Practical and Theoretical Advances in Inference for Partially Identified Models*

Ivan A. Canay
Department of Economics
Northwestern University
iacanay@northwestern.edu

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

November 15, 2016

Abstract

This paper surveys some of the recent literature on inference in partially identified models. After reviewing some basic concepts, including the definition of a partially identified model and the identified set, we turn our attention to the construction of confidence regions in partially identified settings. In our discussion, we emphasize the importance of requiring confidence regions to be uniformly consistent in level over relevant classes of distributions. Due to space limitations, our survey is mainly limited to the class of partially identified models in which the identified set is characterized by a finite number of moment inequalities or the closely related class of partially identified models in which the identified set is a function of a such a set. The latter class of models most commonly arise when interest focuses on a subvector of a vector-valued parameter, whose values are limited by a finite number of moment inequalities. We then rapidly review some important parts of the broader literature on inference in partially identified models and conclude by providing some thoughts on fruitful directions for future research.

KEYWORDS: Partially Identified Model, Confidence Regions, Uniform Asymptotic Validity, Moment Inequalities, Subvector Inference

JEL classification codes: C12, C14

*This paper was prepared for an invited talk at the 11th World Congress of the Econometric Society in Montreal, Canada. We would like to thank Joseph Romano, Andres Santos, and Edward Vytlacil for helpful comments. The research of the first author was supported by National Science Foundation Grants SES-1530534. The research of the second author was supported by National Science Foundation Grants DMS-1308260, SES-1227091, and SES-1530661

1 Introduction

A partially identified model is a model in which the parameter of interest is not uniquely determined by the distribution of the observed data. Instead, as we will explain further below, the parameter of interest is only limited by the distribution of the observed data to a set of possible values, commonly referred to as the identified set. Such models have a surprisingly long history: early contributions include the analysis of linear regressions with mismeasured regressors by Frisch (1934) and the analysis of Cobb-Douglas production functions by Marschak and Andrews (1944). Now, partially identified models are common in virtually all parts of economics and econometrics: measurement error (Klepper and Leamer, 1984; Horowitz and Manski, 1995), missing data (Manski, 1989, 1994; Horowitz and Manski, 1998; Manski and Tamer, 2002), industrial organization (Tamer, 2003; Haile and Tamer, 2003; Ho and Pakes, 2014; Pakes et al., 2015), finance (Hansen and Jagannathan, 1991; Hansen et al., 1995), labor economics (Blundell et al., 2007; Kline et al., 2013; Kline and Tartari, 2015), program evaluation (Manski, 1990, 1997; Manski and Pepper, 2000; Heckman and Vytlacil, 2001; Bhattacharya et al., 2008, 2012; Shaikh and Vytlacil, 2011), and macroeconomics (Faust, 1998; Canova and De Nicolo, 2002; Uhlig, 2005). The references above are far from exhaustive and simply illustrate the widespread popularity these types of models now enjoy in economics. A more detailed account of this history and further references are provided in the excellent summary in Tamer (2012) as well as the accompanying paper in this volume by Ho and Rosen (2016). The rise in the popularity of such models is in no small part due to the advocacy of Charles Manski, who, in a series of books and articles, argued forcefully that partially identified models enable researchers to make more credible inferences by allowing them to relax untenable assumptions used to obtain identification. In addition to aforementioned references, see the book-length treatments in Manski (1995, 2003, 2007, 2013).

In the remainder of this paper, we do not discuss further any examples of partially identified models. Instead, we focus on the problem of conducting inference in such models. Following the literature on this topic, we focus further on the construction of confidence regions in partially identified models. To facilitate our discussion, we begin in Section 2 by defining the notion of a partially identified model and the identified set for a parameter of interest more formally. Using the notation developed in that section, we then discuss in Section 3 two different generalizations of the usual notion of a confidence region in identified settings to partially identified settings – confidence regions for points in the identified set and confidence regions for the identified set itself. In order to keep our discussion of a manageable length, we subsequently restrict attention to confidence regions for points in the identified set, which have thus far received the most attention in the literature, but note that both types of confidence regions may be of interest. Our discussion emphasizes the importance of requiring confidence regions to have asymptotic validity that holds uniformly (in the distribution of the observed data) over relevant sets of distributions. As noted by Imbens and Manski (2004), in partially identified settings, naïve constructions of confidence regions may have asymptotic validity that only holds pointwise (in the distribution of the observed data) over relevant classes of distributions. These confidence regions may, as a consequence, behave poorly in finite samples. In our discussion, we illustrate this phenomenon by means of a simple example adapted from Imbens and Manski (2004).

In Section 4, we turn our attention to specific methods for constructing confidence regions for points in the identified set in a large class of partially identified models. The class of models we consider is the one in which the identified set is defined by (a finite number of) moment inequalities. Such models have numerous applications, including many of those referenced above. Several different approaches to inference in such models have been developed. Following our discussion above, we limit attention to those that exhibit uniform asymptotic validity over a large set of relevant distributions for the observed data. Along the way, we highlight some closely related results in the context of a Gaussian setting that shed light on the finite-sample properties of some of these approaches. We then discuss in Section 5 recently developed methods for inference in the closely related class of partially identified models in which the identified set is a function of a set defined by (a finite number of) moment inequalities. Such models most commonly arise when interest focuses on a subvector of a vector-valued parameter, whose values are limited by (a finite number of) moment inequalities, so we again limit our discussion to this leading special case. A distinguishing feature of our discussion in these two sections is the use of common notation throughout so as to make both the differences and similarities between the various approaches transparent.

The material described above is, of course, only a small subset of the ever-expanding literature on inference in partially identified models. In Section 6, we very briefly highlight several important strands of this literature that, as a consequence of space constraints, we are unable to include in greater detail – the literatures on "many" moment inequalities, conditional moment inequalities, random set-theoretic approaches to inference in partially identified models, and Bayesian approaches to inference in partially identified models. Finally, in Section 7, we conclude by providing some thoughts on fruitful directions for future research.

2 Partially Identified Models

In order to define the notion of a partially identified model more formally, suppose that a researcher observes data with distribution $P \in \mathbf{P} \equiv \{P_\gamma : \gamma \in \Gamma\}$. The set $\mathbf{P}$, consisting of distributions completely characterized by the (possibly infinite-dimensional) parameter γ, constitutes the model for the distribution of the observed data. In this notation, the identified set for γ is defined to be

$$\Gamma_0(P) \equiv \{\gamma \in \Gamma : P_\gamma = P\} \ .$$

In most cases, however, the researcher is not interested in γ itself, but rather a function of γ, say $\theta = \theta(\gamma)$. The identified set for θ is simply defined to be

$$\Theta_0(P) \equiv \theta(\Gamma_0(P)) = \{\theta(\gamma) : \gamma \in \Gamma_0(P)\} \ . \tag{1}$$

If $\Theta_0(P)$ is a singleton for all $P \in \mathbf{P}$, then, similar to Matzkin (2007), the parameter θ is said to be identified (relative to $\mathbf{P}$); if $\Theta_0(P) = \Theta$ for all $P \in \mathbf{P}$, then the parameter θ is said to be unidentified (relative to $\mathbf{P}$); otherwise, θ is said to be partially identified.

While not essential for much of our discussion, in the remainder of the paper, we assume that the observed data consists of n i.i.d. observations $W_i, i = 1, \ldots, n$ and that P denotes the common (marginal) distribution of these observations (rather than the joint distribution of the observations).

Remark 2.1. The identified set for θ defined in (1) is by construction "sharp" in the sense that each possible value for θ in the identified set is in fact compatible with the distribution of the observed data, i.e., for each value of $\theta \in \Theta_0(P)$ there exists $\gamma \in \Gamma$ with $P_\gamma = P$ and $\theta(\gamma) = \theta$. This terminology is sometimes used in the literature in order to draw a distinction with "non-sharp" sets $\tilde{\Theta}_0(P)$ that only satisfy $\tilde{\Theta}_0(P) \supseteq \Theta_0(P)$ for all $P \in \mathbf{P}$. These larger sets may still be useful when it is difficult to characterize $\Theta_0(P)$ directly, but obviously do not exhaust all restrictions on θ in the model $\mathbf{P}$. ■

Remark 2.2. To help make some of these abstract ideas more concrete, it is useful to describe them in the context of the familiar linear model. To this end, suppose

$$Y = \theta' X + \epsilon \ . \tag{2}$$

Here, the distribution of the observed data, P, is the distribution of (Y, X). The model for the distribution of the observed data, $\mathbf{P}$, consists of distributions P_γ for (Y, X) specified by $\gamma = (\theta, P_{X,\epsilon}) \in \Gamma$, where $P_{X,\epsilon}$ is a possible distribution for (X, ϵ), and (2). Often Γ is restricted so that $E_{P_\gamma}[\epsilon X] = 0$ and $E_{P_\gamma}[XX']$ is nonsingular for each $\gamma \in \Gamma$. Under these assumptions, γ is identified (relative to $\mathbf{P}$). In particular, as is well known, $\theta = \theta(\gamma)$ is identified (relative to $\mathbf{P}$) because it may be expressed as $\theta(\gamma) = E_{P_\gamma}[XX']^{-1}E_{P_\gamma}[XY]$, which clearly does not vary over the set $\Gamma_0(P)$. Incidentally, in this case, γ is also identified (relative to $\mathbf{P}$). ■

3 Types of Confidence Regions

The literature on inference in partially identified models has largely focused on the construction of confidence regions, which here we denote by C_n. Some exceptions include Manski (2007) and Manski (2013), which have instead treated the problem from a decision-theoretic perspective. Two distinct notions of confidence regions have been proposed in the literature. The first notion requires that the random set C_n covers each point in the identified set with some pre-specified probability $1 - \alpha$, i.e.,

$$\liminf_{n \to \infty} \inf_{P \in \mathbf{P}} \inf_{\theta \in \Theta_0(P)} P\{\theta \in C_n\} \geq 1 - \alpha \ . \tag{3}$$

For convenience, we henceforth refer to such confidence regions as confidence regions for points in the identified set that are uniformly consistent in level (over $P \in \mathbf{P}$ and $\theta \in \Theta_0(P)$). Such confidence regions have also been referred to as confidence regions for identifiable parameters that are uniformly consistent in level, as in Romano and Shaikh (2008). The second notion requires that the random set C_n covers the entire identified set with some pre-specified probability $1 - \alpha$, i.e.,

$$\liminf_{n \to \infty} \inf_{P \in \mathbf{P}} P\{\Theta_0(P) \subseteq C_n\} \geq 1 - \alpha \ . \tag{4}$$

For convenience, we henceforth refer to such confidence regions as confidence regions for the identified set that are uniformly consistent in level (over $P \in \mathbf{P}$), as in Romano and Shaikh (2010).

Importantly, when θ is identified (relative to $\mathbf{P}$), then both (3) and (4) reduce to

$$\liminf_{n\to\infty} \inf_{P\in\mathbf{P}} P\{\theta_0(P) \in C_n\} \geq 1 - \alpha \ ,$$

where $\theta_0(P)$ is such that $\Theta_0(P) = \{\theta_0(P)\}$ for all $P \in \mathbf{P}$. In this sense, both (3) and (4) generalize the usual notion of a confidence region for parameters that are identified.

As emphasized by Imbens and Manski (2004), confidence region satisfying (4) of course satisfy (3) as well, so confidence regions for points in the identified set are typically smaller than confidence regions for the identified set. Imbens and Manski (2004) argue further that confidence regions for points in the identified set are generally of greater interest than confidence regions for the identified set itself, as there is still only one "true" value for θ in the identified set. Other authors, however, have argued that in some instances confidence regions for the identified set are more desirable. See, for example, Henry and Onatski (2012). Nevertheless, in this review, we focus on confidence regions satisfying (3), which are the type that have received the most attention in the literature on inference in partially identified models. Notable exceptions include Chernozhukov et al. (2007), Bugni (2010), and Romano and Shaikh (2010).

Confidence regions satisfying (3) can be constructed by exploiting the well-known duality between confidence regions and inverting tests of each of the individual null hypotheses

$$H_\theta : \theta \in \Theta_0(P) \tag{5}$$

versus the unrestricted alternative hypothesis that control appropriately the usual probability of a Type 1 error at level $\alpha \in (0, 1)$. More specifically, suppose that for each θ a test of H_θ, $\phi_n(\theta)$, is available that satisfies

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}} \sup_{\theta\in\Theta_0(P)} E_P[\phi_n(\theta)] \leq \alpha \ . \tag{6}$$

It follows that C_n equal to the set of $\theta \in \Theta$ for which H_θ is accepted satisfies (3).

Confidence regions satisfying (4) can also be constructed using the duality between hypothesis testing and constructing confidence regions. However, in this case the tests $\phi_n(\theta)$ for the family of null hypotheses H_θ indexed by $\theta \in \Theta$ are required to control appropriately the familywise error rate, i.e., the probability of even one false rejection under P. See Romano and Shaikh (2010) for further details.

3.1 Uniform vs. Pointwise Consistency in Level

The terminology introduced above for confidence regions satisfying (3) and (4) is intended to distinguish them, respectively, from confidence regions for points in the identified set that are pointwise consistent in level (over $P \in \mathbf{P}$ and $\theta \in \Theta_0(P)$), i.e., those only satisfying

$$\liminf_{n\to\infty} P\{\theta \in C_n\} \geq 1 - \alpha \text{ for all } P \in \mathbf{P} \text{ and } \theta \in \Theta_0(P) \ , \tag{7}$$

and confidence regions for the identified set that are pointwise consistent in level (over $P \in \mathbf{P}$), i.e., those only satisfying

$$\liminf_{n\to\infty} P\{\Theta_0(P) \subseteq C_n\} \geq 1 - \alpha \text{ for all } P \in \mathbf{P} \ . \tag{8}$$

For confidence regions only satisfying (7), the probability of covering some points in the identified set may be too small under certain distributions $P \in \mathbf{P}$ even for arbitrarily large sample sizes. Indeed, the requirement (7) by itself does not even rule out, for example, that for arbitrarily large values of n and every $\epsilon > 0$, there is $P = P(n, \epsilon) \in \mathbf{P}$ and $\theta = \theta(\epsilon, n) \in \Theta_0(P)$ such that

$$P\{\theta \in C_n\} < \epsilon \ .$$

In this sense, inferences based off of confidence regions only satisfying (7) may be very misleading in finite samples. Analogous statements apply to confidence regions that only satisfy (8). These types of behavior are prohibited for confidence regions satisfying the stronger coverage requirements (3) and (4).

In many well-behaved problems, the distinction between confidence regions that are uniformly consistent in level and confidence regions that are pointwise consistent in level is entirely a technical issue: most constructions of confidence regions that are pointwise consistent in level are also uniformly consistent in level provided, for example, that $\mathbf{P}$ is assumed to satisfy some additional moment restrictions. In the case of confidence regions for the mean with i.i.d. data, for instance, the usual constructions are uniformly consistent in level provided that $\mathbf{P}$ satisfies a weak uniform integrability condition instead of simply requiring that a second moment exist. For further discussion, see Bahadur and Savage (1956) and Romano (2004). In less well-behaved problems, however, such as inference in partially identified models, the distinction between confidence regions that are pointwise consistent in level versus those that are uniformly consistent in level is less trivial: some seemingly natural constructions may only achieve uniform consistency in level by restricting $\mathbf{P}$ in ways that rule out relevant distributions. Before proceeding, we illustrate this point in the context of a simple example.

Example 3.1. Let $W_i = (L_i, U_i), i = 1, \ldots, n$ be an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$ on $\mathbf{R}^2$. For ease of exposition, we assume that

$$\mathbf{P} = \{N(\mu, \Sigma) : \mu = (\mu_L, \mu_U) \in \mathbf{R}^2 \text{ with } \mu_L < \mu_U\} ,$$

where Σ is a known covariance matrix with unit variances. The identified set for the parameter of interest θ is assumed to be

$$\Theta_0(P) = [\mu_L(P), \mu_U(P)] .$$

Consider the confidence region

$$C_n = \left[\bar{L}_n - \frac{z_{1-\alpha}}{\sqrt{n}}, \bar{U}_n + \frac{z_{1-\alpha}}{\sqrt{n}}\right] , \tag{9}$$

where $\bar{L}_n$ and $\bar{U}_n$ denote sample averages of $L_i, i = 1, \ldots, n$ and $U_i, i = 1, \ldots, n$, respectively. It follows that

$$P\{\theta \in C_n\} = \Phi(z_{1-\alpha} + \sqrt{n}(\mu_U(P) - \theta)) - \Phi(-z_{1-\alpha} - \sqrt{n}(\theta - \mu_L(P)))$$

for any $P \in \mathbf{P}$ and $\theta \in \Theta_0(P)$, where Φ is the standard normal c.d.f. and $z_{1-\alpha} = \Phi^{-1}(1-\alpha)$. From this expression, it is straightforward to verify that C_n satisfies (7). However, it is also clear that

$$\inf_{P \in \mathbf{P}} \inf_{\theta \in \Theta_0(P)} P\{\theta \in C_n\} = 1 - 2\alpha < 1 - \alpha , \tag{10}$$

so C_n fails to satisfy (3). In other words, for every n, there exists P and $\theta \in \Theta_0(P)$ (with $\mu_L(P)$ and $\mu_U(P)$ close enough together relative to n) under which $P\{\theta \in C_n\} \approx 1 - 2\alpha$. Here, by close enough together relative to n we mean that $\sqrt{n}(\mu_U(P) - \mu_L(P))$ is small. Importantly, this requirement depends on both n and $\mu_U(P) - \mu_L(P)$. In particular, even if one were willing to assume further that

$$\inf_{P \in \mathbf{P}} \mu_U(P) - \mu_L(P) = \underline{\Delta} > 0 ,$$

the sample size required for the approximation to be "good" will still diverge as $\underline{\Delta} \to 0$. ■

Remark 3.1. Imbens and Manski (2004) first highlighted the point made in Example 3.1 in the context of a closely related missing data example. In their setting, $W_i = (X_i Z_i, Z_i), i = 1, \ldots, n$ is an i.i.d. sequence of random variables with distribution $P \in \mathbf{P}$ on $[0,1] \times \{0,1\}$. Because X_i is only observed if $Z_i = 1$, the setting is a simple example of missing data. The parameter of interest is

$$\theta = \theta(Q) \equiv E_Q[X_i] ,$$

where Q is the distribution of (X_i, Z_i). Using the fact that

$$X_i Z_i \leq X_i \leq X_i Z_i + 1 - Z_i ,$$

it is straightforward to show that the identified set for θ is given by

$$\Theta_0(P) = [E_P[X_i Z_i], E_P[X_i Z_i + 1 - Z_i]] .$$

By arguing as in Example 3.1, it is possible to show that the natural counterpart to C_n in (9) satisfies both (7) and (10) provided that

$$\inf_{P \in \mathbf{P}} P\{Z_i = 1\} = 0 ,$$

i.e., the case in which (nearly) all the data is observed is not *a priori* ruled out. ■

4 Inference for Moment Inequalities

In this section, we review several approaches that have been proposed in the literature for inference in models in which the identified set for θ is defined by moment inequalities, i.e.,

$$\Theta_0(P) = \{\theta \in \Theta : E_P[m(W_i, \theta)] \leq 0\} . \tag{11}$$

Here, m is a function taking values in $\mathbf{R}^k$ and the inequality is interpreted component-wise. We focus on the construction of confidence regions satisfying (3). As explained in Section 3, such confidence regions may be constructed by inverting tests of the null hypotheses (5) in a way that satisfies (6). In the setting considered here, these null hypotheses may be written as

$$H_\theta : E_P[m(W_i, \theta)] \le 0 \ . \tag{12}$$

A variety of different tests for (12) have been proposed in the literature. In order to describe these tests succinctly, it is useful to introduce some common notation. To this end, define

$$\begin{aligned} \mu(\theta, P) &= E_P[m(W_i, \theta)] \\ \Omega(\theta, P) &= \text{Corr}_P[m(W_i, \theta)] \\ D(\theta, P) &= \text{diag}(\sigma_j(\theta, P) : 1 \le j \le k) \ , \end{aligned}$$

where

$$\sigma_j^2(\theta, P) = \text{Var}_P[m_j(W_i, \theta)]$$

for $m_j(W_i, \theta)$ equal to the jth component of $m(W_i, \theta)$. Similarly, define

$$\begin{aligned} \bar{m}_n(\theta) &= \mu(\theta, \hat{P}_n) \\ \hat{\Omega}_n(\theta) &= \Omega(\theta, \hat{P}_n) \\ \hat{D}_n(\theta) &= D(\theta, \hat{P}_n) \ , \end{aligned}$$

where $\hat{P}_n$ is the empirical distribution of $W_i, i = 1, \ldots, n$.

All of the tests we discuss below reject H_θ for large values a test statistic

$$T_n(\theta) \equiv T(\hat{D}_n^{-1}(\theta)\sqrt{n}\bar{m}_n(\theta), \hat{\Omega}_n(\theta)) \ , \tag{13}$$

where T is a real-valued function that is weakly increasing in each component of its first argument, continuous in both arguments, and satisfies some additional conditions that mainly control how T behaves when some of the components of its first argument are very large in magnitude and negative. See, for example, Andrews and Soares (2010) for a detailed description of these conditions. Some common examples of statistics satisfying all of the required conditions are the modified method of moments, maximum, and adjusted quasi-likelihood ratio statistics, given by

$$T_n^{\text{mmm}}(\theta) = \sum_{1\le j\le k} \max\left\{\frac{\sqrt{n}\bar{m}_{n,j}(\theta)}{\hat{\sigma}_{n,j}(\theta)}, 0\right\}^2 \tag{14}$$

$$T_n^{\max}(\theta) = \max\left\{\max_{1\le j\le k}\frac{\sqrt{n}\bar{m}_{n,j}(\theta)}{\hat{\sigma}_{n,j}(\theta)}, 0\right\} \tag{15}$$

$$T_n^{\text{ad,qlr}}(\theta) = \inf_{t\in\mathbf{R}^k : t\le 0}\left(\hat{D}_n^{-1}(\theta)\sqrt{n}\bar{m}_n(\theta) - t\right)'\tilde{\Omega}_n(\theta)^{-1}\left(\hat{D}_n^{-1}(\theta)\sqrt{n}\bar{m}_n(\theta) - t\right) , \tag{16}$$

where $\bar{m}_{n,j}(\theta)$ equals the jth component of $\bar{m}_n(\theta)$, $\hat{\sigma}_{n,j}(\theta) \equiv \sigma_j(\theta, \hat{P}_n)$, and

$$\tilde{\Omega}_n(\theta) = \max\{\epsilon - \det(\hat{\Omega}_n(\theta)), 0\}I_k + \hat{\Omega}_n(\theta)$$

for some fixed $\epsilon > 0$, with I_k denoting the k-dimensional identity matrix. The adjustment referred to in the adjusted quasi-likelihood ratio statistic refers to use of $\tilde{\Omega}_n(\theta)$ instead of $\hat{\Omega}_n(\theta)$. The use of this modification stems from the desire to accommodate situations in which $\Omega(\theta, P)$ is (nearly) singular. In the example described in Remark 3.1, this situation arises naturally when the identified set is (nearly) a singleton.

In order to describe different ways of constructing critical values with which to compare $T_n(\theta)$, it is useful to introduce

$$J_n(x, s(\theta), \theta, P) = P\{T(\hat{D}_n^{-1}(\theta)\sqrt{n}(\bar{m}_n(\theta) - \mu(\theta, P)) + \hat{D}_n^{-1}(\theta)s(\theta), \hat{\Omega}_n(\theta)) \le x\} . \tag{17}$$

In terms of (17), the distribution of $T_n(\theta)$ itself is simply

$$J_n(x, \sqrt{n}\mu(\theta, P), \theta, P) = P\{T(\hat{D}_n^{-1}(\theta)\sqrt{n}\bar{m}_n(\theta), \hat{\Omega}_n(\theta)) \le x\} . \tag{18}$$

It is straightforward to derive useful estimators of (17) for a fixed value of $s(\theta)$. For example, one may use the usual nonparametric bootstrap estimator, $J_n(x, s(\theta), \theta, \hat{P}_n)$, or the estimator $J_n(x, s(\theta), \theta, \tilde{P}_n(\theta))$, where $m(W_i, \theta) \sim N(\bar{m}_n(\theta), \hat{\Sigma}_n(\theta))$ under $\tilde{P}_n(\theta)$. On the other hand, it is difficult to derive useful estimators of (18) because it is not possible to estimate $\sqrt{n}\mu(\theta, P)$ consistently. Indeed, its natural estimator, $\sqrt{n}\bar{m}_n(\theta)$, satisfies

$$|\sqrt{n}\bar{m}_n(\theta) - \sqrt{n}\mu(\theta, P)| \xrightarrow{d} |N(0, \Sigma(\theta, P))|$$

under any fixed $\theta \in \Theta_0(P)$ and $P \in \mathbf{P}$, where $\Sigma(\theta, P) = \text{Var}_P[m(W_i, \theta)]$. The different tests we discuss below are mainly distinguished by the way in which they circumvent this difficulty.

Remark 4.1. Even though the first-order or even higher-order asymptotic properties of the tests described below do not depend on whether $J_n(x, s(\theta), \theta, \hat{P}_n)$ or $J_n(x, s(\theta), \theta, \tilde{P}_n(\theta))$ is used as an estimator of (17), Andrews and Barwick (2012) find in a simulation study that $J_n(x, s(\theta), \theta, \hat{P}_n)$ appears to yield some improvements in finite samples over $J_n(x, s(\theta), \theta, \tilde{P}_n(\theta))$ in terms of power. On the other hand, our experience suggests it is much faster to compute $J_n(x, s(\theta), \theta, \tilde{P}_n(\theta))$ than $J_n(x, s(\theta), \theta, \hat{P}_n)$, so, depending on the scale of the problem, it may be convenient to use $J_n(x, s(\theta), \theta, \tilde{P}_n(\theta))$ instead of $J_n(x, s(\theta), \theta, \hat{P}_n)$. ■

Remark 4.2. Note that moment equalities may be included in $\Theta_0(P)$ defined in (11) simply by including both the moment and the negative of the moment as components of m. Doing so, of course, imposes very strong dependence between the corresponding components of m, but it does not affect the validity of the inferential methods discussed below, as none of them impose any restrictions on $\Omega(\theta, P)$. However, in some cases, it may be possible to further improve the procedure by exploiting this additional structure. See, for example, Remark 4.10 below. ■

Remark 4.3. With an appropriate choice of m, the identified set in Example 3.1 may be written in the form given in (11). In particular, it suffices to choose

$$m(W_i, \theta) = \begin{pmatrix} L_i - \theta \\ \theta - U_i \end{pmatrix} .$$

Many other examples of partially identified models can also be accommodated in this framework. See, for example, the discussion of entry models in Ho and Rosen (2016). ■

4.1 Five Methods for Inference for Moment Inequalities

In this section, we describe five different tests for (12). Each of these tests satisfy (6) provided that the following uniform integrability requirement holds:

$$\limsup_{t \to \infty} \sup_{P \in \mathbf{P}} \sup_{\theta \in \Theta_0(P)} E_P \left[\left(\frac{m_j(W_i, \theta) - \mu(\theta, P)}{\sigma_j(\theta, P)} \right)^2 I \left\{ \frac{m_j(W_i, \theta) - \mu(\theta, P)}{\sigma_j(\theta, P)} > t \right\} \right] = 0 . \quad (19)$$

Note that a sufficient condition for (19) is that

$$\sup_{P \in \mathbf{P}} \sup_{\theta \in \Theta_0(P)} E_P \left[\left(\frac{m_j(W_i, \theta) - \mu(\theta, P)}{\sigma_j(\theta, P)} \right)^{2+\delta} \right] < \infty$$

for some $\delta > 0$. This mildly stronger condition has been used in much of the literature in order to achieve uniform consistency in level.

4.1.1 Least Favorable Tests

Least favorable tests are based off of the following observation: since T is increasing in each component of its first argument and

$$\sqrt{n}\mu(\theta, P) \leq 0$$

for any θ and P such that $\theta \in \Theta_0(P)$,

$$J_n^{-1}(1 - \alpha, \sqrt{n}\mu(\theta, P), \theta, P) \leq J_n^{-1}(1 - \alpha, 0_k, \theta, P) \ ,$$

where 0_k is a k-dimensional vector of zeros. In this sense, 0_k is the least favorable value of the nuisance parameter $\sqrt{n}\mu(\theta, P)$. It follows that least favorable tests of the form

$$\phi_n^{\text{lf}}(\theta) \equiv I\{T_n(\theta) > \hat{J}_n^{-1}(1 - \alpha, 0_k, \theta)\} \ ,$$

where $\hat{J}_n(x, 0_k, \theta)$ equals either $J_n(x, 0_k, \theta, \hat{P}_n)$ or $J_n(x, 0_k, \theta, \tilde{P}_n(\theta))$, satisfy (6). Such tests or closely related ones have a long history in the statistics and econometrics literature. See, for example, Kudo (1963), Wolak (1987) and Wolak (1991). For inference in partially identified models, their use has been proposed, for example, by Rosen (2008) and Andrews and Guggenberger (2009).

Such tests are often regarded as being "conservative" in that they compute the critical value under the assumption that all moments are "binding," i.e., $\mu(\theta, P) = 0$. It is therefore worth highlighting the fact that the tests corresponding to these tests in a Gaussian setting are in fact admissible and some are even enjoy certain types of optimality among a restricted class of tests. Before proceeding, we elaborate on these points briefly in the following example.

Example 4.1. Let $X_i, i = 1, \ldots, n$ be an i.i.d. sequence of random variables on $\mathbf{R}^k$ with distribution $N(\mu, \Sigma)$, where $\mu \in \Pi \equiv \mathbf{R}^k$ and Σ is a known, invertible covariance matrix. Consider the problem of testing

$$H_0 : \mu \in \Pi_0 \text{ versus } H_1 : \mu \in \Pi_1 \ , \tag{20}$$

where $\Pi_0 = \{\mu \in \Pi : \mu \leq 0\}$ and $\Pi_1 = \Pi \setminus \Pi_0$. By sufficiency, we may without loss of generality assume that $n = 1$. Hence, the data consists of a single random variable X distributed according to a multivariate Gaussian distribution with unknown mean μ and known, invertible covariance matrix Σ. Here, it is possible to obtain some finite-sample results, so we restrict attention to tests ϕ that satisfy

$$\sup_{\mu \in \Pi_0} E_\mu[\phi] \leq \alpha \ . \tag{21}$$

It is well known that there is no uniformly most powerful test in this setting. This follows, for example, from the derivation in Romano et al. (2014) of the (unique) most powerful test of (20) against the alternative hypothesis that $\mu = a \in \Pi_1$. This most powerful test is a non-trivial function of a and has power equal to

$$1 - \Phi\left(z_{1-\alpha} - \inf_{t \in \mathbf{R}^k : t \leq 0} \sqrt{(t-a)'\Sigma^{-1}(t-a)}\right) \ , \tag{22}$$

where Φ is the standard normal c.d.f. and $z_{1-\alpha} = \Phi^{-1}(1-\alpha)$. Since (22) depends on a, it follows that there is no uniformly most powerful test. Furthermore, restricting attention to unbiased or similar tests is not helpful. See, Lehmann (1952) and Andrews (2012). On the other hand, Romano and Shaikh (2015) show that there are a wide variety of admissible tests. In order to describe these results precisely, it is useful to recall the concepts of α-admissibility and d-admissibility, specialized to the current testing problem. A test ϕ satisfying (21) is α-admissible if for any other test $\tilde{\phi}$ satisfying (21),

$$E_\mu[\tilde{\phi}] \geq E_\mu[\phi] \text{ for all } \mu \in \Pi_1 \tag{23}$$

implies that $E_\mu[\tilde{\phi}] = E_\mu[\phi]$ for all $\mu \in \Pi_1$. A test ϕ satisfying (21) is d-admissible if for any other test $\tilde{\phi}$, (21), (23) and

$$E_\mu[\tilde{\phi}] \leq E_\mu[\phi] \text{ for all } \mu \in \Pi_0$$

imply that $E_\mu[\tilde{\phi}] = E_\mu[\phi]$ for all $\mu \in \Pi$. In this language, Romano and Shaikh (2015) establish that ϕ is d-admissible if it satisfies (21) and its acceptance region is a closed, convex subset of $\mathbf{R}^k$ containing Π_0. If, in addition, the supremum in (21) is attained at some $\mu \in \Pi_0$ and the inequality holds with equality, then ϕ is α-admissible. From these two results, it follows that the counterparts to the least favorable tests based on the three test statistics (14) – (16) are all α-admissible and d-admissible. Romano and Shaikh (2015) go on to show further that the counterpart to the least favorable test based on (15) even enjoys certain types of optimality among a restricted class of tests. The restricted class of tests are non-randomized tests ϕ that satisfy (21) and are monotone in the sense that $\phi(x) = 1$ implies that $\phi(x') = 1$ for all $x' \geq x$. For the testing problem under consideration, the restriction to monotone tests does not at first appear unreasonable. Among this class of tests, the least favorable test based on (15) maximizes

$$\inf_{\mu \in A(\epsilon)} E_\mu[\phi]$$

for any $\epsilon > 0$, where $A(\epsilon) = \{\mu \in \Pi : \mu_j \geq \epsilon \text{ for some } 1 \leq j \leq k\}$. In the case where $k = 2$, this result was established previously by Lehmann (1952). Generalizations to $k > 2$ can be found in Lehmann et al. (2012), but there the joint distribution of X is additionally required to be exchangeable. ■

Despite these seemingly attractive properties, it may not be desirable to construct confidence regions by inverting least favorable tests tests. While it is not possible to construct tests that are unambiguously better than such tests, other tests that incorporate information in the data about $\sqrt{n}\mu(\theta, P)$ when constructing critical values may have much better power over many alternatives at the expense of only marginally worse power at other alternatives. From the perspective of constructing confidence regions that are small, these tests will therefore be more desirable than least favorable tests.

4.1.2 Subsampling

Subsampling-based tests use the subsampling estimate of the distribution of interest (18). In order to define the subsampling estimate of this distribution, we require some further notation. To this end, let $0 < b = b_n < n$ be a sequence of integers such that $b \to \infty$ and $b/n \to 0$. Let $N_n = \binom{n}{b_n}$ and index by $1 \leq \ell \leq N_n$ the distinct subsets of $W_i, i = 1, \ldots, n$ of size b. Denote by $T_{b,\ell}(\theta)$ the quantity $T_b(\theta)$ computed using the ℓth subset of data of size b. Using this notation, the subsampling estimate of $J_n(x, \sqrt{n}\mu(\theta, P), \theta, P)$ is given by

$$L_n(x, \theta) = \frac{1}{N_n} \sum_{1 \leq \ell \leq N_n} I\{T_{b,\ell}(\theta) \leq x\} ,$$

and the corresponding test is given by

$$\phi_n^{\rm sub}(\theta) = I\{T_n(\theta) > L_n^{-1}(1-\alpha,\theta)\} \ .$$

It is possible to show that these tests satisfy (6). In order to gain an appreciation for this result, we sketch some of the main components of a rigorous argument. First note that $L_n(x)$ is a "good" estimator of

$$J_b(x, \sqrt{b}\mu(\theta,P),\theta,P) = P\{T(\hat{D}_b(\theta)^{-1}\sqrt{b}(\bar{m}_b(\theta) - \mu(\theta,P)) + \hat{D}_b(\theta)^{-1}\sqrt{b}\mu(\theta,P), \hat{\Omega}_b(\theta)) \le x\} \ .$$

Indeed, as shown by Romano and Shaikh (2012), for any $\epsilon > 0$, $L_n(x,\theta)$ satisfies

$$\sup_{x\in\mathbf{R}} \sup_{P\in\mathbf{P}} \sup_{\theta\in\Theta_0(P)} P\left\{\sup_{x\in\mathbf{R}} |L_n(x,\theta) - J_b(x,\sqrt{b}\mu(\theta,P),\theta,P)| > \epsilon\right\} \to 0 \ .$$

Next, in order to link $J_b(x,\sqrt{b}\mu(\theta,P),\theta,P)$ with $J_n(x,\sqrt{b}\mu(\theta,P),\theta,P)$, note that

$$\sup_{P\in\mathbf{P}} \sup_{\theta\in\Theta_0(P)} \sup_{s\le 0} |J_b(x,s,\theta,P) - J_n(x,s,\theta,P)| \to 0 \ . \tag{24}$$

The convergence in (24) can be established using a subsequencing argument and Polya's Theorem, as in Romano and Shaikh (2008). Finally, since T is increasing in each component of its first argument and

$$\sqrt{b}\mu(\theta,P) \ge \sqrt{n}\mu(\theta,P)$$

for any θ and P such that $\mu(\theta,P) \le 0$,

$$J_n^{-1}(1-\alpha,\sqrt{n}\mu(\theta,P),\theta,P) \le J_n^{-1}(1-\alpha,\sqrt{b}\mu(\theta,P),\theta,P) \ .$$

As with "least favorable" tests described in the preceding section, the validity of this test hinges on the weak monotonicity of T. This type of result has been established in the literature by Romano and Shaikh (2008) and Andrews and Guggenberger (2009).

Remark 4.4. In practice, implementing subsampling of course requires the choice of a subsample size $b = b_n$. In simulation studies, subsampling appears to work well for an appropriately chosen subsample size, but may behave poorly in finite samples for other choices of the subsample size. See, for example, the simulation study in Bugni (2014). Some data-dependent rules for choosing the subsample size are described in Politis et al. (1999). For an empirical application that make use of subsampling, see the study of the airline industry by Ciliberto and Tamer (2010). ■

4.1.3 Generalized Moment Selection

Generalized moment selection tests are tests of the form

$$\phi_n^{\text{gms}}(\theta) \equiv I\{T_n(\theta) > \hat{J}_n^{-1}(1-\alpha, \hat{s}_n^{\text{gms}}(\theta), \theta)\} ,$$

where $\hat{J}_n(x, \hat{s}_n^{\text{gms}}(\theta), \theta)$ equals either $J_n(x, \hat{s}_n^{\text{gms}}(\theta), \theta, \hat{P}_n)$ or $J_n(x, \hat{s}_n^{\text{gms}}(\theta), \theta, \tilde{P}_n(\theta))$ and

$$\hat{s}_n^{\text{gms}}(\theta) = (\hat{s}_{n,1}^{\text{gms}}(\theta), \ldots, \hat{s}_{n,k}^{\text{gms}}(\theta))'$$

is a function that "selects" which moments are binding. While a wide variety of such functions are possible, below we assume that

$$\hat{s}_{n,j}^{\text{gms}}(\theta) = \begin{cases} 0 & \text{if } \frac{\sqrt{n}\bar{m}_{n,j}(\theta)}{\hat{\sigma}_{n,j}(\theta)} > -\kappa_n \\ -\infty & \text{otherwise} \end{cases} , \tag{25}$$

where

$$0 < \kappa_n \to \infty \text{ and } \kappa_n/\sqrt{n} \to 0 , \tag{26}$$

e.g., $\kappa_n = \log n$. To see why this terminology is appropriate, note that

$$\frac{\sqrt{n}\bar{m}_{n,j}(\theta)}{\hat{\sigma}_{n,j}(\theta)} = \frac{\sqrt{n}(\bar{m}_{n,j}(\theta) - \mu_j(\theta, P))}{\hat{\sigma}_{n,j}(\theta)} + \frac{\sqrt{n}\mu_j(\theta, P)}{\hat{\sigma}_{n,j}(\theta)} . \tag{27}$$

For θ and P such that $\mu_j(\theta, P) \leq 0$, the first term on right-hand side of (27) is $O_P(1)$, whereas the second term either equals zero or diverges in probability to $-\infty$ depending, respectively, on whether $\mu_j(\theta, P) = 0$ or $\mu_j(\theta, P) < 0$. Hence, for such θ and P, with probability approaching one, $\hat{s}_{n,j}^{\text{gms}}(\theta)$ equals 0 or $-\infty$ depending, respectively, on whether $\mu_j(\theta, P) = 0$ or $\mu_j(\theta, P) < 0$. In this sense, $\hat{s}_{n,j}^{\text{gms}}(\theta)$ "selects" whether $\mu_j(\theta, P) = 0$ or $\mu_j(\theta, P) < 0$.

It is possible to show that these tests also satisfy (6). In order to gain an appreciation for this result, we sketch some of the main components of a rigorous argument. As before, the argument hinges on the weak monotonicity of T, but together with the an additional insight related to what features of $\sqrt{n}\mu(\theta, P)$ can be consistently estimated. To this end, first note that to establish (6), it is enough to consider the rejection probability under any sequence $\{(\theta_n, P_n) : \theta_n \in \Theta_0(P_n)\}_{n\geq 1}$. By considering a further subsequence if necessary, one may restrict attention to sequences such that $\hat{\Omega}_n(\theta_n) \stackrel{P}{\to} \Omega^*$ and $\sqrt{n}\mu(\theta_n, P_n) \to s^*$, where s^* may have some components equal to $-\infty$. While consistent estimation of s^* is not possible for the reasons mentioned before, note that $\hat{s}_n^{\text{gms}}(\theta)$ consistently estimates those components that are equal to $-\infty$ and otherwise estimates the components with 0, which is at worst too large, i.e., $\hat{s}_n^{\text{gms}}(\theta_n) \stackrel{P}{\to} s^{\text{gms},*}$, where $s^{\text{gms},*}$ is such that

$$s_j^{\text{gms},*} \equiv \begin{cases} 0 & \text{if } s_j^* > -\infty \\ -\infty & \text{if } s_j^* = -\infty \end{cases} .$$

Furthermore, under such sequences,

$$\begin{pmatrix} T_n(\theta_n) \\ \hat{J}_n^{-1}(1-\alpha, \hat{s}_n^{\text{gms}}(\theta_n), \theta_n) \end{pmatrix} \stackrel{d}{\to} \begin{pmatrix} T(Z + s^*, \Omega^*) \\ J^{-1}(1-\alpha, s^{\text{gms},*}, \Omega^*) \end{pmatrix}$$

where $Z \sim N(0, \Omega^*)$ and $J(x, s, \Omega^*) = P\{T(Z+s, \Omega^*) \leq x\}$. Thus, the limiting rejection probability of the generalized moment selection test under such a sequence equals

$$P\{T(Z + s^*, \Omega^*) > J^{-1}(1-\alpha, s^{\text{gms},*}, \Omega^*)\} .$$

The validity of generalized moment selection tests thus follows from the fact that $s^{\text{gms},*} \geq s^*$. This type of result was first established in the literature by Andrews and Soares (2010), though related results can also be found in Canay (2010) and Bugni (2014).

Remark 4.5. The idea behind generalized moment selection tests bears some resemblance to the idea behind Hodge's superefficient estimator. Indeed, the proposed test would remain valid if (25) were replaced with

$$\hat{s}_{n,j}^{\text{gms,alt}}(\theta) = \begin{cases} 0 & \text{if } \frac{\sqrt{n}\bar{m}_{n,j}(\theta)}{\hat{\sigma}_{n,j}(\theta)} > -\kappa_n \\ \frac{\sqrt{n}\bar{m}_{n,j}(\theta)}{\hat{\sigma}_{n,j}(\theta)} & \text{otherwise} \end{cases} , \tag{28}$$

as this modification would only further increase the critical value. In this form, the connection to Hodge's estimator is clear. It is, of course, well known that Hodge's estimator has undesirable properties from an estimation perspective. More specifically it obtains its improved performance at zero at the expense of arbitrarily worse performance at points near zero. As an estimator of

$$\frac{\sqrt{n}\mu_j(\theta, P)}{\sigma_j(\theta, P)} ,$$

(28) also suffers from these shortcomings, but, due to the weak monotonicity of T, it remarkably leads to tests that satisfy (6) when used in the way described above. ■

Remark 4.6. Andrews and Soares (2010) compare the limiting power of subsampling tests and generalized moment selection tests against certain local alternatives. The comparison depends on the relationship between κ_n and b_n. In particular, if

$$\lim_{n\to\infty} \frac{\kappa_n \sqrt{b_n}}{\sqrt{n}} = 0 , \tag{29}$$

then the limiting power of generalized moment selection tests exceeds the limiting power of subsampling tests against these local alternatives, but this ordering may be reversed when (29) does not hold. Importantly, for b_n that is optimal in terms of the order of the error in rejection probability under the null hypothesis and typical choices of κ_n, (29) holds. ■

4.1.4 Refined Moment Selection

Refined moment selection tests, proposed by Andrews and Barwick (2012), are motivated by the following unattractive feature of generalized moment selection tests: κ_n is only restricted to satisfy (26), so the choice of κ_n for a given sample size n is indeterminate. Unfortunately, given that κ_n satisfies (26), none of the first-order asymptotic properties of generalized moment selection tests depend on κ_n. Indeed, as shown by Bugni (2014), even the higher-order asymptotic properties of generalized moment selection tests do not depend on κ_n. It is therefore difficult to use asymptotic considerations to determine a rule for κ_n. On the other hand, it is clear that for a given sample of size n, a smaller choice of κ_n would be more desirable in terms of power. In an effort to ameliorate this drawback, Andrews and Barwick (2012) study generalized moment selection tests in which κ_n is replaced with a non-negative constant, κ. With this modification, even the first-order asymptotic properties of the test depend on κ, which, as we will explain below, ultimately enables the development of data-dependent choices $\hat{\kappa}_n$ of κ.

To this end, first consider

$$\hat{J}_n^{-1}(1-\alpha, \hat{s}_n^{\text{rms}}(\theta), \theta) ,$$

where $\hat{J}_n(x, \hat{s}_n^{\text{rms}}(\theta), \theta)$ equals $J_n(x, \hat{s}_n^{\text{rms}}(\theta), \theta, \hat{P}_n)$ or $J_n(x, \hat{s}_n^{\text{rms}}(\theta), \theta, \tilde{P}_n(\theta))$ and

$$\hat{s}_n^{\text{rms}}(\theta) = (\hat{s}_{n,1}^{\text{rms}}(\theta), \ldots, \hat{s}_{n,k}^{\text{rms}}(\theta))'$$

is again a function that "selects" which moments are binding. As before, while a wide variety of such functions are possible, below we assume that

$$\hat{s}_{n,j}^{\text{rms}}(\theta) = \begin{cases} 0 & \text{if } \frac{\sqrt{n}\bar{m}_{n,j}(\theta)}{\hat{\sigma}_{n,j}(\theta)} > -\kappa \\ -\infty & \text{otherwise} \end{cases} , \tag{30}$$

where κ is a non-negative constant. Using (27), we see that $\hat{s}_{n,j}^{\text{rms}}(\theta)$ still equals $-\infty$ with probability approaching one for any θ and P such that $\mu_j(\theta, P) < 0$, but no longer equals 0 with probability approaching one for any θ and P such that $\mu_j(\theta, P) = 0$. As a result, the test that simply rejects H_θ when $T_n(\theta)$ exceeds $\hat{J}_n^{-1}(1-\alpha, \hat{s}_n^{\text{rms}}(\theta), \theta)\}$ does not satisfy (6). Indeed, it is over-sized, as can be seen by considering the case where θ and P are such that $\mu(\theta, P) = 0$. For that reason, refined moment selection tests are tests of the form

$$\phi_n^{\text{rms}}(\theta) \equiv I\{T_n(\theta) > \hat{J}_n^{-1}(1-\alpha, \hat{s}_n^{\text{rms}}(\theta), \theta) + \hat{\eta}_n(\theta)\} ,$$

where $\hat{\eta}_n(\theta)$ is an additional size-correction factor.

In order to determine an appropriate size-correction factor, consider the test

$$\tilde{\phi}_n^{\text{rms}}(\theta) \equiv I\{T_n(\theta) > \hat{J}_n^{-1}(1-\alpha, \hat{s}_n^{\text{rms}}(\theta), \theta) + \eta\}$$

for an arbitrary non-negative constant η. Arguing as before in the case of generalized moment selection tests, it is possible to show that the limiting rejection probability of this test under appropriate sequences $\{(\theta_n, P_n) : \theta_n \in \Theta_0(P_n)\}_{n \geq 1}$ equals

$$P\{T(Z + s^*, \Omega^*) > J^{-1}(1-\alpha, s^{\text{rms},*}(Z + s^*), \Omega^*) + \eta\} ,$$

where, again, s^* is the limit of $\sqrt{n}\mu(\theta_n, P_n)$, Ω^* is the limit in probability of $\hat{\Omega}_n(\theta_n)$, $Z \sim N(0, \Omega^*)$,

$J(x, s, \Omega^*) = P\{T(Z + s, \Omega^*) \le x\}$, and $s^{\text{rms},*}(Z + s^*, \kappa)$ is such that

$$s_j^{\text{rms},*}(Z + s^*, \kappa) = \begin{cases} 0 & \text{if } Z_j + s_j^* > \kappa \\ -\infty & \text{otherwise} \end{cases} . \tag{31}$$

The appropriate size-correction factor is thus

$$\eta^*(\Omega^*, \kappa) \equiv \inf \left\{ \eta > 0 : \sup_{s^* \in \mathbf{R}^k : s^* \le 0} P\{T(Z + s^*, \Omega^*) > J^{-1}(1 - \alpha, s^{\text{rms},*}(Z + s^*, \kappa), \Omega^*) + \eta\} \le \alpha \right\} , \tag{32}$$

which may be consistently estimated as $\hat{\eta}_n(\theta) \equiv \eta^*(\hat{\Omega}_n(\theta), \kappa)$.

With $\hat{\eta}_n(\theta)$ computed in this way, it follows immediately that refined moment selection tests satisfy (6). In order to choose κ in a data-dependent fashion, Andrews and Barwick (2012) propose choosing it to maximize some notion of power. Given κ and a finite set of (local) alternatives of interest, A, it follows from the above that the limiting (equally-weighted) average power of the test is

$$\frac{1}{|A|} \sum_{a \in A} P\{T(Z + a, \Omega^*) > J^{-1}(1 - \alpha, s^{\text{rms},*}(Z + a, \kappa), \Omega^*) + \eta^*(\Omega^*, \kappa)\} , \tag{33}$$

where $s^{\text{rms},*}(Z + a, \kappa)$ is defined analogously to (31). As usual, the use of local alternatives is necessary to obtain non-degenerate limiting average power. Denote by $\kappa^*(\Omega^*)$ a (near) maximizer of (33). The data-dependent choice of κ is thus given by its consistent estimator, $\hat{\kappa}_n(\theta) = \kappa^*(\hat{\Omega}_n(\theta))$.

Remark 4.7. The determination of $\eta^*(\Omega^*, \kappa)$ is complicated by the fact that there is no explicit solution to the supremum in (32). Even for a single value of κ, computing $\eta^*(\Omega^*, \kappa)$ is computationally prohibitive for large values of k. Finding, $\kappa^*(\Omega^*)$, a (near) maximizer of (33) is therefore even more computationally demanding, as it involve computing this quantity for many values of κ. In order to alleviate the computational burden, Andrews and Barwick (2012) restrict attention to $k \le 10$ and approximate the supremum in (32) with the maximum over $s^* \in \{-\infty, 0\}^k$. The authors provide extensive numerical evidence, but no proof, in favor of this approximation. While they do not apply to the moment selection function in (30), McCloskey (2015) has recently justified this type of approximation for other moment selection functions. Andrews and Barwick (2012) also provide numerical evidence suggesting that $\kappa^*(\Omega^*)$ and $\eta^*(\Omega^*, \kappa^*(\Omega^*))$ are well approximated by functions that only depend on k and the smallest off-diagonal element of Ω^*. Based on this evidence, they tabulate suggested values of these quantities for $k \leq 10$ and $\alpha = 5\%$. ■

4.1.5 Two-step Tests

Motivated in part by the computational difficulties described in Remark 4.7, Romano et al. (2014) propose a two-step testing procedure that, like refined moment selection, does not rely on an asymptotic framework that can perfectly discriminate between binding and non-binding moments, but remains computationally feasible even for large values of k. Such values of k appear in many applications, including, for example, the applications in Ciliberto and Tamer (2010) and Bajari et al. (2006). Furthermore, as shown in a small simulation study in Romano et al. (2014), the procedure compares favorably in terms of power with the testing procedures described above.

In the first step of the procedure, a confidence region for $\sqrt{n}\mu(\theta, P)$ that is uniformly consistent in level (over $\theta \in \Theta_0(P)$ and $P \in \mathbf{P}$) is constructed, i.e., a random set $M_n(\theta, 1-\beta)$ such that

$$\liminf_{n\to\infty} \inf_{P\in\mathbf{P}} \inf_{\theta\in\Theta_0(P)} P\{\sqrt{n}\mu(\theta, P) \in M_n(\theta, 1-\beta)\} \geq 1-\beta \ . \tag{34}$$

While it is not possible to consistently estimate $\sqrt{n}\mu(\theta, P)$, it is possible to construct such a confidence region. In order to describe a specific construction that satisfies (34) and, as we will see below, also leads to attractive computational features, define

$$K_n(x, \theta, P) \equiv P\left\{\max_{1\leq j\leq k} \frac{\sqrt{n}(\mu_j(\theta, P) - \bar{m}_n(\theta))}{\hat{\sigma}_{n,j}(\theta)} \leq x\right\} \ .$$

Using this notation, an upper rectangular confidence region that satisfies (34) is given by

$$M_n(\theta, 1-\beta) \equiv \left\{\mu \in \mathbf{R}^k : \max_{1\leq j\leq k} \frac{\sqrt{n}(\mu_j - \bar{m}_n(\theta))}{\hat{\sigma}_{n,j}(\theta)} \leq \hat{K}_n^{-1}(1-\beta, \theta)\right\} \ , \tag{35}$$

where $\hat{K}_n(x, \theta)$ equals either $K_n(x, \theta, \hat{P}_n)$ or $K_n(x, \theta, \tilde{P}_n(\theta))$.

In the second step of this procedure, the confidence region is used to restrict the possible values of $\sqrt{n}\mu(\theta, P)$ when constructing the critical value with which to compare the test statistic $T_n(\theta)$.

This idea leads us to consider the critical value

$$\sup_{s \in M_n(\theta, 1-\beta) \cap \mathbf{R}^k_-} \hat{J}_n^{-1}(1 - \alpha + \beta, s, \theta) \ , \tag{36}$$

where $\mathbf{R}^k_- = \{x \in \mathbf{R}^k : x \leq 0\}$ and $\hat{J}_n(x, s, \theta)$ equals $J_n(x, s, \theta, \hat{P}_n)$ or $J_n(x, s, \theta, \tilde{P}_n(\theta))$. The addition of β to the quantile is necessary to account for the possibility that $\sqrt{n}\mu(\theta, P)$ may not lie in $M_n(\theta, 1 - \beta)$. It may be removed by allowing β to tend to zero with the sample size, but this would lead to a testing procedure more akin to generalized moment selection, rather than refined moment selection.

Remarkably, when $M_n(\theta, 1-\beta)$ is given by (35), the calculation of (36) is straightforward even for large values of k. Because T is increasing in each component of its first argument, $\hat{J}_n^{-1}(1-\alpha+\beta, s, \theta)$ is increasing in each component of s. Hence, (36) is simply

$$\hat{J}_n^{-1}(1 - \alpha + \beta, \hat{s}_n^{\text{ts}}(\theta), \theta) \ ,$$

where $\hat{s}_n^{\text{ts}}(\theta) = (\hat{s}_{n,1}^{\text{ts}}(\theta), \ldots, \hat{s}_{n,k}^{\text{ts}}(\theta))'$ is such that

$$\hat{s}_{n,j}^{\text{ts}}(\theta) = \min\{\sqrt{n}\bar{m}_n(\theta) + \hat{\sigma}_{n,j}(\theta)\hat{K}_n^{-1}(1 - \beta, \theta), 0\} \ .$$

With this simplification, two-step tests are tests of the form

$$\phi_n^{\text{ts}}(\theta) \equiv I\{T_n(\theta) > \hat{J}_n^{-1}(1 - \alpha + \beta, \hat{s}_n^{\text{ts}}, \theta)\} \ .$$

Romano et al. (2014) establish that these tests satisfy (6). The result relies on the following inequality:

$$P\{T_n(\theta) > \hat{J}_n^{-1}(1 - \alpha + \beta, \hat{s}_n^{\text{ts}}(\theta), \theta)\} \leq \\ P\{T_n(\theta) > \hat{J}_n^{-1}(1 - \alpha + \beta, \sqrt{n}\mu(\theta, P), \theta)\} + P\{\sqrt{n}\mu(\theta, P) \notin M_n(\theta, 1 - \beta)\} \ .$$

It is straightforward to show that

$$\limsup_{n \to \infty} \sup_{P \in \mathbf{P}} \sup_{\theta \in \Theta_0(P)} P\{T_n(\theta) > \hat{J}_n^{-1}(1 - \alpha + \beta, \sqrt{n}\mu(\theta, P), \theta)\} \leq \alpha - \beta \ .$$

From (34), we have further that

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}} \sup_{\theta\in\Theta_0(P)} P\{\sqrt{n}\mu(\theta,P) \notin M_n(\theta,1-\beta)\} \leq \beta\ .$$

The validity of the two-step test thus follows. It is in fact possible to show further that

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}} \sup_{\theta\in\Theta_0(P)} P\{T_n(\theta) > \hat{J}_n^{-1}(1-\alpha+\beta, \hat{s}_n^{\text{ts}}(\theta),\theta)\} \geq \alpha-\beta\ .$$

Similar methods have been proposed for some parametric testing problems in the statistics literature (Berger and Boos, 1994; Silvapulle, 1996). It has also appeared earlier in the econometrics literature on instrumental variables (Staiger and Stock, 1997) and nonlinear panel data models (Chernozhukov et al., 2013b). Finally, the idea has recently been introduced in a general context by McCloskey (2012).

Remark 4.8. Note that β may be also chosen in a data-dependent way analogous to the way in which κ is chosen in the preceding section. Romano et al. (2014) find, however, that the simple rule of choosing $\beta = \alpha/10$ works well in their simulations. ■

Remark 4.9. In Romano et al. (2014), the test is restricted to only reject H_θ when $M_n(\theta,1-\beta)$ is not contained in $\mathbf{R}^k_-$. If the test statistic is such that $T_n(\theta)=0$ whenever $\bar{m}_n(\theta)\leq 0$, as in the test statistics considered in this paper, then this restriction is redundant. ■

Remark 4.10. Note that in all of the methods discussed so far, moment equalities can be accommodated simply by including both the moment and its negative in the definition of m. In the case of two-step tests discussed in this section, additional improvements are possible when moment equalities are present by exploiting their structure when constructing $M_n(1-\beta)$ in (35). To explain further, write $m=(m^{(\text{eq})}, m^{(\text{ineq})})$, where $m^{(\text{eq})}$ are the pairs of moment inequalities corresponding to moment equalities and $m^{(\text{ineq})}$ are the other moment inequalities. Define $M_n^{(\text{ineq})}(1-\beta)$ according to the right-hand side of (35) with $m^{(\text{ineq})}$ in place of m and let $M_n^{(\text{eq})}(1-\beta)=\{0\}^{\dim(m^{(\text{eq})})}$. Finally, replace $M_n(1-\beta)$ with $M_n^{(\text{eq})}(1-\beta)\times M_n^{(\text{ineq})}(1-\beta)$. The remainder of the procedure may now be followed without modification. ■

5 Subvector Inference for Moment Inequalities

In this section, we review several approaches that have been proposed in the literature for inference in models in which the identified set is given by

$$\Lambda_0(P) = \lambda(\Theta_0(P)) = \{\lambda(\theta) : \theta \in \Theta_0(P)\} ,$$

where $\Theta_0(P)$ is defined as in (11) and $\lambda : \Theta \to \Lambda$. In the most common example, $\Theta \subseteq \mathbf{R}^{d_\theta}$, $\Lambda \subseteq \mathbf{R}^{d_\lambda}$, and λ is a function that selects a subvector of $\theta \in \Theta$, such as a single component. For ease of exposition, we focus on this special case in the remainder of our discussion. As in the preceding section, our aim is to construct confidence regions for points in the identified set that are uniformly consistent in level. Such confidence regions may again be constructed by exploiting the duality between confidence regions and inverting tests of each of the individual null hypotheses

$$H_\lambda : \lambda \in \Lambda_0(P)$$

versus the unrestricted alternative hypothesis that control appropriately the usual probability of a Type I error at level α. In the setting considered here, these null hypotheses may be written as

$$H_\lambda : \exists\ \theta \in \Theta(\lambda) \text{ with } E_P[m(W_i, \theta)] \leq 0 , \tag{37}$$

where

$$\Theta(\lambda) = \{\theta \in \Theta : \lambda(\theta) = \lambda\} .$$

Given tests $\phi_n(\lambda)$ of H_λ for each λ satisfying

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}} \sup_{\lambda\in\Lambda_0(P)} E_P[\phi_n(\lambda)] \leq \alpha , \tag{38}$$

C_n^λ equal to the set of $\lambda \in \Lambda$ for which H_λ is accepted satisfies

$$\liminf_{n\to\infty} \inf_{P\in\mathbf{P}} \inf_{\lambda\in\Lambda_0(P)} P\{\lambda \in C_n^\lambda\} \geq 1 - \alpha . \tag{39}$$

As before, a variety of different tests of (37) have been proposed in the literature. In order to describe these tests succinctly, we will make use of much of the notation introduced in the preceding section. It is additionally useful to define a "profiled" test statistic

$$T_n^{\text{prof}}(\lambda) = \inf_{\theta \in \Theta(\lambda)} T_n(\theta) \ ,$$

where $T_n(\theta)$ is as in (13). Most of the tests we discuss below reject H_λ for large values of $T_n^{\text{prof}}(\lambda)$. It is also useful to introduce

$$J_{n,\lambda}(x, s(\cdot), \lambda, P) = P\left\{ \inf_{\theta \in \Theta(\lambda)} T(\hat{D}_n^{-1}(\theta)\sqrt{n}(\bar{m}_n(\theta) - \mu(\theta, P)) + \hat{D}_n^{-1}(\theta)s(\theta), \hat{\Omega}_n(\theta)) \leq x \right\} \ . \quad (40)$$

In terms of (40), the distribution of $T_n^{\text{proj}}(\lambda)$ itself is simply

$$J_{n,\lambda}(x, \sqrt{n}\mu(\cdot, P), \lambda, P) = P\left\{ \inf_{\theta \in \Theta(\lambda)} T(\hat{D}_n^{-1}(\theta)\sqrt{n}\bar{m}_n(\theta), \hat{\Omega}_n(\theta)) \leq x \right\} \ . \quad (41)$$

When λ is the identity function, (40) and (41) reduce to (17) and (18), respectively. For the same reasons discussed in Section 4, it is difficult to derive useful estimators of (41), but it is straightforward to derive useful estimators of (40) for a fixed value of $s(\cdot)$. An important distinction in the present setting, however, is that $s(\cdot)$ is a function over the set $\Theta(\lambda)$.

Remark 5.1. Note that by choosing the function λ to be a constant function that always equals the same value, the above framework includes as a special case testing the null hypothesis that there exists $\theta \in \Theta$ such that $E_P[m(W_i, \theta)] \leq 0$. A test of such a null hypothesis is typically referred to as a specification test. See Romano and Shaikh (2008), Andrews and Soares (2010), and Bugni et al. (2015) for examples of such tests, which may all be viewed as special cases of the methods we describe below. ■

5.1 Four Methods for Subvector Inference for Moment Inequalities

In this section, we describe three different tests for (37). As before, in order to satisfy (38), each of these tests require the uniform integrability condition (19) to hold, but, in some cases, require further conditions as well. While we refrain from describing these additional conditions in detail, we highlight when they become necessary as we go along.

5.1.1 Projection

Projection tests are tests of the form

$$\phi_n^{\text{proj}}(\lambda) = \inf_{\theta \in \Theta(\lambda)} \phi_n(\theta) \ , \tag{42}$$

where $\phi_n(\theta)$ may be any test satisfying (6), such as those described in Section 4. Provided that $\phi_n(\theta)$ satisfies (6), the projection test $\phi_n^{\text{proj}}(\lambda)$ defined in (42) satisfies (38). To see this, first note that

$$E_P[\phi_n^{\text{proj}}(\lambda)] \leq E_P[\phi_n(\theta)] \tag{43}$$

for any $\theta \in \Theta(\lambda)$. Furthermore, for any $P \in \mathbf{P}$ and $\lambda \in \Lambda_0(P)$, we have that there exists $\theta \in \Theta(\lambda)$ with $E_P[m(W_i, \theta)] \leq 0$. In particular, for any such P and λ, there exists $\theta \in \Theta_0(P)$ such that (43) holds. Hence,

$$\sup_{P \in \mathbf{P}} \sup_{\lambda \in \Lambda_0(P)} E_P[\phi_n^{\text{proj}}(\lambda)] \leq \sup_{P \in \mathbf{P}} \sup_{\theta \in \Theta_0(P)} E_P[\phi_n(\theta)] \ ,$$

from which the desired conclusion follows.

Such tests are discussed in Andrews and Guggenberger (2009) and Andrews and Soares (2010). For examples of their use in empirical research, see Ciliberto and Tamer (2010) and Grieco (2014). As noted by Romano and Shaikh (2008), while they are valid whenever $\phi_n(\theta)$ satisfies (6), a drawback of such tests is that they may be quite conservative in the sense that the left-hand side of (39) is much greater than $1 - \alpha$. In particular, this may be the case even if the left-hand side of (3) equals $1 - \alpha$. Furthermore, the minimization in (42) may be unnecessarily computationally burdensome.

Remark 5.2. The confidence region C_n^λ that results from inverting tests $\phi_n^{\text{proj}}(\lambda)$ in (42) can be described succinctly in terms of the confidence region that results from inverting the corresponding tests $\phi_n(\theta)$ that appear on the right-hand side of (42). In particular,

$$C_n^\lambda = \lambda(C_n) = \{\lambda(\theta) : \theta \in C_n\} \ . \tag{44}$$

From this characterization, the description of these tests as "projection" tests is clear. ■

Remark 5.3. In recent work, Kaido et al. (2016) propose a confidence set C_n for $\theta \in \Theta_0(P)$ with the property that C_n^λ defined in (44) is not conservative in the sense that the left-hand side of (39) equals $1-\alpha$. The construction of such a C_n is obtained by comparing $T_n^{\max}(\theta)$ defined in (15) with an appropriate critical value $\hat{c}_n(\theta)$. The construction of $\hat{c}_n(\theta)$ involves several novel ideas to increase computational tractability. As in generalized moment selection, it also requires the choice of κ_n satisfying (26) as well as the additional choice of $\rho > 0$, which does not depend on n. For further details, we refer the reader to Kaido et al. (2016). ■

5.1.2 Least Favorable Tests

In this section and the following two sections, we consider tests that reject H_λ for large values of $T_n^{\text{prof}}(\lambda)$. As in Section 4.1.1, least favorable tests can again be derived using the observation that T is increasing in each component of its first argument. On the other hand, it is more difficult in the present setting to derive the least favorable value of the nuisance parameter, $\sqrt{n}\mu(\cdot, P)$ as a function over $\Theta(\lambda)$. In particular, 0_k is no longer the least favorable value for the nuisance parameter. In fact, there is no longer a single, least favorable value, but rather a family of least favorable values for the nuisance parameter indexed by $\theta \in \Theta(\lambda)$. To see this, note that for λ and P such that $\lambda \in \Lambda_0(P)$, the only restriction on $\sqrt{n}\mu(\cdot, P)$ as a function over $\Theta(\lambda)$ is that $\sqrt{n}\mu(\theta, P) \leq 0_k$ for some $\theta \in \Theta(\lambda)$. Hence, for each $\theta \in \Theta(\lambda)$, $s_\theta^{\text{lf}}(\cdot)$ defined by

$$s_\theta^{\text{lf}}(\tilde{\theta}) = \begin{cases} 0_k & \text{if } \theta = \tilde{\theta} \\ +\infty_k & \text{otherwise} \end{cases} ,$$

where $+\infty_k$ is a k-dimensional vector whose components all equal $+\infty$, is a least favorable value for the nuisance parameter. Note that there is no "largest" value among these possible least favorable values. As a result, for λ and P such that $\lambda \in \Lambda_0(P)$, we have that

$$\begin{aligned} J_{n,\lambda}^{-1}(1-\alpha, \sqrt{n}\mu(\cdot, P), \lambda, P) &\leq \sup_{\theta \in \Theta(\lambda)} J_{n,\lambda}^{-1}(1-\alpha, s_\theta^{\text{lf}}(\cdot), \lambda, P) \\ &= \sup_{\theta \in \Theta(\lambda)} J_n^{-1}(1-\alpha, 0_k, \theta, P) , \end{aligned}$$

where the equality can be straightforwardly verified by inspection. It follows that least favorable tests of the form

$$\phi_n^{\text{lf}}(\lambda) = I\left\{ T_n^{\text{prof}}(\lambda) > \sup_{\theta \in \Theta(\lambda)} \hat{J}_n^{-1}(1-\alpha, 0_k, \theta) \right\} ,$$

where $\hat{J}_n(x, 0_k, \theta)$ equals either $J_n(x, 0_k, \theta, \hat{P}_n)$ or $J_n(x, 0_k, \theta, \tilde{P}_n(\theta))$, satisfy (38). To the best of our knowledge, such tests have not been explicitly discussed in the literature for this testing problem, though the idea is essentially contained in a number of earlier papers. See, for example, the discussion in Perlman et al. (1999) as well as the more recent applications in Romano and Wolf (2013) and Machado et al. (2013). As with the least favorable tests discussed in Section 4.1.1, such tests might be considered "conservative," but, as before, the tests corresponding to these tests in a Gaussian setting are in fact admissible and enjoy certain types of optimality among a restricted class of tests. Nevertheless, for the same reasons discussed at the end of Section 4.1.1, it may not be desirable to construct confidence regions by inverting such tests. Instead, it is sensible to try to incorporate information in the data about the nuisance parameter when constructing critical values.

5.1.3 Subsampling

Subsampling-based tests use the subsampling estimate of the distribution of interest (41). Using the additional notation introduced in Section 4.1.2, the subsampling estimate of the distribution of $T_n^{\text{prof}}(\lambda)$ is given by

$$L_{n,\text{prof}}(x, \lambda) = \frac{1}{N_n} \sum_{1 \leq \ell \leq N_n} I\left\{ \inf_{\theta \in \Theta(\lambda)} T_{b,\ell}(\theta) \leq x \right\} ,$$

and the corresponding test is given by

$$\phi_n^{\text{sub}}(\lambda) \equiv I\{T_n(\lambda) > L_{n,\text{prof}}^{-1}(1 - \alpha, \lambda)\} . \tag{45}$$

Such tests were first proposed in Romano and Shaikh (2008), who provide high-level conditions under which $\phi_n^{\text{sub}}(\lambda)$ satisfies (38). They additionally verify those conditions in some specific examples, but it is possible to show that $\phi_n^{\text{sub}}(\lambda)$ may fail to satisfy (38) if only (19) is imposed. The failure is related to the fact described in the previous section that 0_k is no longer an upper bound on the nuisance parameter for λ and P such that $\lambda \in \Lambda_0(P)$.

Remark 5.4. As described earlier, implementing subsampling of course requires a choice of subsample size $b = b_n$. The discussion in Remark 4.4 again applies here. ■

5.1.4 Minimum Resampling

As in the preceding two sections, minimum resampling tests, proposed by Bugni et al. (2014), also reject H_λ for large values of $T_n^{\text{prof}}(\lambda)$. In order to describe their construction of a critical value with which to compare $T_n^{\text{prof}}(\lambda)$, it is useful to first note that a naïve application of generalized moment selection does not in general lead to tests satisfying (38) in the present setting. To see this, recall from the discussion in Section 5.1.2 that 0_k is not a least favorable value of the nuisance parameter, $\sqrt{n}\mu(\cdot, P)$ over $\Theta(\lambda)$. In particular, it may be the case that for P and λ such that $\lambda \in \Lambda_0(P)$ that

$$\sqrt{n}\mu_j(\theta, P) > 0 \text{ for some } 1 \leq j \leq k \text{ and } \theta \in \Theta(\lambda) \ . \tag{46}$$

Since $\hat{s}_n^{\text{gms}}(\theta) \leq 0_k$, it follows that (46) may hold with $\hat{s}_{n,j}^{\text{gms}}(\theta)$ in place of 0. As a result, it is not reasonable to expect that using an estimate of

$$J_{n,\lambda}^{-1}(1-\alpha, \hat{s}_n^{\text{gms}}(\cdot), \lambda, P)$$

would lead to a test satisfying (38). In fact, Bugni et al. (2014) provide an explicit counterexample.

Motivated by the failure of a naïve implementation of generalized moment selection in this setting, Bugni et al. (2014) consider two different remedies. Each remedy involves replacing $\hat{s}_n^{\text{gms}}(\cdot)$ with an alternative function that does provide an asymptotic upper bound on the nuisance parameter, $\sqrt{n}\mu(\cdot, P)$ as a function over $\Theta(\lambda)$, in the sense described at the end of Section 4.1.3. In order to describe the first such function, note that $\hat{s}_n^{\text{gms}}(\theta)$ does provide an asymptotic upper bound on the nuisance parameter over the subset of $\Theta(\lambda)$ contained in $\Theta_0(P)$. This observation leads Bugni et al. (2014) to consider the function

$$\hat{s}_n^{(1),\text{bcs}}(\theta) = (\hat{s}_{n,1}^{(1),\text{bcs}}(\theta), \ldots, \hat{s}_{n,k}^{(1),\text{bcs}}(\theta))'$$

where

$$\hat{s}_{n,j}^{(1),\text{bcs}}(\theta) = \begin{cases} \hat{s}_{n,j}^{\text{gms}}(\theta) & \text{if } \theta \in \hat{\Theta}_n \\ +\infty & \text{otherwise} \end{cases} \tag{47}$$

and $\hat{\Theta}_n$ is a set satisfying

$$\inf_{P \in \mathbf{P}} P\{\hat{\Theta}_n \subseteq \Theta(\lambda) \cap \Theta_0(P)^{\log \kappa_n}\} \to 1 \ , \tag{48}$$

where $\Theta_0(P)^{\log \kappa_n}$ is a $\log \kappa_n$ "neighborhood" of $\Theta_0(P)$ and κ_n satisfies (26). Bugni et al. (2014) specifically propose using

$$\hat{\Theta}_n = \{\theta \in \Theta(\lambda) : T_n(\theta) \leq T_n^{\text{prof}}(\lambda)\} \ , \tag{49}$$

i.e., the set of minimizers of $T_n(\theta)$. The second function is simply given by

$$\hat{s}_n^{(2),\text{bcs}}(\theta) = (\hat{s}_{n,1}^{(2),\text{bcs}}(\theta), \ldots, \hat{s}_{n,k}^{(2),\text{bcs}}(\theta))'$$

where

$$\hat{s}_{n,j}^{(2),\text{bcs}}(\theta) = \frac{\kappa_n^{-1}\sqrt{n}\bar{m}_{n,j}(\theta)}{\hat{\sigma}_{n,j}(\theta)}$$

and κ_n satisfies (26). Note that, unlike $\hat{s}_{n,j}^{\text{gms}}(\theta)$, which takes values in $[-\infty, 0]$, both $\hat{s}_{n,j}^{(1),\text{bcs}}(\theta)$ and $\hat{s}_{n,j}^{(2),\text{bcs}}(\theta)$ take values in $[-\infty, +\infty]$.

Using arguments similar to those given in Section 4.1.3, it is possible to provide conditions under which using either $J_{n,\lambda}^{-1}(1-\alpha, \hat{s}_n^{(1),\text{bcs}}(\cdot), \lambda, \hat{P}_n)$ or $J_{n,\lambda}^{-1}(1-\alpha, \hat{s}_n^{(2),\text{bcs}}(\cdot), \lambda, \hat{P}_n)$ leads to tests satisfying (38), but, by combining these two approaches, it is possible to construct an even smaller critical value that also leads to valid tests. In order to describe this improved construction, define

$$J_{n,\text{bcs}}(x, s^{(1)}(\cdot), s^{(2)}(\cdot), \lambda, P) = P\{\min\{R_n(s^{(1)}(\cdot), \lambda, P), R_n(s^{(2)}(\cdot), \lambda, P)\} \leq x\} \ ,$$

where

$$R_n(s(\cdot), \lambda, P) = \inf_{\theta \in \Theta(\lambda)} T(\hat{D}_n^{-1}(\theta)\sqrt{n}(\bar{m}_n(\theta) - \mu(\theta, P)) + \hat{D}_n^{-1}(\theta)s(\theta), \hat{\Omega}_n(\theta)) \ .$$

Using this notation, the "minimum resampling" test Bugni et al. (2014) propose is given by

$$\phi_n^{\text{bcs}}(\lambda) = I\{T_n^{\text{prof}}(\lambda) > J_{n,\text{bcs}}^{-1}(1-\alpha, \hat{s}_n^{(1),\text{bcs}}(\cdot), \hat{s}_n^{(2),\text{bcs}}(\cdot), \lambda, \hat{P}_n)\} \ . \tag{50}$$

Bugni et al. (2014) provide assumptions under which $\phi_n^{\text{bcs}}(\lambda)$ defined in this way satisfies (38). Importantly, in addition to (19), these assumptions include additional conditions under which $\hat{\Theta}_n$ defined in (49) satisfies (48).

Remark 5.5. Bugni et al. (2014) show that the test of H_λ that rejects whenever $T_n^{\text{prof}}(\lambda)$ exceeds $J_{n,\lambda}^{-1}(1-\alpha, \hat{s}_n^{(1),\text{bcs}}(\cdot), \lambda, \hat{P}_n)$ is more powerful in finite samples than the corresponding projection test implemented using the same selection function that appears on the right-hand side of (47). Bugni et al. (2014) also compare the limiting power of the test of H_λ that rejects whenever $T_n^{\text{prof}}(\lambda)$

exceeds $J_{n,\lambda}^{-1}(1-\alpha, \hat{s}_n^{(2),\text{bcs}}(\cdot), \lambda, \hat{P}_n)$ and the subsampling test of H_λ defined in (45) against certain local alternatives. As in Remark 4.6, the comparison depends on the relationship between κ_n and b_n. In particular, if

$$\limsup_{n\to\infty} \frac{\kappa_n \sqrt{b_n}}{\sqrt{n}} \leq 1 \ , \tag{51}$$

then the limiting power of the the test of H_λ that rejects whenever $T_n^{\text{prof}}(\lambda)$ exceeds $J_{n,\lambda}^{-1}(1-\alpha, \hat{s}_n^{(2),\text{bcs}}(\cdot), \lambda, \hat{P}_n)$ exceeds the limiting power of subsampling tests against these local alternatives, but this ordering may be reversed when (51) does not hold. On the other hand, as in Remark 4.6, for b_n that is optimal in terms of the order of the error in rejection probability under the null hypothesis and typical choices of κ_n, (51) holds. Importantly, the test $\phi_n^{\text{bcs}}(\lambda)$ defined in (50), whose critical value is smaller than both $J_{n,\lambda}^{-1}(1-\alpha, \hat{s}_n^{(1),\text{bcs}}(\cdot), \lambda, \hat{P}_n)$ and $J_{n,\lambda}^{-1}(1-\alpha, \hat{s}_n^{(2),\text{bcs}}(\cdot), \lambda, \hat{P}_n)$, inherits both of these power comparisons. ■

6 Important Omissions

Regrettably, our review of inference in partially identified models has necessarily been selective. In this section, we rapidly review some of the most important omissions in our discussion of the literature on inference in partially identified models.

6.1 Inference for "Many" Moment Inequalities

As emphasized above, in many applications k may be large, which motivates considering asymptotic frameworks in which $k = k_n$ tends to infinity with the sample size n. Such an approach, which requires asymptotic approximations for normalized sums with increasing dimensions, was recently developed by Chernozhukov et al. (2013a). More concretely, Chernozhukov et al. (2013a) consider inference in models defined by (12) where the number of moment inequalities k_n is allowed to be of the same order as $\exp(n^\delta)$ for some $\delta > 0$. The tests proposed by the authors could be one-step or two-step and take the form of those described in Section 4. One-step tests involve a "max"-type test statistic

$$\tilde{T}_n^{\max}(\theta) = \max_{1\leq j\leq k} \frac{\sqrt{n}\bar{m}_{n,j}(\theta)}{\hat{\sigma}_{n,j}(\theta)} \ ,$$

and the following critical value

$$\hat{c}^{cck}_{n,k}(1-\alpha,\theta) = \frac{\Phi^{-1}(1-\alpha/k)}{\sqrt{1-\Phi^{-1}(1-\alpha/k)^2/n}} \ .$$

Here, $\Phi(\cdot)$ denotes the distribution function of the standard normal distribution. This critical value arises by using Bonferonni's inequality and a moderate deviation inequality for self-normalized sums. In the same spirit as the tests described in Sections 4.1.2 – 4.1.5, two-step tests improve on one-step tests by incorporating information about $\sqrt{n}\mu(\theta,P)$ when constructing the critical value using a preliminary "selection" step. In the first step, the number of binding moments is estimated to be

$$\hat{k}_n = \sum_{j=1}^{k} \hat{s}^{\text{cck}}_{n,j}(\theta) \ ,$$

where

$$\hat{s}^{\text{cck}}_{n,j}(\theta) = I\left\{\frac{\sqrt{n}\bar{m}_{n,j}(\theta)}{\hat{\sigma}_{n,j}(\theta)} > -2\hat{c}^{cck}_{n,k}(1-\beta,\theta)\right\}$$

and $0<\beta<\frac{\alpha}{3}$; in the second step, $\tilde{T}^{\max}_n(\theta)$ is compared with $\hat{c}^{cck}_{n,\hat{k}_n}(1-\alpha,\theta)$. The formal justification of this test, together with bootstrap counterparts, are in Chernozhukov et al. (2013a).

In a related paper, Menzel (2014) also considers inference for moment inequalities in a context where k grows with the sample size n. However, in contrast to Chernozhukov et al. (2013a), the asymptotic framework is one where k_n is assumed to be smaller than n; specifically, $k_n = O(n^{2/7})$.

6.2 Inference for Conditional Moment Inequalities

In many applications the identified set for θ is determined by conditional moment inequalities, in which case

$$\Theta_0(P) = \{\theta\in\Theta : E_P[m(W_i,\theta)|Z_i] \le 0 \ P\text{-}a.s.\} \ , \tag{52}$$

where, as before, m is a function taking values in $\mathbf{R}^k$ and the inequality is interpreted component-wise. Two seminal papers on inference in conditional moment inequality models are Andrews and Shi (2013) and Chernozhukov et al. (2013d). Below we briefly summarize these two papers.

The approach put forward by Andrews and Shi (2013) consists in transforming the conditional moment inequalities in (52) into an infinite number of unconditional moment inequalities. This can be done by choosing a set of weighting functions $\mathcal{G}$ with the property that

$$\Theta_{0,\mathcal{G}}(P) = \{\theta \in \Theta : E_P[m(W_i,\theta)g(Z_i)] \leq 0 \text{ for all } g \in \mathcal{G}\} \tag{53}$$

is equal to $\Theta_0(P)$ in (52). One of the simplest examples of such a set of functions is the set $\mathcal{G}_{\text{box}}$ of indicators of a rich set of hyperrectangles. In order to define this set of functions more formally, let

$$\mathcal{C}_{\text{box}} \equiv \left\{ \times_{l=1}^{d_z} (z_l - r_l, z_l + r_l] : z \in \mathbf{R}^{d_z}, r \in (0,\bar{r})^{d_z} \right\} ,$$

where $\bar{r} > 0$, i.e., $\mathcal{C}_{\text{box}}$ is the set of hyperrectanges in $\mathbf{R}^{d_z}$ whose sides all have length no greater than $2\bar{r}$. Using this notation,

$$\mathcal{G}_{\text{box}} \equiv \{I\{z \in C\} : C \in \mathcal{C}_{\text{box}}\} . \tag{54}$$

Other examples of collections $\mathcal{G}$ that have the property $\Theta_{0,\mathcal{G}}(P) = \Theta_0(P)$ are discussed in Andrews and Shi (2013).

Similarly to Section 4, confidence regions satisfying (3) may be constructed by inverting tests of the null hypotheses

$$H_\theta : E_P[m(W_i,\theta)g(Z_i)] \leq 0 \text{ for all } g \in \mathcal{G} . \tag{55}$$

The test proposed by Andrews and Shi (2013) represents an extension of generalized moment selection to models with infinitely many unconditional moment inequalities. Specifically, the test rejects H_θ in (55) for large values of

$$T_n(\theta) \equiv \int T_n(\theta,g)dQ(g) ,$$

where $T_n(\theta,g)$ is test statistic for testing $E_P[m(W_i,\theta)g(Z_i)] \leq 0$ analogous to those introduced in 4 and Q is a probability measure on $\mathcal{G}$. The construction of the critical value is conceptually similar to the generalized moment selection approach discussed in Section 4.1.3. For details on the implementation of this test and a proof that it satisfies (6), see Andrews and Shi (2013).

An alternative approach for inference in models where the set $\Theta_0(P)$ is defined by conditional moment inequalities, as in (52), is the "intersection bounds" method proposed by Chernozhukov et al. (2013d). In order to describe the test formally, let

$$\mathcal{V} \equiv \{(z,j) : z \in \mathcal{Z}, 1 \leq j \leq k\} ,$$

where $\mathcal{Z} \subseteq \mathbf{R}^{d_z}$ denotes the compact support of Z. For each $v \in \mathcal{V}$, let

$$\tilde{\mu}(\theta, P, v) = E_P[m_j(W_i, \theta)|Z_i = z]$$

and denote by $\bar{m}_n(\theta, v)$ a suitable estimator of $\tilde{\mu}(\theta, P, v)$. This estimator could be parametric or nonparametric, including series or kernel-type estimators. The choice of an estimator determines an appropriate standard error for $\bar{m}_n(\theta, v)$, which we denote here by $\hat{\sigma}_n(\theta, v)$.

Using this notation, the null hypotheses can be written as

$$H_\theta : \sup_{v \in \mathcal{V}} \tilde{\mu}(\theta, P, v) \leq 0 \ . \tag{56}$$

The test proposed by Chernozhukov et al. (2013d) rejects H_θ when

$$T_n(\theta) = \sup_{v \in \mathcal{V}} \{\bar{m}_n(\theta, v) - \hat{c}_n(\alpha, \theta, v)\hat{\sigma}_n(\theta, v)\} \tag{57}$$

is strictly positive. The construction of the critical value $\hat{c}_n(\alpha, \theta, v)$ is delicate and a clear exposition requires introducing significantly more notation, so we refer the reader to Chernozhukov et al. (2013d). Notably, the intersection bounds approach can be easily implemented in Stata for certain problems using the package developed by Chernozhukov et al. (2013c).

Other papers studying the problem of inference in conditional moment inequality models include Kim (2008); Ponomareva (2010); Lee et al. (2013); Chetverikov (2013); Armstrong (2014b,a, 2015), and Armstrong and Chan (2014). Importantly, the results in Armstrong (2014b) show that the tests in Armstrong and Chan (2014) and Chetverikov (2013) are rate optimal among available procedures in certain smoothness classes. These results also show that the test in Chernozhukov et al. (2013d) is rate optimal under appropriate bandwidth choices for the kernel estimator $\bar{m}_n(\theta, v)$.

6.3 Inference for Partially Identified Models using Random Set Theory

Random set theory provides a mathematical framework to study random objects whose realizations are sets. In the context of partially identified models, this theory has proven to be useful for identification and inference in situations where the object of interest is the identified set $\Theta_0(P)$ in (2), as opposed to points in such set. Even though random set theory is a well-developed area of mathematics that dates back at least to the 1950s, its first application to the problem of inference in partially identified models appeared in Beresteanu and Molinari (2008).

Beresteanu and Molinari (2008) exploit the theory of random sets to carry out inference on the identified set $\Theta_0(P)$. The method applies to situations where $\Theta_0(P)$ is a compact and convex set that is the Aumann expectation of a set-valued random variable, a generalization of the expectation of a random variable to random sets. For a given compact and convex set Ψ, the main inferential problem considered in the paper is testing

$$H_0 : \Theta_0(P) = \Psi$$

versus the unrestricted alternative. The test rejects for large values of the normalized Hausdorff distance between Ψ and a sample analog of $\Theta_0(P)$, and the authors develop bootstrap procedures for constructing critical values. Importantly, inverting such tests leads to confidence regions for the identified set rather than confidence regions for points in the identified set.

More generally, random set theory is especially useful in problems where the identified set $\Theta_0(P)$ is a convex set. This usefulness stems from the fact that in such cases $\Theta_0(P)$ can be unambiguously characterized by its support function, which is defined as

$$\mathrm{SF}_q(\Theta_0(P)) \equiv \sup_{\theta \in \Theta_0(P)} q'\theta \ , \tag{58}$$

for any $q \in \mathbf{R}^{d_\theta}$ such that $||q|| = 1$. Indeed, Kaido (2012) extends the domain of applicability of the approach in Beresteanu and Molinari (2008) by establishing a duality between level set estimators of $\Theta_0(P)$ based on convex criterion functions and the support function of such estimators. This provides a justification for the use of Hausdorff-based statistics beyond those cases where $\Theta_0(P)$ is the Aumann expectation of a random closed set. In subsequent work, Kaido and Santos (2014) examine the efficient estimation of $\Theta_0(P)$ in (11) when $E_P[m(W_i, \theta)]$ are convex functions of θ.

They derive conditions under which the support function admits $\sqrt{n}$-consistent regular estimators and provide a characterization of the semiparametric efficiency bound.

Random set theory has proved to be particularly useful in providing tractable characterizations of (sharp) identified sets in partially identified models. See, for example, Beresteanu et al. (2011), Bontemps et al. (2012), Chesher and Rosen (2014) and Chesher et al. (2013). Incidentally, these tools are also connected to the literature of optimal transportation theory and its application to partially identified models, see Galichon and Henry (2011, 2013). Finally, for a recent comprehensive summary of applications of random set theory to econometrics see Molchanov and Molinari (2014).

6.4 Bayesian Inference for Partially Identified Models

In well-behaved, identified models it is often the case that frequentist confidence sets and Bayesian credible sets about a given parameter of interest coincide, at least asymptotically. Such equivalence breaks down in the context of partially identified models, as shown by Moon and Schorfheide (2012). In fact, prior information about a partially identified parameter θ "influences" the usual posterior inference statements concerning θ even asymptotically. Credible sets for partially identified parameters thus tend to be asymptotically smaller than confidence sets satisfying (3). See Moon and Schorfheide (2012, Corollary 1) for details. It follows that from the Bayesian perspective, frequentist confidence sets are too wide, while from the frequentist perspective, Bayesian credible sets are too narrow.

The lack of asymptotic harmony between the Bayesian and frequentist inference is less severe when the object of interest is the identified set $\Theta_0(P)$ rather than $\theta \in \Theta_0(P)$. In the context of a partially identified model with a well-defined likelihood function, Kitagawa (2012) introduces a robust Bayesian approach with the goal of reconciling Bayesian and frequentist statements. Kitagawa (2012) notes that when parameters are not identified, the prior distribution of the model parameters can be decomposed into a component that is updated by data (revisable prior) and a component that cannot be updated by the data (unrevisable prior). He therefore considers a prior class in such a way that it shares a single prior distribution for the revisable prior, but allows for arbitrary prior distributions for the unrevisable prior. This leads to "bounds" on the posterior distribution for a partially identified parameter due to the presence of a class of priors rather than a single one. Kitagawa (2012, Theorem 5.1) goes further and shows that a credible set from the robust Bayes perspective is also a valid frequentist confidence set for $\Theta_0(P)$, provided the frequentist confidence set is only required to satisfy (8) rather than (4).

More recently, Kline and Tamer (2015) consider a class of partially identified models with the property that the identified set is a known function of identified parameters, as for instance, in Example 3.1. By focusing on this class of models and inference statements about $\Theta_0(P)$, Kline and Tamer (2015) establish a method for Bayesian inference that results in posterior inference statements that do not depend on the prior information asymptotically. As a result, under certain conditions, the $(1-\alpha)$-level credible set for $\Theta_0(P)$ is also a $(1-\alpha)$-level frequentist confidence set for $\Theta_0(P)$ satisfying (8) (see Kline and Tamer, 2015, Theorem 5). Therefore, the results in Kitagawa (2012) and Kline and Tamer (2015) show that inference about the identified set may exhibit asymptotic equivalence between Bayesian and frequentist approaches to inference in partially identified models, as long as the frequentist confidence set is required to satisfy (8) as opposed to (4). As explained in Section 3, such confidence regions share the undesirable feature that the probability of covering $\Theta_0(P)$ may be well below $1-\alpha$ under certain distributions $P \in \mathbf{P}$, even for arbitrarily large sample sizes.

7 Conclusion

As our preceding discussion has hopefully made clear, the literature on inference in partially identified models has advanced significantly in the past decade. Nevertheless, important questions, both theoretical and practical in nature, remain unresolved. Even in simple partially identified models, certain questions of optimality remain open to the best of our knowledge. For instance, in the setting described in Example 3.1, it would be of interest to characterize confidence intervals that are "shortest" in an appropriate sense. Of course, for the reasons discussed in Example 4.1, providing an answer to this question may be challenging. Consider also our discussion of subvector inference for moment inequalities. An obvious omission from our discussion is methods for inference with asymptotic frameworks more like those in Sections 4.1.4 and 4.1.5 for inference for moment inequalities. Furthermore, extensions of these methods to other classes of partially identified models, such as the ones in which the underlying parameter is limited by conditional moment inequalities, remain unavailable. Such problems are the subject of current research. See, for example, Canay et al. (2015), who treat these problems and others as special cases of a more general framework formulated in terms of unions of functional moment inequalities. Finally, while methods for inference in such models have made some inroads in empirical research, their adoption could be facilitated by the development of methods that are computationally feasible in problems of more realistic complexity. For this purpose, it may be useful to consider special classes of partially identified models, where it may be possible to exploit linearity or other structure to gain computational tractability.

References

Andrews, D. W. (2012): "Similar-on-the-boundary tests for moment inequalities exist, but have poor power," Working paper.

Andrews, D. W. K. and P. J. Barwick (2012): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure," *Econometrica*, 80, 2805–2826.

Andrews, D. W. K. and P. Guggenberger (2009): "Validity of Subsampling and "Plug-in Asymptotic" Inference for Parameters Defined by Moment Inequalities," *Econometric Theory*, 25, 669–709.

Andrews, D. W. K. and X. Shi (2013): "Inference Based on Conditional Moment Inequalities," *Econometrica*, 81, 609–666.

Andrews, D. W. K. and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78, 119–158.

Armstrong, T. B. (2014a): "A Note on Minimax Testing and Confidence Intervals in Moment Inequality Models," Manuscript. Yale University.

——— (2014b): "Weighted KS Statistics for Inference on Conditional Moment Inequalities," *Journal of Econometrics*, 181, 92–116.

——— (2015): "Asymptotically exact inference in conditional moment inequality models," *Journal of Econometrics*, 186, 51–65.

Armstrong, T. B. and H. P. Chan (2014): "Multiscale adaptive inference on conditional moment inequalities," Manuscript. Yale University.

Bahadur, R. and L. J. Savage (1956): "The Nonexistence of Certain Statistical Procedures in Nonparametric Problems," *Annals of Mathematical Statistics*, 25, 1115–1122.

Bajari, P., C. L. Benkard, and J. Levin (2006): "Estimating Dynamic Models of Imperfect Competition," Forthcoming *Econometrica*.

Beresteanu, A., I. Molchanov, and F. Molinari (2011): "Sharp Identification Regions in Models with Convex Moment Predictions," *Econometrica*, 79, 1785–1821.

Beresteanu, A. and F. Molinari (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 76, 763–814.

Berger, R. L. and D. D. Boos (1994): "P values maximized over a confidence set for the nuisance parameter," *Journal of the American Statistical Association*, 89, 1012–1016.

Bhattacharya, J., A. M. Shaikh, and E. Vytlacil (2008): "Treatment effect bounds under monotonicity assumptions: An application to swan-ganz catheterization," *The American Economic Review*, 351–356.

——— (2012): "Treatment effect bounds: An application to Swan–Ganz catheterization," *Journal of Econometrics*, 168, 223–243.

Blundell, R., A. Gosling, H. Ichimura, and C. Meghir (2007): "Changes in the distribution of male and female wages accounting for employment composition using bounds," *Econometrica*, 75, 323–363.

Bontemps, C., T. Magnac, and E. Maurin (2012): "Set Identified Linear Models," *Econometrica*, 80, 1129–1156.

Bugni, F., I. A. Canay, and X. Shi (2014): "Inference for Functions of Partially Identified Parameters in Moment Inequality Models," CeMMAP working paper CWP05/14.

Bugni, F. A. (2010): "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," *Econometrica*, 78, 735–753.

——— (2014): "Comparison of Inferential Methods in Partially Identified Models in Terms of Error in Coverage Probability," *Econometric Theory*, 1–56.

Bugni, F. A., I. A. Canay, and X. Shi (2015): "Specification tests for partially identified models defined by moment inequalities," *Journal of Econometrics*, 185, 259–282.

Canay, I. A. (2010): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," *Journal of Econometrics*, 156, 408–425.

Canay, I. A., A. Santos, and A. M. Shaikh (2015): "Inference for Unions of Functional Inequalities," Manuscript. Northwestern University.

Canova, F. and G. De Nicolo (2002): "Monetary disturbances matter for business fluctuations in the G-7," *Journal of Monetary Economics*, 49, 1131–1159.

Chernozhukov, V., D. Chetverikov, and K. Kato (2013a): "Testing Many Moment Inequalities," *arXiv preprint arXiv:1312.7614*.

Chernozhukov, V., I. Fernandez-Val, J. Hahn, and W. Newey (2013b): "Average and Quantile Effects in Nonseparable Panel Models," *Econometrica*, 81, 535–580.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75, 1243–1284.

Chernozhukov, V., W. Kim, S. Lee, and A. Rosen (2013c): "Implementing intersection bounds in Stata," *Stata Journal*, 15, 21–44.

Chernozhukov, V., S. Lee, and A. M. Rosen (2013d): "Intersection Bounds: Estimation and Inference," *Econometrica*, 81, 667–737.

Chesher, A. and A. Rosen (2014): "Generalized Instrumental Variable Models," *CeMMAP working paper CWP04/14.*

Chesher, A., A. Rosen, and K. Smolinski (2013): "An Instrumental Variable Model of Multiple Siscrete Choice," *Quantitative Economics*, 4, 157–196.

Chetverikov, D. (2013): "Adaptive Test of Conditional Moment Inequalities," *arXiv:1201.0167*.

Ciliberto, F. and E. Tamer (2010): "Market Structure and Multiple Equilibria in Airline Industry," *Econometrica*, 77, 1791–1828.

Faust, J. (1998): "The robustness of identified VAR conclusions about money," in *Carnegie-Rochester Conference Series on Public Policy*, Elsevier, vol. 49, 207–244.

Frisch, R. (1934): *Statistical confluence analysis by means of complete regression systems*, vol. 5, Universitetets Økonomiske Instituut.

Galichon, A. and M. Henry (2011): "Set Identification in Models with Multiple Equilibria," *The Review of Economic Studies*, 78, 1264–1298.

——— (2013): "Dilation Bootstrap," *Journal of Econometrics*, 177, 109–115.

Grieco, P. (2014): "Discrete Games with Flexible Information Structures: An Application to Local Grocery Markets," *RAND Journal of Economics*, 45, 303–340.

Haile, P. A. and E. Tamer (2003): "Inference with an Incomplete Model of English Auctions," *Journal of Political Economy*, 111, 1–51.

Hansen, L. P., J. Heaton, and E. G. J. Luttmer (1995): "Econometric evaluation of asset pricing models," *Review of Financial Studies*, 8, 237–274.

Hansen, L. P. and R. Jagannathan (1991): "Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy*, 225–262.

Heckman, J. J. and E. J. Vytlacil (2001): *Instrumental variables, selection models, and tight bounds on the average treatment effect*, Springer.

Henry, M. and A. Onatski (2012): "Set Coverage and Robust Policy," *Economics Letters*, 115, 256–257.

Ho, K. and A. Pakes (2014): "Hospital choices, hospital prices and financial incentives to physicians," *American Economic Review*, 104, 3841–84.

Ho, K. and A. M. Rosen (2016): "Partial Identification in Applied Research: Benefits and Challenges," Manuscript, UCL.

Horowitz, J. L. and C. F. Manski (1995): "Identification and Robustness with Contaminated and Corrupted Data," *Econometrica*, 63, 281–302.

——— (1998): "Censoring of Outcomes and Regressors due to Survey Nonresponse: Identification and Estimation using weights and imputations," *Journal of Econometrics*, 84, 37–58.

Imbens, G. and C. F. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857.

Kaido, H. (2012): "A Dual Approach to Inference for Partially Identified Econometric Models," Manuscript. Boston University.

Kaido, H., F. Molinari, and J. Stoye (2016): "Inference for Projections of Identified Sets," Manuscript. Cornell Univeristy.

Kaido, H. and A. Santos (2014): "Asymptotically Efficient Estimation of Models Defined by Convex Moment Inequalities," *Econometrica*, 82, 387–413.

Kim, K. i. (2008): "Set Estimation and Inference with Models Characterized by Conditional Moment Inequalities," *manuscript. Michigan State University.*

Kitagawa, T. (2012): "Estimation and inference for set-identified parameters using posterior lower probability," Manuscript, UCL.

Klepper, S. and E. E. Leamer (1984): "Consistent sets of estimates for regressions with errors in all variables," *Econometrica: Journal of the Econometric Society*, 163–183.

Kline, B. and E. Tamer (2015): "Bayesian Inference in a Class of Partially Identified Models," *manuscript. Harvard University.*

Kline, P., A. Santos, et al. (2013): "Sensitivity to missing data assumptions: Theory and an evaluation of the US wage structure," *Quantitative Economics*, 4, 231–267.

Kline, P. and M. Tartari (2015): "Bounding the labor supply responses to a randomized welfare experiment: A revealed preference approach," Tech. rep., National Bureau of Economic Research.

Kudo, A. (1963): "A Multivariate Analog of a One-Sided Test," *Biometrica*, 59, 403–418.

Lee, S., K. Song, and Y.-J. Whang (2013): "Testing Functional Inequalities," *Journal of Econometrics*, 172, 14–32.

Lehmann, E. (1952): "Testing multiparameter hypotheses," *The Annals of Mathematical Statistics*, 541–552.

Lehmann, E., J. P. Romano, and J. P. Shaffer (2012): "On optimality of stepdown and stepup multiple test procedures," in *Selected Works of EL Lehmann*, Springer, 693–717.

Machado, C., A. M. Shaikh, and E. J. Vytlacil (2013): "Instrumental Variables and the Sign of the Average Treatment Effect," Tech. rep.

Manski, C. F. (1989): "Anatomy of the Selection Problem," *Journal of Human Resources*, 24, 343–360.

——— (1990): "Nonparametric bounds on treatment effects," *The American Economic Review*, 319–323.

——— (1994): "The selection problem," in *Advances in Econometrics, Sixth World Congress*, vol. 1, 143–70.

——— (1995): *Identification problems in the social sciences*, Harvard University Press.

——— (1997): "Monotone treatment response," *Econometrica: Journal of the Econometric Society*, 1311–1334.

——— (2003): *Partial Identification of Probability Distributions*, Springer-Verlag, New York.

——— (2007): *Identification for Prediction and Decision*, Harvard University Press.

——— (2013): *Public Policy in an Uncertain World: Analysis and Decisions*, Harvard University Press.

Manski, C. F. and J. V. Pepper (2000): "Monotone instrumental variables: with an application to the returns to schooling," *Econometrica*, 68, 997–1010.

Manski, C. F. and E. Tamer (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70, 519–546.

Marschak, J. and W. H. Andrews (1944): "Random simultaneous equations and the theory of production," *Econometrica, Journal of the Econometric Society*, 143–205.

Matzkin, R. (2007): "Nonparametric Identification," in *Handbook of Econometrics*, ed. by J. Heckman and E. Leamer, Elsevier Science, vol. 6B, 5307–5368.

McCloskey, A. (2012): "Bonferroni-based size-correction for nonstandard testing problems," *Available at SSRN 2171912*.

——— (2015): "On the Computation of Size-Correct Power-Directed Tests with Null Hypotheses Characterized by Inequalities,," *manuscript. Brown University.*

Menzel, K. (2014): "Consistent Estimation with Many Moment Inequalities," *Journal of Econometrics*, 182, 329–350.

Molchanov, I. and F. Molinari (2014): "Applications of random set theory in econometrics," *Annu. Rev. Econ.*, 6, 229–251.

Moon, H. R. and F. Schorfheide (2012): "Bayesian and Frequentist Inference in Partially Identified Models," *Econometrica*, 80, 755–782.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2015): "Moment Inequalities and Their Application," *Econometrica*, 83, 315–334.

Perlman, M. D., L. Wu, et al. (1999): "The emperors new tests," *Statistical Science*, 14, 355–369.

Politis, D. N., J. P. Romano, and M. Wolf (1999): *Subsampling*, Springer, New York.

Ponomareva, M. (2010): "Inference in Models Defined by Conditional Moment Inequalities with Continuous Covariates," *manuscript. University of Western Ontario.*

Romano, J. P. (2004): "On Non-parametric Testing, the Uniform Behaviour of the t-test, and Related Problems," *Scandinavian Journal of Statistics*, 31, 567–584.

Romano, J. P. and A. M. Shaikh (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Statistical Planning and Inference*, 138, 2786–2807.

——— (2010): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 78, 169–212.

——— (2012): "On the Uniform Asymptotic Validity of Subsampling and the Bootstrap," *The Annals of Statistics*, 40, 2798–2822.

——— (2015): "On Testing Moment Inequalities," Working paper.

Romano, J. P., A. M. Shaikh, and M. Wolf (2014): "A Practical Two-Step Method for Testing Moment Inequalities," *Econometrica*, 82, 1979–2002.

Romano, J. P. and M. Wolf (2013): "Testing for monotonicity in expected asset returns," *Journal of Empirical Finance*, 23, 93–116.

Rosen, A. M. (2008): "Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities," *Journal of Econometrics*, 146, 107–117.

Shaikh, A. M. and E. J. Vytlacil (2011): "Partial identification in triangular systems of equations with binary dependent variables," *Econometrica*, 949–955.

Silvapulle, M. J. (1996): "A test in the presence of nuisance parameters," *Journal of the American Statistical Association*, 91, 1690–1693.

Staiger, D. and J. H. Stock (1997): "Instrumental Variable Regression with Weak Instruments," *Econometrica*, 50, 1029–1054.

Tamer, E. (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *The Review of Economic Studies*, 70, 147–165.

——— (2012): "Partial Identification in Econometrics," *Annual Review of Economics*, 2, 167–195.

Uhlig, H. (2005): "What are the effects of monetary policy on output? Results from an agnostic identification procedure," *Journal of Monetary Economics*, 52, 381–419.

Wolak, F. (1987): "An Exact Test for Multiple Inequality and Equality Constraints in the Linear Regression Model," *Journal of the American Statistical Association*, 82, 782–793.

——— (1991): "The Local Nature of Hypothesis Tests Involving Inequality Constraints in Nonlinear Models," *Econometrica*, 59, 981–995.