i(\omega) = \left\{ \left( \begin{bmatrix} y_i(\omega) \\ \mathbf{x}_i(\omega) \cdot y_i(\omega) \end{bmatrix} : y_i(\omega) \in Y_i(\omega) \right) \right\},$$

$$(A.3) \quad \begin{aligned} \Theta &= \left[ \begin{array}{cc} 1 & E(\mathbf{x}') \\ E(\mathbf{x}) & E(\mathbf{x}\mathbf{x}') \end{array} \right]^{-1} \mathbb{E}[G] \\ &= \left\{ \boldsymbol{\theta} \in \mathfrak{R}^{d+1}: \boldsymbol{\theta} = \left[ \begin{array}{cc} 1 & E(\mathbf{x}') \\ E(\mathbf{x}) & E(\mathbf{x}\mathbf{x}') \end{array} \right]^{-1} \left[ \begin{array}{c} E(y) \\ E(\mathbf{x}y) \end{array} \right], \left[ \begin{array}{c} y \\ \mathbf{x}y \end{array} \right] \in \mathcal{S}^1(G) \right\}, \end{aligned}$$

$$\Sigma \equiv \left[ \begin{array}{cc} 1 & E(\mathbf{x}') \\ E(\mathbf{x}) & E(\mathbf{x}\mathbf{x}') \end{array} \right], \quad \hat{\Sigma}_n \equiv \left[ \begin{array}{cc} 1 & \bar{\mathbf{x}}' \\ \bar{\mathbf{x}} & \bar{\mathbf{x}\mathbf{x}'} \end{array} \right].$$

and

$$(A.4) \quad \Theta^M = \left\{ \boldsymbol{\theta}^M \in \mathfrak{R}^{d+1}: \boldsymbol{\theta}^M = \arg \min \int (y - \theta_1 - \boldsymbol{\theta}'_2 \mathbf{x})^2 d\eta, \eta \in \mathbf{P}_{y\mathbf{x}} \right\},$$

where

$$(A.5) \quad \mathbf{P}_{y\mathbf{x}} = \left\{ \eta : \Pr(y_L \leq t, \mathbf{x} \leq \mathbf{x}_0) \geq \eta((-\infty, t], (-\infty, \mathbf{x}_0]) \geq \Pr(y_U \leq t, \mathbf{x} \leq \mathbf{x}_0) \quad \forall t \in \mathfrak{R}, \forall \mathbf{x}_0 \in \mathfrak{R}^d, \right. \\ \left. \eta((-\infty, +\infty), (-\infty, \mathbf{x}_0]) = \Pr(\mathbf{x} \leq \mathbf{x}_0) \quad \forall \mathbf{x}_0 \in \mathfrak{R}^d \right\}.$$

and the notation  $\mathbf{x} \leq \mathbf{x}_0$  indicates that each element of  $\mathbf{x}$  is less than or equal to the corresponding element of  $\mathbf{x}_0$ .

**Lemma A.3**  *$G$  and  $\{G_i\}_{i=1}^n$  as defined in equations (A.1)-(A.2) are non-empty compact valued SVRVs.*

**Proof.** By Lemma A.1,  $Y$  is a non-empty, compact valued SVRV. By Theorem 1.2.7 in Li et al. (2002) there is a countable selection  $\{f_i\}_{i=1}^\infty$  such that  $f_i(\omega) \in Y(\omega)$ ,  $f_i \in \mathcal{S}(Y)$  for all  $i$  and  $\text{cl}\{f_i(\omega)\} = Y(\omega)$  for all  $\omega \in \Omega$ . Therefore, it is easy to see that for each  $\lambda \in \mathfrak{R}$ ,  $\lambda Y$  is an SVRV and for each set  $A$ ,  $Y \oplus A$  is an SVRV. Observe that  $\mathbf{x}$  is a random vector defined on the same probability space as  $Y$ . Then  $(\mathbf{x} \cdot Y)(\omega) = \{\mathbf{x}(\omega) f(\omega) : f(\omega) \in Y(\omega)\} = \text{cl}\{\mathbf{x}(\omega) \cdot f_i(\omega)\}$  for each  $\omega \in \Omega$ . Hence,  $\left\{ \begin{matrix} f_i \\ \mathbf{x} \cdot f_i \end{matrix} \right\}$  spans the set  $G$  as defined in (A.1) and therefore  $G$  is an SVRV by Theorem 1.2.7 in Li et al. (2002). The fact that  $G$  is non-empty and compact valued follows from the same arguments as in Lemma A.1. ■

**Lemma A.4** *Under Assumption 5.1,  $\{G_i\}_{i=1}^n$  as defined in equation (A.2) is a sequence of i.i.d. SVRVs.*

**Proof.** (i)  $\{G_i\}_{i=1}^n$  are identically distributed. For  $A \in \mathcal{B}(K(\mathfrak{R}^{d+1}))$ ,

$$\mu_{G_i}(A) = \mu(G_i^{-1}(A)) = \mu \left( \omega : \exists y_i(\omega) \in Y_i(\omega), (a_1, \dots, a_{d+1}) \in A \text{ s.t. } \begin{bmatrix} y_i^{-1}(a_1) = \omega \\ (x_{1i}y_i)^{-1}(a_2) = \omega \\ \vdots \\ (x_{di}y_i)^{-1}(a_{d+1}) = \omega \end{bmatrix} \right).$$

Since  $\{(Y_i, \mathbf{x}_i)\}_{i=1}^n$  are identically distributed,  $\mu_{G_i} = \mu_{G_j}$  for all  $i$  and  $j$ .

(ii)  $\{G_i\}_{i=1}^n$  are independent. For all sets  $C_1, \dots, C_n$  in  $K_k(\mathfrak{R}^{d+1})$ ,

$$\begin{aligned} & \mu(G_1^{-1}(C_1), \dots, G_n^{-1}(C_n)) = \mu(G_1 \cap C_1 \neq \emptyset, \dots, G_n \cap C_n \neq \emptyset) \\ &= \Pr \left\{ \left( \begin{bmatrix} y_1 \\ \mathbf{x}_1 y_1 \end{bmatrix} : y_1 \in [y_{1L}, y_{1U}] \right) \cap C_1 \neq \emptyset, \dots, \left( \begin{bmatrix} y_n \\ \mathbf{x}_n y_n \end{bmatrix} : y_n \in [y_{nL}, y_{nU}] \right) \cap C_n \neq \emptyset \right\} \\ &= \prod_{i=1}^n \Pr \left\{ \left( \begin{bmatrix} y_i \\ \mathbf{x}_i y_i \end{bmatrix} : y_i \in [y_{iL}, y_{iU}] \right) \cap C_i \right\}, \end{aligned}$$

where the last equality comes from the fact that  $\{(y_{iL}, y_{iU}, \mathbf{x}_i)\}_{i=1}^n$  are independent. The result then follows by Proposition 1.1.19 in Molchanov (2005). ■

### A.3.1 Proof of Proposition 5.1

**Proof.** With  $\mathbf{x}$  a column vector in  $\mathfrak{R}^d$ , the sets  $\Theta, \Theta^M$  are non-empty convex and compact subsets of  $\mathfrak{R}^{d+1}$ . For any vector  $\boldsymbol{\theta} \in \Theta$  or  $\boldsymbol{\theta}^M \in \Theta^M$ , let the first entry of such vector correspond to the constant term and be denoted, respectively, by  $\theta_1$  and  $\theta_1^M$ , and the remaining  $d$  entries be denoted, respectively, by  $\boldsymbol{\theta}_2$  and  $\boldsymbol{\theta}_2^M$ . We start by showing that  $\Theta^M \subset \Theta$ . Pick a vector  $\left[ \theta_1^M \quad (\boldsymbol{\theta}_2^M)' \right]' \in \Theta^M$ . Then there exists a distribution for  $(y, \mathbf{x})$  with  $\mathbf{x}$ -marginal  $P(\mathbf{x})$  denoted  $\eta_0^M$  such that  $\left[ \theta_1^M \quad (\boldsymbol{\theta}_2^M)' \right]'$  is a minimizer of the problem in (A.4)-(A.5). Let  $(y^M, \mathbf{x})$  be a random vector with distribution  $\eta_0^M$ . It follows that  $y^M(\omega) \in Y(\omega)$   $\mu - a.s.$ , and therefore  $[y^M(\omega) \quad \mathbf{x}'y^M(\omega)]' \in G(\omega)$   $\mu - a.s.$ , from which  $\left[ \theta_1^M \quad (\boldsymbol{\theta}_2^M)' \right]' \in \Theta$ .

Conversely, pick a vector  $[\theta_1 \quad (\boldsymbol{\theta}_2)']' \in \Theta$ . Then there exists a random vector  $[y \quad y\mathbf{x}]' \in G$  such that

$$\begin{bmatrix} \theta_1 \\ \boldsymbol{\theta}_2 \end{bmatrix} = \begin{bmatrix} 1 & E(\mathbf{x}') \\ E(\mathbf{x}) & E(\mathbf{x}\mathbf{x}') \end{bmatrix}^{-1} \begin{bmatrix} E(y) \\ E(\mathbf{x}y) \end{bmatrix}.$$

We show that the corresponding vector  $(y, \mathbf{x})$  has an admissible probability distribution  $\eta \in \mathbf{P}_{y\mathbf{x}}$ . Because  $(y, \mathbf{x})$  is a selection from  $(Y, \mathbf{X})$ , it follows from Theorem 2.1 in Artstein (1983) that for any  $t \in \mathfrak{R}$ ,  $\mathbf{x}_0 \in \mathfrak{R}^d$

$$\begin{aligned} \eta(y \leq t, \mathbf{x} \leq \mathbf{x}_0) &\leq \mu(Y \cap (-\infty, t] \neq \emptyset, \mathbf{x} \leq \mathbf{x}_0) = \Pr(y_L \leq t, \mathbf{x} \leq \mathbf{x}_0), \\ \eta(y \geq t, \mathbf{x} \leq \mathbf{x}_0) &\leq \mu(Y \cap [t, +\infty) \neq \emptyset, \mathbf{x} \leq \mathbf{x}_0) = \Pr(y_U \geq t, \mathbf{x} \leq \mathbf{x}_0). \end{aligned}$$

The fact that the marginal of  $\eta(y, \mathbf{x})$  is  $P(\mathbf{x})$  follows easily. Hence  $[\theta_1 \quad (\boldsymbol{\theta}_2)']' \in \Theta^M$ . ■

### A.3.2 Proof of Theorem 5.2

Theorem 5.2 can be easily proved by using the following Lemma.

**Lemma A.5** *Let  $\{F, F_i : i \in \mathbb{N}\}$  be i.i.d. non-empty compact valued, integrably bounded SVRVs. Let  $\{\Pi_i : i \in \mathbb{N}\}$  be i.i.d. random  $d \times d$  matrices on  $(\Omega, \mathcal{A}, \mu)$  such that  $\Pi_i \xrightarrow{a.s.} \Pi$  element-by-element, where  $\Pi$  is a  $d \times d$  non-stochastic matrix with finite elements and  $\Pi_i$  has finite elements w.p.1. Then*

$$H \left( \Pi_n \frac{1}{n} \bigoplus_{k=1}^n F_k, \Pi \mathbb{E}[F] \right) \xrightarrow{a.s.} 0.$$

**Proof.** This is a version of Slutsky's theorem for  $d$ -dimensional SVRVs. We interpret vectors in  $\mathfrak{R}^d$  as being column vectors. For simplicity write  $\bar{F}_n = \frac{1}{n} \bigoplus_{k=1}^n F_k$ . Then,

$$\begin{aligned}
& H(\Pi_n \bar{F}_n, \Pi \mathbb{E}[F]) \\
&= \max \left\{ \sup_{\mathbf{f}_n \in \bar{F}_n} \inf_{\mathbf{f} \in \mathbb{E}[F]} \|\Pi_n \mathbf{f}_n - \Pi \mathbf{f}\|, \sup_{\mathbf{f} \in \mathbb{E}[F]} \inf_{\mathbf{f}_n \in \bar{F}_n} \|\Pi_n \mathbf{f}_n - \Pi \mathbf{f}\| \right\} \\
&\leq \max \left\{ \sup_{\mathbf{f}_n \in \bar{F}_n} \inf_{\mathbf{f} \in \mathbb{E}[F]} (\|\Pi_n \mathbf{f}_n - \Pi_n \mathbf{f}\| + \|\Pi_n \mathbf{f} - \Pi \mathbf{f}\|), \sup_{\mathbf{f} \in \mathbb{E}[F]} \inf_{\mathbf{f}_n \in \bar{F}_n} (\|\Pi_n \mathbf{f}_n - \Pi_n \mathbf{f}\| + \|\Pi_n \mathbf{f} - \Pi \mathbf{f}\|) \right\} \\
&\leq \max \left\{ \sup_{\mathbf{f}_n \in \bar{F}_n} \inf_{\mathbf{f} \in \mathbb{E}[F]} (\|\Pi_n\| \|\mathbf{f}_n - \mathbf{f}\| + \|\mathbf{f}\| \|\Pi_n - \Pi\|), \sup_{\mathbf{f} \in \mathbb{E}[F]} \inf_{\mathbf{f}_n \in \bar{F}_n} (\|\Pi_n\| \|\mathbf{f}_n - \mathbf{f}\| + \|\mathbf{f}\| \|\Pi_n - \Pi\|) \right\} \\
&\leq \|\Pi_n\| \max \left\{ \sup_{\mathbf{f}_n \in \bar{F}_n} \inf_{\mathbf{f} \in \mathbb{E}[F]} \|\mathbf{f}_n - \mathbf{f}\|, \sup_{\mathbf{f} \in \mathbb{E}[F]} \inf_{\mathbf{f}_n \in \bar{F}_n} \|\mathbf{f}_n - \mathbf{f}\| \right\} + \|\Pi_n - \Pi\| \sup_{\mathbf{f} \in \mathbb{E}[F]} \|\mathbf{f}\| = o_{a.s.}(1),
\end{aligned}$$

because  $\|\Pi_n\| \xrightarrow{a.s.} \|\Pi\| < \infty$  by Slutsky's theorem,  $\max \left\{ \sup_{\mathbf{f}_n \in \bar{F}_n} \inf_{\mathbf{f} \in \mathbb{E}[F]} \|\mathbf{f}_n - \mathbf{f}\|, \sup_{\mathbf{f} \in \mathbb{E}[F]} \inf_{\mathbf{f}_n \in \bar{F}_n} \|\mathbf{f}_n - \mathbf{f}\| \right\} \xrightarrow{a.s.} 0$  by the LLN for SVRVs,  $\|\Pi_n - \Pi\| \xrightarrow{a.s.} 0$  by the continuous mapping theorem, and  $\sup_{\mathbf{f} \in \mathbb{E}[F]} \|\mathbf{f}\| < \infty$  because all selections of an integrably bounded SVRV are integrable. ■

### Proof. Of Theorem 5.2.

The result follows directly from Lemma A.5, replacing  $\hat{\Sigma}_n^{-1} = \Pi_n$ ,  $\Sigma^{-1} = \Pi$ ,  $F_i = G_i$ , and  $F = G$ . This is because Lemmas A.3 and A.4 show that  $\{G, G_i : i \in \mathbb{N}\}$  are *i.i.d.* non-empty compact valued SVRVs. To verify that these SVRVs are integrably bounded, observe that

$$\int \|G\|_H d\mu \leq E(|y_L|) + E(|y_U|) + \sum_{k=1}^d [E(|x_k y_L|) + E(|x_k y_U|)] < \infty,$$

where the last inequality follows from Assumption 5.2. If in addition Assumption 5.4 is satisfied, then

$$\max \left\{ \sup_{\mathbf{f}_n \in \bar{F}_n} \inf_{\mathbf{f} \in \mathbb{E}[F]} \|\mathbf{f}_n - \mathbf{f}\|, \sup_{\mathbf{f} \in \mathbb{E}[F]} \inf_{\mathbf{f}_n \in \bar{F}_n} \|\mathbf{f}_n - \mathbf{f}\| \right\} = O_p\left(\frac{1}{\sqrt{n}}\right)$$

by the CLT for SVRVs, because

$$\int \|G\|_H^2 d\mu \leq E(|y_L|^2) + E(|y_U|^2) + \sum_{k=1}^d [E(|x_k y_L|^2) + E(|x_k y_U|^2)] < \infty,$$

and  $\|\Pi_n - \Pi\| = O_p\left(\frac{1}{\sqrt{n}}\right)$  by the continuous mapping theorem. ■

### A.3.3 Proof of Theorem 5.3

Before giving this proof, we introduce a Definition and two Lemmas that are used in the last step of the proof of Theorem 5.3.

**Definition 10** Let  $\mathbf{w}$  be a random vector in  $\mathfrak{R}^d$ . Its zonoid,  $\Lambda_{\mathbf{w}}$ , is the Aumann expectation of the random segment in  $\mathfrak{R}^d$  with the endpoints being the origin ( $\mathbf{0}$ ) and the  $d$ -dimensional vector  $\mathbf{w}$ . The lift zonoid,<sup>10</sup>  $\tilde{\Lambda}_{\mathbf{w}}$ , of  $\mathbf{w}$  is the Aumann expectation of the random segment in  $\mathfrak{R}^{d+1}$  with the endpoints being the origin and the  $(d+1)$ -dimensional vector  $(1, \mathbf{w}')$ .

In the following Lemma, for  $p \in [1, \infty)$ , denote by  $\mathbf{L}_{\mathcal{A}}^p = \mathbf{L}^p(\Omega, \mathcal{A}, \mathfrak{R}^d)$  the space of  $\mathcal{A}$ -measurable random variables with values in  $\mathfrak{R}^d$  such that the  $\mathbf{L}^p$ -norm

$$\|\xi\|_p = [E(\|\xi\|^p)]^{\frac{1}{p}}$$

is finite, and for an SVRV  $F$  defined on  $(\Omega, \mathcal{A})$  denote by  $\mathcal{S}_{\mathcal{A}}^p(F)$  the family of all selections of  $F$  from  $\mathbf{L}_{\mathcal{A}}^p$ , so that

$$\mathcal{S}_{\mathcal{A}}^p(F) = \mathcal{S}(F) \cap \mathbf{L}_{\mathcal{A}}^p.$$

**Lemma A.6** Let  $A$  be an integrably bounded SVRV defined on  $(\Omega, \mathcal{A}, \mu)$ . For each  $\sigma$ -algebra  $\mathcal{A}_0 \subset \mathcal{A}$  there exists a unique integrable  $\mathcal{A}_0$ -measurable SVRV  $B$ , denoted by  $B = \mathbb{E}[A | \mathcal{A}_0]$  and called the conditional Aumann expectation of  $A$ , such that

$$\mathcal{S}_{\mathcal{A}_0}^1(B) = cl\{E[\mathbf{a} | \mathcal{A}_0] : \mathbf{a} \in \mathcal{S}_{\mathcal{A}}^1(A)\},$$

where the closure is taken with respect to the norm in  $\mathbf{L}_{\mathcal{A}_0}^1$ . Since  $A$  is integrably bounded, so is  $B$ .

**Proof.** Molchanov (2005), Theorem 2.1.46 ■

Using the definition of a lift zonoid and the properties of conditional expectations of SVRV, we prove the following result:

**Lemma A.7** Define  $\Theta$  as in (5.1). Under Assumptions 5.2 and 5.5,  $\Theta$  is a strictly convex set.

**Proof.** The random set  $G$  can be represented as the convex hull of points with coordinates  $[y \quad \mathbf{x}'y]'$  where  $y(\omega) \in Y(\omega)$ . Define

$$\begin{aligned} \zeta &= y_U - y_L \\ \boldsymbol{\eta} &= \begin{bmatrix} y_L \\ \mathbf{x} \cdot y_L \end{bmatrix} \end{aligned}$$

Then one can represent  $G$  as

$$\begin{aligned} G &= \boldsymbol{\eta} \oplus \tilde{L} \\ \tilde{L} &= co\left\{\mathbf{0}, [\zeta \quad \mathbf{x}' \cdot \zeta]'\right\} \end{aligned}$$

---

<sup>10</sup>The *lift zonoid* is a mathematical concept introduced by Bolker (1969) and used in Economics by Koshvovoy & Mosler (1996) to extend the notions of the Lorenz curve and the Lorenz order to several attributes (for a monograph on the lift zonoid and its applications in Economics see Mosler (2002)).

Hence, by Theorem 2.1.17 in Molchanov (2005),  $\mathbb{E}[G]$  is the sum of the expectation of  $\boldsymbol{\eta}$  (a random vector) and the Aumann expectation of the random set  $\tilde{L}$  being the convex hull of the origin  $(\mathbf{0})$  and  $[\zeta \quad \mathbf{x}' \cdot \zeta]'$  =  $\zeta [1 \quad \mathbf{x}']'$ .

If  $\zeta$  were degenerate and identically equal to 1, then the Aumann expectation  $\mathbb{E}[\tilde{L}]$  would be the lift zonoid of  $\mathbf{x}$ ,  $\tilde{\Lambda}_{\mathbf{x}}$ . In our case  $\zeta$  is a nondegenerate random variable. By Theorem 2.1.47(ii)-(iii) in Molchanov (2005) we can use the properties of the conditional expectation of SVRVs to get

$$\mathbb{E}[G|\zeta] = E(\boldsymbol{\eta}|\zeta) \oplus \mathbb{E}[\tilde{L}|\zeta] = E(\boldsymbol{\eta}|\zeta) \oplus \zeta \mathbb{E}[co(\mathbf{0}, [1 \quad \mathbf{x}']')|\zeta].$$

$\mathbb{E}[co(\mathbf{0}, [1 \quad \mathbf{x}']')|\zeta]$  is a lift zonoid, and therefore  $\mathbb{E}[G|\zeta]$  is a (rescaled by  $\zeta$ ) lift zonoid shifted by the vector  $E(\boldsymbol{\eta}|\zeta)$ . It then follows by Corollary 2.5 in Mosler (2002) that since  $\mathbf{x}$  has an absolutely continuous distribution with respect to Lebesgue measure on  $\mathfrak{R}^d$ ,  $\mathbb{E}[G|\zeta]$  is a strictly convex set  $P(\zeta) - a.s.$  Premultiplying  $G$  by the non-random matrix  $\Sigma^{-1}$ , we obtain that also  $\mathbb{E}[\Sigma^{-1}G|\zeta]$  is a strictly convex set  $P(\zeta) - a.s.$  Therefore by Corollary 1.7.3 in Schneider (1993)  $s(\cdot, \mathbb{E}[\Sigma^{-1}G|\zeta])$  is Fréchet differentiable on  $\mathfrak{R}^{d+1} \setminus \{\mathbf{0}\}$   $P(\zeta) - a.s.$  By Theorem 2.1.47(iv) in Molchanov (2005),  $s(\mathbf{p}, \mathbb{E}[\Sigma^{-1}G|\zeta]) = E[s(\mathbf{p}, \Sigma^{-1}G)|\zeta]$ . Observing that by Corollary 1.7.3 in Schneider (1993) the gradient of  $s(\mathbf{p}, \mathbb{E}[\Sigma^{-1}G|\zeta])$  is given by

$$\nabla s(\mathbf{p}, \mathbb{E}[\Sigma^{-1}G|\zeta]) = \mathbb{E}[\Sigma^{-1}G|\zeta] \cap \left\{ \boldsymbol{\vartheta} \in \mathfrak{R}^{d+1} : \langle \boldsymbol{\vartheta}, \mathbf{p} \rangle = s(\mathbf{p}, \mathbb{E}[\Sigma^{-1}G|\zeta]) \right\},$$

we can apply Theorem 2.1.47(v) in Molchanov (2005) to the sets  $\{\mathbf{0}\}$  and  $\Sigma^{-1}G$  (which is absolutely integrable because by assumption  $G$  is absolutely integrable, see the proof of Theorem 5.2), to obtain

$$\begin{aligned} E \|\nabla s(\mathbf{p}, \mathbb{E}[\Sigma^{-1}G|\zeta])\| &\leq E[\|\mathbb{E}[\Sigma^{-1}G|\zeta]\|_H] = E[H(\mathbb{E}[\Sigma^{-1}G|\zeta], \{\mathbf{0}\})] \\ &\leq E[H(\Sigma^{-1}G, \{\mathbf{0}\})] = E[\|\Sigma^{-1}G\|_H] < \infty. \end{aligned}$$

Hence

$$s(\mathbf{p}, \Theta) = s(\mathbf{p}, \mathbb{E}[\Sigma^{-1}G]) = E[s(\mathbf{p}, \Sigma^{-1}G)] = \int E[s(\mathbf{p}, \Sigma^{-1}G)|\zeta] dP(\zeta)$$

is differentiable at  $\mathbf{p} \in \mathfrak{R}^{d+1} \setminus \{\mathbf{0}\}$ . The final result follows because the support set in each direction  $\mathbf{p} \in \mathbb{S}$  of a set which has a differentiable support function contains only one point (Schneider (1993), Corollary 1.7.3).

■

**Proof. Of Theorem 5.3.** (i) This part of the proof proceeds in steps.

*Step 1. Derivation of the fact that  $\sqrt{n}H(\hat{\Theta}_n, \Theta) \stackrel{A}{\approx} \sqrt{n}H(\Sigma^{-1}\bar{G}_n, \Sigma^{-1}\hat{\Sigma}_n\Theta)$ , where  $\bar{G}_n \equiv \frac{1}{n} \bigoplus_{i=1}^n G_i$ .*

From the definition of  $\hat{\Theta}_n$  and the fact that  $\Theta = \hat{\Sigma}_n^{-1}\hat{\Sigma}_n\Theta$  it follows that

$$H(\hat{\Theta}_n, \Theta) = H(\hat{\Sigma}_n^{-1}\bar{G}_n, \hat{\Sigma}_n^{-1}\hat{\Sigma}_n\Theta).$$

By definition,

$$\begin{aligned}\sqrt{n}H\left(\hat{\Sigma}_n^{-1}\bar{G}_n, \hat{\Sigma}_n^{-1}\hat{\Sigma}_n\Theta\right) &= \sqrt{n}\max\left\{\sup_{\mathbf{g}_n \in \bar{G}_n} \inf_{\boldsymbol{\theta} \in \Theta} \left\|\hat{\Sigma}_n^{-1}\left(\mathbf{g}_n - \hat{\Sigma}_n\boldsymbol{\theta}\right)\right\|, \sup_{\boldsymbol{\theta} \in \Theta} \inf_{\mathbf{g}_n \in \bar{G}_n} \left\|\hat{\Sigma}_n^{-1}\left(\mathbf{g}_n - \hat{\Sigma}_n\boldsymbol{\theta}\right)\right\|\right\}, \\ \sqrt{n}H\left(\Sigma^{-1}\bar{G}_n, \Sigma^{-1}\hat{\Sigma}_n\Theta\right) &= \sqrt{n}\max\left\{\sup_{\mathbf{g}_n \in \bar{G}_n} \inf_{\boldsymbol{\theta} \in \Theta} \left\|\Sigma^{-1}\left(\mathbf{g}_n - \hat{\Sigma}_n\boldsymbol{\theta}\right)\right\|, \sup_{\boldsymbol{\theta} \in \Theta} \inf_{\mathbf{g}_n \in \bar{G}_n} \left\|\Sigma^{-1}\left(\mathbf{g}_n - \hat{\Sigma}_n\boldsymbol{\theta}\right)\right\|\right\}.\end{aligned}$$

Without loss of generality, let

$$\sqrt{n}H\left(\hat{\Sigma}_n^{-1}\bar{G}_n, \hat{\Sigma}_n^{-1}\hat{\Sigma}_n\Theta\right) = \sqrt{n}\sup_{\mathbf{g}_n \in \bar{G}_n} \inf_{\boldsymbol{\theta} \in \Theta} \left\|\hat{\Sigma}_n^{-1}\left(\mathbf{g}_n - \hat{\Sigma}_n\boldsymbol{\theta}\right)\right\|.$$

(A similar argument applies if  $\sqrt{n}H\left(\hat{\Sigma}_n^{-1}\bar{G}_n, \hat{\Sigma}_n^{-1}\hat{\Sigma}_n\Theta\right) = \sqrt{n}\sup_{\boldsymbol{\theta} \in \Theta} \inf_{\mathbf{g}_n \in \bar{G}_n} \left\|\hat{\Sigma}_n^{-1}\left(\mathbf{g}_n - \hat{\Sigma}_n\boldsymbol{\theta}\right)\right\|$ .) Then,

$$\begin{aligned}&\sqrt{n}\sup_{\mathbf{g}_n \in \bar{G}_n} \inf_{\boldsymbol{\theta} \in \Theta} \left\|\hat{\Sigma}_n^{-1}\left(\mathbf{g}_n - \hat{\Sigma}_n\boldsymbol{\theta}\right)\right\| - \sqrt{n}H\left(\Sigma^{-1}\bar{G}_n, \Sigma^{-1}\hat{\Sigma}_n\Theta\right) \\ &\leq \sqrt{n}\sup_{\mathbf{g}_n \in \bar{G}_n} \inf_{\boldsymbol{\theta} \in \Theta} \left\|\hat{\Sigma}_n^{-1}\Sigma\Sigma^{-1}\left(\mathbf{g}_n - \hat{\Sigma}_n\boldsymbol{\theta}\right)\right\| - \sqrt{n}\sup_{\mathbf{g}_n \in \bar{G}_n} \inf_{\boldsymbol{\theta} \in \Theta} \left\|\Sigma^{-1}\left(\mathbf{g}_n - \hat{\Sigma}_n\boldsymbol{\theta}\right)\right\| \\ &\leq \sqrt{n}\sup_{\mathbf{g}_n \in \bar{G}_n} \inf_{\boldsymbol{\theta} \in \Theta} \left\|\hat{\Sigma}_n^{-1}\Sigma\right\| \left\|\Sigma^{-1}\left(\mathbf{g}_n - \hat{\Sigma}_n\boldsymbol{\theta}\right)\right\| - \sqrt{n}\sup_{\mathbf{g}_n \in \bar{G}_n} \inf_{\boldsymbol{\theta} \in \Theta} \left\|\Sigma^{-1}\left(\mathbf{g}_n - \hat{\Sigma}_n\boldsymbol{\theta}\right)\right\| \\ &= \left(\left\|\hat{\Sigma}_n^{-1}\Sigma\right\| - 1\right) \sqrt{n}\sup_{\mathbf{g}_n \in \bar{G}_n} \inf_{\boldsymbol{\theta} \in \Theta} \left\|\Sigma^{-1}\left(\mathbf{g}_n - \hat{\Sigma}_n\boldsymbol{\theta}\right)\right\| = o_p(1) \cdot O_p(1) = o_p(1),\end{aligned}$$

where the last equality follows from the following facts: (1)  $\hat{\Sigma}_n^{-1}\Sigma \xrightarrow{p} I$ , where  $I$  is the identity matrix of appropriate dimensions; (2) the 2-norm of the identity matrix is 1, and the norm is a continuous function; (3) by a similar argument as in the proof of Lemma A.5,  $\sqrt{n}H\left(\Sigma^{-1}\bar{G}_n, \Sigma^{-1}\hat{\Sigma}_n\Theta\right) = O_p(1)$ . Hence,

$$\sqrt{n}H\left(\hat{\Sigma}_n^{-1}\bar{G}_n, \hat{\Sigma}_n^{-1}\hat{\Sigma}_n\Theta\right) - \sqrt{n}H\left(\Sigma^{-1}\bar{G}_n, \Sigma^{-1}\hat{\Sigma}_n\Theta\right) \leq o_p(1).$$

We now need to show that the other direction of the inequality also applies. Without loss of generality, let

$$\sqrt{n}H\left(\Sigma^{-1}\bar{G}_n, \Sigma^{-1}\hat{\Sigma}_n\Theta\right) = \sqrt{n}\sup_{\mathbf{g}_n \in \bar{G}_n} \inf_{\boldsymbol{\theta} \in \Theta} \left\|\Sigma^{-1}\left(\mathbf{g}_n - \hat{\Sigma}_n\boldsymbol{\theta}\right)\right\|.$$

Then,

$$\begin{aligned}&\sqrt{n}\sup_{\mathbf{g}_n \in \bar{G}_n} \inf_{\boldsymbol{\theta} \in \Theta} \left\|\Sigma^{-1}\left(\mathbf{g}_n - \hat{\Sigma}_n\boldsymbol{\theta}\right)\right\| - \sqrt{n}H\left(\hat{\Sigma}_n^{-1}\bar{G}_n, \hat{\Sigma}_n^{-1}\hat{\Sigma}_n\Theta\right) \\ &\leq \sqrt{n}\sup_{\mathbf{g}_n \in \bar{G}_n} \inf_{\boldsymbol{\theta} \in \Theta} \left\|\Sigma^{-1}\hat{\Sigma}_n\hat{\Sigma}_n^{-1}\left(\mathbf{g}_n - \hat{\Sigma}_n\boldsymbol{\theta}\right)\right\| - \sqrt{n}\sup_{\mathbf{g}_n \in \bar{G}_n} \inf_{\boldsymbol{\theta} \in \Theta} \left\|\hat{\Sigma}_n^{-1}\left(\mathbf{g}_n - \hat{\Sigma}_n\boldsymbol{\theta}\right)\right\| \\ &\leq \sqrt{n}\sup_{\mathbf{g}_n \in \bar{G}_n} \inf_{\boldsymbol{\theta} \in \Theta} \left\|\Sigma^{-1}\hat{\Sigma}_n\right\| \left\|\hat{\Sigma}_n^{-1}\left(\mathbf{g}_n - \hat{\Sigma}_n\boldsymbol{\theta}\right)\right\| - \sqrt{n}\sup_{\mathbf{g}_n \in \bar{G}_n} \inf_{\boldsymbol{\theta} \in \Theta} \left\|\hat{\Sigma}_n^{-1}\left(\mathbf{g}_n - \hat{\Sigma}_n\boldsymbol{\theta}\right)\right\| \\ &= \left(\left\|\Sigma^{-1}\hat{\Sigma}_n\right\| - 1\right) \sqrt{n}\sup_{\mathbf{g}_n \in \bar{G}_n} \inf_{\boldsymbol{\theta} \in \Theta} \left\|\hat{\Sigma}_n^{-1}\left(\mathbf{g}_n - \hat{\Sigma}_n\boldsymbol{\theta}\right)\right\| \\ &\leq \left(\left\|\Sigma^{-1}\hat{\Sigma}_n\right\| - 1\right) \left\|\hat{\Sigma}_n^{-1}\right\| \sqrt{n}\sup_{\mathbf{g}_n \in \bar{G}_n} \inf_{\boldsymbol{\theta} \in \Theta} \left\|\left(\mathbf{g}_n - \hat{\Sigma}_n\boldsymbol{\theta}\right)\right\| = o_p(1) \cdot O_p(1) = o_p(1),\end{aligned}$$

where the last inequality follows because of the same facts as above, and because  $\left\|\hat{\Sigma}_n^{-1}\right\| \rightarrow \left\|\Sigma^{-1}\right\| < \infty$  by Slutsky Theorem. Hence,

$$\sqrt{n}H\left(\Sigma^{-1}\bar{G}_n, \Sigma^{-1}\hat{\Sigma}_n\Theta\right) - \sqrt{n}H\left(\hat{\Sigma}_n^{-1}\bar{G}_n, \hat{\Sigma}_n^{-1}\hat{\Sigma}_n\Theta\right) \leq o_p(1),$$

which implies that

$$\left| \sqrt{n}H \left( \hat{\Sigma}_n^{-1} \bar{G}_n, \hat{\Sigma}_n^{-1} \hat{\Sigma}_n \Theta \right) - \sqrt{n}H \left( \Sigma^{-1} \bar{G}_n, \Sigma^{-1} \hat{\Sigma}_n \Theta \right) \right| \xrightarrow{p} 0.$$

*Step 2. Derivation of the asymptotic distribution of  $\sqrt{n} \left[ (s(\mathbf{p}, \Sigma^{-1} \bar{G}_n) - s(\mathbf{p}, \Theta)) \quad (\hat{\Sigma}_n - \Sigma) \Sigma^{-1} \mathbf{p} \right]'$ .*

Lemmas A.3 and A.4 show that  $\{G, G_i : i \in \mathbb{N}\}$  are *i.i.d.* non-empty compact valued SVRVs. Observe that by the assumptions of Theorem 5.3

$$\int \|G\|_H^2 d\mu \leq E(|y_L|^2) + E(|y_U|^2) + \sum_{k=1}^d \left[ E(|x_k y_L|^2) + E(|x_k y_U|^2) \right] < \infty,$$

and  $E(|\mathbf{x}|^4) < \infty$ . Using the same argument as in the proof of Theorem 3.1.8 of Li et al. (2002), it follows by the Central Limit Theorem for  $\mathbb{C}(\mathbb{S})$ -valued random variables of Jain & Marcus (1975) that

$$(A.6) \quad \sqrt{n} \begin{bmatrix} s(\mathbf{p}, \Sigma^{-1} \bar{G}_n) - s(\mathbf{p}, \Theta) \\ (\hat{\Sigma}_n - \Sigma) \Sigma^{-1} \mathbf{p} \end{bmatrix} \xrightarrow{d} \begin{pmatrix} z(\mathbf{p}) \\ \mathbf{u}(\mathbf{p}) \end{pmatrix},$$

where for each  $\mathbf{p} \in \mathbb{S}$ ,  $(z(\mathbf{p}) \quad \mathbf{u}(\mathbf{p}))$  is a  $d+2$  dimensional normal random vector with

$$\begin{aligned} E[z(\mathbf{p})] &= 0, \\ E[\mathbf{u}(\mathbf{p})] &= 0. \end{aligned}$$

Regarding the covariance kernel, denote by  $\{\rho_{ij}\}$ ,  $i, j = 1, \dots, d+1$  the elements of  $\Sigma^{-1}$ . Then,

$$E \left[ \begin{pmatrix} z(\mathbf{p}) \\ \mathbf{u}(\mathbf{p}) \end{pmatrix} \begin{pmatrix} z(\mathbf{p}) \\ \mathbf{u}(\mathbf{p}) \end{pmatrix}' \right] = \begin{bmatrix} E \left[ (s(\mathbf{p}, \Sigma^{-1} G))^2 \right] - E[s(\mathbf{p}, \Sigma^{-1} G)]^2 & \boldsymbol{\kappa}'_{\mathbf{p}, \mathbf{p}} \\ \boldsymbol{\kappa}_{\mathbf{p}, \mathbf{p}} & V_{\mathbf{p}, \mathbf{p}} \end{bmatrix}$$

where  $V_{\mathbf{p}, \mathbf{p}} = E \left[ ((\Pi_{\mathbf{x}} - \Sigma) \Sigma^{-1} \mathbf{p}) ((\Pi_{\mathbf{x}} - \Sigma) \Sigma^{-1} \mathbf{p})' \right]$  and

$$(A.7) \quad \hat{\Sigma}_n = \frac{1}{n} \sum_{i=1}^n \Pi_{\mathbf{x}_i} \quad \text{and} \quad \Pi_{\mathbf{x}_i} \equiv \begin{bmatrix} 1 & x_{1i} & \dots & x_{di} \\ x_{1i} & x_{1i}^2 & \dots & x_{1i} x_{di} \\ \vdots & \vdots & \ddots & \vdots \\ x_{di} & x_{1i} x_{di} & \dots & x_{di}^2 \end{bmatrix},$$

and  $\boldsymbol{\kappa}_{\mathbf{p}, \mathbf{p}}$  is a  $(d+1) \times 1$  vector with first element

$$\kappa_{1, (\mathbf{p}, \mathbf{p})} = \sum_{k=1}^{d+1} \sum_{i=2}^{d+1} \left[ (E[x_{i-1} s(\mathbf{p}, \Sigma^{-1} G)] - s(\mathbf{p}, \Theta) E(x_{i-1})) \right] \rho_{ik} p_k.$$

and  $j$ -th element ( $j = 2, \dots, d+1$ )

$$\begin{aligned} \kappa_{j, (\mathbf{p}, \mathbf{p})} &= \left[ (E[x_{j-1} s(\mathbf{p}, \Sigma^{-1} G)] - s(\mathbf{p}, \Theta) E(x_{j-1})) \sum_{l=1}^{d+1} \rho_{jl} p_l \right. \\ &\quad \left. + \sum_{l=1}^{d+1} \sum_{i=2}^{d+1} (E[x_{j-1} x_{i-1} s(\mathbf{p}, \Sigma^{-1} G)] - s(\mathbf{p}, \Theta) E(x_{j-1} x_{i-1})) \rho_{il} p_l \right]. \end{aligned}$$



with  $x_j$  the  $j$ -th element of  $\mathbf{x}$ .

For the next step in this proof it is also useful to observe that

$$E[z(\mathbf{p})z(\mathbf{q})] = E[s(\mathbf{p}, \Sigma^{-1}G)s(\mathbf{q}, \Sigma^{-1}G)] - E[s(\mathbf{p}, \Sigma^{-1}G)]E[s(\mathbf{q}, \Sigma^{-1}G)],$$

$$(A.8) \quad E[\mathbf{u}(\mathbf{p})\mathbf{u}(\mathbf{q})] = V_{\mathbf{p}, \mathbf{q}} = E\left[\left((\Pi_{\mathbf{x}} - \Sigma)\Sigma^{-1}\mathbf{p}\right)\left((\Pi_{\mathbf{x}} - \Sigma)\Sigma^{-1}\mathbf{q}\right)'\right]$$

and

$$E[z(\mathbf{p})\mathbf{u}(\mathbf{q})] = \boldsymbol{\kappa}_{\mathbf{p}, \mathbf{q}},$$

where  $\boldsymbol{\kappa}_{\mathbf{p}, \mathbf{q}}$  is a  $(d+1) \times 1$  vector with 1-st and  $j$ -th element ( $j = 2, \dots, d+1$ ) given respectively by

$$(A.9) \quad \begin{aligned} \kappa_{1,(\mathbf{p}, \mathbf{q})} &= \sum_{k=1}^{d+1} \sum_{i=2}^{d+1} \left[ (E[x_{i-1}s(\mathbf{p}, \Sigma^{-1}G)] - s(\mathbf{p}, \Theta)E(x_{i-1})) \right] \rho_{ik}q_k, \\ \kappa_{j,(\mathbf{p}, \mathbf{q})} &= \left[ (E[x_{j-1}s(\mathbf{p}, \Sigma^{-1}G)] - s(\mathbf{p}, \Theta)E(x_{j-1})) \sum_{l=1}^{d+1} \rho_{jl}q_l \right. \\ &\quad \left. + \sum_{l=1}^{d+1} \sum_{i=2}^{d+1} (E[x_{j-1}x_{i-1}s(\mathbf{p}, \Sigma^{-1}G)] - s(\mathbf{p}, \Theta)E(x_{j-1}x_{i-1})) \rho_{il}q_l \right]. \end{aligned}$$

*Step 3. Derivation of the asymptotic distribution of  $\sqrt{n} \left[ s(\mathbf{p}, \Sigma^{-1}\bar{G}_n) - s(\mathbf{p}, \Sigma^{-1}\hat{\Sigma}_n\Theta) \right]$ .*

Observe that

$$\sqrt{n} \left[ s(\mathbf{p}, \Sigma^{-1}\bar{G}_n) - s(\mathbf{p}, \Sigma^{-1}\hat{\Sigma}_n\Theta) \right] = \sqrt{n} \left\{ \left[ s(\mathbf{p}, \Sigma^{-1}\bar{G}_n) - s(\mathbf{p}, \Theta) \right] - \left[ s(\mathbf{p}, \Sigma^{-1}\hat{\Sigma}_n\Theta) - s(\mathbf{p}, \Theta) \right] \right\}.$$

Recall that by Lemma A.7, since  $\mathbf{x}$  has an absolutely continuous distribution with respect to Lebesgue measure on  $\mathfrak{R}^d$ ,  $\Theta$  is a strictly convex set. Hence, its support set in the direction  $\mathbf{p} \in \mathfrak{S}$  is a singleton given by

$$\boldsymbol{\xi}_{\mathbf{p}} = \Theta \cap \{\boldsymbol{\vartheta} \in \mathfrak{R}^{d+1} : \langle \boldsymbol{\vartheta}, \mathbf{p} \rangle = s(\mathbf{p}, \Theta)\}.$$

By Corollary 1.7.3 of Schneider (1993) it follows that  $s(\mathbf{p}, \Theta)$  is (Fréchet) differentiable at  $\mathbf{p} \in \mathfrak{R}^{d+1} \setminus \{\mathbf{0}\}$  with gradient equal to  $\boldsymbol{\xi}_{\mathbf{p}}$ , and therefore the Functional Delta Method (Van der Vaart (2000), Theorem 20.8) implies that

$$\begin{aligned} \sqrt{n} \left[ s(\mathbf{p}, \Sigma^{-1}\hat{\Sigma}_n\Theta) - s(\mathbf{p}, \Theta) \right] &= \sqrt{n} \left[ s(\hat{\Sigma}_n\Sigma^{-1}\mathbf{p}, \Theta) - s(\mathbf{p}, \Theta) \right] \\ &= \sqrt{n} \left[ s\left(\mathbf{p} + (\hat{\Sigma}_n - \Sigma)\Sigma^{-1}\mathbf{p}, \Theta\right) - s(\mathbf{p}, \Theta) \right] \\ &= \frac{s\left(\mathbf{p} + \frac{1}{\sqrt{n}} \left[ \sqrt{n}(\hat{\Sigma}_n - \Sigma)\Sigma^{-1}\mathbf{p} \right], \Theta\right) - s(\mathbf{p}, \Theta)}{\frac{1}{\sqrt{n}}} \\ &\xrightarrow{d} \langle \boldsymbol{\xi}_{\mathbf{p}}, \mathbf{u}(\mathbf{p}) \rangle, \end{aligned}$$

where  $\mathbf{u}(\mathbf{p})$  was given in equation (A.6). It then follows by another application of the Functional Delta Method that

$$\sqrt{n} \left[ s(\mathbf{p}, \Sigma^{-1}\bar{G}_n) - s(\mathbf{p}, \Sigma^{-1}\hat{\Sigma}_n\Theta) \right] \xrightarrow{d} z(\mathbf{p}) - \langle \boldsymbol{\xi}_{\mathbf{p}}, \mathbf{u}(\mathbf{p}) \rangle,$$

with  $z(\mathbf{p})$  and  $\mathbf{u}(\mathbf{p})$  given in equation (A.6) are such that

$$E[z(\mathbf{p}) - \langle \boldsymbol{\xi}_{\mathbf{p}}, \mathbf{u}(\mathbf{p}) \rangle] = 0$$

and

$$\begin{aligned} & E [(z(\mathbf{p}) - \langle \boldsymbol{\xi}_{\mathbf{p}}, \mathbf{u}(\mathbf{p}) \rangle) (z(\mathbf{q}) - \langle \boldsymbol{\xi}_{\mathbf{q}}, \mathbf{u}(\mathbf{q}) \rangle)] \\ &= E [s(\mathbf{p}, \Sigma^{-1}G) s(\mathbf{q}, \Sigma^{-1}G)] - E [s(\mathbf{p}, \Sigma^{-1}G)] E [s(\mathbf{q}, \Sigma^{-1}G)] - \langle \boldsymbol{\xi}_{\mathbf{p}}, \boldsymbol{\kappa}_{\mathbf{q}, \mathbf{p}} \rangle - \langle \boldsymbol{\kappa}_{\mathbf{p}, \mathbf{q}}, \boldsymbol{\xi}_{\mathbf{q}} \rangle + \langle \boldsymbol{\xi}_{\mathbf{p}}, V_{\mathbf{p}, \mathbf{q}} \boldsymbol{\xi}_{\mathbf{q}} \rangle, \end{aligned}$$

where  $V_{\mathbf{p}, \mathbf{q}}$  and  $\boldsymbol{\kappa}_{\mathbf{p}, \mathbf{q}}$  were given in (A.8)-(A.9).

Finally, by Hörmander's Embedding Theorem,

$$\sqrt{n}H(\Sigma^{-1}\bar{G}_n, \Sigma^{-1}\hat{\Sigma}_n\Theta) = \sqrt{n} \sup_{\mathbf{p}: \|\mathbf{p}\|=1} \left| s(\mathbf{p}, \Sigma^{-1}\bar{G}_n) - s(\mathbf{p}, \Sigma^{-1}\hat{\Sigma}_n\Theta) \right|.$$

Letting  $v(\mathbf{p}) \equiv z(\mathbf{p}) - \langle \boldsymbol{\xi}_{\mathbf{p}}, \mathbf{u}(\mathbf{p}) \rangle$  for each  $\mathbf{p} \in \mathbb{S}$ , the result follows by the continuous mapping theorem.

(ii) The fact that  $\Sigma$  is of full rank implies that  $\Sigma^{-1}\mathbf{p} \neq \mathbf{0}$  for each  $\mathbf{p} \in \mathbb{S}$ . Let

$$(A.10) \quad f(\mathbf{x}_i) = \left\langle \mathbf{p}, \Sigma^{-1} \begin{bmatrix} 1 \\ \mathbf{x}_i \end{bmatrix} \right\rangle.$$

Then

$$\begin{aligned} s(\mathbf{p}, \Sigma^{-1}G_i) &= \sup \left\{ \left\langle \mathbf{p}, \Sigma^{-1} \begin{bmatrix} y_i \\ \mathbf{x}_i \cdot y_i \end{bmatrix} \right\rangle : \begin{bmatrix} y_i \\ \mathbf{x}_i \cdot y_i \end{bmatrix} \in \left( \begin{bmatrix} y_i \\ \mathbf{x}_i \cdot y_i \end{bmatrix}, y_i(\omega) \in Y_i(\omega) \right) \right\} \\ &= y_{iL} \mathbb{I}(f(\mathbf{x}_i) < 0) f(\mathbf{x}_i) + y_{iU} \mathbb{I}(f(\mathbf{x}_i) \geq 0) f(\mathbf{x}_i), \end{aligned}$$

where  $\mathbb{I}(\cdot)$  is the indicator function of the event in brackets. Because  $s(\mathbf{p}, \Theta)$  is Fréchet differentiable for any  $\mathbf{p} \in \mathfrak{R}^{d+1} \setminus \{\mathbf{0}\}$  with gradient  $\boldsymbol{\xi}_{\mathbf{p}}$ , it follows that for each  $\mathbf{p} \in \mathbb{S}$ ,

$$\begin{aligned} \sqrt{n} [s(\mathbf{p}, \Sigma^{-1}\bar{G}_n) - s(\mathbf{p}, \Sigma^{-1}\hat{\Sigma}_n\Theta)] &= \sqrt{n} [s(\mathbf{p}, \Sigma^{-1}\bar{G}_n) - s(\mathbf{p}, \Theta) + s(\mathbf{p}, \Theta) - s(\mathbf{p}, \Sigma^{-1}\hat{\Sigma}_n\Theta)] \\ &= \sqrt{n} [s(\mathbf{p}, \Sigma^{-1}\bar{G}_n) - s(\mathbf{p}, \Theta)] - \langle \boldsymbol{\xi}_{\mathbf{p}}, \sqrt{n}(\hat{\Sigma}_n - \Sigma) \Sigma^{-1}\mathbf{p} \rangle + o_p(1) \\ &= \sqrt{n} \left[ \frac{1}{n} \sum_{i=1}^n s(\mathbf{p}, \Sigma^{-1}G_i) - s(\mathbf{p}, \Theta) \right] - \frac{1}{n} \sum_{i=1}^n \sqrt{n} \langle \boldsymbol{\xi}_{\mathbf{p}}, (\Pi_{\mathbf{x}_i} - \Sigma) \Sigma^{-1}\mathbf{p} \rangle + o_p(1) \\ &= \sqrt{n} \frac{1}{n} \sum_{i=1}^n [s(\mathbf{p}, \Sigma^{-1}G_i) - s(\mathbf{p}, \Theta) - \langle \boldsymbol{\xi}_{\mathbf{p}}, (\Pi_{\mathbf{x}_i} - \Sigma) \Sigma^{-1}\mathbf{p} \rangle] + o_p(1), \end{aligned}$$

where  $\Pi_{\mathbf{x}_i}$  and  $f(\mathbf{x}_i)$  were defined in (A.7) and (A.10), and  $s(\mathbf{p}, \Sigma^{-1}G_i) - s(\mathbf{p}, \Theta) - \langle \boldsymbol{\xi}_{\mathbf{p}}, (\Pi_{\mathbf{x}_i} - \Sigma) \Sigma^{-1}\mathbf{p} \rangle$  is a sequence of i.i.d.  $\mathbb{C}(\mathbb{S})$ -valued random variables satisfying the conditions of the Central Limit Theorem of Jain & Marcus (1975) (see also Theorem 3.1.8 of Li et al. (2002)). Hence, for each  $\mathbf{p} \in \mathbb{S}$ ,

$$\begin{aligned} \text{Var}(v(\mathbf{p})) &= \text{Var}(s(\mathbf{p}, \Sigma^{-1}G_i) - s(\mathbf{p}, \Theta) - \langle \boldsymbol{\xi}_{\mathbf{p}}, (\Pi_{\mathbf{x}_i} - \Sigma) \Sigma^{-1}\mathbf{p} \rangle) \\ &= \text{Var}(y_{iL} \mathbb{I}(f(\mathbf{x}_i) < 0) f(\mathbf{x}_i) + y_{iU} \mathbb{I}(f(\mathbf{x}_i) \geq 0) f(\mathbf{x}_i) - \langle \boldsymbol{\xi}_{\mathbf{p}}, \Pi_{\mathbf{x}_i} \Sigma^{-1}\mathbf{p} \rangle) \end{aligned}$$

It follows from the Law of Iterated Expectations that

$$\begin{aligned}
\text{Var}(v(\mathbf{p})) &\geq E \left[ \text{Var} \left( y_{iL} \mathbb{I}(f(\mathbf{x}_i) < 0) f(\mathbf{x}_i) + y_{iU} \mathbb{I}(f(\mathbf{x}_i) \geq 0) f(\mathbf{x}_i) - \langle \boldsymbol{\xi}_{\mathbf{p}}, \Pi_{\mathbf{x}_i} \Sigma^{-1} \mathbf{p} \rangle \mid \mathbf{x}_i \right) \right] \\
&= E \left[ (f(\mathbf{x}_i))^2 \mathbb{I}(f(\mathbf{x}_i) < 0) \text{Var}(y_{iL} \mid \mathbf{x}_i) + (f(\mathbf{x}_i))^2 \mathbb{I}(f(\mathbf{x}_i) \geq 0) \text{Var}(y_{iU} \mid \mathbf{x}_i) \right] \\
&\geq \sigma^2 E \left[ (f(\mathbf{x}_i))^2 \mathbb{I}(f(\mathbf{x}_i) < 0) + (f(\mathbf{x}_i))^2 \mathbb{I}(f(\mathbf{x}_i) \geq 0) \right] \\
&= \sigma^2 E \left[ \left\langle \mathbf{p}, \Sigma^{-1} \begin{bmatrix} 1 \\ \mathbf{x}_i \end{bmatrix} \right\rangle^2 \right] = \sigma^2 \langle \mathbf{p}, \Sigma^{-1} \Sigma \Sigma^{-1} \mathbf{p} \rangle > 0 \text{ for each } \mathbf{p} \in \mathbb{S},
\end{aligned}$$

where the second inequality follows because by assumption  $\text{Var}(y_L \mid \mathbf{x}), \text{Var}(y_U \mid \mathbf{x}) \geq \sigma^2 > 0$   $P(\mathbf{x}) - a.s.$ , and the last equality follows because  $\Sigma$  is of full rank by assumption. Hence  $\text{Var}(v(\mathbf{p})) > 0$  for each  $\mathbf{p} \in \mathbb{S}$ , and by the same argument as in the proof of Proposition 3.1 the law of  $\|v\|_{\mathbb{C}(\mathbb{S})}$  is absolutely continuous with respect to Lebesgue measure on  $\mathfrak{R}$ .

■

#### A.3.4 Proof of Corollary 5.4

**Proof.** By triangle inequality,

$$\begin{aligned}
&\sqrt{m_n} \left| H \left( \hat{\Sigma}_n^{-1} \bar{G}_{m_n}, \Sigma^{-1} \mathbb{E}[G] \right) - H \left( \Sigma^{-1} \bar{G}_{m_n}, \Sigma^{-1} \mathbb{E}[G] \right) \right| \\
&\leq \sqrt{m_n} H \left( \hat{\Sigma}_n^{-1} \bar{G}_{m_n}, \Sigma^{-1} \bar{G}_{m_n} \right) \leq \sqrt{m_n} \left\| \hat{\Sigma}_n^{-1} - \Sigma^{-1} \right\| \sup_{\mathbf{g}_{m_n} \in \bar{G}_{m_n}} \|\mathbf{g}_{m_n}\| \\
&= \sqrt{m_n} O_p \left( n^{-\frac{1}{2}} \right) \sup_{\mathbf{p}: \|\mathbf{p}\|=1} \left| \frac{1}{m_n} \sum_{i=1}^{m_n} s(\mathbf{p}, G_i) \right| = o_p(1),
\end{aligned}$$

where the last equality follows because the  $G_i^s$  are integrably bounded. The result then follows directly from Theorem 2.3. Absolute continuity with respect to Lebesgue measure on  $\mathfrak{R}$  of the law of  $\|v\|_{\mathbb{C}(\mathbb{S})}$  follows by the same argument as in the proof of Theorem 5.3-(ii). ■

**Lemma A.8** *Let  $\Theta$  be defined as in (A.3) and  $\hat{\Theta}_n = \hat{\Sigma}_n^{-1} \frac{1}{n} \bigoplus_{i=1}^n G_i$ . Let the Assumptions of Theorem 5.2 and Assumption 5.5 be satisfied. Let  $\boldsymbol{\xi}_{\mathbf{p}} = \Theta \cap \{\boldsymbol{\vartheta} \in \mathfrak{R}^{d+1} : \langle \boldsymbol{\vartheta}, \mathbf{p} \rangle = s(\mathbf{p}, \Theta)\}$ ,  $\mathbf{p} \in \mathbb{S}$ , and denote by  $\hat{\boldsymbol{\xi}}_{n,\mathbf{p}} = \hat{\Theta}_n \cap \{\boldsymbol{\vartheta} \in \mathfrak{R}^{d+1} : \langle \boldsymbol{\vartheta}, \mathbf{p} \rangle = s(\mathbf{p}, \hat{\Theta}_n)\}$ ,  $\mathbf{p} \in \mathbb{S}$ . Then  $H(\hat{\boldsymbol{\xi}}_{n,\mathbf{p}}, \{\boldsymbol{\xi}_{\mathbf{p}}\}) \xrightarrow{a.s.} 0$ .*

**Proof.** Theorem 5.2 guarantees that  $H(\hat{\Theta}_n, \Theta) \xrightarrow{a.s.} 0$ . This implies that  $s(\mathbf{p}, \hat{\Theta}_n) \xrightarrow{a.s.} s(\mathbf{p}, \Theta)$  uniformly in  $\mathbf{p} \in \mathbb{S}$ , and therefore  $\{\boldsymbol{\vartheta} \in \mathfrak{R}^{d+1} : \langle \boldsymbol{\vartheta}, \mathbf{p} \rangle = s(\mathbf{p}, \hat{\Theta}_n)\}$  converges to  $\{\boldsymbol{\vartheta} \in \mathfrak{R}^{d+1} : \langle \boldsymbol{\vartheta}, \mathbf{p} \rangle = s(\mathbf{p}, \Theta)\}$  with respect to the Hausdorff distance for each  $\mathbf{p} \in \mathbb{S}$ . Observe that  $\hat{\boldsymbol{\xi}}_{n,\mathbf{p}} \neq \emptyset$  for each  $n \in \mathbb{N}$  and each  $\mathbf{p} \in \mathbb{S}$ . Moreover, if  $\hat{\boldsymbol{\xi}}_{n,\mathbf{p}}$  contains a convergent subsequence  $\hat{\boldsymbol{\xi}}_{n_j,\mathbf{p}}$ ,  $j \in \mathbb{N}$ , then  $\hat{\boldsymbol{\xi}}_{n_j,\mathbf{p}} \xrightarrow{a.s.} \boldsymbol{\xi}_{\mathbf{p}}$ . Such a convergent subsequence almost surely exists. This is because any sequence  $\hat{\boldsymbol{\xi}}_{n_j,\mathbf{p}} \in \hat{\boldsymbol{\xi}}_{n_j,\mathbf{p}}$  is a sequence in  $\hat{\Theta}_{n_j}$ . The sequence of sets  $\hat{\Theta}_{n_j}$  converges to the compact set  $\Theta$ , and therefore  $\hat{\boldsymbol{\xi}}_{n_j,\mathbf{p}}$  is almost surely a bounded sequence.

Hence  $\hat{\xi}_{n\mathbf{p}}$  almost surely contains a convergent subsequence. The result then follows from Theorem 1.8.7 in Schneider (1993). ■

### A.3.5 Proof of Proposition 5.5

**Proof.** By a similar argument to the one in Step 1 of the proof of Theorem 5.3, since  $\hat{\Sigma}_n^{-1*}$  is a consistent estimator of  $\Sigma^{-1}$ ,

$$\sqrt{n}H \left( \hat{\Sigma}_n^{-1*} \frac{1}{n} \bigoplus_{i=1}^n G_i^*, \hat{\Theta}_n \right) \stackrel{A}{=} \sqrt{n}H \left( \Sigma^{-1} \frac{1}{n} \bigoplus_{i=1}^n G_i^*, \Sigma^{-1} \hat{\Sigma}_n^* \hat{\Theta}_n \right),$$

so that by Hörmander's Embedding Theorem

$$r_n^* \stackrel{A}{=} \sqrt{n} \left\| \frac{1}{n} \sum_{i=1}^n s(\cdot, \Sigma^{-1} G_i^*) - s(\cdot, \Sigma^{-1} \hat{\Sigma}_n^* \hat{\Theta}_n) \right\|_{\mathbb{C}(\mathbb{S})}.$$

Observe that

$$\begin{aligned} & \sqrt{n} \left[ \frac{1}{n} \sum_{i=1}^n s(\mathbf{p}, \Sigma^{-1} G_i^*) - s(\mathbf{p}, \Sigma^{-1} \hat{\Sigma}_n^* \hat{\Theta}_n) \right] \\ = & \sqrt{n} \left( \left[ \frac{1}{n} \sum_{i=1}^n s(\mathbf{p}, \Sigma^{-1} G_i^*) - \frac{1}{n} \sum_{i=1}^n s(\mathbf{p}, \Sigma^{-1} G_i) \right] - \left[ s(\mathbf{p}, \Sigma^{-1} \hat{\Sigma}_n^* \hat{\Theta}_n) - \frac{1}{n} \sum_{i=1}^n s(\mathbf{p}, \Sigma^{-1} G_i) \right] \right). \end{aligned}$$

Recalling that  $\hat{\Theta}_n = \hat{\Sigma}_n^{-1} \bar{G}_n$ , where  $\bar{G}_n = \frac{1}{n} \bigoplus_{i=1}^n G_i$ , we get that

$$s(\mathbf{p}, \Sigma^{-1} \hat{\Sigma}_n^* \hat{\Theta}_n) - \frac{1}{n} \sum_{i=1}^n s(\mathbf{p}, \Sigma^{-1} G_i) = s(\mathbf{p} + \Sigma \hat{\Sigma}_n^{-1} (\hat{\Sigma}_n^* - \hat{\Sigma}_n) \Sigma^{-1} \mathbf{p}, \Sigma^{-1} \bar{G}_n) - s(\mathbf{p}, \Sigma^{-1} \bar{G}_n).$$

By Theorem 3.3.1 and the discussion on p. 160 of Schneider (1993), any non-empty, convex and compact subset of  $\mathfrak{R}^{d+1}$  can be approximated arbitrarily accurately by a non-empty, strictly convex and compact subset of  $\mathfrak{R}^{d+1}$ . Hence, for any  $\varepsilon_n = o_p\left(\frac{1}{\sqrt{n}}\right)$  we can find a sequence of non-empty, strictly convex and compact sets  $\bar{\Gamma}_n$  such that  $H(\Sigma^{-1} \bar{G}_n, \Sigma^{-1} \bar{\Gamma}_n) < \varepsilon_n = o_p\left(\frac{1}{\sqrt{n}}\right)$ , which implies that  $|s(\mathbf{p}, \Sigma^{-1} \bar{G}_n) - s(\mathbf{p}, \Sigma^{-1} \bar{\Gamma}_n)| = o_p\left(\frac{1}{\sqrt{n}}\right)$  uniformly in  $\mathbf{p} \in \mathfrak{R}^{d+1}$ . Under the assumptions of Theorem 5.3,  $s(\mathbf{p}, \Theta)$  is Fréchet differentiable at  $\mathbf{p} \in \mathfrak{R}^{d+1} \setminus \{\mathbf{0}\}$  with gradient equal to  $\xi_{\mathbf{p}}$ . Similarly, because  $\Sigma^{-1} \bar{\Gamma}_n$  is a strictly convex set,  $s(\mathbf{p}, \Sigma^{-1} \bar{\Gamma}_n)$  is Fréchet differentiable at  $\mathbf{p} \in \mathfrak{R}^{d+1} \setminus \{\mathbf{0}\}$  with gradient equal to  $\bar{\xi}_{n,\mathbf{p}}$ . Hence

$$\begin{aligned} & \sqrt{n} \left| \left[ s(\mathbf{p} + \Sigma \hat{\Sigma}_n^{-1} (\hat{\Sigma}_n^* - \hat{\Sigma}_n) \Sigma^{-1} \mathbf{p}, \Sigma^{-1} \bar{G}_n) - s(\mathbf{p}, \Sigma^{-1} \bar{G}_n) \right] \right. \\ & \quad \left. - \left[ s(\mathbf{p} + \Sigma \hat{\Sigma}_n^{-1} (\hat{\Sigma}_n^* - \hat{\Sigma}_n) \Sigma^{-1} \mathbf{p}, \Theta) - s(\mathbf{p}, \Theta) \right] \right| \\ \leq & \sqrt{n} \left| \left[ s(\mathbf{p} + \Sigma \hat{\Sigma}_n^{-1} (\hat{\Sigma}_n^* - \hat{\Sigma}_n) \Sigma^{-1} \mathbf{p}, \Sigma^{-1} \bar{G}_n) - s(\mathbf{p}, \Sigma^{-1} \bar{G}_n) \right] \right. \\ & \quad \left. - \left[ s(\mathbf{p} + \Sigma \hat{\Sigma}_n^{-1} (\hat{\Sigma}_n^* - \hat{\Sigma}_n) \Sigma^{-1} \mathbf{p}, \Sigma^{-1} \bar{\Gamma}_n) - s(\mathbf{p}, \Sigma^{-1} \bar{\Gamma}_n) \right] \right| \\ & + \sqrt{n} \left| \left[ s(\mathbf{p} + \Sigma \hat{\Sigma}_n^{-1} (\hat{\Sigma}_n^* - \hat{\Sigma}_n) \Sigma^{-1} \mathbf{p}, \Sigma^{-1} \bar{\Gamma}_n) - s(\mathbf{p}, \Sigma^{-1} \bar{\Gamma}_n) \right] \right. \\ & \quad \left. - \left[ s(\mathbf{p} + \Sigma \hat{\Sigma}_n^{-1} (\hat{\Sigma}_n^* - \hat{\Sigma}_n) \Sigma^{-1} \mathbf{p}, \Theta) - s(\mathbf{p}, \Theta) \right] \right| \\ \leq & \sqrt{n} \cdot o_p\left(\frac{1}{\sqrt{n}}\right) + \sqrt{n} \left| \left( \Sigma \hat{\Sigma}_n^{-1} (\hat{\Sigma}_n^* - \hat{\Sigma}_n) \Sigma^{-1} \mathbf{p} \right) [\bar{\xi}_{n,\mathbf{p}} - \xi_{\mathbf{p}}] \right| + o_p(1) \\ = & o_p(1), \end{aligned}$$

where the last inequality follows from the fact that  $s(\mathbf{p}, \Sigma^{-1}\bar{\Gamma}_n)$  and  $s(\mathbf{p}, \Theta)$  are Fréchet differentiable for any  $\mathbf{p} \in \mathfrak{R}^{d+1} \setminus \{\mathbf{0}\}$ . The last equality follows because  $(\Sigma \hat{\Sigma}_n^{-1} (\hat{\Sigma}_n^* - \hat{\Sigma}_n) \Sigma^{-1} \mathbf{p}) = O_p\left(\frac{1}{\sqrt{n}}\right)$  and because by an argument similar to the one used to prove Lemma A.8,  $\bar{\boldsymbol{\zeta}}_{n,\mathbf{p}} - \boldsymbol{\xi}_{\mathbf{p}} = o_p(1)$ .

By Theorem 2.4 of Giné & Zinn (1990) and standard arguments

$$\sqrt{n} \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n s(\mathbf{p}, \Sigma^{-1} G_i^*) - \frac{1}{n} \sum_{i=1}^n s(\mathbf{p}, \Sigma^{-1} G_i) \\ \Sigma \hat{\Sigma}_n^{-1} (\hat{\Sigma}_n^* - \hat{\Sigma}_n) \Sigma^{-1} \mathbf{p} \end{bmatrix} \xrightarrow{d} \begin{pmatrix} z(\mathbf{p}) \\ \mathbf{u}(\mathbf{p}) \end{pmatrix}$$

where  $(z(\mathbf{p}) \ \mathbf{u}(\mathbf{p}))$  is the  $d+2$  dimensional normal random vector in (A.6). Therefore an application of the Delta Method for the bootstrap (Van der Waart & Wellner (2000), Theorem 3.9.11) implies that

$$\sqrt{n} \left[ \frac{1}{n} \sum_{i=1}^n s(\mathbf{p}, \Sigma^{-1} G_i^*) - s(\mathbf{p}, \Sigma^{-1} \hat{\Sigma}_n^* \hat{\Theta}_n) \right] \xrightarrow{d} z(\mathbf{p}) - \langle \boldsymbol{\xi}_{\mathbf{p}}, \mathbf{u}(\mathbf{p}) \rangle.$$

By the continuous mapping theorem

$$r_n^* \xrightarrow{d} \|v\|_{\mathcal{C}(\mathbb{S})},$$

where  $v(\mathbf{p}) \equiv z(\mathbf{p}) - \langle \boldsymbol{\xi}_{\mathbf{p}}, \mathbf{u}(\mathbf{p}) \rangle$  for each  $\mathbf{p} \in \mathbb{S}$ . It then follows by the same argument as in the proof of Proposition 3.1 that since  $\text{Var}(v(\mathbf{p})) > 0$  for each  $\mathbf{p} \in \mathbb{S}$ , the critical values of the simulated distribution consistently estimate the critical values of  $\|v\|_{\mathcal{C}(\mathbb{S})}$ , that is  $\hat{c}_{\alpha n}^{BLP} = c_{\alpha}^{BLP} + o_p(1)$ . ■

### A.3.6 Proof of Proposition 5.6

**Proof.** For each  $\mathbf{p}, \mathbf{q} \in \mathbb{S}$ , by Lemma A.8  $\hat{\boldsymbol{\xi}}_{\mathbf{p}} \xrightarrow{p} \boldsymbol{\xi}_{\mathbf{p}}$ , and by Theorem 5.2

$$\frac{1}{n} \sum_{i=1}^n s(\mathbf{p}, \hat{\Sigma}_n^{-1} G_i) \frac{1}{n} \sum_{i=1}^n s(\mathbf{q}, \hat{\Sigma}_n^{-1} G_i) \xrightarrow{a.s.} E[s(\mathbf{p}, \Sigma^{-1} G)] E[s(\mathbf{q}, \Sigma^{-1} G)].$$

Next we want to show that

$$\frac{1}{n} \sum_{i=1}^n s(\mathbf{p}, \hat{\Sigma}_n^{-1} G_i) s(\mathbf{q}, \hat{\Sigma}_n^{-1} G_i) \xrightarrow{p} E[s(\mathbf{p}, \Sigma^{-1} G) s(\mathbf{q}, \Sigma^{-1} G)].$$

Observe that

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n s(\mathbf{p}, \hat{\Sigma}_n^{-1} G_i) s(\mathbf{q}, \hat{\Sigma}_n^{-1} G_i) - E[s(\mathbf{p}, \Sigma^{-1} G) s(\mathbf{q}, \Sigma^{-1} G)] \\ (A.11) \quad &= \frac{1}{n} \sum_{i=1}^n s(\mathbf{p}, \hat{\Sigma}_n^{-1} G_i) s(\mathbf{q}, \hat{\Sigma}_n^{-1} G_i) - \frac{1}{n} \sum_{i=1}^n s(\mathbf{p}, \Sigma^{-1} G_i) s(\mathbf{q}, \Sigma^{-1} G_i) \\ & \quad + \frac{1}{n} \sum_{i=1}^n s(\mathbf{p}, \Sigma^{-1} G_i) s(\mathbf{q}, \Sigma^{-1} G_i) - E[s(\mathbf{p}, \Sigma^{-1} G) s(\mathbf{q}, \Sigma^{-1} G)]. \end{aligned}$$

By the Strong Law of Large Numbers in Banach spaces of Mourier (1955),

$$\frac{1}{n} \sum_{i=1}^n s(\mathbf{p}, \Sigma^{-1} G_i) s(\mathbf{q}, \Sigma^{-1} G_i) - E[s(\mathbf{p}, \Sigma^{-1} G) s(\mathbf{q}, \Sigma^{-1} G)] = o_p(1),$$

so that we can concentrate on the first two terms in (A.11). Observe that given any two sets  $A$  and  $B$ , if  $A \subset B$  then  $s(\mathbf{p}, A) \leq s(\mathbf{p}, B)$ . Moreover,

$$\begin{aligned}\hat{\Sigma}_n^{-1}G_i &= \left(\Sigma^{-1} + \left(\hat{\Sigma}_n^{-1} - \Sigma^{-1}\right)\right)G_i \subset \Sigma^{-1}G_i \oplus \left(\hat{\Sigma}_n^{-1} - \Sigma^{-1}\right)G_i, \\ \Sigma^{-1}G_i &= \left(\hat{\Sigma}_n^{-1} + \left(\Sigma^{-1} - \hat{\Sigma}_n^{-1}\right)\right)G_i \subset \hat{\Sigma}_n^{-1}G_i \oplus \left(\Sigma^{-1} - \hat{\Sigma}_n^{-1}\right)G_i.\end{aligned}$$

Hence,

$$\left|s\left(\mathbf{p}, \hat{\Sigma}_n^{-1}G_i\right) - s\left(\mathbf{p}, \Sigma^{-1}G_i\right)\right| \leq \left|s\left(\mathbf{p}, \left(\hat{\Sigma}_n^{-1} - \Sigma^{-1}\right)G_i\right)\right| + \left|s\left(\mathbf{p}, \left(\Sigma^{-1} - \hat{\Sigma}_n^{-1}\right)G_i\right)\right|,$$

and therefore

$$\begin{aligned}& \left|\frac{1}{n} \sum_{i=1}^n s\left(\mathbf{p}, \hat{\Sigma}_n^{-1}G_i\right) s\left(\mathbf{q}, \hat{\Sigma}_n^{-1}G_i\right) - \frac{1}{n} \sum_{i=1}^n s\left(\mathbf{p}, \Sigma^{-1}G_i\right) s\left(\mathbf{q}, \Sigma^{-1}G_i\right)\right| \\ &= \left|\frac{1}{n} \sum_{i=1}^n s\left(\mathbf{p}, \hat{\Sigma}_n^{-1}G_i\right) \left[s\left(\mathbf{q}, \hat{\Sigma}_n^{-1}G_i\right) - s\left(\mathbf{q}, \Sigma^{-1}G_i\right)\right] + \frac{1}{n} \sum_{i=1}^n \left[s\left(\mathbf{p}, \hat{\Sigma}_n^{-1}G_i\right) - s\left(\mathbf{p}, \Sigma^{-1}G_i\right)\right] s\left(\mathbf{q}, \Sigma^{-1}G_i\right)\right| \\ &\leq \frac{1}{n} \sum_{i=1}^n \left|s\left(\mathbf{p}, \hat{\Sigma}_n^{-1}G_i\right)\right| \left|s\left(\mathbf{q}, \hat{\Sigma}_n^{-1}G_i\right) - s\left(\mathbf{q}, \Sigma^{-1}G_i\right)\right| + \frac{1}{n} \sum_{i=1}^n \left|s\left(\mathbf{p}, \hat{\Sigma}_n^{-1}G_i\right) - s\left(\mathbf{p}, \Sigma^{-1}G_i\right)\right| \left|s\left(\mathbf{q}, \Sigma^{-1}G_i\right)\right| \\ &\leq \frac{1}{n} \sum_{i=1}^n \left|s\left(\mathbf{p}, \hat{\Sigma}_n^{-1}G_i\right)\right| \left(\left|s\left(\mathbf{q}, \left(\hat{\Sigma}_n^{-1} - \Sigma^{-1}\right)G_i\right)\right| + \left|s\left(\mathbf{q}, \left(\Sigma^{-1} - \hat{\Sigma}_n^{-1}\right)G_i\right)\right|\right) \\ &\quad + \frac{1}{n} \sum_{i=1}^n \left|s\left(\mathbf{q}, \Sigma^{-1}G_i\right)\right| \left(\left|s\left(\mathbf{p}, \left(\hat{\Sigma}_n^{-1} - \Sigma^{-1}\right)G_i\right)\right| + \left|s\left(\mathbf{p}, \left(\Sigma^{-1} - \hat{\Sigma}_n^{-1}\right)G_i\right)\right|\right) \\ &\leq \frac{1}{n} \sum_{i=1}^n \left(\|\mathbf{p}\| \left\|\hat{\Sigma}_n^{-1}\right\| \|G_i\|_H\right) \left(\|\mathbf{q}\| \left\|\hat{\Sigma}_n^{-1} - \Sigma^{-1}\right\| \|G_i\|_H + \|\mathbf{q}\| \left\|\Sigma^{-1} - \hat{\Sigma}_n^{-1}\right\| \|G_i\|_H\right) \\ &\quad + \frac{1}{n} \sum_{i=1}^n \left(\|\mathbf{q}\| \left\|\Sigma^{-1}\right\| \|G_i\|_H\right) \left(\|\mathbf{p}\| \left\|\hat{\Sigma}_n^{-1} - \Sigma^{-1}\right\| \|G_i\|_H + \|\mathbf{p}\| \left\|\Sigma^{-1} - \hat{\Sigma}_n^{-1}\right\| \|G_i\|_H\right) \\ &= 2 \left(\left\|\hat{\Sigma}_n^{-1}\right\| + \left\|\Sigma^{-1}\right\|\right) \left\|\hat{\Sigma}_n^{-1} - \Sigma^{-1}\right\| \frac{1}{n} \sum_{i=1}^n \|G_i\|_H^2 \\ &= o_p(1)\end{aligned}$$

where the third inequality follows because for any matrix  $\Pi$  of proper dimensions, by the Cauchy-Schwartz inequality

$$\left|s\left(\mathbf{p}, \Pi G_i\right)\right| \leq \|\mathbf{p}\| \|\Pi G_i\|_H \leq \|\mathbf{p}\| \|\Pi\| \|G_i\|_H;$$

the second to last equality follows because  $\|\mathbf{p}\| = \|\mathbf{q}\| = 1$ ; and the last equality follows by Slutsky Theorem, since  $\frac{1}{n} \sum_{i=1}^n \|G_i\|_H^2 \xrightarrow{p} E\left[\|G_i\|_H^2\right]$  by the Strong Law of Large Numbers in Banach spaces and  $E\left[\|G_i\|_H^2\right] < \infty$  by assumption,  $\left\|\hat{\Sigma}_n^{-1}\right\| \xrightarrow{p} \left\|\Sigma^{-1}\right\| < \infty$  by assumption, and  $\left\|\hat{\Sigma}_n^{-1} - \Sigma^{-1}\right\| = o_p(1)$ .

A similar argument guarantees that  $\hat{\kappa}_{\mathbf{p}, \mathbf{q}} \xrightarrow{p} \kappa_{\mathbf{p}, \mathbf{q}}$ . The fact that  $\hat{V}_{\mathbf{p}, \mathbf{q}} \xrightarrow{p} V_{\mathbf{p}, \mathbf{q}}$  follows by standard arguments. It then follows by the same argument as in the proof of Proposition 3.1 that since  $Var(v(\mathbf{p})) > 0$  for each  $\mathbf{p} \in \mathbb{S}$ , the critical values of the simulated distribution consistently estimate the critical values of  $\|v\|_{C(\mathbb{S})}$ . ■

## References

- Anderson, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities, *Proceedings of the American Mathematical Society* **6**: 170–176.
- Andrews, D. W. K. (1997). A conditional kolmogorov test, *Econometrica* **65**: 1097–1128.
- Andrews, D. W. K., Berry, S. & Jia, P. (2004). Confidence regions for parameters in discrete games with multiple equilibria, with an application to discount chain store location, *Technical report*, Yale University.
- Artstein, Z. (1972). Set-valued measures, *Transactions of the American Mathematical Society* **165**: 103–125.
- Artstein, Z. (1974). On the calculus of closed set-valued functions, *Indiana University Mathematics Journal* **24**(5): 433–441.
- Artstein, Z. (1983). Distributions of random sets and random selections, *Israel Journal of Mathematics* **46**(4): 313–324.
- Artstein, Z. & Vitale, R. A. (1975). A strong law of large numbers for random compact sets, *The Annals of Probability* **3**(5): 879–882.
- Aumann, R. J. (1965). Integrals of set valued functions, *Journal of Mathematical Analysis and Applications* **12**: 1–12.
- Bolker, E. D. (1969). A class of convex bodies, *Transactions of the American Mathematical Society* **145**: 323–345.
- Castaing, C. & Valadier, M. (1977). *Convex Analysis and Measurable Multifunctions*, Vol. 580 of *Lecture Notes in Mathematics*, Springer Verlag, Berlin.
- Chernozhukov, V., Hahn, J. & Newey, W. (2005). Bound analysis in panel models with correlated random effects, *Technical report*, MIT, UCLA and MIT.
- Chernozhukov, V., Hong, H. & Tamer, E. (2004). Parameter set inference in a class of econometric models, *Technical report*, MIT, Duke University and Northwestern University.

- DeBlasi, F. S. & Iervolino, F. (1969). Equazioni differenziali con soluzioni a valore compatto convesso, *Bollettino della Unione Matematica Italiana* **4**(2): 491–501. In Italian.
- Galichon, A. & Henry, M. (2006). Inference in incomplete models, *Technical report*, Harvard and Columbia University.
- Giné, E., Hahn, M. G. & Zinn, J. (1983). *Limit Theorems for Random Sets: Application of Probability in Banach Space Results*, Vol. 990 of *Lecture Notes in Mathematics*, Springer, Berlin, pp. 112–135.
- Giné, E. & Zinn, J. (1990). Bootstrapping general empirical measures, *Annals of Probability* **18**(2): 851–869.
- Guggenberger, P., Hahn, J. & Kim, K. (2006). Specification testing under moment inequalities, *Technical report*, UCLA.
- Honoré, B. E. & Tamer, E. (2004). Bounds on parameters in dynamic discrete choice models, *Technical report*, Princeton University.
- Hörmander, L. (1954). Sur la fonction d'appui des ensembles convexes dans un espace localement convexe, *Arkiv för Matematik* **3**: 181–186.
- Horowitz, J. L. & Manski, C. F. (1997). What can be learned about population parameters when the data are contaminated, in G. S. Maddala & C. R. Rao (eds), *Handbook of Statistics, Vol. 15*, Elsevier Science B. V., North Holland, pp. 439–466.
- Horowitz, J. L. & Manski, C. F. (1998). Censoring of outcomes and regressors due to survey nonresponse: Identification and estimation using weights and imputations, *Journal of Econometrics* **84**: 37–58.
- Horowitz, J. L. & Manski, C. F. (2000). Nonparametric analysis of randomized experiments with missing covariate and outcome data, *Journal of the American Statistical Association* **95**(449): 77–84. With Comments.
- Imbens, G. W. & Manski, C. F. (2004). Confidence intervals for partially identified parameters, *Econometrica* **72**(6): 1845–1857.



- Jain, N. C. & Marcus, M. B. (1975). Central limit theorems for C(S)-valued random variables, *Journal of Functional Analysis* **19**: 216–231.
- Koshevoy, G. & Mosler, K. (1996). The lorenz zonoid of a multivariate distribution, *Journal of the American Statistical Association* **91**(434): 873–882.
- Lewbel, A. (2000). Identification of the binary choice model with misclassification, *Econometric Theory* **16**: 603–609.
- Li, S., Ogura, Y. & Kreinovich, V. (2002). *Limit Theorems and Applications of Set-Valued and Fuzzy Set-Valued Random Variables*, Kluwer Academic Publishers.
- Lifshits, M. A. (1982). On the absolute continuity of distributions of functionals of random processes, *Theory of Probability and Its Applications* **27**: 600–607.
- Magnac, T. & Maurin, E. (2005). Partial identification in monotone binary models: Discrete regressors and interval data, *Technical Report 280*, IDEI.
- Manski, C. F. (1989). Anatomy of the selection problem, *Journal of Human Resources* **24**: 343–360.
- Manski, C. F. (2003). *Partial Identification of Probability Distributions*, Springer Series in Statistics, Springer-Verlag, New York.
- Manski, C. F. & Tamer, E. (2002). Inference on regressions with interval data on a regressor or outcome, *Econometrica* **70**(2): 519–546.
- Mas-Collel, A., Whinston, M. D. & Green, J. R. (1995). *Microeconomic Theory*, Oxford University Press, Oxford, UK.
- Meyer, C. D. (2000). *Matrix Analysis and Applied Linear Algebra*, SIAM, Philadelphia, PA.
- Molchanov, I. S. (2005). *Theory of Random Sets*, Springer Verlag, London.
- Molinari, F. (2005). Partial identification of probability distributions with misclassified data, *Technical report*, Cornell University.
- Mosler, K. (2002). *Multivariate Dispersion, Central Regions and Depth. The Lift Zonoid Approach*, Vol. 165 of *Lecture Notes in Statistics*, Springer Verlag, Berlin.

- Mourier, E. (1955).  $l$ -random elements and  $l^*$ -random elements in banach spaces, *Proceedings of the Third Berkeley Symposium in Mathematical Statistics and Probability* **2**: 231–242.
- Newey, W. K. & McFadden, D. L. (1994). *Large Sample Estimation and Hypothesis Testing*, Handbook of Econometrics, Vol. IV, Elsevier Science B.V., chapter 36, pp. 2111–2245.
- Pakes, A., Porter, J., Ho, K. & Ishii, J. (2005). Moment inequalities and their application, *Technical report*, Harvard and Wisconsin University.
- Politis, D. N., Romano, J. P. & Wolf, M. (1999). *Subsampling*, Springer Verlag, New York.
- Rockafellar, R. T. (1970). *Convex Analysis*, Princeton University Press, Princeton, New Jersey.
- Rosen, A. (2006). Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities, *Technical report*, Northwestern University.
- Schneider, R. (1993). *Convex Bodies: The Brunn-Minkowski Theory*, Encyclopedia of Mathematics and Its Applications, Cambridge University Press, Cambridge, UK.
- Shaikh, A. M. (2005). Inference for a class of partially identified econometric models, *Technical report*, Stanford University.
- Starr, R. M. (1969). Quasi-equilibria in markets with non-convex preferences, *Econometrica* **37**: 25–38.
- Tsirel'son, V. S. (1975). The density of the distribution of the maximum of a gaussian process, *Theory of Probability and Its Applications* **24**: 847–856.
- Van der Vaart, A. W. (2000). *Asymptotic Statistics*, Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, UK.
- Van der Vaart, A. W. & Wellner, J. A. (2000). *Weak Convergence and Empirical Processes*, Springer Verlag, New York.
- Weil, W. (1982). An application of the central limit theorem for banach-space-valued random variables to the theory of random sets, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **60**: 203–208.