ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

An International Society for the Advancement of Economic Theory in its Relation to Statistics and Mathematics

http://www.econometricsociety.org/

Econometrica, Vol. 76, No. 4 (July, 2008), 763–814

ASYMPTOTIC PROPERTIES FOR A CLASS OF PARTIALLY IDENTIFIED MODELS

ARIE BERESTEANU
Duke University, Durham, NC 27708-0097, U.S.A.

FRANCESCA MOLINARI
Cornell University, Ithaca, NY 14853-7601, U.S.A.

Econometrica, Vol. 76, No. 4 (July, 2008), 763–814

ASYMPTOTIC PROPERTIES FOR A CLASS OF PARTIALLY IDENTIFIED MODELS

BY ARIE BERESTEANU AND FRANCESCA MOLINARI[1]

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 interval-identified scalar parameters, 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 population identification region with respect to the Hausdorff distance. We then show that the Hausdorff distance and the directed Hausdorff distance between the population identification region and the estimator, when properly normalized by $\sqrt{n}$, converge in distribution to functions of a Gaussian process whose covariance kernel depends on parameters of the population identification region. We provide consistent bootstrap procedures to approximate these limiting distributions. Using similar arguments as those applied for vector valued random variables, we develop a methodology to test assumptions about the true identification region and its subsets. We show that these results can be used to construct a *confidence collection* and a *directed confidence collection*. Those are (respectively) collection of sets that, when specified as a null hypothesis for the true value (a subset of values) of the population identification region, cannot be rejected by our tests.

KEYWORDS: Partial identification, confidence collections, set valued random variables, support functions.

1. INTRODUCTION

THIS PAPER CONTRIBUTES to the growing literature on inference for partially identified population features. These features include vectors of parameters or

[1]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. We thank the co-editor, Whitney Newey, and three anonymous reviewers for comments that greatly improved the paper. We benefited from comments by L. Barseghyan, V. Chernozhukov, A. Guerdjikova, B. Hansen, S. Hoderlein, H. Hong, B. Honoré, J. Horowitz, N. Kiefer, E. Mammen, C. Manski, M. Marinacci, K. Mosler, U. Müller, M. Nielsen, S. Resnick, G. Samorodnitsky, E. Tamer, T. Tatur, T. Vogelsang, M. Wells, seminar participants at Columbia, Cornell (Economics and Statistics), Duke, Harvard–MIT, Mannheim, NYU, Princeton, Syracuse, Torino, UCL, Vanderbilt, Wisconsin, Yale, Northwestern University's 2005 Conference on Incomplete Econometric Models, the 2005 Triangle Econometrics Conference, and the 2006 NY Econometrics Camp. We thank Jenny Law, Liyan Yang, and especially Bikramjit Das for able research assistance. Beresteanu gratefully acknowledges financial support from the Micro Incentives Research Center at Duke University and from NSF Grant SES-0617559. Molinari gratefully acknowledges financial support from the Center for Analytic Economics at Cornell University and from NSF Grant SES-0617482.

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, and Tamer (2007)). 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 and Manski (2004)) or the entire identification region (Chernozhukov, Hong, and Tamer (2007)) 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 compact and convex identification region[2] 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 SVRVs, a particular expectation operator needs to be used—the *Aumann expectation*; similarly, a particular summation operator needs to be used—the *Minkowski summation*—so as to get the set analog of the sample average, which is called the *Minkowski sample average*.[3] 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 those widely used in the space of vectors. Moreover, convex compact sets can be represented as elements of a functional space through their *support function* (Rockafellar (1970, Chap. 13)). The support function of a compact convex SVRV is a continuous valued random variable indexed on the unit sphere, and the support function of a Minkowski sample average of SVRVs is the sample average of the support functions of the SVRVs. Hence

[2]When the population identification region is not a convex set, our analysis applies to its convex hull.

[3]These concepts are formalized in Section 2.

we can derive the asymptotic properties of set estimators using laws of large numbers and central limit theorems for stochastic processes that are of familiar use in econometrics. This enables us to conduct inference for the entire identification region of a population feature of interest and for its subsets in a way that is completely analogous to how inference would be conducted if this feature were point identified.

Overview

Our methodology easily applies to the entire class of interval-identified scalar parameters. Moreover, it applies to the class of partially identified population features which have a compact and convex identification region that 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 of semiparametric binary models with interval regressor data (under the assumptions of Magnac and Maurin (2008)).

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. These limit theorems are obtained by exploiting the isometric embedding of the family of convex compact subsets of $\Re^d$ with the Hausdorff metric into the Banach space of continuous functions on the unit sphere with the uniform metric, which is provided by the support function. This allows one to represent random sets as elements of a functional space, for which limit theorems are available. 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.[4]

We then introduce two Wald-type statistics which, respectively, allow one to test (i) hypothesis about the entire identification region and (ii) its subsets. The first statistic is based on the Hausdorff distance. Because the Hausdorff distance between two nonempty compact sets A and B is equal to zero if and only if $A = B$, this distance is ideal to test a hypothesis which specifies the entire

[4]The choice of the Hausdorff distance to establish consistency results is a natural one, as this distance is a generalization of the Euclidean distance, and it is widely used in the literature on estimation of partially identified models (e.g., Manski and Tamer (2002)).

identification region of a vector of parameters.[5] When the Hausdorff distance between our estimator and the population identification region is properly normalized by $\sqrt{n}$, it converges in distribution to a function of a Gaussian process whose covariance kernel depends on parameters of the population identification region which can be consistently estimated. Hence, our first test statistic is given by the Hausdorff distance between the estimated set and the hypothesized population identification region. The null hypothesis is rejected if this statistic exceeds a critical value. We show that this critical value can be consistently estimated using bootstrap methods.

The second statistic is based on the directed Hausdorff distance. Because the directed Hausdorff distance from a nonempty compact set A to a nonempty compact set B is equal to zero if and only if $A \subseteq B$, this distance is ideal to test an inclusion hypothesis which specifies a subset of the identification region. Similarly to the test based on the Hausdorff distance, normalization by $\sqrt{n}$ of the directed Hausdorff distance from the population identification region to the estimator converges in distribution to a function of a Gaussian process. Hence, our second test statistic is given by the directed Hausdorff distance from the hypothesized subset of the population identification region to the estimator. Similarly to the previous case, the null hypothesis is rejected if this test statistic exceeds a critical value that can be consistently estimated using the bootstrap.

We show that the tests that we propose are consistent against any fixed alternative. We then extend the notion of local alternatives to partially identified models, derive the asymptotic distribution of our tests against these local alternatives, and show that the tests are locally asymptotically unbiased.

Our tests can be inverted to yield confidence statements about the true identification region or its subsets. In the case of inversion of the test based on the Hausdorff distance, we obtain what we call a *confidence collection* which 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 $(1 - \alpha)$. In the case of inversion of the test based on the directed Hausdorff distance, we obtain what we call a *directed confidence collection* which is given by the collection of all sets that, when specified as a null hypothesis for a subset of values of the population identification region, cannot be rejected by our test. Also in this case the population identification region is (asymptotically) one of its elements with a prespecified confidence level $(1 - \alpha)$. Additionally, the union of the sets in the directed confidence collection has the property of covering (asymptotically) the true identification region with a prespecified confidence level $(1 - \alpha)$. This result establishes a clear connection between our Wald-type

[5]This statistic can also be used to test hypotheses which specify the entire identification region of linear combinations of the partially identified population features.

test statistic based on the directed Hausdorff distance and the quasi-likelihood ratio (QLR)-type inferential statistic proposed by Chernozhukov, Hong, and Tamer (2007), which allows one to construct confidence sets which asymptotically cover the true identification region with confidence level $(1-\alpha)$. In the special case of interval identified scalar parameters, we establish the asymptotic equivalence between the square of our test statistic based on the directed Hausdorff distance, and the inferential statistic of Chernozhukov, Hong, and Tamer (2007).

Hence, for the class of problems addressed in this paper, 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 (and the directed Hausdorff distance) between the population identification region and its estimator 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 statistics 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 wants to test hypotheses and construct confidence sets for the "true" value of the population feature of interest, following the insight of Imbens and Manski (2004). For this case, our methodology based on the directed Hausdorff distance provides conservative confidence sets that asymptotically cover each point in the identification region with a prespecified probability.

Structure of the Paper

In Section 2 we propose our test statistics, establish their properties, provide a consistent bootstrap procedure to estimate their limiting distributions, show how the test statistics can be inverted to obtain the confidence collection and the directed confidence collection, and provide a simple characterization of the collections. In Section 3 we apply our results to the problem of inference for interval-identified scalar parameters and give a step-by-step description of how our method can be implemented in practice. In Section 4 we apply our results to conduct inference for the best linear predictor parameters when only interval data are available for the outcome variable, but the covariates are perfectly observed. We show how to test linear restrictions on these parameters and we prove that the estimator of the identification region for a single component of this parameter vector can be computed using standard statistical packages. Section 5 presents Monte Carlo results evaluating the finite sample performance of our estimators and test statistics, and Section 6 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 and Tamer (2002), Chernozhukov, Hong, and Tamer (2002, 2007), Honore and Tamer (2006), Andrews, Berry, and Jia (2004), and Chernozhukov, Hahn, and 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) considered 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 proposed confidence intervals that asymptotically cover the entire identification region with fixed probability. For the same class of problems, Imbens and Manski (2004) suggested 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, Hong, and Tamer (2002) were 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 provided 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 developed resampling methods to implement these procedures. Romano and Shaikh (2006) analyzed the same problem and proposed various subsampling procedures. Andrews, Berry, and Jia (2004) considered 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 suggested 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, and Ishii (2005) considered single agent and multiple agent structural models in which again equilibrium conditions impose moment inequality restrictions on the parameters of interest. They suggested a conservative specification test for the value of the estimated parameters. Rosen

(2006) considered related models in which the parameter of interest is partially identified by a finite number of moment inequalities. He established a connection between these models and the literature on multivariate one-sided hypothesis tests and showed 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, and Kim (2006) further explored specification tests for parameter vectors obtained as the solution of moment inequalities. Molinari (2008b) considered 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 proposed conservative confidence sets given by the union of the identification regions obtained by replacing the estimated parameters with the elements of their Wald confidence ellipsoid. Galichon and Henry (2006) proposed a specification test for partially identified structural models based on an extension of the Kolmogorov–Smirnov test for Choquet capacities.

2. HYPOTHESIS TESTING AND CONFIDENCE COLLECTIONS

2.1. *Preliminaries*

We begin this section with some preliminary definitions that prove useful in what follows.[6] Throughout the paper (with a few exceptions), we reserve the use of capital Latin letters to sets and SVRVs; we use lowercase Latin letters for random variables and boldface lowercase Latin letters for random vectors. We denote sets of parameters by capital Greek letters, scalar valued parameters by lowercase Greek letters, and vector valued parameters by boldface lowercase Greek letters. We denote by $\|\cdot\|$ the Euclidean norm, by $\langle\cdot,\cdot\rangle$ the inner product in $\Re^d$, by $\mathbb{S}^{d-1}=\{\mathbf{p}\in\Re^d:\|\mathbf{p}\|=1\}$ the unit sphere in $\Re^d$, and by $\mathbb{C}(\mathbb{S}^{d-1})$ the set of continuous functions from $\mathbb{S}^{d-1}$ to $\Re$. For $h\in\mathbb{C}(\mathbb{S}^{d-1})$, we let $\|h\|_{\mathbb{C}(\mathbb{S}^{d-1})}=\sup_{\mathbf{p}\in\mathbb{S}^{d-1}}|h(\mathbf{p})|$ be the $\mathbb{C}(\mathbb{S}^{d-1})$ norm. We denote $(g)_+\equiv\max(0,g)$ and $(g)_-\equiv\max(0,-g)$.

Set Valued Random Variables, Distance Functions, and Support Functions

Let $(\Omega,\mathcal{A},\mu)$ be a probability space. Let $\Re^d$ denote the Euclidean space, equipped with the Euclidean norm, and let $K(\Re^d)$ be the collection of all nonempty closed subsets of $\Re^d$.[7] Denote by $K_k(\Re^d)$ the set of nonempty compact

[6]Beresteanu and Molinari (2006) provided 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, and Kreinovich (2002)). The first self-contained treatment of the mathematical theory of SVRVs was given by Matheron (1975).

[7]The theory of SVRVs generally applies to $K(\mathfrak{X})$, the space of closed subsets of a Banach space $\mathfrak{X}$ (e.g., Molchanov (2005)). For the purposes of this paper it suffices to consider $\mathfrak{X}=\Re^d$, which simplifies the exposition.

subsets of $\Re^d$, and by $K_{kc}(\Re^d)$ the set of nonempty compact and convex subsets of $\Re^d$.

An SVRV is a measurable mapping $F:\Omega \to K(\Re^d)$ that associates a set to each point in the sample space. We can think of an SVRV as a bundle of random variables—its selections. A formal definition follows.

DEFINITION 1: A map $F:\Omega \to K(\Re^d)$ is called a *set-valued random variable* (SVRV) if for each closed subset C of $\Re^d$, $F^{-1}(C) = \{\omega \in \Omega : F(\omega) \cap C \neq \emptyset\} \in \mathcal{A}$. For any SVRV F, a (measurable) *selection* of F is a random vector $\mathbf{f}$ (taking values in $\Re^d$) such that $\mathbf{f}(\omega) \in F(\omega)$ μ-a.s. We denote by $\mathcal{S}(F)$ the set of all selections from F.

To measure the distance from one set to another, the distance between sets, and the norm of a set, we use the following:

DEFINITION 2: Let A and B be two subsets of $\Re^d$.

(a) The *directed Hausdorff distance* from A to B (or *lower Hausdorff hemimetric*) is denoted

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

(b) The *Hausdorff distance*[8] between A and B (or *Hausdorff metric*) is denoted

$$H(A, B) = \max\{d_H(A, B), d_H(B, A)\}.$$

(c) The *Hausdorff norm* of a set is denoted

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

To represent sets as elements of a functional space, we use the support function:

DEFINITION 3: Let $F \in K(\Re^d)$. Then the *support function* of F at $\mathbf{p} \in \Re^d$, denoted $s(\mathbf{p}, F)$, is given by $s(\mathbf{p}, F) = \sup_{\mathbf{f}\in F} \langle \mathbf{p}, \mathbf{f}\rangle$.

Expectation of SVRVs

Denote by $\mathbf{L}^1 = \mathbf{L}^1(\Omega, \Re^d)$ the space of $\mathcal{A}$-measurable random variables with values in $\Re^d$ such that the $\mathbf{L}^1$-norm $\|\boldsymbol{\xi}\|_1 = E[\|\boldsymbol{\xi}\|]$ is finite. For an SVRV F defined on $(\Omega, \mathcal{A})$ the family of all integrable selections of F is given by

[8]If A and B are unbounded subsets of $\Re^d$, $H(A, B)$ may be infinite. However, $(K_k(\Re^d), H(\cdot, \cdot))$ is a complete and separable metric space (Li, Ogura, and Kreinovich (2002, Theorems 1.1.2 and 1.1.3)). The same conclusion holds for $(K_{kc}(\Re^d), H(\cdot, \cdot))$.

$\mathcal{S}^1(F) = \mathcal{S}(F) \cap \mathbf{L}^1$. Below we define integrably bounded SVRVs, and we define the Aumann expectation (Aumann (1965)) of an SVRV F, denoted $\mathbb{E}[F]$. We reserve the notation $E[\cdot]$ for the expectation of random variables and random vectors.

DEFINITION 4: An SVRV $F : \Omega \to K(\Re^d)$ is called *integrably bounded* if $\|F\|_H = \sup\{\|\mathbf{f}\| : \mathbf{f} \in F\}$ has a finite expectation.

DEFINITION 5: 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 coordinatewise. 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 nonempty (Aumann (1965); see also Li, Ogura, and Kreinovich (2002, Theorem 1.2.6)), the Aumann expectation of an integrably bounded SVRV is nonempty. Moreover, if F is an integrably bounded random compact set on a nonatomic probability space or if F is an integrably bounded random convex compact set, then $\mathbb{E}[F]$ is convex and coincides with $\mathbb{E}[\mathrm{co}\, F]$, and $E[s(\mathbf{p}, F)] = s(\mathbf{p}, \mathbb{E}[F])$ (Artstein (1974)).

Linear Transformations

Sections 3 and 4 discuss examples of partially identified parameters whose identification region is given by an Aumann expectation of SVRVs that can be constructed from observable random variables. Below we introduce novel general procedures based on the Hausdorff distance and on the directed Hausdorff distance to conduct inference for these identification regions and their subsets. Before getting to the details of our procedure, we observe that given a nonrandom finite matrix $\mathcal{R}$ of dimension $l \times d$, if $\{F, F_i : i \in \mathbb{N}\}$ are independent and identically distributed (i.i.d.) nonempty, compact valued SVRVs in $K_k(\Re^d)$ with $E[\|F\|_H^2] < \infty$ and $\mathcal{R}F = \{\mathbf{t} \in \Re^l : \mathbf{t} = \mathcal{R}\mathbf{f}, \mathbf{f} \in \mathcal{S}(F)\}$, then $\{\mathcal{R}F, \mathcal{R}F_i : i \in \mathbb{N}\}$ are i.i.d. nonempty, compact valued SVRVs in $K_k(\Re^l)$ with $E[\|\mathcal{R}F\|_H^2] < \infty$. It then follows from Theorem A.2 and Lemma A.1 in the Appendix that

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

$$\sqrt{n} d_H\left(\frac{1}{n} \bigoplus_{i=1}^n \mathcal{R}F_i, \mathbb{E}[\mathcal{R}F] \right) \xrightarrow{d} \sup_{\mathbf{p} \in \mathbb{S}^{l-1}} \{z^{\mathcal{R}}(\mathbf{p})\}_+,$$

$$\sqrt{n}d_H\left(\mathbb{E}[\mathcal{R}F], \frac{1}{n}\bigoplus_{i=1}^{n}\mathcal{R}F_i\right) \xrightarrow{d} \sup_{\mathbf{p}\in\mathbb{S}^{l-1}}\{-z^{\mathcal{R}}(\mathbf{p})\}_+,$$

where $z^{\mathcal{R}}$ is a Gaussian random system with $E[z^{\mathcal{R}}(\mathbf{p})] = 0$ for all $\mathbf{p} \in \mathbb{S}^{l-1}$ and $E[z^{\mathcal{R}}(\mathbf{p})z^{\mathcal{R}}(\mathbf{q})] = E[s(\mathcal{R}'\mathbf{p}, F)s(\mathcal{R}'\mathbf{q}, F)] - E[s(\mathcal{R}'\mathbf{p}, F)]E[s(\mathcal{R}'\mathbf{q}, F)]$ for all $\mathbf{p}, \mathbf{q} \in \mathbb{S}^{l-1}$. In particular, $\mathcal{R}$ can be chosen to select specific components of the partially identified parameter vector (or linear functions of such components) so as to conduct inference on them. Below, for simplicity, we assume that $\mathcal{R} = I$ (the identity matrix). The results for $\mathcal{R} \neq I$ involve a straightforward modification.

2.2. *Inference Based on the Hausdorff Distance*

Hypothesis Testing

Given a nonempty, compact, and convex (nonrandom) set Ψ_0, to test the hypothesis

$$\mathfrak{H}_0 : \mathbb{E}[F] = \Psi_0, \quad \mathfrak{H}_A : \mathbb{E}[F] \neq \Psi_0$$

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

$$\text{reject } \mathfrak{H}_0 \quad \text{if } \sqrt{n}H\left(\frac{1}{n}\bigoplus_{i=1}^{n}\operatorname{co}F_i, \Psi_0\right) > c_\alpha,$$

$$\text{do not reject } \mathfrak{H}_0 \quad \text{if } \sqrt{n}H\left(\frac{1}{n}\bigoplus_{i=1}^{n}\operatorname{co}F_i, \Psi_0\right) \leq c_\alpha,$$

where c_α is chosen so that

$$\Pr\left\{\|z\|_{\mathbb{C}(\mathbb{S}^{d-1})} > c_\alpha\right\} = \alpha \tag{2.1}$$

and z is a Gaussian random system with $E[z(\mathbf{p})] = 0$ for all $\mathbf{p} \in \mathbb{S}^{d-1}$ and $E[z(\mathbf{p})z(\mathbf{q})] = E[s(\mathbf{p}, F)s(\mathbf{q}, F)] - E[s(\mathbf{p}, F)]E[s(\mathbf{q}, F)]$ for all $\mathbf{p}, \mathbf{q} \in \mathbb{S}^{d-1}$.

Observe that the test statistic uses the Hausdorff distance between Ψ_0 and $\frac{1}{n}\bigoplus_{i=1}^{n}\operatorname{co}F_i$, rather than the Hausdorff distance between Ψ_0 and $\frac{1}{n}\bigoplus_{i=1}^{n}F_i$. Using $\bigoplus_{i=1}^{n}\operatorname{co}F_i$ to construct the test statistic greatly simplifies the computations, because the calculation of the Hausdorff distance between convex sets has been widely studied in computational geometry, and fast algorithms are easily implementable. At the same time, by Shapley–Folkman's theorem (Starr (1969)), $H(\frac{1}{n}\bigoplus_{i=1}^{n}\operatorname{co}F_i, \frac{1}{n}\bigoplus_{i=1}^{n}F_i)$ goes to zero[9] at the rate $\frac{1}{n}$. Hence, by Theorem A.1 in the Appendix, $\frac{1}{n}\bigoplus_{i=1}^{n}\operatorname{co}F_i$ is a consistent estimator of $\mathbb{E}[F]$, and by

[9]Minkowski averaging SVRVs which are compact valued but not necessarily convex is asymptotically "convexifying," as noted by Artstein and Vitale (1975).

Theorem A.2, $\sqrt{n}H(\frac{1}{n}\bigoplus_{i=1}^{n}\operatorname{co}F_i, \mathbb{E}[F])$ and $\sqrt{n}H(\frac{1}{n}\bigoplus_{i=1}^{n}F_i, \mathbb{E}[F])$ have the same asymptotic distribution. In what follows, we use $\bar{F}_n$ to denote $\frac{1}{n}\bigoplus_{i=1}^{n}\operatorname{co}F_i$.

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

ALGORITHM 2.1:

1. Generate a bootstrap sample of size n, $\{F_i^*: i = 1, \dots, n\}$, by drawing a random sample from the joint empirical distribution of $\{F_i: i = 1, \dots, n\}$ with replacement.
2. Compute

$$r_n^* \equiv \sqrt{n}H(\bar{F}_n^*, \bar{F}_n). \tag{2.2}$$

3. Use the results of b repetitions of Steps 1 and 2 to compute the empirical distribution function of r_n^* at a point t, denoted by $J_n(t)$.
4. Estimate the quantile c_α defined in equation (2.1) by

$$\hat{c}_{\alpha n} = \inf\{t: J_n(t) \geq 1 - \alpha\}. \tag{2.3}$$

The results of Giné and 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 2.1: *Let $\{F, F_i: i \in \mathbb{N}\}$ be i.i.d. nonempty, compact valued SVRVs such that $E[\|F\|_H^2] < \infty$. Then $r_n^* \xrightarrow{d} \|z\|_{\mathbb{C}(\mathbb{S}^{d-1})}$, where r_n^* is defined in equation (2.2) and z is a Gaussian random system with $E[z(\mathbf{p})] = 0$ for all $\mathbf{p} \in \mathbb{S}^{d-1}$ and with covariance kernel $E[z(\mathbf{p})z(\mathbf{q})] = E[s(\mathbf{p}, F)s(\mathbf{q}, F)] - E[s(\mathbf{p}, F)]E[s(\mathbf{q}, F)]$ for all $\mathbf{p}, \mathbf{q} \in \mathbb{S}^{d-1}$. If in addition* $\operatorname{Var}(z(\mathbf{p})) > 0$ *for each $\mathbf{p} \in \mathbb{S}^{d-1}$, then $\hat{c}_{\alpha n} = c_\alpha + o_p(1)$, where $\hat{c}_{\alpha n}$ is defined in (2.3).*

A consistent estimator of the critical value c_α can be alternatively obtained using the following procedure. Simulate the distribution of the supremum of a Gaussian random system with mean function equal to zero for each $\mathbf{p} \in \mathbb{S}^{d-1}$ and covariance kernel equal to

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

for all $\mathbf{p}, \mathbf{q} \in \mathbb{S}^{d-1}$. Observe that the supremum is to be taken over $\mathbf{p} \in \mathbb{S}^{d-1}$, 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

$$\hat{\gamma}(\cdot, \cdot) \xrightarrow{\text{a.s.}} E[s(\cdot, F)s(\cdot, F)] - E[s(\cdot, F)]E[s(\cdot, F)].$$

By the same argument as in the proof of Proposition 2.1, if $\mathrm{Var}(z(\mathbf{p}))>0$ for each $\mathbf{p}\in\mathbb{S}^{d-1}$, the critical values of the simulated distribution consistently estimate the critical values of $\|z\|_{\mathbb{C}(\mathbb{S}^{d-1})}$.

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

THEOREM 2.2: *Let* $\{F, F_i : i\in\mathbb{N}\}$ *be i.i.d. nonempty, compact valued SVRVs such that* $E[\|F\|_H^2]<\infty$, *and let* Ψ_A *be a nonempty, compact, and convex valued set such that* $\mathbb{E}[F]=\Psi_A\neq\Psi_0$. *Let* $\mathrm{Var}(z(\mathbf{p}))>0$ *for each* $\mathbf{p}\in\mathbb{S}^{d-1}$. *Then*

$$\lim_{n\to\infty}\Pr\{\sqrt{n}H(\bar{F}_n,\Psi_0)>\hat{c}_{\alpha n}\}=1.$$

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 nonempty, compact, and convex alternative sets $\{\Psi_{An}\}$,

$$\Psi_{An}\in K_{kc}(\Re^d):\quad \Psi_{An}\oplus\frac{1}{\sqrt{n}}\Delta_1=\Psi_0\oplus\frac{1}{\sqrt{n}}\Delta_2, \tag{2.4}$$

where Δ_1 and Δ_2 are nonrandom nonempty, compact, and convex sets for which there exists a nonempty, compact, and convex set Δ_3 such that $\Psi_0=\Delta_1\oplus\Delta_3$. Let $\kappa\equiv H(\Delta_1,\Delta_2)<\infty$ and observe that by the properties of the Hausdorff distance between two sets[10]

$$\sqrt{n}H(\Psi_{An},\Psi_0)=\sqrt{n}H\left(\Psi_{An}\oplus\frac{1}{\sqrt{n}}\Delta_1,\Psi_0\oplus\frac{1}{\sqrt{n}}\Delta_1\right)=\kappa. \tag{2.5}$$

Clearly, for larger values of κ, the local alternatives get farther away from the null, with a resulting increase in the power of the test.

This choice of local alternatives allows us to consider a large class of deviations from the null hypothesis. It encompasses the case in which $\Delta_1=\{0\}$, $\Delta_2\neq\{0\}$, and the sets Ψ_{An} shrink (i.e., the null is a subset of the true identification region) and/or shift to become equal to Ψ_0, and the case in which $\Delta_1\neq\{0\}$, $\Delta_2=\{0\}$, and the sets Ψ_{An} enlarge (i.e., the null is a superset of the true identification region) and/or shift to become equal to Ψ_0.

The following theorem gives the asymptotic distribution of our test under these local alternatives and establishes its local asymptotic unbiasedness.

THEOREM 2.3: *Let* $\{F, F_i : i\in\mathbb{N}\}$ *be i.i.d. nonempty, compact valued SVRVs such that* $E[\|F\|_H^2]<\infty$, *let* Ψ_0 *be a nonempty, compact, and convex valued set, and let* $\{\Psi_{An}\}$ *be the sequence of sets defined in* (2.4). *Then*

$$\sqrt{n}H(\bar{F}_n,\Psi_0)\xrightarrow{d}\|w\|_{\mathbb{C}(\mathbb{S}^{d-1})}$$

[10]See, for example, DeBlasi and Iervolino (1969).

under Ψ_{An}, where w is a Gaussian random system with $E[w(\mathbf{p})] = s(\mathbf{p}, \Delta_2) - s(\mathbf{p}, \Delta_1)$ for all $\mathbf{p} \in \mathbb{S}^{d-1}$ and $E[w(\mathbf{p})w(\mathbf{q})] = E[s(\mathbf{p}, F)s(\mathbf{q}, F)] - E[s(\mathbf{p}, F)]E[s(\mathbf{q}, F)]$ for all $\mathbf{p}, \mathbf{q} \in \mathbb{S}^{d-1}$. Moreover, under Ψ_{An} the test is asymptotically locally unbiased, that is,

$$\lim_{n\to\infty} \Pr\{\sqrt{n}H(\bar{F}_n, \Psi_0) > \hat{c}_{\alpha n}\} \geq \alpha.$$

Confidence Collections

The "confidence collection" 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 $CC_{n,1-\alpha}$. 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 $CC_{n,1-\alpha}$ as

$$CC_{n,1-\alpha} = \{\tilde{\Psi} \in K_{kc}(\Re^d) : \sqrt{n}H(\bar{F}_n, \tilde{\Psi}) \leq \hat{c}_{\alpha n}\},$$

where $\hat{c}_{\alpha n}$ is defined in equation (2.3). Hence if the law of $\|z\|_{\mathbb{C}(\mathbb{S}^{d-1})}$ is continuous,

$$\lim_{n\to\infty} \Pr\{\mathbb{E}[F] \in CC_{n,1-\alpha}\} = 1 - \alpha.$$

In practice, it can be difficult to characterize all the sets that belong to the confidence collection. However, the union of all sets in $CC_{n,1-\alpha}$ can be calculated in a particularly simple way. The following theorem shows how.

THEOREM 2.4: *Let $\mathcal{U}_n = \bigcup\{\tilde{\Psi} : \tilde{\Psi} \in CC_{n,1-\alpha}\}$ and $B_{\hat{c}_{\alpha n}} = \{\mathbf{b} \in \Re^d : \|\mathbf{b}\| \leq \frac{\hat{c}_{\alpha n}}{\sqrt{n}}\}$. Then $\mathcal{U}_n = \bar{F}_n \oplus B_{\hat{c}_{\alpha n}}$.*

An interesting consequence of Theorem 2.4 is that the union of all sets in $CC_{n,1-\alpha}$ is also included in $CC_{n,1-\alpha}$ and thus represents the largest set that cannot be rejected as a null hypothesis.

2.3. *Inference Based on the Directed Hausdorff Distance*

Hypothesis Testing

Suppose that a researcher wants to test whether a certain set of values is contained in the identification region. To accommodate this case, we propose the following hypothesis, which specifies a subset of the identification region[11]:

$$\mathfrak{H}_0 : \Psi_0 \subseteq \mathbb{E}[F], \quad \mathfrak{H}_A : \Psi_0 \nsubseteq \mathbb{E}[F],$$

where Ψ_0 is a nonempty, compact, and convex (nonrandom) set. We call this an inclusion test. Using the triangle inequality and the fact that under the null hypothesis $d_H(\Psi_0, \mathbb{E}[F]) = 0$, we obtain

$$\begin{aligned} d_H(\Psi_0, \bar{F}_n) &\leq d_H(\Psi_0, \mathbb{E}[F]) + d_H(\mathbb{E}[F], \bar{F}_n) \\ &= \sup_{\mathbf{p} \in \mathbb{S}^{d-1}} \left\{ -\left(\frac{1}{n} \sum_{i=1}^{n} s(\mathbf{p}, \operatorname{co} F_i) - E[s(\mathbf{p}, F)] \right) \right\}_+, \end{aligned}$$

where the equality follows from Lemma A.1 in the Appendix. Under the same assumptions of Theorem A.2, it follows that

$$\sqrt{n} d_H(\mathbb{E}[F], \bar{F}_n) \xrightarrow{d} \sup_{\mathbf{p} \in \mathbb{S}^{d-1}} \{-z(\mathbf{p})\}_+, \tag{2.6}$$

where z is a Gaussian random system with $E[z(\mathbf{p})] = 0$ for all $\mathbf{p} \in \mathbb{S}^{d-1}$ and $E[z(\mathbf{p})z(\mathbf{q})] = E[s(\mathbf{p}, F)s(\mathbf{q}, F)] - E[s(\mathbf{p}, F)]E[s(\mathbf{q}, F)]$ for all $\mathbf{p}, \mathbf{q} \in \mathbb{S}^{d-1}$.

Hence we can use the criterion

$$\begin{aligned} &\text{reject } \mathfrak{H}_0 \quad \text{if } \sqrt{n} d_H(\Psi_0, \bar{F}_n) > \tilde{c}_\alpha, \\ &\text{do not reject } \mathfrak{H}_0 \quad \text{if } \sqrt{n} d_H(\Psi_0, \bar{F}_n) \leq \tilde{c}_\alpha, \end{aligned}$$

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

$$\Pr\left\{ \sup_{\mathbf{p} \in \mathbb{S}^{d-1}} \{-z(\mathbf{p})\}_+ > \tilde{c}_\alpha \right\} = \alpha. \tag{2.7}$$

By construction, denoting by $\widehat{\tilde{c}}_{\alpha n}$ an estimator of $\tilde{c}_\alpha$ obtained through a bootstrap procedure similar to the one in Algorithm 2.1,

$$\lim_{n \to \infty} \Pr\{\sqrt{n} d_H(\Psi_0, \bar{F}_n) > \widehat{\tilde{c}}_{\alpha n}\} \leq \lim_{n \to \infty} \Pr\left\{\sqrt{n} d_H(\mathbb{E}[F], \bar{F}_n) > \widehat{\tilde{c}}_{\alpha n}\right\} = \alpha.$$

Because $\Psi_0 = \mathbb{E}[F]$ is contained in $\mathfrak{H}_0$, our test's significance level is asymptotically equal to α.

Our testing procedure preserves the property of rejecting a false null hypothesis with probability approaching 1 as the sample size increases. In particular, we have the following result:

[11] For the case that one wants to test $\mathfrak{H}_0 : \mathbb{E}[F] \subseteq \Psi_0$ against $\mathfrak{H}_A : \mathbb{E}[F] \nsubseteq \Psi_0$, similar algebra to what follows in this section gives that we can use the criterion reject $\mathfrak{H}_0$ if $\sqrt{n} d_H(\bar{F}_n, \Psi_0) > \check{c}_\alpha$, do not reject $\mathfrak{H}_0$ if $\sqrt{n} d_H(\bar{F}_n, \Psi_0) \leq \check{c}_\alpha$, where $\check{c}_\alpha$ is chosen so that $\Pr\{\sup_{\mathbf{p} \in \mathbb{S}^{d-1}} \{z(\mathbf{p})\}_+ > \check{c}_\alpha\} = \alpha$ and $\sup_{\mathbf{p} \in \mathbb{S}^{d-1}} \{z(\mathbf{p})\}_+$ is the limiting distribution of $\sqrt{n} d_H(\bar{F}_n, \mathbb{E}[F])$.

COROLLARY 2.5: *Let* $\{F, F_i : i \in \mathbb{N}\}$ *be i.i.d. nonempty, compact valued SVRVs such that* $E[\|F\|_H^2] < \infty$ *and suppose that* $\Psi_0 \nsubseteq \mathbb{E}[F]$. *Let* $\alpha < \frac{1}{2}$ *and* $\mathrm{Var}(z(\mathbf{p})) > 0$ *for each* $\mathbf{p} \in \mathbb{S}^{d-1}$, *where* $z(\mathbf{p})$ *is defined in equation* (2.6). *Then*

$$\lim_{n\to\infty} \Pr\{\sqrt{n} d_H(\Psi_0, \bar{F}_n) > \widehat{c}_{\alpha n}\} = 1.$$

The asymptotic distribution of this test under local alternatives at distance (proportional to) $1/\sqrt{n}$ from the null hypothesis follows from Theorem 2.3. In particular, suppose we are interested in the power of our test of $\mathfrak{H}_0$ against a sequence of nonempty, compact, and convex alternative sets $\{\Psi_{An}\}$,

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

where Δ_1 is a nonrandom nonempty, compact, and convex set for which there exists a nonempty, compact and convex set Δ_3 such that $\Psi_0 = \Delta_1 \oplus \Delta_3$. Then

$$\sqrt{n} d_H(\Psi_0, \bar{F}_n) \xrightarrow{d} \sup_{\mathbf{p}\in\mathbb{S}^{d-1}} \{-w(\mathbf{p})\}_+ \tag{2.8}$$

under Ψ_{An}, where w is a Gaussian random system with $E[w(\mathbf{p})] = -s(\mathbf{p}, \Delta_1)$ for all $\mathbf{p} \in \mathbb{S}^{d-1}$ and $E[w(\mathbf{p})w(\mathbf{q})] = E[s(\mathbf{p}, F)s(\mathbf{q}, F)] - E[s(\mathbf{p}, F)]E[s(\mathbf{q}, F)]$ for all $\mathbf{p}, \mathbf{q} \in \mathbb{S}^{d-1}$. When additionally Δ_1 is such that $s(\mathbf{p}, \Delta_1) \geq 0$ for all $\mathbf{p} \in \mathbb{S}^{d-1}$, so that $\Psi_{An} \subseteq \Psi_0$, it is easy to see that the test is asymptotically locally unbiased. We return to the implications of this result when we discuss the inversion of this test to obtain confidence sets.

Directed Confidence Collections

The "directed confidence collection" is the collection of all sets that, when specified as a null hypothesis for a possible subset of the population identification region, cannot be rejected by our test. The term "directed" is to emphasize the connection with the test statistic that uses the directed Hausdorff distance. We denote this collection by $DCC_{n,1-\alpha}$:

$$DCC_{n,1-\alpha} = \{\tilde{\Psi} \in K_{kc}(\Re^d) : \sqrt{n} d_H(\tilde{\Psi}, \bar{F}_n) \leq \widehat{c}_{\alpha n}\}.$$

If the law of $\sup_{\mathbf{p}\in\mathbb{S}^{d-1}}\{-z(\mathbf{p})\}_+$ is continuous, then for each $\tilde{\Psi} \subseteq \mathbb{E}[F]$,

$$\lim_{n\to\infty} \Pr\{\tilde{\Psi} \in DCC_{n,1-\alpha}\} \geq 1 - \alpha$$

with equality for $\tilde{\Psi} = \mathbb{E}[F]$. Similarly to the result presented in Theorem 2.4, the union of the sets in $DCC_{n,1-\alpha}$ can be calculated in a particularly simple way and it represents the largest set that cannot be rejected as a null hypothesis in the inclusion test.

COROLLARY 2.6: *Let* $\mathcal{DU}_n = \bigcup\{\tilde{\Psi} : \tilde{\Psi} \in DCC_{n,1-\alpha}\}$ *and* $B_{\widehat{\bar{c}}_{\alpha n}} = \{\mathbf{b} \in \Re^d : \|\mathbf{b}\| \leq \frac{\widehat{\bar{c}}_{\alpha n}}{\sqrt{n}}\}$. *Then* $\mathcal{DU}_n = \bar{F}_n \oplus B_{\widehat{\bar{c}}_{\alpha n}} \subset \mathcal{U}_n$.

Confidence Sets

Chernozhukov, Hong, and Tamer (2007) proposed criterion-function based (QLR-type) confidence sets that cover the true identification region with probability asymptotically equal to $1-\alpha$. We show here that for the class of problems addressed in this paper, our Wald-type confidence set $\mathcal{DU}_n$ possesses (asymptotically) the same coverage property.

PROPOSITION 2.7: *Let* $\{F, F_i : i \in \mathbb{N}\}$ *be i.i.d. nonempty, compact valued SVRVs such that* $E[\|F\|_H^2] < \infty$. *Then*

$$\lim_{n\to\infty} \Pr\{\mathbb{E}[F] \subseteq \mathcal{DU}_n\} = 1-\alpha.$$

This proposition also implies that our results for the local power of the inclusion test based on the directed Hausdorff distance can be translated into results for coverage of a false local region. In particular, let

$$\Psi_{0n} = \mathbb{E}[F] \oplus \frac{1}{\sqrt{n}}\Delta_1.$$

Then $\sqrt{n}d_H(\Psi_{0n}, \bar{F}_n) \xrightarrow{d} \sup_{\mathbf{p}\in\mathbb{S}^{d-1}}\{-w(\mathbf{p})\}_+$, where the Gaussian process $w(\cdot)$ is defined after equation (2.8). Hence it follows from the discussion after equation (2.8) that whenever Δ_1 is such that $s(\mathbf{p}, \Delta_1) \geq 0$ for all $\mathbf{p} \in \mathbb{S}^{d-1}$, $\lim_{n\to\infty} \Pr\{\sqrt{n}d_H(\Psi_{0n}, \bar{F}_n) > \widehat{\bar{c}}_{\alpha n}\} \geq \alpha$, and in this case

$$\lim_{n\to\infty} \Pr\{\Psi_{0n} \subseteq \mathcal{DU}_n\} \leq 1-\alpha.$$

Testing Procedure and Confidence Sets for Points in the Identification Region

The inferential approach in this section targets the entire identification region of a partially identified population feature, and provides asymptotically exact significance level with which to test hypotheses and construct confidence collections. However, there are applications in which the researcher is interested in testing hypotheses and constructing confidence sets for the "true" value of the population feature, following the insight of Imbens and Manski (2004). For this case, our test statistic based on the directed Hausdorff distance allows one to conduct conservative tests of hypotheses and construct conservative confidence sets that asymptotically cover each point in the identification region with a prespecified probability. In particular, given a real valued vector ψ_0 and a prespecified significance level $\alpha \in (0,1)$, suppose that one wants to test $\mathfrak{H}_0 : \psi_0 \in \mathbb{E}[F]$ (i.e., $\{\psi_0\} \subseteq \mathbb{E}[F]$) against $\mathfrak{H}_A : \psi_0 \notin \mathbb{E}[F]$ (i.e., $\{\psi_0\} \nsubseteq \mathbb{E}[F]$).

Then one can reject $\mathfrak{H}_0$ against $\mathfrak{H}_A$ if $\sqrt{n}d_H(\{\psi_0\}, \bar{F}_n) > \tilde{c}_\alpha$, and fail to reject otherwise, where $\tilde{c}_\alpha$ is given in equation (2.7). This test preserves the property of rejecting a false null hypothesis with probability approaching 1 as the sample size increases. The confidence set $\mathcal{DU}_n$ introduced in Corollary 2.6 has the property that for each $\tilde{\psi} \in \mathbb{E}[F]$, $\lim_{n\to\infty} \Pr\{\tilde{\psi} \in \mathcal{DU}_n\} \geq 1-\alpha$. Clearly this confidence set is conservative, as illustrated by Imbens and Manski (2004, Lemma 1), because $\mathcal{DU}_n$ covers the entire identification region with probability asymptotically equal to $1-\alpha$.

3. INFERENCE FOR INTERVAL-IDENTIFIED PARAMETERS

The results of Section 2 can be extended to the entire class of models that give interval bounds (in $\Re$) for the parameter of interest (Manski (2003) gave a comprehensive presentation of a large group of such problems). The only requirement is that there is joint asymptotic normality of the endpoints of the interval and that the covariance matrix of the limiting distribution can be consistently estimated. This result is useful in practice, as many applications of partially identified models fall into this category.[12] It also helps us relate the theory of SVRVs to the well-known laws of large numbers (LLNs) and central limit theorems (CLTs) for scalar valued random variables. Here we specialize the limiting distribution of our test statistics for the case of interval identification, and detail a step-by-step procedure to test hypothesis and construct confidence collections.

Suppose that the population identification region for a scalar valued parameter of interest is given by the interval $\Psi \equiv [\psi_L, \psi_U]$. Let $\bar{Y}_n \equiv [\bar{y}_{nL}, \bar{y}_{nU}]$ denote the estimated identification region and denote by a_n a growing sequence (at a possibly nonparametric rate). Then if

$$(3.1)\qquad a_n \begin{pmatrix} \bar{y}_{nL} - \psi_L \\ \bar{y}_{nU} - \psi_U \end{pmatrix} \xrightarrow{d} \begin{pmatrix} z_{-1} \\ z_1 \end{pmatrix} \sim N(\mathbf{0}, \Pi),$$

simple algebra and the continuous mapping theorem imply that

$$(3.2)\qquad a_n H(\bar{Y}_n, \Psi) = a_n \max\{|\bar{y}_{nL} - \psi_L|, |\bar{y}_{nU} - \psi_U|\} \xrightarrow{d} \max\{|z_{-1}|, |z_1|\},$$

$$(3.3)\qquad a_n d_H(\Psi, \bar{Y}_n) = a_n \max\{(\bar{y}_{nL} - \psi_L)_+, (\bar{y}_{nU} - \psi_U)_-\} \xrightarrow{d} \max\{(z_{-1})_+, (z_1)_-\}.$$

The result in equation (3.3) allows us to establish the asymptotic equivalence of the square of our test statistic based on the directed Hausdorff distance and

[12] A few examples include Manski, Sandefur, McLanahan, and Powers (1992), Hotz, Mullin, and Sanders (1997), Manski and Nagin (1998), Ginther (2000), Manski and Pepper (2000), Pepper (2000, 2003), Haile and Tamer (2003), Scharfstein, Manski, and Anthony (2004), Gonzalez-Luna (2005), Molinari (2008a), Dominitz and Sherman (2006).

the criterion-function based inferential statistic proposed by Chernozhukov, Hong, and Tamer (2007). While the result that we give is mathematically simple and does not need a formal proof, we view it as conceptually important. It establishes, for the interval-identified case, an asymptotic equivalence result which corresponds, in the point-identified case, to the asymptotic equivalence between the distributions of the QLR and the Wald test statistics. We therefore place this result in a formal theorem.

THEOREM 3.1: *Let $\Psi \equiv [\psi_L, \psi_U]$ and $\bar{Y}_n \equiv [\bar{y}_{nL}, \bar{y}_{nU}]$. Denote by a_n a growing sequence. Then if equation* (3.1) *holds*,

$$a_n^2 [d_H(\Psi, \bar{Y}_n)]^2 \xrightarrow{d} \max\{(z_{-1})_+^2, (z_1)_-^2\}, \tag{3.4}$$

$$a_n^2 \sup_{\psi \in \Psi} [(\bar{y}_{nL} - \psi)_+^2 + (\bar{y}_{nU} - \psi)_-^2] \xrightarrow{d} \max\{(z_{-1})_+^2, (z_1)_-^2\}. \tag{3.5}$$

The result in equation (3.5) *is due to* Chernozhukov, Hong, and Tamer (2002).

Replacing Π with a consistent estimator $\hat{\Pi}$, the quantiles of the cumulative distribution function of each of the random variables appearing on the right-hand side of the limits in equations (3.2) and (3.3) can be easily estimated using the following procedure, which also details how to test the hypothesis about Ψ.

Algorithm for Estimation of the Critical Values and Hypothesis Testing

1. Suppose $\mathfrak{H}_0 : \Psi = \Psi_0$, $\mathfrak{H}_A : \Psi \neq \Psi_0$. Then:

(a) Draw a large random sample of pairs $(\hat{z}_{-1}, \hat{z}_1)$ from the distribution $N(0, \hat{\Pi})$. For each pair compute $r^* = \max\{|\hat{z}_{-1}|, |\hat{z}_1|\}$.

(b) Use the results of step (a) to compute the empirical distribution function of r^*, $\hat{J}(\cdot)$.

(c) Estimate the quantile $c_\alpha : \Pr\{\max\{|z_{-1}|, |z_1|\} > c_\alpha\} = \alpha$ by

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

(d) Calculate $a_n H(\bar{Y}_n, \Psi_0) = a_n \max\{|\bar{y}_{nL} - \psi_{0L}|, |\bar{y}_{nU} - \psi_{0U}|\}$.

(e) If $a_n H(\bar{Y}_n, \Psi_0) > \hat{c}_{\alpha n}$, reject $\mathfrak{H}_0$ at the $100\alpha\%$ level; otherwise fail to reject.

2. Suppose $\mathfrak{H}_0 : \Psi_0 \subseteq \Psi$, $\mathfrak{H}_A : \Psi \nsubseteq \Psi_0$. Then:

(a) Draw a large random sample of pairs $(\hat{z}_{-1}, \hat{z}_1)$ from the distribution $N(0, \hat{\Pi})$. For each pair compute $\tilde{r}^* = \max\{(\hat{z}_{-1})_+, (\hat{z}_1)_-\}$.

(b) Use the results of step (a) to compute the empirical distribution function of $\tilde{r}^*$, $\widehat{\tilde{J}}(\cdot)$.

(c) Estimate the quantile $\tilde{c}_\alpha : \Pr\{\max\{(z_{-1})_+, (z_1)_-\} > \tilde{c}_\alpha\} = \alpha$ by

$$\widehat{\tilde{c}}_{\alpha n} = \inf\{t : \widehat{\tilde{J}}(t) \geq 1 - \alpha\}.$$

(d) Calculate $a_n d_H(\Psi_0, \bar{Y}_n) = a_n \max\{(\bar{y}_{nL} - \psi_{0L})_+, (\bar{y}_{nU} - \psi_{0U})_-\}$.

(e) If $a_n d_H(\Psi_0, \bar{Y}_n) > \widehat{\tilde{c}}_{\alpha n}$, reject $\mathfrak{H}_0$ at the $100\alpha\%$ level, otherwise fail to reject.

Construction of $CC_{n,1-\alpha}$, $DCC_{n,1-\alpha}$, $\mathcal{U}_n$, and $\mathcal{DU}_n$, and Choice Among Them

In the case of interval-identified parameters, $CC_{n,1-\alpha}$, $DCC_{n,1-\alpha}$, $\mathcal{U}_n$, and $\mathcal{DU}_n$ are extremely easy to construct. In particular

$$CC_{n,1-\alpha} = \left\{[\psi_{0L}, \psi_{0U}] : \psi_{0L} \leq \psi_{0U}, \psi_{0L} \in \left[\bar{y}_{nL} - \frac{\hat{c}_{\alpha n}}{a_n}, \bar{y}_{nL} + \frac{\hat{c}_{\alpha n}}{a_n}\right], \psi_{0U} \in \left[\bar{y}_{nU} - \frac{\hat{c}_{\alpha n}}{a_n}, \bar{y}_{nU} + \frac{\hat{c}_{\alpha n}}{a_n}\right]\right\},$$

$$DCC_{n,1-\alpha} = \left\{[\psi_{0L}, \psi_{0U}] : \bar{y}_{nL} - \frac{\widehat{\tilde{c}}_{\alpha n}}{a_n} \leq \psi_{0L} \leq \psi_{0U} \leq \bar{y}_{nU} + \frac{\widehat{\tilde{c}}_{\alpha n}}{a_n}\right\},$$

$$\mathcal{U}_n = \left[\bar{y}_{nL} - \frac{\hat{c}_{\alpha n}}{a_n}, \bar{y}_{nU} + \frac{\hat{c}_{\alpha n}}{a_n}\right],$$

$$\mathcal{DU}_n = \left[\bar{y}_{nL} - \frac{\widehat{\tilde{c}}_{\alpha n}}{a_n}, \bar{y}_{nU} + \frac{\widehat{\tilde{c}}_{\alpha n}}{a_n}\right].$$

When the researcher is not comfortable conjecturing the entire identification region of the parameter of interest, but only a subset of it, $\mathcal{DU}_n$ is the proper confidence set to report. This set, obtained through the inversion of the Wald statistic based on the directed Hausdorff distance, answers the question "What are the values that cannot be rejected as subsets of the identification region, given the available data and the maintained assumptions?"

When the researcher is interested in answering the question "What are the sets that can be equal to the entire identification region of the parameter of interest, given the available data and the maintained assumptions?," the proper confidence statement to make is based on $CC_{n,1-\alpha}$. This is the collection of sets obtained through the inversion of the Wald statistic based on the Hausdorff distance. While $\mathcal{DU}_n$ is a smaller confidence set than $\mathcal{U}_n$ (the union of the intervals in $CC_{n,1-\alpha}$), it is important to observe that not all proper subsets of $\mathcal{DU}_n$ are elements of $CC_{n,1-\alpha}$. A proper subset of $\mathcal{DU}_n$ might be rejected as a null hypothesis for the entire identification region, for example, because it might be too small and therefore not be an element of $CC_{n,1-\alpha}$.

EXAMPLE—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 $\Pr\{y_L \leq y \leq y_U\} = 1$. Manski (1989) showed that $[E(y_L), E(y_U)]$ is the sharp bound for

$E[y]$. Hence the results in equations (3.2) and (3.3) directly apply here, with $\bar{Y}_n = [\frac{1}{n}\sum_{i=1}^n y_{iL}, \frac{1}{n}\sum_{i=1}^n y_{iU}]$. To exemplify our approach in the simplest possible setting, we rederive the result in equations (3.2) and (3.3) using the language of SVRVs.

Let $Y = [y_L, y_U]$. Assume that $\{(y_{iL}, y_{iU})\}_{i=1}^n$ are i.i.d. random vectors, let $Y_i = [y_{iL}, y_{iU}]$, and denote $\bar{Y}_n \equiv \frac{1}{n}\bigoplus_{i=1}^n Y_i$. Then we have the following result:

THEOREM 3.2: *Let* $\{(y_{iL}, y_{iU})\}_{i=1}^n$ *be i.i.d. real valued random vectors such that* $\Pr\{y_{iL} \le y_i \le y_{iU}\} = 1$ *and* $E(|y_L|) < \infty$, $E(|y_U|) < \infty$.

(i) *Then* $\mathbb{E}[Y] = [E(y_L), E(y_U)]$ *and* $H(\bar{Y}_n, \mathbb{E}[Y]) \overset{a.s.}{\to} 0$ *as* $n \to \infty$.

(ii) *If* $E(y_L^2) < \infty$, $E(y_U^2) < \infty$, *then*

$$\sqrt{n}H(\bar{Y}_n, \mathbb{E}(Y)) \overset{d}{\to} \max\{|z_{-1}|, |z_1|\},$$

$$\sqrt{n}d_H(\mathbb{E}(Y), \bar{Y}_n) \overset{d}{\to} \max\{(z_{-1})_+, (z_1)_-\},$$

where

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

4. 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 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 $\Pr\{y_L \le y \le y_U\} = 1$. We remark that best linear prediction finds the linear function that minimizes square loss, but does not impose that such linear function is best nonparametric (Manski (2003, pp. 56–58)). In other words, we are not restricting the conditional expectation of y given $\mathbf{x}$ to be linear.

Let $Y = [y_L, y_U]$. Throughout this section, we maintain the following assumption:

ASSUMPTION 4.1: *Let* $(y, y_L, y_U, \mathbf{x})$ *be a random vector in* $\Re \times \Re \times \Re \times \Re^d$ *such that* $\Pr\{y_L \le y \le y_U\} = 1$. *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 that $\mathbf{x} \in \Re^d$. To simplify the notation, in this section we introduce ideas restricting attention to the case that $x \in \Re$ (though our assumptions are written for the general case $d \ge 1$). Let

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

where $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$, $\overline{x^2} = \frac{1}{n}\sum_{i=1}^n x_i^2$. Assume the following:

ASSUMPTION 4.2: *$E(|y_L|) < \infty$, $E(|y_U|) < \infty$, $E(|y_L x_k|) < \infty$, and $E(|y_U \times x_k|) < \infty$, $k = 1, \dots, d$.*

ASSUMPTION 4.3: *Σ is of full rank.*

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

$$E(y) = \theta_1 + \theta_2 E(x),$$
$$E(xy) = \theta_1 E(x) + \theta_2 E(x^2),$$

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

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

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

$$G(\omega) = \left\{ \begin{bmatrix} y(\omega) \\ x(\omega)y(\omega) \end{bmatrix} : y(\omega) \in Y(\omega) \right\}. \tag{4.1}$$

Let G_i be the mapping defined as in (4.1) using (y_{iL}, y_{iU}, x_i). Lemma A.4 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

$$\begin{aligned} \Theta &= \Sigma^{-1}\mathbb{E}[G] \\ &= \left\{ \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} : \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} 1 & E(x) \\ E(x) & E(x^2) \end{bmatrix}^{-1} \begin{bmatrix} E(y) \\ E(xy) \end{bmatrix}, \begin{bmatrix} y \\ xy \end{bmatrix} \in \mathcal{S}^1(G) \right\}. \end{aligned} \tag{4.2}$$

Before proceeding to deriving the estimator of Θ and its asymptotic properties, we show that the identification region Θ defined in (4.2) is identical to the identification region for the BLP obtained following the approach in Manski (2003) and denoted by Θ^M:

$$\Theta^M = \left\{ \begin{bmatrix} \theta_1^M \\ \theta_2^M \end{bmatrix} : \begin{bmatrix} \theta_1^M \\ \theta_2^M \end{bmatrix} = \arg\min \int (y - \theta_1 - \theta_2 x)^2 \, d\eta, \eta \in \mathbf{P}_{yx} \right\}, \tag{4.3}$$

where

$$\text{(4.4)}\qquad \mathbf{P}_{yx} = \big\{\eta : \Pr(y_L \le t, x \le x_0) \ge \eta((-\infty, t], (-\infty, x_0]) \ge \Pr(y_U \le t, x \le x_0) \forall t \in \Re, \forall x_0 \in \Re, \eta((-\infty, +\infty), (-\infty, x_0]) = \Pr(x \le x_0)\, \forall x_0 \in \Re\big\}.$$

PROPOSITION 4.1: *Let Assumptions 4.1–4.3 hold. Let Θ and Θ^M be defined, respectively, as in (4.2) and (4.3)–(4.4). Then $\Theta = \Theta^M$.*

Given these preliminaries, we apply the analogy principle and define the sample analog of Θ as

$$\text{(4.5)}\qquad \hat{\Theta}_n = \hat{\Sigma}_n^{-1}\bar{G}_n,$$

where $\bar{G}_n = \frac{1}{n}\bigoplus_{i=1}^n G_i$.

Observe that by construction, Θ is a convex set. Lemma A.8 in the Appendix describes further its geometry, showing that when x has an absolutely continuous distribution and G is integrably bounded, Θ is a strictly convex set, that is, it does not 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 $\Re^d$.

Theorem 4.2 below shows that under mild regularity conditions on the moments of (y_L, y_U, x) as in Assumption 4.2, $\hat{\Theta}_n$ is a consistent estimator of Θ. Under the additional Assumption 4.4, this convergence occurs at the rate $O_p(\frac{1}{\sqrt{n}})$.

ASSUMPTION 4.4: $E(|y_L|^2) < \infty$, $E(|y_U|^2) < \infty$, $E(|y_L x_k|^2) < \infty$, *and* $E(|y_U \times x_k|^2) < \infty$, $E(|x_k|^4) < \infty$, $k = 1, \dots, d$.

THEOREM 4.2: *Let Assumptions 4.1, 4.2, and 4.3 hold. Define Θ and $\hat{\Theta}_n$ as in (4.2) and (4.5), respectively. Then $H(\hat{\Theta}_n, \Theta) \xrightarrow{a.s.} 0$. If in addition Assumption 4.4 holds, then $H(\hat{\Theta}_n, \Theta) = O_p(\frac{1}{\sqrt{n}})$.*

The proof of Theorem 4.2 is based on an extension of Slutsky's theorem to SVRVs, which we provide in Lemma A.6 in the Appendix. The rate of convergence that we obtain is $1/\sqrt{n}$, irrespective of whether x has a continuous or a discrete distribution. To derive the asymptotic distribution of $H(\hat{\Theta}_n, \Theta)$, we impose an additional condition[13] on the distribution of x:

[13]When x has a discrete distribution, the population identification region Θ is a polytope, given by a Minkowski sum of segments. As a result, when x is discrete, the support function of Θ is not differentiable and the functional delta method (which we use when x is absolutely continuous) cannot be applied. In this case, tedious calculations allow one to express the extreme points of Θ as functions of moments of (y_L, y_U, x). Hence the exact asymptotic distribution of $H(\hat{\Theta}_n, \Theta)$ can be derived.

ASSUMPTION 4.5: *The distribution of* $\mathbf{x}$ *is absolutely continuous with respect to Lebesgue measure on* $\Re^d$.

THEOREM 4.3: *Let Assumptions* 4.1, 4.3, 4.4, *and* 4.5 *hold*.

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

$$(4.6)\qquad E[v(\mathbf{p})v(\mathbf{q})] = E[s(\mathbf{p}, \Sigma^{-1}G)s(\mathbf{q}, \Sigma^{-1}G)] - E[s(\mathbf{p}, \Sigma^{-1}G)]E[s(\mathbf{q}, \Sigma^{-1}G)] - \langle \boldsymbol{\xi}_{\mathbf{p}}, \boldsymbol{\kappa}_{\mathbf{q},\mathbf{p}}\rangle - \langle \boldsymbol{\kappa}_{\mathbf{p},\mathbf{q}}, \boldsymbol{\xi}_{\mathbf{q}}\rangle + \langle \boldsymbol{\xi}_{\mathbf{p}}, V_{\mathbf{p},\mathbf{q}}\boldsymbol{\xi}_{\mathbf{q}}\rangle$$

for all $\mathbf{p}, \mathbf{q} \in \mathbb{S}^d$. *The singleton* $\boldsymbol{\xi}_{\mathbf{p}} = \Theta \cap \{\boldsymbol{\vartheta} \in \Re^{d+1} : \langle \boldsymbol{\vartheta}, \mathbf{p}\rangle = s(\mathbf{p}, \Theta)\}$, *and the matrix* $V_{\mathbf{p},\mathbf{q}}$ *and the vector* $\boldsymbol{\kappa}_{\mathbf{p},\mathbf{q}}$ *are given*, *respectively*, *in equations* (A.7) *and* (A.8).

(ii) *Assume in addition that* $\mathrm{Var}(y_L|\mathbf{x})$, $\mathrm{Var}(y_U|\mathbf{x}) \geq \sigma^2 > 0$ $P(\mathbf{x})$-*a.s. Then* $\mathrm{Var}(v(\mathbf{p})) > 0$ *for each* $\mathbf{p} \in \mathbb{S}^d$, *and therefore the laws of* $\|v\|_{\mathbb{C}(\mathbb{S}^d)}$ *and* $\sup_{\mathbf{p}\in\mathbb{S}^d}\{-v(\mathbf{p})\}_+$ *are absolutely continuous with respect to Lebesgue measure*, *respectively*, *on* $\Re_+$ *and* $\Re_{++}$.

The difficulty in obtaining the asymptotic distribution of the statistics in Theorem 4.3 can be explained by recalling how one would proceed in the point-identified case. When $\hat{\Theta}_n$ and Θ are singletons,

$$\sqrt{n}(\hat{\theta}_n - \theta) = \hat{\Sigma}_n^{-1}\sqrt{n}\left(\begin{bmatrix} \bar{y} \\ \overline{xy} \end{bmatrix} - \hat{\Sigma}_n \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}\right),$$

with the second expression giving the product of a random matrix converging in probability to a nonsingular matrix, and the sample average of a mean zero i.i.d. random vector converging to a multivariate normal distribution. In this case, an application of Slutsky theorem delivers the desired result. In our case, the Aumann expectation of the random set given by $\hat{\Sigma}_n\Theta$ is not equal to $\Sigma\Theta$, and therefore a simple Slutsky-type result which extends the one that would be applied in the point-identified case (and which we provide in the Appendix, Lemma A.9) does not suffice for obtaining the limiting distribution of our statistics.[14] Hence, to prove Theorem 4.3 we proceed in steps. We start by looking at the difference between the support functions of $\hat{\Theta}_n$ and Θ. We rewrite these support functions as the sum of two (not necessarily independent) elements, and use Skorokhod representation theorem and an application of the delta method for $\mathbb{C}(\mathbb{S}^d)$ valued random variables to derive their joint asymptotic distribution. However, the use of this functional delta method requires

[14]For a scalar random variable $\hat{\sigma}_n$ such that $\Pr(\hat{\sigma}_n \geq 0) = 1$, $\mathbb{E}[\hat{\sigma}_n\Theta] = E[\hat{\sigma}_n]\Theta$ (Molchanov (2005, Theorem 2.1.48)) and therefore standard arguments can be applied easily.

differentiability of the support function of Θ; Lemma A.8 in the Appendix establishes this condition when x has an absolutely continuous distribution.[15]

Consider now the case that one is interested in a subset of the parameters of the BLP and in testing linear restrictions on Θ. Let $\mathcal{R}$ denote a matrix of linear restrictions, including the special case that $\mathcal{R}$ projects the set Θ on one of its components.

COROLLARY 4.4: *Let Assumptions 4.1, 4.3, 4.4, and 4.5 hold. Let $\mathcal{R}$ be a nonrandom full-rank finite matrix of dimension $l \times (d+1)$ with $l < (d+1)$.*

(i) *Then* $\sqrt{n}H(\mathcal{R}\hat{\Theta}_n, \mathcal{R}\Theta) \xrightarrow{d} \|v^{\mathcal{R}}\|_{\mathbb{C}(\mathbb{S}^{l-1})}$ *and* $\sqrt{n}d_H(\mathcal{R}\Theta, \mathcal{R}\hat{\Theta}_n) \xrightarrow{d} \sup_{\mathbf{p}\in\mathbb{S}^{l-1}}\{-v^{\mathcal{R}}(\mathbf{p})\}_+$, *where $v^{\mathcal{R}}$ is a linear function of a vector Gaussian random system with $E[v^{\mathcal{R}}(\mathbf{p})] = 0$ for all $\mathbf{p} \in \mathbb{S}^{l-1}$ and*

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

where for $\mathbf{p}, \mathbf{q} \in \mathbb{S}^{l-1}$ the singleton $\boldsymbol{\xi}_{\mathcal{R}'\mathbf{p}} = \Theta \cap \{\boldsymbol{\vartheta} \in \Re^{d+1} : \langle \boldsymbol{\vartheta}, \mathcal{R}'\mathbf{p}\rangle = s(\mathcal{R}'\mathbf{p}, \Theta)\}$, and the matrix $V_{\mathcal{R}'\mathbf{p},\mathcal{R}'\mathbf{q}}$ and the vector $\boldsymbol{\kappa}_{\mathcal{R}'\mathbf{p},\mathcal{R}'\mathbf{q}}$ are given, respectively, in equations (A.7) *and* (A.8).

(ii) *Let $\mathcal{R}$ project the set Θ on one of its components. Without loss of generality, assume $\mathcal{R} = [0 \quad 0 \quad \cdots \quad 0 \quad 1]$. Then*

$$
\sqrt{n}H(\mathcal{R}\hat{\Theta}_n, \mathcal{R}\Theta) \xrightarrow{d} \max\{|v(-\mathcal{R})|, |v(\mathcal{R})|\},
$$

$$
\sqrt{n}d_H(\mathcal{R}\Theta, \mathcal{R}\hat{\Theta}_n) \xrightarrow{d} \max\{(v(-\mathcal{R}))_+, (v(\mathcal{R})_-)\},
$$

where v is defined in Theorem 4.3.

Often empirical researchers are particularly interested in estimation and inference for a single component of the BLP parameter vector. Corollary 4.4 establishes the asymptotic distribution of our test statistics for this case. In terms of computation of the bounds, our methodology provides substantial advantages. In particular, without loss of generality, for $i = 1, \ldots, n$ denote by $\mathbf{x}_i = [1, x_{i1}, \ldots, x_{id-1}, x_{id}] = [\mathbf{x}_{i1}, x_{id}]$ and by $\tilde{x}_{id}$ the residuals obtained after projecting x_d on the other covariates $\mathbf{x}_1$. Then the following result holds:

[15] Beresteanu and Molinari (2006, Corollary 5.4) gave an extremely simple approximation to the distribution in Theorem 4.3 based on sample size adjustments which allow one to ignore the randomness in $\hat{\Sigma}_n$ at the cost of a slower rate of convergence.

COROLLARY 4.5: *Let Assumptions 4.1–4.3 hold. Let* $\hat{\Theta}_{d+1,n} = \{\theta_{d+1} \in \Re: [\boldsymbol{\theta}_1 \quad \theta_{d+1}]' \in \hat{\Theta}_n\}$. *Then*

$$\hat{\Theta}_{d+1,n} = \frac{1}{\sum_{i=1}^{n} \tilde{x}_{id}^2} \left[\sum_{i=1}^{n} \min\{\tilde{x}_{id} y_{iL}, \tilde{x}_{id} y_{iU}\}, \sum_{i=1}^{n} \max\{\tilde{x}_{id} y_{iL}, \tilde{x}_{id} y_{iU}\} \right].$$

The practical relevance of this result is that the calculation of the estimator of the identification region of the BLP coefficient for a single variable can be carried out using standard statistical packages.[16] This further facilitates implementation of the bootstrap procedure that we propose in Algorithm 4.2 below.

Similarly to the discussion in Section 2.2, the asymptotic distributions of our test statistics depend 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 of $\sqrt{n}H(\hat{\Theta}_n, \Theta)$. A similar procedure can be applied for $\sqrt{n}d_H(\Theta, \hat{\Theta}_n)$.

ALGORITHM 4.2:

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

2. Compute

$$r_n^* \equiv \sqrt{n}H(\hat{\Sigma}_n^{-1*}\bar{G}_n^*, \hat{\Theta}_n). \tag{4.7}$$

3. Use the results of b repetitions of Steps 1 and 2 to compute the empirical distribution of r_n^* at a point t, denoted by $J_n(t)$.

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

$$\hat{c}_{\alpha n}^{\mathrm{BLP}} = \inf\{t: J_n(t) \geq 1 - \alpha\}. \tag{4.8}$$

The asymptotic validity of this procedure follows by an application of the delta method for the bootstrap. In particular, the following result holds:

PROPOSITION 4.6: *Let the assumptions of Theorem 4.3*(i) *hold. Then* $r_n^* \xrightarrow{d} \|v\|_{\mathbb{C}(\mathbb{S}^d)}$, *where* r_n^* *is defined in equation* (4.7) *and the random variable* v *is given in Theorem 4.3. If in addition the assumptions of Theorem 4.3*(ii) *hold, then* $\hat{c}_{\alpha n}^{\mathrm{BLP}} = c_\alpha^{\mathrm{BLP}} + o_p(1)$, *where* $\hat{c}_{\alpha n}^{\mathrm{BLP}}$ *is defined in* (4.8).

[16]We are grateful to an anonymous referee for suggesting this result. The proof of this result is based on the Frisch–Waugh–Lovell theorem (see Magnac and Maurin (2008) for a related use of this theorem). If one is interested in a subset of the parameters of the BLP of dimension $k \geq 2$, this theorem can again be applied, and it again yields computational advantages. Stoye (2007) proposed computationally equivalent estimators to those in Corollary 4.5, without using the Frisch–Waugh–Lovell theorem.

Alternatively we could obtain consistent estimates of c_{α}^{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 $\mathbf{p} \in \mathbb{S}^d$ and with a covariance kernel which consistently estimates the covariance kernel in (4.6). Beresteanu and Molinari (2006, Proposition 5.6) established the asymptotic validity of this procedure.

5. MONTE CARLO RESULTS

In this section we conduct a series of Monte Carlo experiments to evaluate the properties of the test proposed in Section 2.2, 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, Hong, and Tamer (2002) 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, Hong, and Tamer (2002, Sec. 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 3). 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," and use the statistic based on the Hausdorff distance to test $\mathfrak{H}_0 : \mathbb{E}[Y] = \Psi_0$, $\mathfrak{H}_A : \mathbb{E}[Y] \neq \Psi_0$. For each sample, we estimate the critical value c_α, $\alpha = 0.05$, implementing Algorithm 2.1 with 2000 bootstrap repetitions.[17] We build the local alternatives using equation (2.4) with $\Delta_1 = \{0\}$ and $\Delta_2 = \delta \frac{\{0.5\}}{\sqrt{n}}$. More specifically, the local alternatives are defined as $\Psi_{A_n}(\delta) = \Psi_0 \bigoplus \delta \frac{\{0.5\}}{\sqrt{n}}$ 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 $\mathbb{E}(Y)$.

[17]Fortran 90 code for computing the Minkowski sample average of intervals in $\Re$ and $\Re^2$, the Hausdorff distance, and the directed Hausdorff distance between polytopes in $\Re^2$, and for implementing Algorithm 2.1 and Algorithm 4.2 is available upon request from the authors.

TABLE I

REJECTION RATES OF THE TEST BASED ON THE DIRECTED HAUSDORFF DISTANCE, AND OF THE TEST BASED ON THE HAUSDORFF DISTANCE AGAINST LOCAL ALTERNATIVES (NOMINAL LEVEL = 0.05) − $E(y)$

		$\sqrt{n}H(\bar{Y}_n, \Psi_0)$						
Sample Size	$\sqrt{n}d_H(\Psi_0, \bar{Y}_n)$	$\delta = 0$	0.5	1	2	4	8	16
$n = 100$	0.091	0.046	0.061	0.092	0.212	0.580	0.979	1.000
$n = 200$	0.076	0.047	0.058	0.087	0.200	0.577	0.985	1.000
$n = 500$	0.065	0.049	0.059	0.085	0.190	0.570	0.989	1.000
$n = 1000$	0.061	0.049	0.059	0.087	0.190	0.565	0.989	1.000
$n = 2000$	0.060	0.051	0.061	0.086	0.190	0.568	0.990	1.000

In terms of equation (2.5), 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 I, columns 2–8.[18] The empirical size of our test (column 2) is quite close to the nominal level of 0.05. As δ (and therefore κ) increases, the rejection rates increase for each given sample size, and for a given δ, the rejection rates are stable across sample sizes. These results are invariant with the width of $\mathbb{E}(Y)$.

We then run a similar bootstrap procedure to estimate $\tilde{c}_\alpha$, $\alpha = 0.05$, and use the statistic based on the directed Hausdorff distance to test $\mathfrak{H}_0 : \Psi_0 \subseteq \mathbb{E}[Y]$, $\mathfrak{H}_A : \Psi_0 \nsubseteq \mathbb{E}[Y]$. Table I, column 1, reports the rejection rate of this test for different sample sizes, when the nominal level is 0.05. For small samples (e.g., $n = 100$) the test is a little oversized. However, as n increases, this size distortion decreases and the rejection rate gets close to its nominal level.[19]

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 based on the Hausdorff distance at the 0.05 level for the best linear predictor of the logarithm of wage given education. Figure 1 depicts Θ, the population identification region of the parameters of the BLP given in equation (4.2).

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 $\Theta_{A_n}(\delta) = \Theta_0 \oplus \delta \frac{\{(3,0.3)\}}{\sqrt{n}}$ for $\delta \in \{0, \frac{1}{2}, 1, 2, 4, 8, 16\}$, where $\Theta_0 = \Theta$ is the polytope in Figure 1, and the vector $(3, 0.3)$ gives the width (approximated to the

[18]We also conducted Monte Carlo experiments in which the critical value c_α 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 I.

[19]We observe that inversion of this test leads to the construction of the confidence set $\mathcal{DU}_n$, and therefore the results in Table I, column 1, can be used to infer the coverage properties of $\mathcal{DU}_n$.

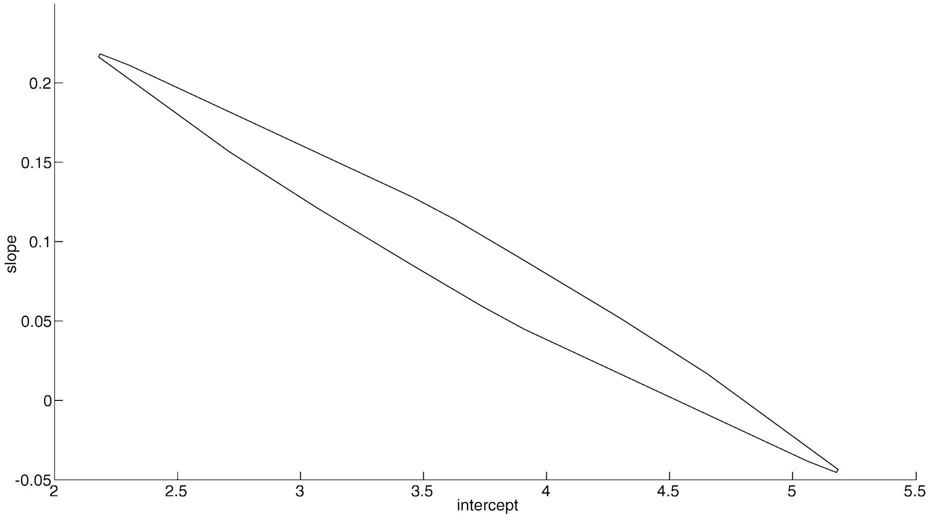

FIGURE 1.—The population identification region of the parameters of the BLP.

first decimal point) of the projection of Θ on each axis. In terms of equation (2.5), the distance between the local alternative and the null is $\kappa = 3.015\delta$. 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_α, $\alpha = 0.05$, implementing Algorithm 4.2 with 2000 bootstrap repetitions. The rejection rates of the null and local alternatives are shown in Table II, columns 2–8. As the second 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 δ (and therefore κ) increases, the rejection rates increase for each given sample size.

TABLE II

REJECTION RATES OF THE TEST BASED ON THE DIRECTED HAUSDORFF DISTANCE AND OF THE TEST BASED ON THE HAUSDORFF DISTANCE AGAINST LOCAL ALTERNATIVES (NOMINAL LEVEL = 0.05) − BLP

		$\sqrt{n}H(\hat{\Theta}_n, \Theta_0)$						
Sample Size	$\sqrt{n}d_H(\Theta_0, \hat{\Theta}_n)$	$\delta = 0$	0.5	1	2	4	8	16
$n = 100$	0.077	0.062	0.074	0.094	0.151	0.352	0.867	1.000
$n = 200$	0.076	0.068	0.078	0.095	0.150	0.349	0.880	1.000
$n = 500$	0.066	0.062	0.070	0.085	0.135	0.336	0.889	1.000
$n = 1000$	0.062	0.059	0.067	0.081	0.129	0.330	0.896	1.000
$n = 2000$	0.060	0.057	0.061	0.072	0.121	0.321	0.900	1.000

We then run a similar bootstrap procedure to estimate $\tilde{c}_{\alpha}^{\mathrm{BLP}}$, $\alpha = 0.05$, and use the statistic based on the directed Hausdorff distance to test $\mathfrak{H}_0 : \Theta_0 \subseteq \Theta$, $\mathfrak{H}_A : \Theta_0 \nsubseteq \Theta$. Table II, column 1, reports the rejection rate of this test for different sample sizes, when the nominal level is 0.05. Again, for small samples (e.g., $n = 100$) the test is a little oversized. However, as n increases, this size distortion decreases and the rejection rate gets close to its nominal level.[20]

In summary, our Monte Carlo experiments show that the test described in Section 2.2 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. Similarly, the rejection rate of the test described in Section 2.3 is close to its nominal level even with small samples.

6. CONCLUSIONS

This paper has introduced a methodology to conduct estimation and inference for partially identified population features in a completely analogous way to how estimation and inference would be conducted if the population features were point identified. We have shown that for a certain class of partially identified population features, which include means and best linear predictors with interval outcome data, and can be easily extended to parameter vectors characterizing semiparametric binary models with interval regressor data, the identification region is given by a transformation of the Aumann expectation of an SVRV. Extending the analogy principle to SVRVs, we proved 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 and the directed Hausdorff distance between the population identification region and the proposed estimator are normalized by $\sqrt{n}$, these statistics converge in distribution to different functions of the same Gaussian process, whose covariance kernel depends on parameters of the population identification region. We introduced consistent bootstrap procedures to estimate the quantiles of these distributions.

The asymptotic distribution results have allowed us to introduce procedures to test, respectively, whether the population identification region is equal to a particular set or whether a certain set of values is included in the population identification region. These test procedures are consistent against any fixed alternative and are locally asymptotically unbiased. A Monte Carlo exercise showed that our tests perform well even in small samples.

The test statistic based on the Hausdorff distance 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

[20]As before, the results in Table II, column 1, can be used to infer the coverage properties of $\mathcal{DU}_n$.

a 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 $1 - \alpha$. The test statistic based on the directed Hausdorff distance can be inverted to construct a directed confidence collection, given by the collection of all sets that, when specified as a null hypothesis for a subset of values of the population identification region, cannot be rejected by our test. Its main property is that (asymptotically) the union of the sets comprising it covers the population identification region with a prespecified confidence level $1-\alpha$. This result establishes a clear connection, within the class of models studied in this paper, between our Wald-type confidence sets, and the QLR-type confidence sets proposed by Chernozhukov, Hong, and Tamer (2007).

Dept. of Economics, Duke University, Durham, NC 27708-0097, U.S.A.; arie@econ.duke.edu

and

Dept. of Economics, Cornell University, 492 Uris Hall, Ithaca, NY 14853-7601, U.S.A.; fm72@cornell.edu.

Manuscript received June, 2006; final revision received November, 2007.

APPENDIX: LEMMAS AND PROOFS

Preliminaries

We first define notation and state limit theorems that we use throughout the appendix. In this appendix, we use the capital Greek letters Π and Σ 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 Π and a set $A \subset \Re^d$, let $\Pi A = \{\mathbf{r} \in \Re^d : \mathbf{r} = \Pi \mathbf{a}, \mathbf{a} \in A\}$ and observe that $s(\mathbf{p}, \Pi A) = s(\Pi' \mathbf{p}, A)$, where Π' 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 $\Pi'\Pi$. This matrix norm is compatible with its underlying vector norm (i.e., the Euclidean norm), so that $\|\Pi \mathbf{a}\| \leq \|\Pi\| \|\mathbf{a}\|$, and is a continuous function of the elements of the matrix. We let $\Longrightarrow$ denote weak convergence.

An SVRV $F : \Omega \to K_{kc}(\Re^d)$ can be represented through the support function of its realizations.[21] Sublinearity of the support function implies that when considering the support function of a set, it suffices to restrict attention to vectors $\mathbf{p} \in \mathbb{S}^{d-1}$. This results in a $\mathbb{C}(\mathbb{S}^{d-1})$ valued random variable. Hörmander's

[21] Similarly, a Minkowski average of SVRVs $F_i : \Omega \to K_{kc}(\Re^d)$, $i = 1, \dots, n$, can be represented through the sample average of the corresponding support functions $s(\mathbf{p}, F_i)$, $i = 1, \dots, n$.

embedding theorem (Li, Ogura, and Kreinovich (2002, Theorem 1.1.12)) ensures that $(K_{kc}(\Re^d), H(\cdot,\cdot))$ can be isometrically embedded into a closed convex cone in $\mathbb{C}(\mathbb{S}^{d-1})$. In particular, for any $F_1, F_2 : \Omega \to K_{kc}(\Re^d)$,

$$H(F_1, F_2) = \|s(\cdot, F_1) - s(\cdot, F_2)\|_{\mathbb{C}(\mathbb{S}^{d-1})}. \tag{A.1}$$

Let $\mathcal{B}(K(\Re^d))$ be the Borel field of $K(\Re^d)$ with respect to the Hausdorff metric H. Then it follows from Definition 1 that an SVRV $F : \Omega \to K(\Re^d)$ is $\mathcal{B}(K(\Re^d))$-measurable (this result is stated in Molchanov (2005, Theorem 1.2.3); for a proof see Li, Ogura, and Kreinovich (2002, Theorem 1.2.3)). Define $\mathcal{A}_F = \sigma\{F^{-1}(A) : A \in \mathcal{B}(K(\Re^d))\}$ to be the σ-algebra generated by inverse images of sets in $\mathcal{B}(K(\Re^d))$. Also let $\mu_F(A) = \mu(F^{-1}(A))$ for any $A \in \mathcal{B}(K(\Re^d))$ denote the distribution of F.

DEFINITION 6: Let F_1 and F_2 be two $\mathcal{B}(K(\Re^d))$-measurable SVRVs defined on the same measurable space Ω. 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 μ_{F_1} and μ_{F_2} are identical.

Many of the results that we obtain in this paper are based on a strong law of large numbers and on a central limit theorem for SVRVs which parallel the classic results for random variables. We state these results here for the reader's convenience.

THEOREM A.1—(SLLN): *Let $\{F, F_i : i \in \mathbb{N}\}$ be i.i.d. nonempty, integrably bounded, compact valued SVRVs. Then*

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

For the proof, see Artstein and Vitale (1975).

THEOREM A.2—(CLT): *Let $\{F, F_i : i \in \mathbb{N}\}$ be i.i.d. nonempty, compact valued SVRVs such that $E[\|F\|_H^2] < \infty$. Then*

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

where z is a Gaussian random system with $E[z(\mathbf{p})] = 0$ for all $\mathbf{p} \in \mathbb{S}^{d-1}$ and $E[z(\mathbf{p})z(\mathbf{q})] = E[s(\mathbf{p}, F)s(\mathbf{q}, F)] - E[s(\mathbf{p}, F)]E[s(\mathbf{q}, F)]$ for all $\mathbf{p}, \mathbf{q} \in \mathbb{S}^{d-1}$.

For the proof, see Giné, Hahn, and Zinn (2006).

LEMMA A.1: *Given any two compact convex sets* $K, L \subset \Re^d$,

$$d_H(K, L) = \sup_{\mathbf{p}\in\mathbb{B}^d}\{s(\mathbf{p}, K) - s(\mathbf{p}, L)\}$$
$$= \max\Big\{0, \sup_{\mathbf{p}\in\mathbb{S}^{d-1}}\{s(\mathbf{p}, K) - s(\mathbf{p}, L)\}\Big\},$$

where $\mathbb{B}^d = \{\mathbf{p} \in \Re^d : \|\mathbf{p}\| \le 1\}$.

PROOF: By definition, $d_H(K, L) = \inf\{c > 0 : K \subset L \oplus c\mathbb{B}^d\}$. Let $d_H(K, L) \le \alpha$. Then $K \subset L \oplus \alpha\mathbb{B}^d$, which implies that $s(\mathbf{p}, K) \le s(\mathbf{p}, L \oplus \alpha\mathbb{B}^d) = s(\mathbf{p}, L) + \alpha$ for all $\mathbf{p} \in \mathbb{B}^d$. This argument can be reversed to show that if $\sup_{\mathbf{p}\in\mathbb{B}^d}\{s(\mathbf{p}, K) - s(\mathbf{p}, L)\} \le \alpha$, then $d_H(K, L) \le \alpha$. Hence, $d_H(K, L) = \sup_{\mathbf{p}\in\mathbb{B}^d}\{s(\mathbf{p}, K) - s(\mathbf{p}, L)\}$. To see that $\sup_{\mathbf{p}\in\mathbb{B}^d}\{s(\mathbf{p}, K) - s(\mathbf{p}, L)\} = \max\{0, \sup_{\mathbf{p}\in\mathbb{S}^{d-1}}\{s(\mathbf{p}, K) - s(\mathbf{p}, L)\}\}$, observe that if $s(\mathbf{p}, K) - s(\mathbf{p}, L) \le 0$ for all $\mathbf{p} \in \mathbb{B}^d$, then $\sup_{\mathbf{p}\in\mathbb{B}^d}\{s(\mathbf{p}, K) - s(\mathbf{p}, L)\} = 0$ and $\max\{0, \sup_{\mathbf{p}\in\mathbb{S}^{d-1}}\{s(\mathbf{p}, K) - s(\mathbf{p}, L)\}\} = 0$. (Observe incidentally that $d_H(K, L) = 0$ and $s(\mathbf{p}, K) \le s(\mathbf{p}, L)$ for every $\mathbf{p} \in \mathbb{B}^d$ if and only if $K \subseteq L$.) Suppose now that there exists at least a $\mathbf{p} \in \mathbb{B}^d$ such that $s(\mathbf{p}, K) - s(\mathbf{p}, L) > 0$. Let $\tilde{\mathbf{p}} \in \arg\sup_{\mathbf{p}\in\mathbb{B}}\{s(\mathbf{p}, K) - s(\mathbf{p}, L)\}$. Then $\tilde{\mathbf{p}} \in \mathbb{S}^{d-1}$. To see why this is the case, suppose by contradiction that $\tilde{\mathbf{p}} \notin \mathbb{S}^{d-1}$. Let $\mathbf{p}^\star = \frac{\tilde{\mathbf{p}}}{\|\tilde{\mathbf{p}}\|} \in \mathbb{S}^{d-1}$ and observe that we are assuming $\|\tilde{\mathbf{p}}\| < 1$. Then $0 < s(\tilde{\mathbf{p}}, K) - s(\tilde{\mathbf{p}}, L) < \frac{1}{\|\tilde{\mathbf{p}}\|}[s(\tilde{\mathbf{p}}, K) - s(\tilde{\mathbf{p}}, L)] = s(\mathbf{p}^\star, K) - s(\mathbf{p}^\star, L)$, which leads to a contradiction. *Q.E.D.*

A.1. *Proofs for Section 2.2*

PROOF OF PROPOSITION 2.1: To establish the asymptotic validity of this procedure, observe that by Theorem 2.4 in Giné and Zinn (1990),

$$\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} s(\cdot, \operatorname{co} F_i^*) - \frac{1}{n}\sum_{i=1}^{n} s(\cdot, \operatorname{co} F_i)\right] \Longrightarrow z(\cdot)$$

as a sequence of processes indexed by $\mathbf{p} \in \mathbb{S}^{d-1}$. Observing that for each $g \in \mathbb{C}(\Re)$, the functional on $\mathbb{C}(\mathbb{S}^{d-1})$ defined by $h(x) = g(\|x\|_{\mathbb{C}(\mathbb{S}^{d-1})})$ belongs to $\mathbb{C}(\mathbb{C}(\mathbb{S}^{d-1}), \|\cdot\|_{\mathbb{C}(\mathbb{S}^{d-1})})$, the result follows by the continuous mapping theorem using standard arguments (e.g., Politis, Romano, and Wolf (1999, Chap. 1)).

The processes considered in this paper are separable with bounded realizations (a consequence of the fact that the support function of a bounded set $F \in K_k(\Re^d)$ is Lipschitz with Lipschitz constant $\|F\|_H$, (Molchanov (2005, Theorem F.1))). Hence, it follows from Theorem 1 in Tsirel'son (1975) and from the corollary in Lifshits (1982) that $\operatorname{Var}(z(\mathbf{p})) > 0$ for each $\mathbf{p} \in \mathbb{S}^{d-1}$ is a sufficient condition for the law of $\|z\|_{\mathbb{C}(\mathbb{S}^{d-1})}$ to be absolutely continuous with respect to Lebesgue measure on $\Re_+$. The result then follows by standard arguments. *Q.E.D.*

PROOF OF THEOREM 2.2: By triangle inequality, $H(\Psi_A, \Psi_0) \leq H(\bar{F}_n, \Psi_0) + H(\bar{F}_n, \Psi_A)$. Hence,

$$\begin{aligned}
&\lim_{n\to\infty} \Pr\{\sqrt{n}H(\bar{F}_n, \Psi_0) > \hat{c}_{\alpha n}\} \\
&\quad \geq \lim_{n\to\infty} \Pr\{\sqrt{n}H(\Psi_A, \Psi_0) - \sqrt{n}H(\bar{F}_n, \Psi_A) > \hat{c}_{\alpha n}\} \\
&\quad = \lim_{n\to\infty} \Pr\left\{H(\Psi_A, \Psi_0) > H(\bar{F}_n, \Psi_A) + \frac{c_\alpha + o_p(1)}{\sqrt{n}}\right\} \\
&\quad = \Pr\{H(\Psi_A, \Psi_0) > 0\} = 1,
\end{aligned}$$

where the second to last equality follows from Theorem A.1 and because $\hat{c}_{\alpha n} \xrightarrow{p} c_\alpha < \infty$ (by Proposition 2.1), and the last equality follows because $H(\Psi_A, \Psi_0)$ is a positive constant. *Q.E.D.*

PROOF OF THEOREM 2.3: By Hörmander's embedding theorem, $\sqrt{n}H(\bar{F}_n, \Psi_0) = \sqrt{n}\sup_{\mathbf{p}\in\mathbb{S}^{d-1}} |s(\mathbf{p}, \bar{F}_n) - s(\mathbf{p}, \Psi_0)|$. Let

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

Similarly, $h_k(\cdot) = s(\cdot, F_k) - s(\cdot, \Psi_0) = s(\cdot, F_k) - s(\cdot, \Psi_{An}) + s(\cdot, \frac{1}{\sqrt{n}}\Delta_2) - s(\cdot, \frac{1}{\sqrt{n}}\Delta_1)$. Hence under the local alternative, $h, h_1, h_2, \ldots$ are $C(\mathbb{S}^{d-1})$ 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 limiting distribution of $\sqrt{n}H(\bar{F}_n, \Psi_0)$ then follows from Proposition 3.1.9 in Li, Ogura, and Kreinovich (2002).

The above result and Proposition 2.1 imply that

$$\begin{aligned}
\lim_{n\to\infty} \Pr(\sqrt{n}H(\bar{F}_n, \Psi_0) > \hat{c}_{\alpha n}) &= \lim_{n\to\infty} \Pr(\sqrt{n}H(\bar{F}_n, \Psi_0) + o_p(1) > c_\alpha) \\
&= \Pr\{\|w\|_{\mathbb{C}(\mathbb{S}^{d-1})} > c_\alpha\}.
\end{aligned}$$

Recall that z is a Gaussian random system with $E[z(\mathbf{p})] = 0$ $\forall \mathbf{p} \in \mathbb{S}^{d-1}$ and $\mathrm{Cov}(z(\mathbf{p}), z(\mathbf{q})) = E[s(\mathbf{p}, F)s(\mathbf{q}, F)] - E[s(\mathbf{p}, F)]E[s(\mathbf{q}, F)]$. Also, $w(\cdot)$ is a Gaussian random system with $w \stackrel{d}{=} z + \tau$, where $\tau(\mathbf{p}) = s(\mathbf{p}, \Delta_2) - s(\mathbf{p}, \Delta_1)$ $\forall \mathbf{p} \in \mathbb{S}^{d-1}$ and $\stackrel{d}{=}$ means equivalent in distribution. So by Anderson's lemma (see, e.g., Molchanov (2005, Theorem 3.1.12)), for any finite set S_j,

$$\Pr\Big(\sup_{\mathbf{p} \in S_j} |w(\mathbf{p})| > c_\alpha\Big) \geq \Pr\Big(\sup_{\mathbf{p} \in S_j} |z(\mathbf{p})| > c_\alpha\Big).$$

The metric space $(\mathbb{S}^{d-1}, \|\cdot,\cdot\|)$, where $\|\cdot,\cdot\|$ denotes the usual Euclidean distance in $\Re^d$, is separable. Also $z(\cdot)$ and $w(\cdot)$ are uniformly continuous in probability with respect to $\|\cdot,\cdot\|$ (a consequence of the fact that the support function of a closed bounded set F is Lipschitz with Lipschitz constant $\|F\|_H$, (Molchanov (2005, Theorem F.1))). Thus both have separable versions. Without loss of generality we may assume that $z(\cdot)$ and $w(\cdot)$ are separable. The result then follows from the same argument as in Section 6 of Andrews (1997, p. 1114). *Q.E.D.*

PROOF OF THEOREM 2.4: We first show that $\bar{F}_n \oplus B_{\hat{c}_{\alpha n}} \subset \mathcal{U}_n$. This follows because $\bar{F}_n \oplus B_{\hat{c}_{\alpha n}}$ is a convex compact set (the Minkowski sum of convex sets is convex) and because $\sqrt{n}H(\bar{F}_n, \bar{F}_n \oplus B_{\hat{c}_{\alpha n}}) = \sqrt{n}H(\{0\}, B_{\hat{c}_{\alpha n}}) = \hat{c}_{\alpha n}$, where the last equality follows from the definition of $B_{\hat{c}_{\alpha n}}$. Hence $\bar{F}_n \oplus B_{\hat{c}_{\alpha n}} \in CC_{n,1-\alpha}$. We now show that $\mathcal{U}_n \subset \bar{F}_n \oplus B_{\hat{c}_{\alpha n}}$. Take any $\boldsymbol{\psi} \in \mathcal{U}_n$. Then by definition $\exists \tilde{\Psi} \in CC_{n,1-\alpha}$ such that $\boldsymbol{\psi} \in \tilde{\Psi}$ and $\sqrt{n}H(\bar{F}_n, \tilde{\Psi}) \leq \hat{c}_{\alpha n}$. Choose $\tilde{\mathbf{f}} \in \bar{F}_n$ such that $\tilde{\mathbf{f}} = \arg\inf_{\mathbf{f} \in \bar{F}_n} \|\boldsymbol{\psi} - \mathbf{f}\|$. Let $\tilde{\mathbf{b}} = \boldsymbol{\psi} - \tilde{\mathbf{f}}$; then by construction $\boldsymbol{\psi} = \tilde{\mathbf{f}} + \tilde{\mathbf{b}}$, and $\tilde{\mathbf{b}} \in B_{\hat{c}_{\alpha n}}$ because

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

where the last inequality follows because $\sqrt{n}H(\bar{F}_n, \tilde{\Psi}) \leq \hat{c}_{\alpha n}$. Hence $\boldsymbol{\psi} \in \bar{F}_n \oplus B_{\hat{c}_{\alpha n}}$. *Q.E.D.*

A.2. *Proofs for Section 2.3*

PROOF OF COROLLARY 2.5: By the triangle inequality, $d_H(\Psi_0, \mathbb{E}[F]) \leq d_H(\Psi_0, \bar{F}_n) + d_H(\bar{F}_n, \mathbb{E}[F])$, and the result follows from a similar argument as in the proof of Theorem 2.2 because $d_H(\Psi_0, \mathbb{E}[F])$ is a positive constant when $\Psi_0 \not\subseteq \mathbb{E}[F]$, $d_H(\bar{F}_n, \mathbb{E}[F]) \xrightarrow{p} 0$ by Theorem A.1, and by Theorem 2 in Lifshits (1982), if $\mathrm{Var}(z(\mathbf{p})) > 0$ for each $\mathbf{p} \in \mathbb{S}^{d-1}$, the law of $\sup_{\mathbf{p} \in \mathbb{S}^{d-1}} \{-z(\mathbf{p})\}_+$ is absolutely continuous with respect to Lebesgue measure on $\Re_{++}$, so that $\widehat{\tilde{c}}_{\alpha n} = \tilde{c}_\alpha + o_p(1)$. *Q.E.D.*

PROOF OF COROLLARY 2.6: The fact that $\mathcal{DU}_n = \bar{F}_n \oplus B_{\widehat{\tilde{c}}_{\alpha n}}$ follows from a similar argument as in the proof of Theorem 2.4. The fact that $\mathcal{DU}_n \subset \mathcal{U}_n$ is an obvious consequence of the fact that for any two compact sets A and B, $H(A, B) \geq d_H(A, B)$. *Q.E.D.*

PROOF OF PROPOSITION 2.7: We show that $\mathbb{E}[F] \in DCC_{n,1-\alpha}$ if and only if $\mathbb{E}[F] \subseteq \mathcal{DU}_n$. Trivially, $\mathbb{E}[F] \in DCC_{n,1-\alpha} \Longrightarrow \mathbb{E}[F] \subseteq \mathcal{DU}_n$. Next, given any two subsets A and B of $\Re^d$, let B^c denote the c-envelope (or parallel set) of B, that is, $B^c = \{\tilde{\mathbf{b}} \in \Re^d : \inf_{\mathbf{b} \in B} \|\mathbf{b} - \tilde{\mathbf{b}}\| \leq c\}$. Because $d_H(A, B) = \inf\{c > 0 : A \subset B^c\}$, it follows that

$$\begin{aligned} \mathbb{E}[F] \subseteq \mathcal{DU}_n \quad &\Longleftrightarrow \quad \mathbb{E}[F] \subseteq \bar{F}_n \oplus B_{\widehat{\tilde{c}}_{\alpha n}} \quad \Longrightarrow \quad d_H(\mathbb{E}[F], \bar{F}_n) \leq \frac{\widehat{\tilde{c}}_{\alpha n}}{\sqrt{n}} \\ &\Longrightarrow \quad \mathbb{E}[F] \in DCC_{n,1-\alpha}. \end{aligned}$$

Therefore,

$$\lim_{n\to\infty} \Pr\{\mathbb{E}[F] \subseteq \mathcal{DU}_n\} = \lim_{n\to\infty} \Pr\{\mathbb{E}[F] \in DCC_{n,1-\alpha}\} = 1 - \alpha. \qquad Q.E.D.$$

A.3. *Proofs for Section 3*

PROOF OF THEOREM 3.2:
This result can be easily proved by using the following lemmas.

LEMMA A.2: *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 nonempty SVRV.*

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

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

From the fact that y_L and y_U are random variables, also $s(p, Y)$ 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 nonempty follows because $\Pr(y_L \leq y_U) = 1$. *Q.E.D.*

LEMMA A.3: *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 $A \in \mathcal{B}(K(\Re))$, $\mu_{Y_i}(A) = \mu(Y_i^{-1}(A)) = \Pr\{[y_{iL}, y_{iU}] \cap A \neq \emptyset\}$. 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 $K_k(\Re)$,

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

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). *Q.E.D.*

To complete the proof of Theorem 3.2:

(i) Let $y(\omega) \in Y(\omega) = [y_L(\omega), y_U(\omega)]$ μ-a.s. Then $\int_\Omega y\, d\mu \in [E(y_L), E(y_U)]$. This implies $E[Y] \subset [E(y_L), E(y_U)]$. 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 $y_b(\omega) = \alpha y_L(\omega) + (1-\alpha) y_U(\omega)$. Then $\int_\Omega y_b\, d\mu = b$ and $y_b \in \mathcal{S}(Y)$, and therefore $[E(y_L), E(y_U)] \subset E[Y]$.

Lemmas A.2 and A.3 show that $\{Y, Y_i : i \in \mathbb{N}\}$ are i.i.d. compact valued nonempty SVRVs. To verify that the SVRVs are integrably bounded, observe that $\int \|Y\|_H\, d\mu \le E(|y_L|) + E(|y_U|) < \infty$, where the last inequality follows from the assumptions in Theorem 3.2. All the conditions of Theorem A.1 are therefore satisfied.

(ii) Our random sets are i.i.d., compact valued (i.e., in $K_k(\Re)$), and nonempty by Lemmas A.2 and A.3. We know that $\int_\Omega \|Y(\omega)\|_H^2\, d\mu \le \int_\Omega y_L(\omega)^2\, d\mu + \int_\Omega y_U(\omega)^2\, d\mu$, which is finite by assumption. Therefore, all the conditions in Theorem A.2 are satisfied. By the definition of the support function, $s(p, Y)(\omega) = y_U(\omega)$ if $p = 1$, $s(p, Y)(\omega) = -y_L(\omega)$ if $p = -1$. The covariance kernel in the theorem follows by simple algebra. *Q.E.D.*

A.4. *Proofs for Section 4*

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

$$\Sigma \equiv \begin{bmatrix} 1 & E(\mathbf{x}') \\ E(\mathbf{x}) & E(\mathbf{x}\mathbf{x}') \end{bmatrix}, \qquad \hat{\Sigma}_n \equiv \begin{bmatrix} 1 & \overline{\mathbf{x}}' \\ \overline{\mathbf{x}} & \overline{\mathbf{x}\mathbf{x}'} \end{bmatrix},$$

$$G(\omega) = \left\{ \left(\begin{bmatrix} y(\omega) \\ \mathbf{x}(\omega) y(\omega) \end{bmatrix} : y(\omega) \in Y(\omega) \right) \right\}, \tag{A.2}$$

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

$$\text{(A.3)} \qquad \Theta = \Sigma^{-1}\mathbb{E}[G] = \left\{ \boldsymbol{\theta} \in \Re^{d+1} : \boldsymbol{\theta} = \begin{bmatrix} 1 & E(\mathbf{x}') \\ E(\mathbf{x}) & E(\mathbf{x}\mathbf{x}') \end{bmatrix}^{-1} \begin{bmatrix} E(y) \\ E(\mathbf{x}y) \end{bmatrix}, \begin{bmatrix} y \\ \mathbf{x}y \end{bmatrix} \in \mathcal{S}^1(G) \right\},$$

and

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

where

$$\begin{aligned} \mathbf{P}_{y\mathbf{x}} = \{ \eta : & \Pr(y_L \le t, \mathbf{x} \le \mathbf{x}_0) \ge \eta((-\infty, t], (-\infty, \mathbf{x}_0]) \\ & \ge \Pr(y_U \le t, \mathbf{x} \le \mathbf{x}_0) \, \forall t \in \Re, \; \forall \mathbf{x}_0 \in \Re^d, \\ & \eta((-\infty, +\infty), (-\infty, \mathbf{x}_0]) = \Pr(\mathbf{x} \le \mathbf{x}_0) \, \forall \mathbf{x}_0 \in \Re^d \}, \end{aligned}$$

$\boldsymbol{\theta} = [\,\theta_1 \quad \boldsymbol{\theta}_2'\,]'$, and the notation $\mathbf{x} \le \mathbf{x}_0$ indicates that each element of $\mathbf{x}$ is less than or equal to the corresponding element of $\mathbf{x}_0$.

LEMMA A.4: *G and $\{G_i\}_{i=1}^n$ as defined in equation (A.2) are nonempty compact valued SVRVs.*

PROOF: By Lemma A.2, Y is a nonempty, compact valued SVRV. By Theorem 1.2.7 in Li, Ogura, and Kreinovich (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 $\text{cl}\{f_i(\omega)\} = Y(\omega)$ μ-a.s. Therefore, it is easy to see that for each $\lambda \in \Re$, λY 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 $(\mathbf{x}Y)(\omega) = \{\mathbf{x}(\omega) f(\omega) : f(\omega) \in Y(\omega)\} = \text{cl}\{\mathbf{x}(\omega) f_i(\omega)\}$ μ-a.s. Hence, $\left\{ \begin{smallmatrix} f_i \\ \mathbf{x} f_i \end{smallmatrix} \right\}$ spans the set G as defined in (A.2) and therefore G is an SVRV by Theorem 1.2.7 in Li, Ogura, and Kreinovich (2002). The fact that G is nonempty and compact valued follows from the same arguments as in Lemma A.2. *Q.E.D.*

LEMMA A.5: *Under Assumption 4.1, $\{G_i\}_{i=1}^n$ 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 $A \in \mathcal{B}(K(\Re^{d+1}))$,

$$\mu_{G_i}(A) = \mu(G_i^{-1}(A))$$

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

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 $C_1, \ldots, C_n$ in $K_k(\Re^{d+1})$,

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

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). *Q.E.D.*

PROOF OF PROPOSITION 4.1: With $\mathbf{x}$ a column vector in $\Re^d$, the sets Θ and Θ^M are convex and compact subsets of $\Re^{d+1}$. By assumption, Θ and Θ^M are also nonempty, because the set G is integrably bounded. 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 θ_1 and θ_1^M, and let 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 $[\,\theta_1^M \quad (\boldsymbol{\theta}_2^M)'\,]' \in \Theta^M$. Then there exists a distribution for $(y, \mathbf{x})$ with x-marginal $P(\mathbf{x})$ denoted η_0^M such that $[\,\theta_1^M \quad (\boldsymbol{\theta}_2^M)'\,]'$ is a minimizer of the problem in (A.4). Let $(y^M, \mathbf{x})$ be a random vector with distribution η_0^M. It follows that $y^M(\omega) \in Y(\omega)$ μ-a.s., and therefore $[\,y^M(\omega) \quad \mathbf{x}'y^M(\omega)\,]' \in G(\omega)$ μ-a.s., from which $[\,\theta_1^M \quad (\boldsymbol{\theta}_2^M)'\,]' \in \Theta$.

Conversely, pick a vector $[\,\theta_1 \quad (\boldsymbol{\theta}_2)'\,]' \in \Theta$. Then there exists a random vector $[\,y \quad y\mathbf{x}'\,]' \in \mathcal{S}^1(G)$ 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 $\eta \in P_{y\mathbf{x}}$. Because $(y, \mathbf{x})$ is a selection from $(Y, \mathbf{x})$, it follows from Theorem 2.1 in Artstein (1983) that for any $t \in \Re$, $x_0 \in \Re^d$,

$$\eta(y \le t, \mathbf{x} \le \mathbf{x}_0) \le \mu(Y \cap (-\infty, t] \ne \emptyset, \mathbf{x} \le \mathbf{x}_0) = \Pr(y_L \le t, \mathbf{x} \le \mathbf{x}_0),$$
$$\eta(y \ge t, \mathbf{x} \le \mathbf{x}_0) \le \mu(Y \cap [t, +\infty) \ne \emptyset, \mathbf{x} \le \mathbf{x}_0) = \Pr(y_U \ge t, \mathbf{x} \le \mathbf{x}_0).$$

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

PROOF OF THEOREM 4.2: Theorem 4.2 can be easily proved by using the following lemma.

LEMMA A.6: *Let $\{F, F_i : i \in \mathbb{N}\}$ be i.i.d. nonempty 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 $\Pi_i \overset{a.s.}{\to} \Pi$ element-by-element, where Π is a $d \times d$ nonstochastic matrix with finite elements and Π_i has finite elements with probability* 1. *Then*

$$H(\Pi_n \bar{F}_n, \Pi \mathbb{E}[F]) \overset{a.s.}{\to} 0.$$

PROOF: This is a version of Slutsky's theorem for d-dimensional SVRVs. We interpret vectors in $\Re^d$ as being column vectors. Then

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

because $\|\Pi_n\| \overset{\text{a.s.}}{\to} \|\Pi\| < \infty$ by Slutsky's theorem, $\max\{\sup_{\mathbf{f}_n \in \bar{F}_n} \inf_{\mathbf{f} \in \mathbb{E}[F]} \|\mathbf{f}_n - \mathbf{f}\|, \sup_{\mathbf{f} \in \mathbb{E}[F]} \inf_{\mathbf{f}_n \in \bar{F}_n} \|\mathbf{f}_n - \mathbf{f}\|\} \overset{\text{a.s.}}{\to} 0$ by the LLN for SVRVs, $\|\Pi_n - \Pi\| \overset{\text{a.s.}}{\to} 0$ 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. *Q.E.D.*

The proof of Theorem 4.2 follows directly from Lemma A.6, replacing $\hat{\Sigma}_n^{-1} = \Pi_n$, $\Sigma^{-1} = \Pi$, $F_i = G_i$, and $F = G$. This is because Lemmas A.4 and A.5

show that $\{G, G_i : i \in \mathbb{N}\}$ are i.i.d. nonempty compact valued SVRVs. To verify that these SVRVs are integrably bounded, observe that $\int \|G\|_H \, d\mu \le E(|y_L|) + E(|y_U|) + \sum_{k=1}^d [E(|x_k y_L|) + E(|x_k y_U|)] < \infty$, where the last inequality follows from the assumptions in Theorem 4.2. If in addition $E(|y_L|^2) < \infty$, $E(|y_U|^2) < \infty$, $E(|x_k y_L|^2) < \infty$, $E(|x_k y_U|^2) < \infty$, and $E(|x_k|^4) < \infty$, $k = 1, \dots, d$, then

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

by the CLT for SVRVs, because $\int \|G\|_H^2 \, d\mu \le E(|y_L|^2) + E(|y_U|^2) + \sum_{k=1}^d [E(|x_k y_L|^2) + E(|x_k y_U|^2)] < \infty$ and $\|\Pi_n - \Pi\| = O_p(\frac{1}{\sqrt{n}})$ by the continuous mapping theorem. *Q.E.D.*

PROOF OF THEOREM 4.3: Before giving this proof, we introduce a definition and three lemmas.

DEFINITION 7: Let $\mathbf{w}$ be a random vector in $\Re^d$. Its zonoid, $\Lambda_{\mathbf{w}}$, is the Aumann expectation of the random segment in $\Re^d$ with the endpoints being the origin $(\mathbf{0})$ and the d-dimensional vector $\mathbf{w}$. The lift zonoid, $\tilde{\Lambda}_{\mathbf{w}}$, of $\mathbf{w}$ is the Aumann expectation of the random segment in $\Re^{d+1}$ 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 $\mathbf{L}^p_{\mathcal{A}} = \mathbf{L}^p(\Omega, \mathcal{A}, \Re^d)$ the space of $\mathcal{A}$-measurable random variables with values in $\Re^d$ such that the $\mathbf{L}^p$-norm $\|\boldsymbol{\xi}\|_p = [E(\|\boldsymbol{\xi}\|^p)]^{1/p}$ is finite, and for an SVRV F defined on $(\Omega, \mathcal{A})$ denote by $\mathcal{S}^p_{\mathcal{A}}(F)$ the family of all selections of F from $\mathbf{L}^p_{\mathcal{A}}$, so that $\mathcal{S}^p_{\mathcal{A}}(F) = \mathcal{S}(F) \cap \mathbf{L}^p_{\mathcal{A}}$.

LEMMA A.7: *Let A be an integrably bounded SVRV defined on $(\Omega, \mathcal{A}, \mu)$. For each σ-algebra $\mathcal{A}_0 \subset \mathcal{A}$ there exists a unique integrable $\mathcal{A}_0$-measurable SVRV B, denoted by $B = E[A|\mathcal{A}_0]$ and called the conditional Aumann expectation of A, such that*

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

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

For the proof, see 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.8: *Define Θ as in (4.2). Under Assumptions 4.2, 4.3, and 4.5, Θ 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

$$\zeta = y_U - y_L, \quad \boldsymbol{\eta} = [\,y_L \quad \mathbf{x}'y_L\,]'.$$

Then one can represent G as

$$G = \boldsymbol{\eta} \oplus \tilde{L}, \quad \tilde{L} = \mathrm{co}\{\mathbf{0}, [\,\zeta \quad \mathbf{x}'\zeta\,]'\}.$$

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 $[\,\zeta \quad \mathbf{x}'\zeta\,]' = \zeta[\,1 \quad \mathbf{x}'\,]'$.

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

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

$\mathbb{E}[\mathrm{co}(\mathbf{0}, [\,1 \quad \mathbf{x}'\,]')|\zeta]$ is a lift zonoid, and therefore $\mathbb{E}[G|\zeta]$ is a (rescaled by ζ) 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 $\Re^d$, $\mathbb{E}[G|\zeta]$ is a strictly convex set $P(\zeta)$-a.s. Premultiplying G by the nonrandom matrix Σ^{-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 $\Re^{d+1}\backslash\{\mathbf{0}\}$ $P(\zeta)$-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

$$\begin{aligned}&\nabla s(\mathbf{p}, \mathbb{E}[\Sigma^{-1}G|\zeta])\\ &\quad = \mathbb{E}[\Sigma^{-1}G|\zeta] \cap \big\{\boldsymbol{\vartheta} \in \Re^{d+1} : \langle \boldsymbol{\vartheta}, \mathbf{p}\rangle = s(\mathbf{p}, \mathbb{E}[\Sigma^{-1}G|\zeta])\big\},\end{aligned}$$

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 4.2) to obtain

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

Hence

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

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

LEMMA A.9: *Let $\{F, F_i : i \in \mathbb{N}\}$ be i.i.d. nonempty compact valued SVRVs such that $E[\|F\|_H^2] < \infty$. Let $\{\Pi_i : i \in \mathbb{N}\}$ be i.i.d. random $d \times d$ matrices on $(\Omega, \mathcal{A}, \mu)$ such that $\Pi_i \xrightarrow{p} \Pi$ element-by-element, where Π is a $d \times d$ nonstochastic matrix with finite elements and Π_i has finite elements with probability* 1. *Then*

$$\sqrt{n}\big(s(\cdot, \Pi_n \bar{F}_n) - s(\cdot, \Pi_n \mathbb{E}[F])\big) \Longrightarrow z^{\Pi}(\cdot),$$

where z^{Π} is a Gaussian random system with $E[z^{\Pi}(\mathbf{p})] = 0$ for all $\mathbf{p} \in \mathbb{S}^{d-1}$, and $E[z^{\Pi}(\mathbf{p}) z^{\Pi}(\mathbf{q})] = E[s(\mathbf{p}, \Pi F) s(\mathbf{q}, \Pi F)] - E[s(\mathbf{p}, \Pi F)] E[s(\mathbf{q}, \Pi F)]$ for all $\mathbf{p}, \mathbf{q} \in \mathbb{S}^{d-1}$.

PROOF: It follows from the assumptions of the lemma that the sets $\{\Pi F, \Pi F_i : i \in \mathbb{N}\}$ are i.i.d. nonempty, compact valued SVRVs such that $E[\|\Pi F\|_H^2] < \infty$. By the central limit theorem for $\mathbb{C}(\mathbb{S}^{d-1})$-valued random variables in Li, Ogura, and Kreinovich (2002, Proposition 3.1.9), the assumptions of which are verified in Li, Ogura, and Kreinovich (2002, Proof of Theorem 3.1.8), it follows that

$$\sqrt{n}\big(s(\cdot, \Pi \bar{F}_n) - s(\cdot, \Pi \mathbb{E}[F])\big) \equiv z_n^{\Pi}(\cdot) \Longrightarrow z^{\Pi}(\cdot)$$

in $\mathbb{C}(\mathbb{S}^{d-1})$. By Skorokhod representation theorem there exist random elements $z_n^{\Pi\star}(\cdot)$ and $z^{\Pi\star}(\cdot)$ with $z_n^{\Pi\star}(\cdot) \stackrel{d}{=} z_n^{\Pi}(\cdot)$ and $z^{\Pi\star}(\cdot) \stackrel{d}{=} z^{\Pi}(\cdot)$, such that $z_n^{\Pi\star}(\cdot) \to z^{\Pi\star}(\cdot)$ a.s. By Theorem 1.8.12 of Schneider (1993), this convergence is uniform. Since $\Pi_n \xrightarrow{p} \Pi$, the result follows by standard arguments. *Q.E.D.*

The proof of Theorem 4.3 proceeds in steps.

Step 1—Derivation of the fact that:

$$\begin{aligned}
&\sqrt{n}[s(\cdot, \hat{\Theta}_n) - s(\cdot, \Theta)] \\
&\quad \stackrel{A}{=} \sqrt{n}\big[(s(\cdot, \Sigma^{-1} \bar{G}_n) - s(\cdot, \Theta)) + \big(s(\cdot, \hat{\Sigma}_n^{-1} \mathbb{E}[G]) - s(\cdot, \Theta)\big)\big],
\end{aligned}$$

where $\stackrel{A}{=}$ means asymptotically equivalent in distribution: Using the definition of $\hat{\Theta}_n$,

$$\begin{aligned}
&\sqrt{n}[s(\cdot, \hat{\Theta}_n) - s(\cdot, \Theta)] \\
&\quad = \sqrt{n}\big[\big(s(\cdot, \hat{\Sigma}_n^{-1} \bar{G}_n) - s(\cdot, \hat{\Sigma}_n^{-1} \mathbb{E}[G])\big) + \big(s(\cdot, \hat{\Sigma}_n^{-1} \mathbb{E}[G]) - s(\cdot, \Theta)\big)\big].
\end{aligned}$$

By Lemma A.9,

$$\sqrt{n}\big(s(\cdot,\hat{\Sigma}_n^{-1}\bar{G}_n)-s(\cdot,\hat{\Sigma}_n^{-1}\mathbb{E}[G])\big)\stackrel{A}{=}\sqrt{n}(s(\cdot,\Sigma^{-1}\bar{G}_n)-s(\cdot,\Theta)).$$

Step 2—Derivation of the asymptotic distribution of $\sqrt{n}[(s(\cdot,\Sigma^{-1}\bar{G}_n)-s(\cdot,\Theta))\quad \mathbf{u}_n(\cdot)]'$, where $\mathbf{u}_n(\mathbf{p})\equiv(\hat{\Sigma}_n-\Sigma)\hat{\Sigma}_n^{-1}\mathbf{p}$, $\mathbf{p}\in\mathbb{S}^d$:

Lemmas A.4 and A.5 show that $\{G, G_i : i\in\mathbb{N}\}$ are i.i.d. nonempty compact valued SVRVs. Observe that by the assumptions of Theorem 4.3, $\int\|G\|_H^2\,d\mu\le E(|y_L|^2)+E(|y_U|^2)+\sum_{k=1}^d[E(|x_k y_L|^2)+E(|x_k y_U|^2)]<\infty$ and $E(|\mathbf{x}|^4)<\infty$. Using the same argument as in the proof of Lemma A.9, it follows by the central limit theorem for $C(\mathbb{S}^d)$ valued random variables in Li, Ogura, and Kreinovich (2002, Proposition 3.1.9) and by the Cramér–Wold device that

$$\sqrt{n}\begin{bmatrix} s(\cdot,\Sigma^{-1}\bar{G}_n)-s(\cdot,\Theta)\\ \mathbf{u}_n(\cdot)\end{bmatrix}\Longrightarrow\begin{pmatrix} z^{\Sigma^{-1}}(\cdot)\\ \mathbf{u}(\cdot)\end{pmatrix}\tag{A.5}$$

as a sequence of processes indexed by $\mathbf{p}\in\mathbb{S}^d$, where for each $\mathbf{p}\in\mathbb{S}^d$, $(z^{\Sigma^{-1}}(\mathbf{p})\quad\mathbf{u}(\mathbf{p}))$ is a $(d+2)$-dimensional normal random vector with $E[z^{\Sigma^{-1}}(\mathbf{p})]$ and $E[\mathbf{u}(\mathbf{p})]=0$. Regarding the covariance kernel, denote by $\{\rho_{ij}\}$, $i,j=1,\dots,d+1$, the elements of Σ^{-1} and let

$$\hat{\Sigma}_n=\frac{1}{n}\sum_{i=1}^n\Pi_i\quad\text{with}\quad \Pi_i\equiv\begin{bmatrix}1 & x_{1i} & \cdots & x_{di}\\ x_{1i} & x_{1i}^2 & \cdots & x_{1i}x_{di}\\ \vdots & \vdots & \ddots & \vdots\\ x_{di} & x_{1i}x_{di} & \cdots & x_{di}^2\end{bmatrix}.\tag{A.6}$$

Then

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

where $V_{\mathbf{p},\mathbf{p}}=E[((\Pi_i-\Sigma)\Sigma^{-1}\mathbf{p})((\Pi_i-\Sigma)\Sigma^{-1}\mathbf{p})']$ and $\boldsymbol{\kappa}_{\mathbf{p},\mathbf{p}}$ is a $(d+1)\times 1$ vector with first element

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

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

$$\kappa_{j,(\mathbf{p},\mathbf{p})}=\Bigg[\big(E[x_{j-1}s(\mathbf{p},\Sigma^{-1}G)]-s(\mathbf{p},\Theta)E(x_{j-1})\big)\sum_{l=1}^{d+1}\rho_{1l}p_l$$

$$+\sum_{l=1}^{d+1}\sum_{i=2}^{d+1}\big(E[x_{j-1}x_{i-1}s(\mathbf{p},\Sigma^{-1}G)] - s(\mathbf{p},\Theta)E(x_{j-1}x_{i-1})\big)\rho_{il}p_l\Bigg]$$

with x_j the jth element of $\mathbf{x}$.

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

$$\text{(A.7)}\qquad \begin{aligned} &E\big[z^{\Sigma^{-1}}(\mathbf{p})z^{\Sigma^{-1}}(\mathbf{q})\big] \\ &\quad = E[s(\mathbf{p},\Sigma^{-1}G)s(\mathbf{q},\Sigma^{-1}G)] - E[s(\mathbf{p},\Sigma^{-1}G)]E[s(\mathbf{q},\Sigma^{-1}G)], \\ &E[\mathbf{u}(\mathbf{p})\mathbf{u}(\mathbf{q})] = V_{\mathbf{p},\mathbf{q}} = E\big[((\Pi_i-\Sigma)\Sigma^{-1}\mathbf{p})((\Pi_i-\Sigma)\Sigma^{-1}\mathbf{q})'\big], \end{aligned}$$

and $E[z^{\Sigma^{-1}}(\mathbf{p})\mathbf{u}(\mathbf{q})] = \boldsymbol{\kappa}_{\mathbf{p},\mathbf{q}}$, where $\boldsymbol{\kappa}_{\mathbf{p},\mathbf{q}}$ is a $(d+1)\times 1$ vector with 1st and jth element $(j=2,\dots,d+1)$ given, respectively, by

$$\text{(A.8)}\qquad \begin{aligned} \kappa_{1,(\mathbf{p},\mathbf{q})} &= \sum_{k=1}^{d+1}\sum_{i=2}^{d+1}\big[\big(E[x_{i-1}s(\mathbf{p},\Sigma^{-1}G)] - s(\mathbf{p},\Theta)E(x_{i-1})\big)\big]\rho_{ik}q_k, \\ \kappa_{j,(\mathbf{p},\mathbf{q})} &= \Bigg[\big(E[x_{j-1}s(\mathbf{p},\Sigma^{-1}G)] - s(\mathbf{p},\Theta)E(x_{j-1})\big)\sum_{l=1}^{d+1}\rho_{1l}q_l \\ &\quad + \sum_{l=1}^{d+1}\sum_{i=2}^{d+1}\big(E[x_{j-1}x_{i-1}s(\mathbf{p},\Sigma^{-1}G)] \\ &\quad - s(\mathbf{p},\Theta)E(x_{j-1}x_{i-1})\big)\rho_{il}q_l\Bigg]. \end{aligned}$$

Step 3—Derivation of the asymptotic distribution of $\sqrt{n}[s(\cdot,\hat{\Sigma}_n^{-1}\bar{G}_n) - s(\cdot,\Theta)]$: Step 1 gives that

$$\begin{aligned} &\sqrt{n}[s(\cdot,\hat{\Sigma}_n^{-1}\bar{G}_n) - s(\cdot,\Theta)] \\ &\quad \stackrel{A}{=} \sqrt{n}\big\{(s(\cdot,\Sigma^{-1}\bar{G}_n) - s(\cdot,\Theta)) + \big(s(\cdot,\hat{\Sigma}_n^{-1}\mathbb{E}[G]) - s(\cdot,\Theta)\big)\big\}. \end{aligned}$$

Recall that by Lemma A.8, since $\mathbf{x}$ has an absolutely continuous distribution with respect to Lebesgue measure on $\Re^d$, Θ is a strictly convex set. Hence, its support set in the direction $\mathbf{p}\in\mathbb{S}^d$ is the singleton

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

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

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

where $\mathbf{u}(\mathbf{p})$ is given in equation (A.5). It then follows by another application of the functional delta method that

$$\begin{aligned}
&\sqrt{n}\big[(s(\cdot, \Sigma^{-1}\bar{G}_n) - s(\cdot, \Theta)) + \big(s(\cdot, \hat{\Sigma}_n^{-1}\mathbb{E}[G]) - s(\cdot, \Theta)\big)\big] \\
&\quad \Longrightarrow z^{\Sigma^{-1}}(\cdot) - \langle \boldsymbol{\xi}_{\cdot}, \mathbf{u}(\cdot)\rangle,
\end{aligned}$$

as a sequence of processes indexed by $\mathbf{p} \in \mathbb{S}^d$, with $z^{\Sigma^{-1}}(\mathbf{p})$ and $\mathbf{u}(\mathbf{p})$ given in equation (A.5) and such that $E[z^{\Sigma^{-1}}(\mathbf{p}) - \langle \boldsymbol{\xi}_{\mathbf{p}}, \mathbf{u}(\mathbf{p})\rangle] = 0$ and

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

where $V_{\mathbf{p},\mathbf{q}}$ and $\boldsymbol{\kappa}_{\mathbf{p},\mathbf{q}}$ are given in (A.7) and (A.8).

By Hörmander's embedding theorem and by Lemma A.1, $\sqrt{n}H(\hat{\Theta}_n, \Theta) = \sqrt{n}\sup_{\mathbf{p}\in\mathbb{S}^d} |s(\mathbf{p}, \hat{\Sigma}_n^{-1}\bar{G}_n) - s(\mathbf{p}, \Theta)|$ and $\sqrt{n}d_H(\Theta, \hat{\Theta}_n) = \sqrt{n}\sup_{\mathbf{p}\in\mathbb{S}^d}[-(s(\mathbf{p}, \hat{\Sigma}_n^{-1}\bar{G}_n) - s(\mathbf{p}, \Theta))]_+$. Letting $v(\mathbf{p}) \equiv z^{\Sigma^{-1}}(\mathbf{p}) - \langle \boldsymbol{\xi}_{\mathbf{p}}, \mathbf{u}(\mathbf{p})\rangle$ for each $\mathbf{p} \in \mathbb{S}^d$, the result follows by the continuous mapping theorem.

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

$$f(\mathbf{x}_i) = \left\langle \mathbf{p}, \Sigma^{-1}\begin{bmatrix} 1 \\ \mathbf{x}_i \end{bmatrix}\right\rangle. \tag{A.9}$$

Then

$$\begin{aligned}
s(\mathbf{p}, \Sigma^{-1}G_i) = \sup\Bigg\{&\left\langle \mathbf{p}, \Sigma^{-1}\begin{bmatrix} y_i \\ \mathbf{x}_i y_i \end{bmatrix}\right\rangle : \\
&\begin{bmatrix} y_i \\ \mathbf{x}_i y_i \end{bmatrix} \in \left(\begin{bmatrix} y_i \\ \mathbf{x}_i y_i \end{bmatrix}, y_i(\omega) \in Y_i(\omega)\right)\Bigg\}
\end{aligned}$$

$$= y_{iL}\mathbb{I}(f(\mathbf{x}_i) < 0)f(\mathbf{x}_i) + y_{iU}\mathbb{I}(f(\mathbf{x}_i) \geq 0)f(\mathbf{x}_i),$$

where $\mathbb{I}(\cdot)$ is the indicator function of the event in brackets. From the arguments in Steps 1–3 it follows that for each $\mathbf{p} \in \mathbb{S}^d$,

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

where $f(\mathbf{x}_i)$ is defined in (A.9). It follows from the law of iterated expectations that

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

where the second inequality follows because by assumption, $\operatorname{Var}(y_L|\mathbf{x})$, $\operatorname{Var}(y_U|\mathbf{x}) \geq \sigma^2 > 0$ $P(\mathbf{x})$-a.s., and the last inequality follows because Σ is of full rank by assumption. Hence $\operatorname{Var}(v(\mathbf{p})) > 0$ for each $\mathbf{p} \in \mathbb{S}^d$, and by the same argument as in the proof of Proposition 2.1, the law of $\|v\|_{\mathbb{C}(\mathbb{S}^d)}$ is absolutely continuous with respect to Lebesgue measure on $\Re_+$. Theorem 2 in Lifshits (1982) assures that the law of $\sup_{\mathbf{p}\in\mathbb{S}^d}\{-v(\mathbf{p})\}_+$ is absolutely continuous with respect to Lebesgue measure on $\Re_{++}$.

Q.E.D.

PROOF OF COROLLARY 4.4: (i) By definition, $\sqrt{n}[s(\mathbf{p}, \mathcal{R}\hat{\Theta}_n) - s(\mathbf{p}, \mathcal{R}\Theta)] = \sqrt{n}[s(\mathcal{R}'\mathbf{p}, \hat{\Theta}_n) - s(\mathcal{R}'\mathbf{p}, \Theta)]$. Let $\tilde{\mathbf{p}} \equiv \mathcal{R}'\mathbf{p}$. A simple normalization makes $\tilde{\mathbf{p}}$ an element of $\mathbb{S}^{l-1}$. Using the definition of $\hat{\Theta}_n$,

$$\begin{aligned}\text{(A.10)}\quad &\sqrt{n}[s(\tilde{\mathbf{p}}, \hat{\Theta}_n) - s(\tilde{\mathbf{p}}, \Theta)] \\ &= \sqrt{n}[s(\tilde{\mathbf{p}}, \hat{\Sigma}_n^{-1}\bar{G}_n) - s(\tilde{\mathbf{p}}, \Theta)] \\ &= \sqrt{n}\big[\big(s(\tilde{\mathbf{p}}, \hat{\Sigma}_n^{-1}\bar{G}_n) - s(\tilde{\mathbf{p}}, \hat{\Sigma}_n^{-1}\mathbb{E}[G])\big) \\ &\qquad + \big(s(\tilde{\mathbf{p}}, \hat{\Sigma}_n^{-1}\mathbb{E}[G]) - s(\tilde{\mathbf{p}}, \Theta)\big)\big].\end{aligned}$$

By Lemma A.9, $\sqrt{n}(s(\cdot,\hat{\Sigma}_n^{-1}\bar{G}_n)-s(\cdot,\hat{\Sigma}_n^{-1}\mathbb{E}[G]))\Longrightarrow z^{\Sigma^{-1}}(\cdot)$. Regarding the other portion of expression (A.10), using a similar argument as in the proof of Theorem 4.3, we obtain

$$\begin{aligned}
&\sqrt{n}\big(s(\tilde{\mathbf{p}},\hat{\Sigma}_n^{-1}\mathbb{E}[G])-s(\tilde{\mathbf{p}},\Theta)\big)\\
&\quad=\sqrt{n}(s(\Sigma\hat{\Sigma}_n^{-1}\tilde{\mathbf{p}},\Theta)-s(\tilde{\mathbf{p}},\Theta))\\
&\quad=\sqrt{n}\big[s(\tilde{\mathbf{p}}-(\hat{\Sigma}_n-\Sigma)\hat{\Sigma}_n^{-1}\tilde{\mathbf{p}},\Theta)-s(\tilde{\mathbf{p}},\Theta)\big]\\
&\quad\overset{d}{\to}-\langle\boldsymbol{\xi}_{\tilde{\mathbf{p}}},\mathbf{u}(\tilde{\mathbf{p}})\rangle,
\end{aligned}$$

where $\boldsymbol{\xi}$ and $\mathbf{u}$ are defined in Theorem 4.3. The result follows.

(ii) Observing that when $\mathcal{R}=[0\quad 0\quad\cdots\quad 0\quad 1]$,

$$\begin{aligned}
&\sqrt{n}H(\mathcal{R}\hat{\Theta}_n,\mathcal{R}\Theta)\\
&\quad=\sqrt{n}\max\big\{|s(-\mathcal{R},\hat{\Theta}_n)-s(-\mathcal{R},\Theta)|,|s(\mathcal{R},\hat{\Theta}_n)-s(\mathcal{R},\Theta)|\big\},\\
&\sqrt{n}d_H(\mathcal{R}\Theta,\mathcal{R}\hat{\Theta}_n)\\
&\quad=\sqrt{n}\max\big\{\big[(s(-\mathcal{R},\hat{\Theta}_n)-s(-\mathcal{R},\Theta))\big]_+,\\
&\qquad\qquad\big[(s(\mathcal{R},\hat{\Theta}_n)-s(\mathcal{R},\Theta))\big]_-\big\},
\end{aligned}$$

the result follows easily. *Q.E.D.*

PROOF OF COROLLARY 4.5: Without loss of generality, let us write, for $i=1,\dots,n$, $\mathbf{x}_i=[1\quad x_{i1}\quad\cdots\quad x_{id-1}\quad x_{id}]=[\mathbf{x}_{i1}\quad x_{id}]$ and $G_i=\{\mathbf{x}_i'y_i:y_i\in Y_i\}$. In this proof we deviate from the notation used in the previous theorems and let $X=[X_1\quad X_d]$, $X_1=[\mathbf{x}_{11}'\quad\cdots\quad\mathbf{x}_{n1}']'$, $X_d=[x_{1d}\quad\cdots\quad x_{nd}]'$. Define $P=X(X'X)^{-1}X'$, $M=I-P$, $P_k=X_k(X_k'X_k)^{-1}X_k'$, and $M_k=I-P_k$ for $k=1,\dots,d$. We have $\hat{\Sigma}_n=\frac{1}{n}(X'X)$ and $\hat{\Theta}_n=\hat{\Sigma}_n^{-1}\frac{1}{n}\bigoplus_{i=1}^n G_i=(X'X)^{-1}\bigoplus_{i=1}^n G_i$. Let $\tilde{X}_d=M_1X_d=\tilde{x}_{id}$ and $\tilde{G}_i=\{\tilde{x}_{id}y_i:y_i\in Y_i\}$. Let $\hat{\Theta}_{d+1,n}=\{\theta_{d+1}\in\Re:[\boldsymbol{\theta}_1\quad\theta_{d+1}]'\in\hat{\Theta}_n\}$. Define $\tilde{\Theta}_{d+1,n}=(\tilde{X}_d'\tilde{X}_d)^{-1}\bigoplus_{i=1}^n\tilde{G}_i$. Note that $(\tilde{X}_d'\tilde{X}_d)=\sum_{i=1}^n\tilde{x}_{id}^2$ is a scalar. Also note that

$$\begin{aligned}
\bigoplus_{i=1}^n\tilde{G}_i&=\bigoplus_{i=1}^n\{\tilde{x}_{id}y_i:y_i\in Y_i=[y_{iL},y_{iU}]\}\\
&=\bigoplus_{i=1}^n[\min\{\tilde{x}_{id}y_{iL},\tilde{x}_{id}y_{iU}\},\max\{\tilde{x}_{id}y_{iL},\tilde{x}_{id}y_{iU}\}]\\
&=\left[\sum_{i=1}^n\min\{\tilde{x}_{id}y_{iL},\tilde{x}_{id}y_{iU}\},\sum_{i=1}^n\max\{\tilde{x}_{id}y_{iL},\tilde{x}_{id}y_{iU}\}\right]
\end{aligned}$$

from which

$$\tilde{\Theta}_{d+1,n} = \frac{1}{\sum_{i=1}^{n} \tilde{x}_{id}^2} \left[\sum_{i=1}^{n} \min\{\tilde{x}_{id} y_{iL}, \tilde{x}_{id} y_{iU}\}, \sum_{i=1}^{n} \max\{\tilde{x}_{id} y_{iL}, \tilde{x}_{id} y_{iU}\} \right].$$

Let $\hat{\theta}_d \in \hat{\Theta}_{d+1,n}$. Then there exists $\hat{\boldsymbol{\theta}} = [\hat{\boldsymbol{\theta}}_1 \quad \hat{\theta}_d]' \in \hat{\Theta}_n$, where $\hat{\boldsymbol{\theta}}_1 \in \Re^d$. Therefore, there exists $\mathbf{y} = [y_1 \quad \cdots \quad y_n]'$ s.t. $y_i \in Y_i$, $\forall i = 1, \dots, n$ and $\hat{\boldsymbol{\theta}} = (X'X)^{-1}X'\mathbf{y}$. Now by using the Frisch–Waugh–Lovell (FWL) theorem, we can deduce that $\hat{\theta}_d = (\tilde{X}_d'\tilde{X}_d)^{-1}\tilde{X}_d'\mathbf{y} \in (\tilde{X}_d'\tilde{X}_d)^{-1} \bigoplus_{i=1}^{n} \tilde{G}_i = \tilde{\Theta}_{d+1,n}$. Thus $\tilde{\Theta}_{d+1,n} \supseteq \hat{\Theta}_{d+1,n}$.

Now take $\theta_d \in \tilde{\Theta}_{d+1,n}$. Then we can write $\theta_d = (\tilde{X}_d'\tilde{X}_d)^{-1}\tilde{X}_d'\mathbf{y}$, where $\mathbf{y} = [y_1 \quad \cdots \quad y_n]'$ s.t. $y_i \in Y_i$, $\forall i = 1, \dots, n$. A simple manipulation gives

$$\begin{aligned} \mathbf{y} &= P\mathbf{y} + M\mathbf{y} = X(X'X)^{-1}X'\mathbf{y} + M\mathbf{y} \\ &= X\bar{\theta} + M\mathbf{y} \quad (\text{where } \bar{\boldsymbol{\theta}} = (X'X)^{-1}X'\mathbf{y} \in \hat{\Theta}_n) \\ &= X_1\bar{\boldsymbol{\theta}}_1 + X_d\bar{\theta}_d + M\mathbf{y}. \end{aligned}$$

Now again by using FWL theorem, we get that $\bar{\theta}_d = (\tilde{X}_d'\tilde{X}_d)^{-1}\tilde{X}_d'\mathbf{y} = \theta_d \in \hat{\Theta}_{d+1,n}$. Hence $\tilde{\Theta}_{d+1,n} \subseteq \hat{\Theta}_{d+1,n}$. *Q.E.D.*

PROOF OF PROPOSITION 4.6: Using Lemma A.9 and Theorem 2.4 in Giné and Zinn (1990), since $\hat{\Sigma}_n^{-1*}$ is a consistent estimator of Σ^{-1}, we have that

$$\begin{aligned} r_n^* &= \sqrt{n} H(\hat{\Sigma}_n^{-1*}\bar{G}_n^*, \hat{\Theta}_n) \\ &\stackrel{A}{=} \sqrt{n} \sup_{\mathbf{p}\in\mathbb{S}^d} \big| (s(\mathbf{p}, \Sigma^{-1}\bar{G}_n^*) - s(\mathbf{p}, \Sigma^{-1}\bar{G}_n)) \\ &\qquad\qquad + (s(\mathbf{p}, \hat{\Sigma}_n^{-1*}\bar{G}_n) - s(\mathbf{p}, \hat{\Theta}_n)) \big|. \end{aligned}$$

Looking at the second portion of the above expression, we get

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

By Theorem 3.3.1 and the discussion on page 160 of Schneider (1993), any nonempty, convex, and compact subset of $\Re^{d+1}$ can be approximated arbitrarily accurately by a nonempty, strictly convex, and compact subset of $\Re^{d+1}$. Hence, for any $\varepsilon_n = o_p(\frac{1}{\sqrt{n}})$ we can find a sequence of nonempty, strictly convex, and compact sets $\bar{\Gamma}_n$ such that $H(\hat{\Theta}_n, \bar{\Gamma}_n) < \varepsilon_n = o_p(\frac{1}{\sqrt{n}})$, which implies that $|s(\mathbf{p}, \hat{\Theta}_n) - s(\mathbf{p}, \bar{\Gamma}_n)| = o_p(\frac{1}{\sqrt{n}})$ uniformly in $\mathbf{p} \in \Re^{d+1}$. Under the assumptions of Theorem 4.3, $s(\mathbf{p}, \Theta)$ is Fréchet differentiable at $\mathbf{p}$ with gradient equal to $\boldsymbol{\xi}_{\mathbf{p}}$.

Similarly, because $\bar{\Gamma}_n$ is a strictly convex set, $s(\mathbf{p}, \bar{\Gamma}_n)$ is Fréchet differentiable at $\mathbf{p}$ with gradient $\bar{\boldsymbol{\varsigma}}_{n,\mathbf{p}} = \bar{\Gamma}_n \cap \{\boldsymbol{\gamma} \in \Re^{d+1} : \langle \boldsymbol{\gamma}, \mathbf{p}\rangle = s(\mathbf{p}, \bar{\Gamma}_n)\}$. Hence

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

where the last inequality follows from the fact that $s(\mathbf{p}, \bar{\Gamma}_n)$ and $s(\mathbf{p}, \Theta)$ are Fréchet differentiable for any $\mathbf{p} \in \Re^{d+1}\backslash\{\mathbf{0}\}$. The last equality follows because $((\hat{\Sigma}_n - \hat{\Sigma}_n^*)\hat{\Sigma}_n^{-1*}\mathbf{p}) = O_p(\frac{1}{\sqrt{n}})$ and because $\bar{\boldsymbol{\varsigma}}_{n,\mathbf{p}} - \boldsymbol{\xi}_{\mathbf{p}} = o_p(1)$ due to the fact that $|s(\mathbf{p}, \Theta) - s(\mathbf{p}, \bar{\Gamma}_n)| = o_p(1)$ uniformly in $\mathbf{p} \in \Re^{d+1}$. Hence

$$\begin{aligned}
r_n^* \stackrel{A}{=} \sqrt{n} \sup_{\mathbf{p}\in\mathbb{S}^d} &\big|(s(\mathbf{p}, \Sigma^{-1}\bar{G}_n^*) - s(\mathbf{p}, \Sigma^{-1}\bar{G}_n)) \\
&+ \big(s(\mathbf{p}+(\hat{\Sigma}_n - \hat{\Sigma}_n^*)\hat{\Sigma}_n^{-1*}\mathbf{p}, \Theta) - s(\mathbf{p}, \Theta)\big)\big|.
\end{aligned}$$

By Theorem 2.4 of Giné and Zinn (1990) and standard arguments, denoting by $\mathbf{u}_n^*(\mathbf{p}) \equiv (\hat{\Sigma}_n^* - \hat{\Sigma}_n)\hat{\Sigma}_n^{-1*}\mathbf{p}$,

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

as a sequence of processes indexed by $\mathbf{p} \in \mathbb{S}^d$, where $(\, z^{\Sigma^{-1}}(\mathbf{p}) \quad \mathbf{u}(\mathbf{p})\,)$ is the $(d+2)$-dimensional normal random vector in (A.5). Therefore an application of the delta method for the bootstrap (Van der Vaart and Wellner (2000, Theorem 3.9.11)) implies that

$$\begin{aligned}
&\sqrt{n}\big[(s(\cdot, \Sigma^{-1}\bar{G}_n^*) - s(\cdot, \Sigma^{-1}\bar{G}_n)) + (s(\cdot, \hat{\Sigma}_n^{-1*}\bar{G}_n) - s(\cdot, \hat{\Theta}_n))\big] \\
&\stackrel{d}{\to} z^{\Sigma^{-1}}(\cdot) - \langle \boldsymbol{\xi}_{\mathbf{p}}, \mathbf{u}(\cdot)\rangle
\end{aligned}$$

as a sequence of processes indexed by $\mathbf{p} \in \mathbb{S}^d$. By the continuous mapping theorem $r_n^* \xrightarrow{d} \|v\|_{\mathbb{C}(\mathbb{S}^d)}$, where $v(\mathbf{p}) \equiv z^{\Sigma^{-1}}(\mathbf{p}) - \langle \boldsymbol{\xi}_{\mathbf{p}}, \mathbf{u}(\mathbf{p}) \rangle$ for each $\mathbf{p} \in \mathbb{S}^d$. It then follows by the same argument as in the proof of Proposition 2.1 that since $\mathrm{Var}(v(\mathbf{p})) > 0$ for each $\mathbf{p} \in \mathbb{S}^d$, the critical values of the simulated distribution consistently estimate the critical values of $\|v\|_{\mathbb{C}(\mathbb{S}^d)}$, that is, $\hat{c}_{\alpha n} = c_\alpha + o_p(1)$. *Q.E.D.*

REFERENCES

ANDREWS, D. W. K. (1997): "A Conditional Kolmogorov Test," *Econometrica*, 65, 1097–1128. [796]

ANDREWS, D. W. K., S. BERRY, AND P. JIA (2004): "Confidence Regions for Parameters in Discrete Games With Multiple Equilibria, With an Application to Discount Chain Store Location," Discussion Paper, Yale University. [768]

ARTSTEIN, Z. (1974): "On the Calculus of Closed Set-Valued Functions," *Indiana University Mathematics Journal*, 24, 433–441. [771]

——— (1983): "Distributions of Random Sets and Random Selections," *Israel Journal of Mathematics*, 46, 313–324. [801]

ARTSTEIN, Z., AND R. A. VITALE (1975): "A Strong Law of Large Numbers for Random Compact Sets," *The Annals of Probability*, 3, 879–882. [772,793]

AUMANN, R. J. (1965): "Integrals of Set Valued Functions," *Journal of Mathematical Analysis and Applications*, 12, 1–12. [764,771]

BERESTEANU, A., AND F. MOLINARI (2006): "Asymptotic Properties for a Class of Partially Identified Models," Working Paper CWP10/06, CeMMAP. [769,786,788]

CHERNOZHUKOV, V., J. HAHN, AND W. NEWEY (2005): "Bound Analysis in Panel Models With Correlated Random Effects," Discussion Paper, UCLA and MIT. [768]

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2002): "Parameter Set Inference in a Class of Econometric Models," Discussion Paper, MIT, Duke University, and Northwestern University. [768,780,788]

——— (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75, 1243–1284. [764,767,768,778,780,792]

DEBLASI, F. S., AND F. IERVOLINO (1969): "Equazioni Differenziali con Soluzioni a Valore Compatto Convesso," *Bollettino della Unione Matematica Italiana*, 4, 491–501. (In Italian.) [774]

DOMINITZ, J., AND R. P. SHERMAN (2006): "Identification and Estimation of Bounds on School Performance Measures: A Nonparametric Analysis of a Mixture Model With Verification," *Journal of Applied Econometrics*, 21, 1295–1326. [779]

GALICHON, A., AND M. HENRY (2006): "Inference in Incomplete Models," Discussion Paper, Harvard University and Columbia University. [769]

GINÉ, E., AND J. ZINN (1990): "Bootstrapping General Empirical Measures," *The Annals of Probability*, 18, 851–869. [773,794,810,811]

GINÉ, E., M. G. HAHN, AND J. ZINN (1983): "Limit Theorems for Random Sets: Application of Probability in Banach Space Results," in *Probability in Banach Spaces*, Vol. IV, Lecture Notes in Mathematics, Vol. 990. Berlin: Springer, 112–135. [793]

GINTHER, D. (2000): "Alternative Estimates of the Effect of Schooling on Earnings," *Review of Economics and Statistics*, 82, 103–116. [779]

GONZALEZ-LUNA, M. L. (2005): "Nonparametric Bounds on the Returns to Language Skills," *Journal of Applied Econometrics*, 20, 771–795. [779]

GUGGENBERGER, P., J. HAHN, AND K. KIM (2006): "Specification Testing Under Moment Inequalities," Discussion Paper, UCLA. [769]

HAILE, P., AND E. TAMER (2003): "Inference With an Incomplete Model of English Auctions," *Journal of Political Economy*, 111, 1–51. [779]

HONORE, B. E., AND E. TAMER (2006): "Bounds on Parameters in Panel Dynamic Discrete Choice Models," *Econometrica*, 74, 611–631. [768]

HOROWITZ, J. L., AND C. F. MANSKI (1997): "What Can Be Learned About Population Parameters When the Data Are Contaminated," in *Handbook of Statistics*, Vol. 15, ed. by G. S. Maddala and C. R. Rao. Amsterdam: Elsevier Science, 439–466. [768]

——— (1998): "Censoring of Outcomes and Regressors Due to Survey Nonresponse: Identification and Estimation Using Weights and Imputations," *Journal of Econometrics*, 84, 37–58. [768]

——— (2000): "Nonparametric Analysis of Randomized Experiments With Missing Covariate and Outcome Data," *Journal of the American Statistical Association*, 95, 77–84. [768]

HOTZ, J. V., C. H. MULLIN, AND S. G. SANDERS (1997): "Bounding Casual Effects Using Data From a Contaminated Natural Experiment: Analyzing the Effects of Teenage Childbearing," *Review of Economic Studies*, 64, 575–603. [779]

IMBENS, G. W., AND C. F. MANSKI (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857. [764,767,768,778,779]

LI, S., Y. OGURA, AND V. KREINOVICH (2002): *Limit Theorems and Applications of Set-Valued and Fuzzy Set-Valued Random Variables*. Dordrecht: Kluwer. [769-771,793,795,799,804,805]

LIFSHITS, M. A. (1982): "On the Absolute Continuity of Distributions of Functionals of Random Processes," *Theory of Probability and Its Applications*, 27, 600–607. [794,796,808]

MAGNAC, T., AND E. MAURIN (2008): "Partial Identification in Monotone Binary Models: Discrete Regressors and Interval Data," *Review of Economic Studies*, forthcoming. [765,787]

MANSKI, C. F. (1989): "Anatomy of the Selection Problem," *Journal of Human Resources*, 24, 343–360. [764,768,781]

——— (2003): *Partial Identification of Probability Distributions*. New York: Springer Verlag. [764,779,782,783]

MANSKI, C. F., AND D. S. NAGIN (1998): "Bounding Disagreements About Treatment Effects: A Case Study of Sentencing and Recidivism," *Sociological Methodology*, 28, 99–137. [779]

MANSKI, C. F., AND J. V. PEPPER (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68, 997–1010. [779]

MANSKI, C. F., AND E. TAMER (2002): "Inference on Regressions With Interval Data on a Regressor or Outcome," *Econometrica*, 70, 519–546. [765,768]

MANSKI, C. F., G. D. SANDEFUR, S. MCLANAHAN, AND D. POWERS (1992): "Alternative Estimates of the Effect of Family Structure During Adolescence on High School Graduation," *Journal of the American Statistical Association*, 87, 25–37. [779]

MATHERON, G. (1975): *Random Sets and Integral Geometry*. New York: Wiley. [769]

MOLCHANOV, I. S. (2005): *Theory of Random Sets*. London: Springer Verlag. [769,785,793,794, 796-798,800,802,803]

MOLINARI, F. (2008a): "Missing Treatments," *Journal of Business & Economic Statistics*, forthcoming. [779]

——— (2008b): "Partial Identification of Probability Distributions With Misclassified Data," *Journal of Econometrics*, forthcoming. [769]

MOSLER, K. (2002): *Multivariate Dispersion, Central Regions and Depth. The Lift Zonoid Approach*, Lecture Notes in Statistics, Vol. 165. Berlin: Springer Verlag. [803]

MOURIER, E. (1955): "L-Random Elements and L^*-Random Elements in Banach Spaces," in *Proceedings of the Third Berkeley Symposium in Mathematical Statistics and Probability*, Vol. 2. Berkeley: University of California Press, 231–242. [773]

PAKES, A., J. PORTER, K. HO, AND J. ISHII (2005): "Moment Inequalities and Their Application," Discussion Paper, Harvard University and Wisconsin University. [768]

PEPPER, J. (2000): "The Intergenerational Transmission of Welfare Receipt: A Nonparametric Bounds Analysis," *Review of Economics and Statistics*, 82, 472–488. [779]

——— (2003): "Using Experiments to Evaluate Performance Standards: What Do Welfare-to-Work Demonstrations Reveal to Welfare Reformers?" *Journal of Human Resources*, 38, 860–880. [779]

POLITIS, D. N., J. P. ROMANO, AND M. WOLF (1999): *Subsampling*. New York: Springer Verlag. [794]

ROCKAFELLAR, R. (1970): *Convex Analysis*. Princeton, NJ: Princeton University Press. [764]

ROMANO, J. P., AND A. M. SHAIKH (2006): "Inference for the Identified Set in Partially Identified Econometric Models," Discussion Paper, Stanford University. [768]

ROSEN, A. (2006): "Confidence Sets for Partially Identified Parameters That Satisfy a Finite Number of Moment Inequalities," Discussion Paper, UCL. [768]

SCHARFSTEIN, D. O., C. F. MANSKI, AND J. C. ANTHONY (2004): "On the Construction of Bounds in Prospective Studies With Missing Ordinal Outcomes: Application to the Good Behavior Game Trial," *Biometrics*, 60, 154–164. [779]

SCHNEIDER, R. (1993): *Convex Bodies: The Brunn–Minkowski Theory*, Encyclopedia of Mathematics and Its Applications. Cambridge, U.K.: Cambridge University Press. [803,804,807,810]

STARR, R. M. (1969): "Quasi-Equilibria in Markets With Non-Convex Preferences," *Econometrica*, 37, 25–38. [772]

STOYE, J. (2007): "Bounds on Generalized Linear Predictors With Incomplete Outcome Data," *Reliable Computing*, 13, 293–302. [787]

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. [794]

VAN DER VAART, A. W. (2000): *Asymptotic Statistics*, Statistical and Probabilistic Mathematics. Cambridge, U.K.: Cambridge University Press. [807]

VAN DER VAART, A. W., AND J. A. WELLNER (2000): *Weak Convergence and Empirical Processes*. New York: Springer Verlag. [811]