Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities

Adam M. Rosen

Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities

Adam M. Rosen[*][†]

Department of Economics, University College London

Gower Street, London WC1E 6BT, UK

and

Institute for Fiscal Studies

7 Ridgmount Street, London WC1E 7AE, UK

First Version: November 2005
This Version: June 2006

Abstract

This paper proposes a new way to construct confidence sets for a parameter of interest in models comprised of finitely many moment inequalities. Building on results from the literature on multivariate one-sided tests, I show how to test the hypothesis that any particular parameter value is logically consistent with the maintained moment inequalities. The associated test statistic has an asymptotic chi-bar-square distribution, and can be inverted to construct an asymptotic confidence set for the parameter of interest, even if that parameter is only partially identified. The confidence sets are easily computed, and Monte Carlo simulations demonstrate good finite sample performance.

JEL classification: C3, C12

Keywords: Partial identification, Inference, Moment inequalities

1 Introduction

When the assumptions of an econometric model are not restrictive enough to point identify the parameters of interest, but nonetheless impose meaningful restrictions on the values these parameters may take, the parameters are said to be partially identified.[1] Much of the early research on partial identification has not focused on issues of statistical inference, and for good reason. First, sufficient characterization of the identified set for partially identified parameters is a necessary precursor for statistical inference. Second, in some cases, the size of the identified set is significantly larger than the imprecision of estimates due to sampling variation.[2] However, in order to build confidence regions, perform hypothesis tests, or compare set-identified parameters to point estimates derived from more restrictive models, sampling variation must be taken into account.

Until recently, much of the literature on partial identification has sought to build "bounds" for univariate parameters. That is, if the parameter of interest, θ_0, is univariate, the identification region can often be characterized by just two numbers, the lower and upper bounds of an interval in $\mathbb{R}$. In this case, an asymptotically valid bootstrap procedure can be used to build confidence intervals for the entire identified set, such as those constructed by Manski and Nagin (1998) and Horowitz and Manski (2000). In these cases, the population identification region can be written as an interval $[L, U]$, and this procedure yields a confidence interval $\left[\hat{L}_\alpha, \hat{U}_\alpha\right]$ such that $\lim_{n \to \infty} \Pr\left([L, U] \subseteq \left[\hat{L}_\alpha, \hat{U}_\alpha\right]\right) = \alpha$, for a pre-specified value of $\alpha \in [0, 1]$. Also in the case where the parameter of interest is univariate, Imbens and Manski (2004) more recently show how to construct confidence intervals with pre-specified asymptotic coverage for the parameter θ_0 itself, rather than for the entire identification region Θ^*. That is, they show how to compute $\left[\hat{L}_\alpha, \hat{U}_\alpha\right]$ such that $\inf_{\theta \in \Theta^*} \lim_{n \to \infty} \Pr\left(\theta \in \left[\hat{L}_\alpha, \hat{U}_\alpha\right]\right) = \alpha$, where Θ^* denotes the identification region for the true model parameter θ_0.[3] If the economist wishes to perform inference on θ_0 rather than Θ^*, their technique yields a strictly smaller confidence interval for any coverage level.

Moving beyond the realm of univariate parameters, Chernozhukov, Hong, and Tamer (2004) (CHT) develop a subsampling procedure to build asymptotically valid confidence sets of a pre-specified level for the identified set in any model in which the identified set can be written as those values that minimize an objective function. They further show in their appendix how to modify their procedure to build confidence sets that cover just the parameter of interest, rather than the whole identification region, with the desired asymptotic coverage probability. Shaikh

---

[1] Manski (2003) offers a vast survey of models in which parameters of interest are partially identified. I adopt the term "partial identification" from this text.

[2] See Manski and Nagin (1998), for example.

[3] Imbens and Manski (2004) also show how to construct confidence intervals $\left[\hat{L}_\alpha, \hat{U}_\alpha\right]$ such that $\lim_{N \to \infty} \inf_{\theta \in \Theta^*} \Pr\left(\theta \in \left[\hat{L}_\alpha, \hat{U}_\alpha\right]\right) = \alpha$. These confidence regions are larger than the former kind, but have the additional quality of uniform asymptotic coverage over Θ^*.

(2005) also employs subsampling to construct confidence sets for both the entire identified set or the parameter of interest. In particular he justifies the use of an iterative procedure to construct confidence regions for the entire identified set. While the inferential approaches of both CHT and Shaikh are applicable in a very general class of models, their reliance on subsampling may in some cases be computationally burdensome.

Other recent research on building confidence regions for partially identified parameters and/or their identification regions includes Andrews, Berry, and Jia (2004), Pakes, Porter, Ho, and Ishii (2006), henceforth (PPHI), and Beresteanu and Molinari (2006). PPHI study the use of moment inequalities to perform inference on θ_0 in models with agents who make optimal, or approximately optimal, decisions from an either discrete or continuous choice set. They develop conservative confidence sets for model parameters, and apply their technique to two applications in Industrial Organization: an investigation of how banks choose their ATM locations, and an analysis of the determination of HMO hospital networks in the United States.[4] To perform inference, PPHI use simulations from a multivariate normal distribution to approximate the distribution of the moments in their model. Their method for building confidence sets for model parameters is straightforward to implement, but admittedly conservative in that asymptotic coverage is at least as high as the nominal level. Exactly how conservative their confidence sets are depends on the particular application, and this remains a topic for future research. Andrews, Berry, and Jia (2004) develop a means of inference on Θ^* in incomplete models of firms' entry and exit decisions. Their estimation procedure makes use of the necessary conditions for Nash Equilibrium, which are typically moment inequality restrictions. To perform inference, they simulate these inequalities for different parameter values, and use a bootstrap procedure to construct confidence sets for model parameters. They provide an application to the location decisions of Wal-mart, Kmart, and other discount chain stores. Beresteanu and Molinari (2006) use the theory of set-valued random variables (SVRVs) to analyze the asymptotic behavior of a class of set-valued estimators for partially identified parameters. In particular, they show how to build confidence collections for the identified set in these models. Another related recent paper is that of Moon and Schorfheide (2006), who focus on models that are comprised of both moment equalities and inequalities, in which the moment equalities are sufficient for point identification. They show that incorporating the additional inequality restrictions in an empirical likelihood estimator can improve the efficiency of their estimates.

In this paper, I focus explicitly on models that are comprised of a finite number of moment inequalities, $\mathbb{E}[m(y, x, \theta_0)] \geq 0$, where $m(y, x, \theta_0)$ is a vector-valued function of random variables (y, x), which is known up to the value of the possibly multivariate parameter θ_0. This class of models includes many examples from the econometrics literature, dating back at least to Frisch (1934), who derived bounds for β in the simple linear regression model with measurement error.

---

[4] The applications are explored in further detail in Ishii (2005) and Ho (2005).

More recently, Klepper and Leamer (1984) extend Frisch's result to the multivariate linear regression model with errors in all variables. Another example of bounds that can be cast in terms of moment inequalities are the Frechet bounds (Frechet (1951)) on the value of the joint CDF of two random variables evaluated at any point based on knowledge of only the marginal CDFs. More recent examples of models based on finitely many moment inequalities include the case of interval data on outcomes studied by Manski and Tamer (2002) when the covariate space is discrete, bounds on treatment effects[5], and the case of inference on the mean of a univariate distribution with missing data, studied by Manski (1989) and Imbens and Manski (2004).

I build confidence regions for the parameter of interest through pointwise testing. The technique employed in this paper makes use of prior results from the literature on multivariate hypothesis testing, such as Kudo (1963), Perlman (1969), Gourieroux, Holly, and Monfort (1982), and Wolak (1991).[6] Specifically, I construct a test statistic that, under sufficient regularity conditions, and when evaluated at θ_0, has an asymptotic distribution that is a mixture of chi-square distributions, the chi-bar-square distribution. This test statistic can then be used to construct approximate confidence sets for θ_0 with pre-specified asymptotic coverage. For this purpose, I restrict attention to models where knowledge of the parameter of interest is obtained by means of a finite set of moment inequality restrictions. As discussed above, many models previously studied in the literature can be written in such a form. A careful distinction must be made, however, between models that achieve partial identification from a finite set of unconditional moment restrictions, and those that rely on an infinite set of such restrictions. The latter case offers significant complication, and is not covered in this paper.[7]

The procedure for building confidence sets amounts to a test for each value of the parameter θ in the underlying parameter space. I characterize the asymptotic distribution of a test statistic $\hat{Q}_n(\theta)$ under the null hypothesis that θ is an element of the identified set Θ^*, or, equivalently, that $\mathbb{E}[m(y, x, \theta)] \geq 0$. The confidence set for θ_0 is then simply the set of θ that are not rejected by this hypothesis test. Because pointwise testing is based on fixed θ, the theory needed to guarantee proper asymptotic coverage relies completely on the distribution of observables.[8] The procedure I employ is relatively straightforward and easy to implement in practice in many cases of interest, which is demonstrated with two specific examples. A computational drawback is that for some models, the cutoff value for the test statistic $\hat{Q}_n(\theta)$ differs for different values of $\theta \in \Theta^*$. That is, the test statistic $\hat{Q}_n(\theta)$ is not asymptotically pivotal because its asymptotic distribution depends

---

[5] Some specific examples include Manski and Nagin (1998), Molinari (2005), Balke and Pearl (1997), Manski and Pepper (2000) and Hotz, Mullin, and Sanders (1997).

[6] Sen and Silvapulle (2004) offer a thorough compendium of this body of research.

[7] Also falling into the latter class are models based on conditional moment inequalities where the conditioning variables have continuous support. In this case, one could potentially instead use a finite number of the implied unconditional moment inequalities to build conservative confidence sets as described in this paper.

[8] Hu (2002) uses a conceptually similar approach to building confidence sets in a GMM framework in which a subset of model parameters might not be point-identified.

on the variance of those components of $m(y, x, \theta)$ that have expected value zero. This problem is overcome by building conservative confidence sets for θ_0 by using an upper bound on the number of such components. The dimension of $m(y, x, \theta)$, J, is clearly an upper bound, but in models with partially identified parameters there is often a smaller upper bound. This is indeed the case in both examples of this paper.

The paper proceeds as follows. In section 2 I present the moment inequality model on which I focus, as well as my strategy for building confidence regions. I state precisely what the properties of my confidence sets are, and distinguish them from two other types of confidence sets in the literature. In section 3 I describe the pointwise hypothesis testing procedure. Section 4 then presents two easy ways to construct conservative confidence sets based on the hypothesis test of section 3. The methodology is computationally simple and no simulation, subsampling, or bootstrapping is required. In section 5 I present two examples and investigate the performance of confidence sets in these models via Monte Carlo simulations. This includes a comparison, via both simulation and analytically, to the confidence sets of Imbens and Manski (2004) for the case of the mean with missing data. It is shown that when there is positive probability of missing data, the two types of confidence sets are asymptotically identical for any fixed confidence level. Section 6 concludes and offers avenues for continued research. All proofs are in the Appendix.

2   The Model

Let $Z \equiv \{(x_i, y_i) : i = 1, ..., n\}$ be a random sample of observations drawn from population $\{\Omega, \mathcal{F}, P\}$, where $\mathcal{F}$ is the Borel sets on Ω. (x, y) is used to denote a representative draw of a random variable according to distribution P. Let $\mathcal{X}$, $\mathcal{Y}$ denote the support of the random variables x, y, respectively, where $\mathcal{X} \subseteq \mathbb{R}^s$ and $\mathcal{Y} \subseteq \mathbb{R}^p$. I take y to be the outcome variables and x covariates. Each observation (x_i, y_i) represents *all* information directly observed by the econometrician for each $i = 1, ..., n$. If partial identification is a result of missing data, for example, then (x_i, y_i) excludes those features of individual i in the population that are missing.[9] I use θ to denote a representative value of the parameter of interest, where θ is known to belong to the compact space $\Theta \subseteq \mathbb{R}^k$. The set Θ^* is the set of values of $\theta \in \Theta$ that satisfy the restrictions of the model, i.e. Θ^* is the identified set for θ. The "true" underlying value of θ in the model is denoted θ_0, but in general θ_0 might not be identified by the restrictions of the model.

In this paper I study models based on moment inequalities. In full generality, the model is

---

[9]This is made more explicit in the missing data example of section 5.1.

summarized by the restrictions

$$\mathbb{E}\left[m\left(y, x, \theta_0\right)\right] = \mathbb{E} \begin{bmatrix} m_1\left(y, x, \theta_0\right) \\ \vdots \\ m_J\left(y, x, \theta_0\right) \end{bmatrix} \geq \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{1}$$

$J < \infty$ is the number of moment inequalities of the model. This rules out, for example, the model $E\left[m\left(y, x, \theta_0\right) | x\right] \geq 0$, where $\mathcal{X}$ has infinite cardinality. In summary, the model is given by the following three assumptions.

Assumption 1 *(A1: random sampling) $Z \equiv \{(x_i, y_i) : i = 1, ..., N\}$ are i.i.d. observations distributed $\mathbb{P}$.*

Assumption 2 *(A2: compact parameter space) θ_0 is an element of the compact space $\Theta \subseteq \mathbb{R}^k$.*

Assumption 3 *(A3: moment inequalities) $\mathbb{E}\left[m\left(y, x, \theta_0\right)\right] \geq 0$, where $m\left(\cdot, \cdot, \cdot\right) : \mathbb{R}^p \times \mathbb{R}^s \times \Theta \rightarrow \mathbb{R}^J$.*

The assumptions above yield the following identified set for θ_0.

Definition 1 *Given assumptions (A1)-(A3), the identified set for θ_0 is*

$$\Theta^* = \{\theta \in \Theta : \mathbb{E}\left[m\left(y, x, \theta\right)\right] \geq 0\}.$$

The identified set for θ_0, Θ^*, is the set of parameter values θ that satisfy the restrictions of the model, and thus θ_0 is necessarily an element of this set. If Θ^* is a singleton, then $\Theta^* = \{\theta_0\}$ and θ_0 is point identified. If Θ^* is empty, the model is rejected, and if Θ^* is neither empty nor singleton, then θ_0 is only partially identified. In this case, the model is informative even though θ_0 is not point identified. By definition of the identified set, there is no way to distinguish between any of the elements of Θ^* being the true θ_0 on the basis of observables; any element of the identified set is a plausible value for θ_0, as all elements of Θ^* are observationally equivalent by definition.

The confidence sets of this paper are based on a test of the hypothesis that $\theta \in \Theta^*$ against the alternative $\theta \notin \Theta^*$, or equivalently, the test

$$\begin{aligned} H_0 &: \quad \mathbb{E}\left[m\left(y, x, \theta\right)\right] \geq 0 \\ H_1 &: \quad \mathbb{E}\left[m\left(y, x, \theta\right)\right] \ngeq 0, \end{aligned} \tag{2}$$

for any fixed candidate value of $\theta \in \Theta$. First, I show how to perform this test with pre-specified asymptotic size α. Once the testing procedure is established for fixed θ, a $1 - \alpha$ level confidence set for θ_0 is constructed by taking the set of θ that are not rejected by this hypothesis test.

Formally, the constructed confidence set obtains the desired asymptotic coverage probability by means of standard arguments given by Lehmann (1986, pp. 90-91.). The only difference here is that if θ_0 is not identified, the null hypothesis is composite rather than simple. In Lehmann's notation, let $A(\theta)$ denote the "acceptance" region for the null hypothesis. That is, the test yields a region $A(\theta)$ such that one fails to reject H_0 if the sample point $Z = \{(x_i, y_i) : i = 1, ..., n\} \in A(\theta)$. To guarantee asymptotic type 1 error of no more than α, the acceptance region is constructed to satisfy

$$\inf_{\theta \in \Theta^*} \lim_{n \to \infty} \Pr\{Z \in A(\theta)\} = 1 - \alpha. \tag{3}$$

An asymptotic $1 - \alpha$ confidence set for θ_0 is then given by

$$CR_{1-\alpha} = \{\theta \in \Theta : Z \in A(\theta)\},$$

since

$$\lim_{n \to \infty} \Pr\{\theta_0 \in CR_{1-\alpha}\} = \lim_{n \to \infty} \Pr\{Z \in A(\theta_0)\} \geq \inf_{\theta \in \Theta^*} \lim_{n \to \infty} \Pr\{Z \in A(\theta)\} = 1 - \alpha.$$

If one were instead interested in a confidence set for Θ^*, then to obtain the same asymptotic level of coverage $1 - \alpha$, a (weakly) larger confidence set is needed. This follows from the observation that

$$\Pr\{\Theta^* \subseteq CR_{1-\alpha}\} = \Pr\{\cap_{\theta \in \Theta^*} (\theta \in CR_{1-\alpha})\} = \Pr\{\cap_{\theta \in \Theta^*} (Z \in A(\theta))\} \leq \inf_{\theta \in \Theta^*} \Pr\{Z \in A(\theta)\}.$$

Combining this with the fact that

$$\lim_{n \to \infty} \inf_{\theta \in \Theta^*} \Pr\{Z \in A(\theta)\} \leq \inf_{\theta \in \Theta^*} \lim_{n \to \infty} \Pr\{Z \in A(\theta)\}, \tag{4}$$

it follows that

$$\lim_{n \to \infty} \Pr\{\Theta^* \subseteq CR_{1-\alpha}\} \leq \inf_{\theta \in \Theta^*} \lim_{n \to \infty} \Pr\{Z \in A(\theta)\},$$

so that $CR_{1-\alpha}$ is not guaranteed to have asymptotic coverage of $1 - \alpha$ for the entire identification region Θ^*, even though it is guaranteed to have the desired coverage for the true parameter θ_0. To obtain an asymptotically valid confidence region for Θ^* one needs first to construct an acceptance region $A^*(\theta)$ with the property that

$$\lim_{n \to \infty} \Pr\{\cap_{\theta \in \Theta^*} (Z \in A^*(\theta))\} = 1 - \alpha, \tag{5}$$

with associated confidence set

$$CR^*_{1-\alpha} = \{\theta \in \Theta : Z \in A^*(\theta)\},$$

since

$$\lim_{n\to\infty} \Pr\left\{\Theta^* \subseteq CR^*_{1-\alpha}\right\} = \lim_{n\to\infty} \Pr\left\{\cap_{\theta\in\Theta^*}\left(\theta \in CR^*_{1-\alpha}\right)\right\} = \lim_{n\to\infty} \Pr\left\{\cap_{\theta\in\Theta^*}\left(Z \in A^*\left(\theta\right)\right)\right\} = 1 - \alpha.$$

The difference between the two types of confidence sets has been discussed previously by Imbens and Manski (2004) and Chernozhukov, Hong, and Tamer (2004, Appendix G).

The pointwise testing approach I employ here relies on a test of the hypothesis $\theta \in \Theta^*$, and can thus be used to build confidence regions of the former type. That is, this paper focuses entirely on constructing confidence sets $CR_{1-\alpha}$ for the parameter of interest θ_0, that satisfy

$$\inf_{\theta\in\Theta^*} \lim_{n\to\infty} \Pr\left\{\theta \in CR_{1-\alpha}\right\} = 1 - \alpha, \tag{6}$$

since

$$\lim_{n\to\infty} \Pr\left\{\theta_0 \in CR_{1-\alpha}\right\} \geq \inf_{\theta\in\Theta^*} \lim_{n\to\infty} \Pr\left\{\theta \in CR_{1-\alpha}\right\}.$$

As long as (6) holds, $CR_{1-\alpha}$ must have at least $1 - \alpha$ asymptotic coverage for θ_0.

The asymptotic coverage so obtained will not, however, be uniform over $\theta \in \Theta^*$. If uniformity is desired, one must instead construct a confidence region $\widetilde{CR}_{1-\alpha}$ such that

$$\lim_{n\to\infty} \inf_{\theta\in\Theta^*} \Pr\left\{\theta \in \widetilde{CR}_{1-\alpha}\right\} = 1 - \alpha.$$

Imbens and Manski (2004) also show how to construct such confidence intervals for univariate θ_0 in their model. Confidence regions of this type require more stringent regularity conditions.[10]

3 Testing the Hypothesis that $\theta \in \Theta^*$

In this section, I consider a test of the hypothesis (2) for a fixed candidate value of θ. To test this hypothesis, I construct a test statistic, $\hat{Q}_n\left(\theta\right)$ whose asymptotic distribution, when scaled by n, is chi-bar-square (a mixture of chi-square random variables) under the null hypothesis. The test statistic is in general not asymptotically pivotal, but can still be used to construct conservative confidence sets for θ_0. Given the asymptotic distribution of $n\hat{Q}_n\left(\theta\right)$, for any fixed θ, I compute a cutoff value $C_{1-\alpha}$ such that

$$\inf_{\theta\in\Theta^*} \lim_{n\to\infty} \Pr\left\{n\hat{Q}_n\left(\theta\right) \leq C_{1-\alpha}\right\} \geq 1 - \alpha.$$

---

[10]In the model of Imbens and Manski (2004), sufficient regularity conditions are required to guarantee a uniform central limit theorem over Θ^* holds when confidence sets of this nature are constructed. When the infimum is taken outside the limit, a uniform CLT is not necessary.

A nominal α-level confidence set is then given by

$$CR_{1-\alpha} \equiv \left\{ \theta \in \Theta : n\hat{Q}_n(\theta) \leq C_{1-\alpha} \right\}.$$

To construct the test statistic $\hat{Q}_n(\theta)$, $\mathbb{E}[m(y,x,\theta)]$ is first estimated nonparametrically by its sample analog $\hat{E}_n[m(y,x,\theta)]$. Then, under mild regularity conditions, it is straightforward to characterize the asymptotic distribution of $\hat{E}_n[m(y,x,\theta)]$ about $\mathbb{E}[m(y,x,\theta)]$. However, the goal is to test whether θ is contained in the identified set implied by the restrictions (1). As such, I consider the asymptotic distribution of the following minimum Wald-type statistic:

$$\hat{Q}_n(\theta) = \min_{t \geq 0} \; \left[\hat{E}_n[m(y,x,\theta)] - t \right]' \hat{V}_\theta^{-1} \left[\hat{E}_n[m(y,x,\theta)] - t \right],$$

where $\hat{V}_\theta$ is the sample variance of $m(y,x,\theta)$. If the moment restrictions $\mathbb{E}[m(y,x,\theta)] \geq 0$ are true, i.e. if $\theta \in \Theta^*$, then $\hat{Q}_n(\theta)$ should be small. In this case, violations of $\hat{E}_n[m(y,x,\theta)] \geq 0$ are attributable to no more than sampling variation. It is worth noting that this statistic, taken as a function of the parameter θ, is in fact the sample analog of a modified minimum distance (MMD) objective function, as defined by Manski and Tamer (2002). The population version of $\hat{Q}_n(\theta)$ is

$$Q(\theta) = \min_{t \geq 0} \; [\mathbb{E}[m(y,x,\theta)] - t]' V_\theta^{-1} [\mathbb{E}[m(y,x,\theta)] - t],$$

where V_θ is the variance of $m(y,x,\theta)$. This is because $Q(\theta) = 0$ if and only if $\mathbb{E}[m(y,x,\theta)] \geq 0$. While Manski and Tamer (2002) derive conditions for consistency of MMD estimators, characterization of the asymptotic distribution of MMD estimators was left as a topic for future research.

Outside the context of estimating partially identified parameters, test statistics of this form have been used previously in the literature on multivariate one-sided hypothesis testing, e.g. Perlman (1969), Kodde and Palm (1986), and Wolak (1991). In these prior studies, however, the distribution of unobservables is modeled parametrically, and θ_0 is point identified and can be consistently estimated. Here, there is no parametric specification for unobservables and θ_0 need not be point identified. The formulation that is closest to that considered here is that of Wolak (1991). Wolak shows that the limiting distribution of test statistics of the form $\hat{Q}_n(\theta)$ depends only on those constraints that are satisfied with equality at the least favorable value of θ satisfying the null hypothesis, here that $\mathbb{E}[m(y,x,\theta)] \geq 0$. In his model, however, there is a known function $h(\theta)$ in place of $\mathbb{E}[m(y,x,\theta)]$. In the setting of this paper, aside from the complication that here θ_0 is only partially identified, it is also the case that $\mathbb{E}[m(y,x,\theta)]$ is not a known function, but rather must be estimated.

This is a substantial complication because, as I show in Proposition 1, the asymptotic distribution of $\hat{Q}_n(\theta)$, sufficiently scaled, is degenerate except on the boundary of the null hypothesis.[11]

---

[11] Andrews (2001) considers hypothesis tests when a parameter is on the boundary of the maintained hypothesis,

Thus, the cutoff value of $\hat{Q}_n(\theta)$ used to compute the critical region is driven entirely by the subset of $\mathbb{E}[m(y,x,\theta)] \geq 0$ such that $\mathbb{E}[m(y,x,\theta)]$ is on the boundary of $\mathbb{R}_+^J$, i.e. the set of θ such that $\mathbb{E}[m_j(y,x,\theta)] = 0$ for at least one $j \in \{1,...,J\}$. In Wolak's model, this complication also arises, but in that setting h is a known function, and the boundary of the set $\{\theta : h(\theta) \geq 0\}$ is known.

To derive asymptotics for $\hat{Q}_n(\theta)$, I impose the following two additional assumptions.

Assumption 4 *(A4: finite variance of m on Θ^*)* $\sup_{\theta \in \Theta^*} \mathbb{E}[m(y,x,\theta) m(y,x,\theta)'] < \infty$, *i.e. each element of the matrix* $\mathbb{E}[m(y,x,\theta) m(y,x,\theta)']$ *is finite for all* $\theta \in \Theta^*$.

Assumption 5 *(A5: positive definite variance) For each* $\theta \in \Theta^*$, V_θ *is positive definite.*

Assumption (A4), along with (A1), guarantees that the strong law of large numbers and a central limit theorem hold for $\mathbb{E}[m(y,x,\theta)]$, while assumption (A5) guarantees that V_θ is invertible. Under (A1) and (A4), it follows that for all $\theta \in \Theta^*$,

$$\hat{E}_n[m(y,x,\theta)] = \frac{1}{n}\sum_{i=1}^n m(y_i,x_i,\theta) \overset{a.s.}{\to} \mathbb{E}[m(y,x,\theta)], \tag{7}$$

$$\hat{V}_n[m(y,x,\theta)] = \frac{1}{n}\sum_{i=1}^n \left(m(y,x,\theta) - \hat{E}_n[m(y,x,\theta)]\right)\left(m(y,x,\theta) - \hat{E}_n[m(y,x,\theta)]\right)' \tag{8}$$

$$\overset{a.s.}{\to} var\{m(y,x,\theta)\} \equiv V_\theta,$$

and

$$\sqrt{n}\left\{\hat{E}_n[m(y,x,\theta)] - \mathbb{E}[m(y,x,\theta)]\right\} \overset{d}{\to} N(0,V_\theta). \tag{9}$$

The validity of assumption (A4) depends on the problem at hand. In the absence of (A4), what is needed for the asymptotic results of this section are the three conditions written above; the consistency of the sample mean and variance for $\mathbb{E}[m(y,x,\theta)]$ over Θ^*, and a central limit theorem for $\sqrt{n}\left\{\hat{E}_n[m(y,x,\theta)] - \mathbb{E}[m(y,x,\theta)]\right\}$ for each $\theta \in \Theta^*$.[12] Because the goal here is construction of a confidence set $CR_{1-\alpha}$ such that $\inf_{\theta \in \Theta^*} \lim_{N \to \infty} \Pr\{\theta \in CR_{1-\alpha}\} = 1-\alpha$, it is enough for these conditions to hold pointwise over Θ^*. If instead the researcher's goal was to construct a confidence set with uniform coverage over Θ^*, i.e. sets such that $\lim_{N \to \infty} \inf_{\theta \in \Theta^*} \Pr\{\theta \in CR_{1-\alpha}\} = 1-\alpha$, then stronger conditions would be needed, as discussed at the end of section 2.

There are two approaches to guaranteeing the validity of assumption (A4). Because the researcher cannot distinguish between θ_0 and any other element of Θ^*, one way is to show that

---

rather than the null. However, the hypothesis test (2) can be recast so that θ_0 does in fact lie on the boundary of the maintained hypothesis under the null. This point is elaborated in Appendix A.

[12]Both the assumption that the observations are iid and that the rate of convergence of $\hat{E}_n[m(y,x,\theta)]$ to $\mathbb{E}_n[m(y,x,\theta)]$ is $\sqrt{n}$ can be relaxed, as long as (7), (8), and (9) can be shown to hold at each $\theta \in \Theta^*$ for some sequence of constants $a_n \to \infty$ replacing $\sqrt{n}$.

this assumption holds at each value of $\theta \in \Theta^*$. This is what is done in both examples of this paper. The second approach is to simply impose stricter regularity at θ_0 itself a priori. That is, if the researcher knows that $\mathbb{E}\left[m\left(y,x,\theta_0\right)m\left(y,x,\theta_0\right)'\right] < \infty$, i.e. that m has finite variance at θ_0, then the identified set Θ^* can be taken to be the set of $\theta \in \Theta$ such that $\mathbb{E}\left[m\left(y,x,\theta\right)\right] \geq 0$ *and* $\mathbb{E}\left[m\left(y,x,\theta\right)m\left(y,x,\theta\right)'\right] < \infty$. Then, by assumption, (7), (8), and (9) hold at each $\theta \in \Theta^*$.

The first step to deriving asymptotics for $n\hat{Q}_n\left(\theta\right)$ under H_0 shows formally that only those components of $\mathbb{E}\left[m\left(y,x,\theta\right)\right]$ exactly equal to zero have a non-negligible contribution asymptotically. Before proceeding with the first result, I define some necessary notation. For expositional convenience, I refer to the subset of the J constraints that hold with equality at any particular value of θ as the set of binding constraints. Without loss of generality, let the first $b\left(\theta\right)$ constraints be the subset of binding constraints at θ, so that $\mathbb{E}\left[m_j\left(y,x,\theta\right)\right] = 0$, $j = 1,...,b\left(\theta\right)$, and $\mathbb{E}\left[m_j\left(y,x,\theta\right)\right] > 0$, $j = b\left(\theta\right)+1,...,J$. Let $m^*\left(y,x,\theta\right) = \left(m_1\left(y,x,\theta\right),...,m_{b\left(\theta\right)}\left(y,x,\theta\right)\right)'$ denote the subvector of moments that have mean zero, and let $V_\theta^* = var\left(m^*\left(y,x,\theta\right)\right)$. Let $b \equiv b\left(\theta_0\right)$, $V \equiv var\left(m\left(y,x,\theta_0\right)\right)$, and $V^* \equiv var\left(m^*\left(y,x,\theta_0\right)\right)$. Finally, I use $\Pr\left\{\chi_j^2 \geq c\right\}$ to denote the probability that a chi-square random variable with degrees of freedom j is at least as great as the constant c, where χ_0^2 denotes a point mass as zero. The following proposition characterizes the limiting distribution of $n\hat{Q}_n\left(\theta\right)$ under the hypothesis that $\theta \in \Theta^*$.

Proposition 1 *Under assumptions (A1)-(A5)*

$$\lim_{n\to\infty} \Pr\left\{n\hat{Q}_n\left(\theta\right) \geq c\right\} = \sum_{j=0}^{b(\theta)} w\left(b\left(\theta\right),b\left(\theta\right)-j,V_\theta^*\right)\Pr\left\{\chi_j^2 \geq c\right\}, \tag{10}$$

where $w\left(b,b-j,V^*\right)$ is the weights function defined by Wolak (1987) and Kudo (1963) evaluated at $\left(b,b-j,V^*\right)$.

Corollary 1 *Suppose that $\mathbb{E}\left[m\left(y,x,\theta\right)\right] > 0$. Then*

$$\lim_{n\to\infty} \Pr\left\{n\hat{Q}_n\left(\theta\right) > 0\right\} = 0.$$

Proposition 1 closely follows Lemma 1 of Wolak (1991). The first step to the proof shows that the limiting distribution of $n\hat{Q}_n\left(\theta\right)$ is determined only by those terms that correspond to components of $\mathbb{E}\left[m\left(y,x,\theta\right)\right]$ that are exactly equal to 0. The contribution of the other components vanishes in the limit as $n \to \infty$. The first corollary is an immediate implication; when $\mathbb{E}\left[m\left(y,x,\theta\right)\right] > 0$, $n\hat{Q}_n\left(\theta\right)$ is $o_p\left(1\right)$.

The weights function $w\left(b,j,V\right)$ has arisen repeatedly in research on multivariate one-sided hypothesis testing. It is the probability that t_0 has exactly j positive components, where t_0 is the

minimizer of the probability limit of the objective function. That is,

$$t_0 = \arg\min_{t \geq 0} \left[z - t \right]' V^{-1} \left[z - t \right],$$

where $z \sim N(0, V)$ is a multivariate normal random variable of dimension b, and

$$
\begin{aligned}
w(b, j, V) &= \Pr\{t_0 \text{ has exactly } j \text{ positive components}\} \\
&= \Pr\{t_0 \text{ has exactly } b - j \text{ components equal to zero}\}.
\end{aligned}
$$

These weights are referred to as "level probabilities" of a chi-bar-square distribution. Closed form expressions for the weights are given by Wolak (1987) for the case where $b \leq 4$, or where V is diagonal. More generally, closed-form expressions for the weights have not been obtained, but they can be approximated with arbitrary accuracy by means of simulation.[13]

If V_θ^* and $b(\theta)$ were known, then it would be straightforward using previously developed techniques to compute the cutoff value C_α such that $\sum_{j=0}^{b(\theta)} w(b(\theta), b(\theta) - j, V_\theta^*) \Pr\left\{\chi_j^2 \geq c\right\} = \alpha$. Unfortunately, V_θ^* and $b(\theta)$ are not known in this case. An intuitive solution would be to plug consistent estimates into the weights function, but unfortunately this approach won't work. It is straightforward to consistently estimate $V_\theta = var\{m(y, x, \theta)\}$, but the CDF of the limit distribution given by (10) is discontinuous in $b(\theta)$, so that the problem cannot be overcome by merely plugging in consistent estimates for V_θ^* and $b(\theta)$. This problem can, however, be overcome by taking a conservative approach. Section 4 details how this can be done by using an upper bound for $b(\theta)$ to construct a cutoff value C_α^{b*} such that

$$\inf_{\theta \in \Theta^*} \lim_{n \to \infty} \Pr\left\{ n\hat{Q}_n(\theta) \leq C_{1-\alpha}^{b*} \right\} \geq 1 - \alpha.$$

4 Computing Cutoff Values For Confidence Sets

In this section I illustrate a way to perform inference despite not knowing $b(\theta)$, the number of binding moment inequalities at any particular $\theta \in \Theta^*$. I provide two ways to compute cutoff values to build confidence sets that cover θ_0 with at least probability $1 - \alpha$ asymptotically. Both approaches have the advantage that the cutoff values are extremely easy to compute with any software package that provides values of chi-square CDFs. The first method is generally applicable, while the second shows how a smaller, but still conservative, cutoff value can be computed when it is known that V_θ^* is diagonal, which is in fact the case in both examples of section 5. Both approaches require that the researcher impose an upper bound on $b(\theta)$ for $\theta \in \Theta^*$; an obvious upper bound is the total number of moment inequalities, J. In some settings, it may be credible to

---

[13]Sen and Silvapulle (2004, pp. 78-80).

impose a smaller upper bound; more generally, I use b^* to denote the chosen upper bound. In fact, both examples considered in this paper are settings in which it is known that strictly fewer than J of the constraints can bind at any given value of θ. This happens because the model implies both upper and lower bounds on the expectation of a function of θ. This is not an uncommon occurrence in models with partially identified parameters.

4.1 Cutoff values for general V^*

The asymptotic distribution of $n\hat{Q}_n(\theta)$ obtained in Proposition 1 for any fixed $\theta \in \Theta^*$ is discontinuous in the unknowns $b(\theta)$ and V_θ^*. However, whatever V_θ^*, an upper bound on $b(\theta)$ can be used to construct a cutoff value that can used to perform the hypothesis test (2). This cutoff value can then be used to build conservative, asymptotically valid confidence sets for θ_0. The following corollary provides the main result.

Corollary 2 *Let (A1)-(A5) hold. Let $\sup_{\theta \in \Theta^*} b(\theta) \leq b^*$. Then for any $c > 0$,*

$$\sup_{\theta \in \Theta^*} \lim_{n \to \infty} \Pr\left\{ n\hat{Q}_n(\theta) \geq c \right\} \leq \frac{1}{2} \Pr\left\{ \chi_{b^*}^2 \geq c \right\} + \frac{1}{2} \Pr\left\{ \chi_{b^*-1}^2 \geq c \right\}.$$

The proof follows from the fact that the weights function satisfies the properties $0 \leq w(b, j, V^*) \leq 1/2$, $\sum_{j=0}^{b} w(b, j, V^*) = 1$, and $\Pr\left\{ \chi_j^2 > c \right\}$ is increasing in j, for any $c > 0$. The upper bound on the tail probability of the limit distribution of $n\hat{Q}_n(\theta)$ is obtained by putting as much weight as possible on the highest terms of the chi-bar-square summation of (10). Results on the upper bound on chi-bar-square tail probabilities have been used in prior research, going back at least to Perlman (1969).[14] Exactly how slack the inequality is depends on the feasible values of the variance matrix V_θ^* over $\theta \in \Theta^*$. Wolak (1991) discusses the possible slackness of such bounds in his framework. I leave investigation of the slackness of this bound in this framework to future work.

This corollary gives a way to construct asymptotically valid confidence sets for θ_0. This is because an implication of the corollary is that if $C_\alpha^{b^*}$ solves

$$\frac{1}{2} \Pr\left\{ \chi_{b^*}^2 \geq C_\alpha^{b^*} \right\} + \frac{1}{2} \Pr\left\{ \chi_{b^*-1}^2 \geq C_\alpha^{b^*} \right\} = \alpha, \tag{11}$$

Then

$$CR_{1-\alpha} = \left\{ \theta \in \Theta : n\hat{Q}_n(\theta) \leq C_\alpha^{b^*} \right\}$$

---

[14]Perlman derives upper bounds on tail probabilities of mixtures F distributions that employ the same weights function.

has asymptotic coverage probability of at least $1 - \alpha$ for θ_0 since

$$
\begin{aligned}
\lim_{n\to\infty} \Pr\left\{ n\hat{Q}_n(\theta_0) \leq C_\alpha^{b^*} \right\} &= 1 - \lim_{n\to\infty} \Pr\left\{ n\hat{Q}_n(\theta_0) \geq C_\alpha^{b^*} \right\} \\
&\geq 1 - \sup_{\theta\in\Theta^*} \lim_{n\to\infty} \Pr\left\{ n\hat{Q}_n(\theta) \geq C_\alpha^{b^*} \right\} \\
&\geq 1 - \frac{1}{2}\Pr\left\{ \chi^2_{b^*} \geq C_\alpha^{b^*} \right\} + \frac{1}{2}\Pr\left\{ \chi^2_{b^*-1} \geq C_\alpha^{b^*} \right\} = 1 - \alpha.
\end{aligned}
$$

The cutoff value $C_\alpha^{b^*}$ is trivial to compute using standard statistical software that can compute values of the chi-square CDF.

4.2 Computing conservative cutoff values when V^* is diagonal

When V_θ^* is a diagonal, then $w\left(b\left(\theta\right), j, V_\theta^*\right)$ only depends on $b\left(\theta\right)$ and j, but not V_θ^*. This is because the weights function depends only on the correlation matrix associated with V_θ^*. When all of the off diagonal elements of V_θ^* are zero, the weights function takes the simple form given by the following corollary. This result also provides a smaller cutoff value for the hypothesis test (2) and thus a smaller confidence region for θ_0 when V_θ^* is diagonal.

Corollary 3 *Let (A1)-(A5) hold. Suppose that V_θ^* is diagonal and that $\sup_{\theta\in\Theta^*} b\left(\theta\right) \leq b^*$. Then*

$$
w\left(b\left(\theta\right), j, V_\theta^*\right) = w\left(b\left(\theta\right), j, I_b\right) = 2^{-b(\theta)}\binom{b\left(\theta\right)}{j}, \tag{12}
$$

and $\forall c \geq 0$

$$
\sup_{\theta\in\Theta^*} \lim_{n\to\infty} \Pr\left\{ n\hat{Q}_n(\theta) \geq c \right\} \leq \sum_{j=0}^{b^*} 2^{-b^*}\binom{b^*}{j}\Pr\left\{ \chi_j^2 \geq c \right\}. \tag{13}
$$

Just as Corollary 2 provides a way to construct conservative confidence sets for θ_0 so does Corollary 3 when V^* is diagonal. By the same reasoning as in section 4.1, if $C_\alpha^{b^*}$ solves

$$
\sum_{j=0}^{b^*} 2^{-b^*}\binom{b^*}{j}\Pr\left\{ \chi_j^2 \geq C_\alpha^{b^*} \right\} = \alpha, \tag{14}
$$

then

$$
CR_{1-\alpha} = \left\{ \theta \in \Theta : n\hat{Q}_n(\theta) \leq C_\alpha^{b^*} \right\}
$$

has asymptotic coverage probability for θ_0 of at least $1 - \alpha$.

4.3 Computing Confidence Sets

In this subsection, I briefly outline the steps required to compute a confidence set $CR_{1-\alpha}$ for θ_0 with asymptotic coverage of at least $1 - \alpha$, when $\sup_{\theta\in\Theta^*} b\left(\theta\right) \leq b^*$ and assumptions (A1)-(A4)

hold.

1. Compute the unique value of $C_{1-\alpha}$ such that

$$\sup_{\theta \in \Theta^*} \sum_{j=0}^{b(\theta)} w\left(b\left(\theta\right), b\left(\theta\right) - j, V_\theta^*\right) \Pr\left\{\chi_j^2 \geq C_{1-\alpha}\right\} = \alpha.$$

- If V^* is diagonal, this is the value of $C_{1-\alpha}$ that solves

$$\sum_{j=0}^{b^*} 2^{-b^*} \binom{b^*}{j} \Pr\left\{\chi_j^2 \geq C_{1-\alpha}\right\} = \alpha.$$

- If V^* is not diagonal, this is the value of $C_{1-\alpha}$ that solves

$$\frac{1}{2} \Pr\left\{\chi_{b^*}^2 \geq C_{1-\alpha}\right\} + \frac{1}{2} \Pr\left\{\chi_{b^*-1}^2 \geq C_{1-\alpha}\right\} = \alpha.$$

2. Choose a fine grid G of candidate values of θ over the parameter space Θ^*. For each $\theta \in G$, compute $n\hat{Q}_n(\theta)$. If $n\hat{Q}_n(\theta) \leq C_{1-\alpha}$, then $\theta \in CR_{1-\alpha}$. If $n\hat{Q}_n(\theta) > C_{1-\alpha}$, then $\theta \notin CR_{1-\alpha}$.

Appropriate choice of grid values G depends on the particular application. How fine the grid should be depends on the desired level of precision for $CR_{1-\alpha}$. If Θ^* is known to be sufficiently regular (e.g. closed and convex), certain values of θ may be able to be included or discarded without explicitly evaluating $n\hat{Q}_n(\theta)$.

5   Examples

In this section I provide two specific examples of moment inequality models that have appeared previously in the literature. I demonstrate how to build confidence sets for model parameters, and I perform Monte Carlo simulations to evaluate the finite sample properties of the confidence sets in these two cases.

5.1   Example 1: Estimating the Mean of a Univariate Random Variable with Missing Data

Consider the setup of Imbens and Manski (2004): Let $\{(x_i, z_i) : i = 1, ...n\}$ be a random sample from a population of (x, z) pairs with support $[0, 1] \times \{0, 1\}$, where $z = 1$ indicates that x is observed, while if $z = 0$, x is not observed. The probability that x is observed, $p = \Pr\{z = 1\}$, is assumed to be less than one, and is not known to researcher, but is consistently estimated by its sample analog. The goal is inference on $\theta_0 \equiv \mathbb{E}[x]$. Let $\mu_1 = \mathbb{E}[x|z = 1]$, which is identified by the

sampling process. This model yields two moment inequalities:

$$\theta \geq \theta_L \equiv p \cdot \mu_1,$$
$$\theta \leq \theta_U \equiv p \cdot \mu_1 + 1 - p,$$

or, in the form of (1),

$$E\left[m_1\left(x, z, \theta\right)\right] = E\left[\theta - xz\right] \geq 0, \tag{15}$$
$$E\left[m_2\left(x, z, \theta\right)\right] = E\left[1 - z + xz - \theta\right] \geq 0.$$

The identified set for θ_0 in this model is

$$\Theta^* = \left[\theta_L, \theta_U\right],$$

and the variance of $m\left(x, z, \theta\right)$ is

$$V_\theta = V = var\left(-xz, xz - z\right) = \begin{pmatrix} \sigma_l^2 & \sigma_{lu} \\ \sigma_{lu} & \sigma_u^2 \end{pmatrix},$$

where

$$\sigma_l^2 = var\left(xz\right),$$
$$\sigma_u^2 = var\left(xz - z\right),$$

and

$$\sigma_{lu} = cov\left(xz, z\right) - var\left(xz\right).$$

$\hat{Q}_n\left(\theta\right)$ is given by

$$\hat{Q}_n\left(\theta\right) = \min_{t_1, t_2 \geq 0} \begin{pmatrix} \hat{E}_n\left[\theta - xz\right] - t_1 \\ \hat{E}_n\left[1 - z + xz - \theta\right] - t_2 \end{pmatrix}' \hat{V}^{-1} \begin{pmatrix} \hat{E}_n\left[\theta - xz\right] - t_1 \\ \hat{E}_n\left[1 - z + xz - \theta\right] - t_2 \end{pmatrix},$$

where $\hat{V}$ is the sample analog of V. In this case, the required assumptions are satisfied due to the observations being i.i.d., and the fact that x and z both have bounded support. Thus $m\left(x, z, \theta\right)$ must have finite expectation and variance for each θ that satisfies (15). Since $p < 1$, only at most one of $E\left[m_1\left(x, z, \theta\right)\right]$ or $E\left[m_2\left(x, z, \theta\right)\right]$ can be equal to zero. Thus, the maximum number of binding constraints is one, and V^* is just a number, and is therefore diagonal so that corollary 3 applies.[15] Applying this result, the cutoff value for $n\hat{Q}_n\left(\theta\right)$ needed to build a confidence set for

---

[15] In fact, because in this case the limit distribution of $n\hat{Q}_n\left(\theta\right)$ is a sum of only two terms, the weights are known exactly. Each of the two terms of the summation must have weight $\frac{1}{2}$.

θ_0 with at least $1 - \alpha$ asymptotic coverage is the unique value of c_α that solves

$$\frac{1}{2} \Pr \left\{ \chi_0^2 \geq c_\alpha \right\} + \frac{1}{2} \Pr \left\{ \chi_1^2 \geq c_\alpha \right\} = \alpha.$$

Since $c_\alpha > 0$, $\Pr \left\{ \chi_0^2 \geq c_\alpha \right\} = 0$, and this equation simplifies to

$$\frac{1}{2} \Pr \left\{ \chi_1^2 \geq c_\alpha \right\} = \alpha.$$

5.1.1 Simulations

I simulate iid draws of (x, z) in order to compare confidence regions constructed according to the moment inequality approach to those of Imbens and Manski (2004). The two approaches yield nearly identical results. Let the moment inequality confidence set of level α be denoted C_α^{MI}, for moment inequalities, and the Imbens/Manski confidence set C_α^{IM}. The sets C_α^{IM} are constructed as described in section 4 of their paper. Their sets have the added benefit that their coverage is uniform over all $\theta \in [p \cdot \mu_1, p \cdot \mu_1 + 1 - p]$.

I run simulations under two different specifications for the distribution of (x, z). For the first specification, I draw x from the uniform$(0, 1)$ distribution and z from the Bernoulli(p) distribution, independently of each other, inducing joint distribution F_1. Under this specification, x is missing completely at random. The second distribution, denoted F_2, is one in which (x, z) are not independent of each other, so that missingness is not at random. In this case, x is distributed beta$(4, 2)$ conditional on $z = 0$, and beta$(2, 4)$ when $z = 1$. In this case, x tends to be higher when it is not observed; the conditional distribution of x given $z = 0$ stochastically dominates that of x given $z = 0$, with $\mathbb{E}[x|z = 0] = 2/3$ and $\mathbb{E}[x|z = 1] = 1/3$. For each simulation, for the specified values of p and n, I draw a dataset from the specified population distribution of (x, z). The simulated sample data is then $\{(\tilde{x}_i, z_i) : i = 1, ..., n, \ \tilde{x}_i = x_i \text{ if } z_i = 1, \ \tilde{x}_i = \emptyset \text{ if } z_i = 0\}$. To evaluate the empirical coverage probability of the confidence regions, I compute the bounds for the population identification region $[\theta_L, \theta_U]$ and check to see if each of these points is contained in the two confidence regions. I keep track of how often these points are in the identification regions over many simulations. Formally, the procedure is as follows:

1. Specify the number of simulations to draw, R, the sample size for each simulation, n, p, and α.

2. Define $REJ_L^{IM}, REJ_U^{IM}, REJ_L^{MI}$, and REJ_U^{MI}, and set them all equal to 0. These variables will keep track of the number of times each of the two procedures reject $\theta_L \in \Theta^*$ and $\theta_U \in \Theta^*$.

3. Perform the following procedure R times.

   (a) Draw a random sample of $(\tilde{x}, z)$ of size n from the population.

17

(b) Compute C_α^{IM} and C_α^{MI}. C_α^{MI} is given by the set of θ that satisfy $n\hat{Q}_n(\theta) \leq c_\alpha$, where $\alpha = \frac{1}{2}\Pr\{\chi_1^2 \geq c_\alpha\}$ and χ_1^2 denotes a random variable with the chi-squared distribution with 1 degree of freedom.

    i. If $\theta_L \notin C_\alpha^{IM}$ increment REJ_L^{IM}, and if $\theta_U \notin C_\alpha^{IM}$ increment REJ_U^{IM}.

    ii. If $\theta_L \notin C_\alpha^{MI}$ increment REJ_L^{MI}, and if $\theta_U \notin C_\alpha^{MI}$ increment REJ_U^{MI}.

4. From the R simulations, compute $\widehat{CP}_\alpha^{IM} = \min\left\{\hat{P}\left(\theta_L \in C_\alpha^{IM}\right), \hat{P}\left(\theta_U \in C_\alpha^{IM}\right)\right\}$ and $\widehat{CP}_\alpha^{MI} = \min\left\{\hat{P}\left(\theta_L \in C_\alpha^{MI}\right), \hat{P}\left(\theta_U \in C_\alpha^{MI}\right)\right\}$. This is the observed probability with which the two confidence sets were guaranteed to cover θ_0.

Note that even though a particular value of θ_0 was used for the simulations, *any* value of θ_0 in the interval $[\theta_L, \theta_U]$ could generate the same distribution of observables for some data generation process consistent with the maintained modeling assumptions. Thus, a confidence set for the true underlying model parameter θ_0 must achieve the desired asymptotic coverage for *each* $\theta_0 \in [\theta_L, \theta_U]$. The procedure above measures the observed frequency with which this occurs because for either of the confidence sets CI_α,

$$\min_{\theta \in [\theta_L, \theta_U]} \hat{P}(\theta \in CI_\alpha) = \min_{\theta \in \{\theta_L, \theta_U\}} \hat{P}(\theta \in CI_\alpha),$$

since each type of confidence set is a closed interval.

Tables 1 and 2 compare the empirical coverage of each of the two confidence sets for different choices of n, p, α when $(x, z) \sim F_1$, while tables 3 and 4 do the same for $(x, z) \sim F_2$. The number of repetitions fixed at $R = 5000$ in all cases. For the results reported in Tables 1 and 3, $p = 0.7$, while for those in Tables 2 and 4, $p = 0.9$. The empirical coverage probabilities for both types of regions are very close to each other and approximate the desired target coverage probability rather well. The case where the observed coverage probabilities of the two types of differ most are those sets with nominal level 0.99. In this case, the coverage from moment inequality approach is always slightly less than the coverage of Imbens and Manski's confidence sets, though both are very close to the nominal level in all cases. As discussed in the subsequent section, this is an consequence of the modification that Imbens and Manski make in order to achieve uniform asymptotic coverage of all θ over Θ^*.

5.1.2   Analytical Comparison to Imbens and Manski (2004)

Straightforward examination of the boundaries of the confidence intervals obtained by the method of Imbens and Manski (2004) and the moment inequality method of this paper shows that the nearly identical empirical coverage probabilities of the two procedures is no accident. Indeed, as

Table 1: Observed coverage probabilities for p=0.7 when x is uniformly distributed on the unit interval and missing completely at random.

Target Coverage ($p = 0.7$)	0.75		0.85		0.95		0.99	
Actual Coverage for θ_0:	C^{IM}	C^{MI}	C^{IM}	C^{MI}	C^{IM}	C^{MI}	C^{IM}	C^{MI}
N								
100	0.7496	0.7496	0.8514	0.8514	0.9514	0.9514	0.9982	0.9888
500	0.7520	0.7520	0.8498	0.8498	0.9516	0.9514	0.9986	0.9896
1000	0.7514	0.7514	0.8516	0.8516	0.9504	0.9504	0.9978	0.9888

Table 2: Observed coverage probabilities for p=0.9 when x is uniformly distributed on the unit interval and missing completely at random.

Target Coverage ($p = 0.9$)	0.75		0.85		0.95		0.99	
Actual Coverage for θ_0:	C^{IM}	C^{MI}	C^{IM}	C^{MI}	C^{IM}	C^{MI}	C^{IM}	C^{MI}
N								
100	0.7540	0.7510	0.8554	0.8544	0.9498	0.9494	0.9956	0.9884
500	0.7492	0.7492	0.8484	0.8484	0.9460	0.9460	0.9974	0.9882
1000	0.7482	0.7482	0.8484	0.8484	0.9454	0.9454	0.9978	0.9906

Table 3: Observed coverage probabilities for p=0.7 when x|z=1 is distributed beta(2,4) and x|z=0 is distributed beta(4,2).

Target Coverage ($p = 0.7$)	0.75		0.85		0.95		0.99	
Actual Coverage for θ_0:	C^{IM}	C^{MI}	C^{IM}	C^{MI}	C^{IM}	C^{MI}	C^{IM}	C^{MI}
N								
100	0.7470	0.7470	0.8464	0.8464	0.9480	0.9480	0.9960	0.9854
500	0.7430	0.7430	0.8458	0.8458	0.9464	0.9464	0.9968	0.9882
1000	0.7474	0.7474	0.8502	0.8502	0.9484	0.9484	0.9972	0.9904

Table 4: Observed coverage probabilities for p=0.9 when x|z=1 is distributed beta(2,4) and x|z=0 is distributed beta(4,2).

Target Coverage ($p = 0.9$)	0.75		0.85		0.95		0.99	
Actual Coverage for θ_0:	C^{IM}	C^{MI}	C^{IM}	C^{MI}	C^{IM}	C^{MI}	C^{IM}	C^{MI}
N								
100	0.7352	0.7352	0.8296	0.8292	0.9346	0.9340	0.9916	0.9890
500	0.7566	0.7566	0.8488	0.8488	0.9452	0.9452	0.9978	0.9890
1000	0.7358	0.7358	0.8374	0.8374	0.9446	0.9446	0.9954	0.9878

the sample size increases, the boundaries of the two types of confidence intervals converge to one another in probability for any fixed confidence level.

The confidence intervals of Imbens and Manski have upper and lower CI bounds of

$$CI^{IM} = \left[L_n^{IM}, U_n^{IM} \right],$$

where $\bar{C}_n$ satisfies[16]

$$\Phi \left(\bar{C}_n + \sqrt{n} \frac{\hat{\theta}_U - \hat{\theta}_L}{\max \left(\hat{\sigma}_u, \hat{\sigma}_l \right)} \right) - \Phi \left(-\bar{C}_n \right) = 1 - \alpha, \tag{16}$$

and the boundaries of the region are given by

$$L_n^{IM} = \hat{\theta}_L - \bar{C}_n \cdot \hat{\sigma}_l / \sqrt{n},$$

and

$$U_n^{IM} = \hat{\theta}_U + \bar{C}_n \cdot \hat{\sigma}_u / \sqrt{n}.$$

When there is a strictly positive probability of missingness ($p < 1$), then $\sqrt{n} \left(\hat{\theta}_U - \hat{\theta}_L \right)$ tends to ∞ an $n \to \infty$. Thus, the term $\Phi \left(\bar{C}_n + \sqrt{n} \frac{\hat{\theta}_U - \hat{\theta}_L}{\max(\hat{\sigma}_u, \hat{\sigma}_l)} \right)$ of (16) converges to one for any finite $\bar{C}_n$, so that

$$\bar{C}_n \xrightarrow{p} z_{1-\alpha},$$

where $z_{1-\alpha}$ denotes the $1 - \alpha$ quantile of the standard normal distribution. This is the first step in showing the asymptotic equivalence of the two confidence sets, and is stated formally in the following proposition.

Proposition 2 *Let the assumptions of section 5.1 hold. Then*

$$\bar{C}_n \xrightarrow{p} z_{1-\alpha},$$

$$L_n^{IM} - \left(\hat{\theta}_L - z_{1-\alpha} \cdot \hat{\sigma}_l / \sqrt{n} \right) \xrightarrow{p} 0,$$

and

$$U_n^{IM} - \left(\hat{\theta}_U + z_{1-\alpha} \cdot \hat{\sigma}_u / \sqrt{n} \right) \xrightarrow{p} 0.$$

The confidence intervals proposed in this paper for Example 1 are given by

$$CI^{MI} = \left\{ \theta \in \Theta : nQ_n \left(\theta \right) \leq c_\alpha \right\},$$

---

[16] Imbens and Manski (2004) consider three different types of confidence intervals. The one used here is the only one given for the case where p is a parameter to be estimated, which is the case in Example 1. This confidence interval provides uniform asymptotic coverage over Θ^*, rather than just for the true θ_0.

where c_α solves

$$\frac{1}{2}\Pr\left\{\chi_1^2 \geq c_\alpha\right\} = \alpha. \tag{17}$$

Proposition 3 shows that this region corresponds to a closed interval with endpoints L_n^{MI} and U_n^{MI}, defined below. In conjunction with Proposition 2, it follows that $L_n^{IM} - L_n^{MI} \xrightarrow{p} 0$ and $U_n^{IM} - U_n^{MI} \xrightarrow{p} 0$, so that under the stated conditions, the two confidence intervals are asymptotically equivalent.

Proposition 3 *Let the assumptions of section 5.1 hold. Then*

$$CI^{MI} = \left[L_n^{MI}, U_n^{MI}\right],$$

where

$$L_n^{MI} = \hat{\theta}_L - z_{1-\alpha} \cdot \hat{\sigma}_l/\sqrt{n},$$

and

$$U_n^{MI} = \hat{\theta}_U + z_{1-\alpha} \cdot \hat{\sigma}_u/\sqrt{n}.$$

For the model studied in example 1, these propositions show that the two confidence sets are asymptotically identical. However, they differ in the excluded case of point identification (i.e. when there is no missing data). The first term of the IM inequality is included in order to guarantee uniform asymptotic coverage of θ_0 even as $p \to 1$, i.e. over all possible lengths for the identification region, including a point for the case where $p = 1$. The moment inequality approach requires that the researcher impose an upper bound on the number of binding constraints; when there is missing data, so that $p < 1$, then only one constraint may bind. In the case of no missing data, both constraints are satisfied with equality, so the correct upper bound on the number of binding constraints on this case is two, not one. When the maximal number of binding constraints is two rather than one, application of Proposition 1 gives a larger value for c_α, so as to ensure asymptotic coverage of at least $1 - \alpha$. The correction that Imbens and Manski use to guarantee uniform coverage in this model, and which can apply to either case, does not appear to have a straightforward analog to models with arbitrarily many moment inequalities.

The cost of Imbens and Manski's uniformity correction is a slightly larger confidence set for any fixed nominal level. This is straightforward from algebraic inspection of (16) and (17). The latter condition can be reformulated as $1 - \Phi(-c_\alpha) = 1 - \alpha$, by which it can be seen that relative to (16) the latter condition replaces the term $\Phi\left(\bar{C}_n + \sqrt{n}\frac{\hat{\theta}_U - \hat{\theta}_L}{\max(\hat{\sigma}_u, \hat{\sigma}_l)}\right)$ with 1. For fixed $\alpha \in (0,1)$, this requires that $\bar{C}_n > c_\alpha$, and in turn that $CI^{MI} \subset C^{IM}$. However, as n grows large, the difference between the two confidence intervals becomes negligible. For any sample size, the difference is greater for higher confidence levels due to the "flatness" of the normal CDF at its upper tail; in this case the difference in $\bar{C}_n$ in c_α is higher because Φ is nearly flat at its $1 - \alpha$ quantile. This

is why the coverage of the two types of confidence sets differs the most for $1 - \alpha = 0.99$ in the simulations.

5.2 Example 2: Mean Regression with Interval Outcome Data

In this subsection I consider one of the models studied by Manski and Tamer (2002) as another simple example. Let a random sample of size n of (y_1, y_0, x) be observed by the econometrician, where:[17]

$$
\begin{aligned}
y^* &= \beta_0 + \beta_1 x + u, \\
y_0 &= \text{floor}\,(y^*), \\
y_1 &= \text{ceil}\,(y^*).
\end{aligned}
$$

The econometrician does not observe y^*, but knows that $P\{y_0 \leq y^* \leq y_1\} = 1$, $P\{y_0 = y_1\} < 1$, and that $E[u|x] = 0$ and $E[u^2|x] < \infty$. The econometrician's goal is inference on the model parameters $\beta \equiv (\beta_0, \beta_1)$, and I use B^* to denote the identified set for β.[18] Thus the conditional moment restrictions

$$
\begin{aligned}
E\left[-y_0 + \beta_0 + \beta_1 x | x\right] &\geq 0, \\
E\left[y_1 - \beta_0 - \beta_1 x | x\right] &\geq 0,
\end{aligned}
$$

are satisfied for all $x \in \mathcal{X}$. If $\mathcal{X}$ is finite, then this yields a finite number of unconditional moment inequalities, two for every element of $\mathcal{X}$. The moments all have finite mean and variance because of the restrictions on u.

Suppose, for example, that $\mathcal{X} = \{1, 2\}$. Then (1) is

$$
\mathbb{E}\left[m\left(y_1, y_0, x, \beta\right)\right] = \begin{pmatrix} E\left[-y_0|x=1\right] + \beta_0 + \beta_1 \\ E\left[y_1|x=1\right] - \beta_0 - \beta_1 \\ E\left[-y_0|x=2\right] + \beta_0 + 2\beta_1 \\ E\left[y_1|x=2\right] - \beta_0 - 2\beta_1 \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{18}
$$

As in example 1, the variance of $m(y_1, y_0, x, \beta)$ does not depend on β, and can be consistently

---

[17] The functions floor and ceil round their argument down and up, respectively, to the nearest integer value.

[18] For this section, since the goal is inference on model parameters in a linear model, I use β to denote the parameter of interest rather than θ_0.

estimated by

$$\hat{V} = \begin{pmatrix} \hat{\sigma}_{01}^2 & -\hat{c}_1 & 0 & 0 \\ -\hat{c}_1 & \hat{\sigma}_{11}^2 & 0 & 0 \\ 0 & 0 & \hat{\sigma}_{02}^2 & -\hat{c}_2 \\ 0 & 0 & -\hat{c}_2 & \hat{\sigma}_{12}^2 \end{pmatrix},$$

where $\hat{\sigma}_{ij}^2 = \widehat{var}(y_i | x = j)$, and $\hat{c}_j = \widehat{cov}(y_1, y_0 | x = j)$. $\hat{Q}_n(\beta)$ is given by

$$\hat{Q}_n(\beta) = \min_{t \geq 0} (\mathbb{E}[m(y_1, y_0, x, \beta)] - t)' \hat{V}^{-1} (\mathbb{E}[m(y_1, y_0, x, \beta)] - t).$$

Furthermore, because $E[y_1 | x] > E[y_0 | x]$, at most only one of the first two components and one of the last two components of $\mathbb{E}[m(y_1, y_0, x, \beta)]$ can equal zero for any value of β. Thus, at most 2 of the inequalities can bind at any β, and the variance of the binding inequalities, V^* is diagonal. As a result, the method for constructing confidence sets when V^* is diagonal is applicable.

5.2.1 Simulations

In this section I simulate the model described above, i.e.

$$
\begin{aligned}
y^* &= \beta_0 + \beta_1 x + u, \\
y_0 &= \text{floor}(y), \\
y_1 &= \text{ceil}(y),
\end{aligned}
$$

where it is known by the econometrician that $E[u|x] = 0$, $E[u^2|x] < \infty$, and only a random sample of (y_0, y_1, x) are observed. The econometrician knows that $y^* \in [y_0, y_1]$, but does not observe y^*. In particular, but unbeknownst to the econometrician, the following parameter values and distributions comprise the data generation process:

- x takes the values 1 or 2, each with equal probability.

- u is distributed according to the standard normal distribution.

- x and u are iid and independent of each other.

- $(\beta_0, \beta_1) = (1, 1)$.

10,000 draws were made from this DGP, comprising the "population". Simulated data were then drawn as random samples from this population. The population identified set for $\beta = (\beta_0, \beta_1)$, B^*, is shown in Figure 1.

This is the set of values for β that are consistent with the distribution of (y_0, y_1, x) and the knowledge that $P\{y_0 \leq y^* \leq y_1\} = 1$ and $E[u|x] = 0$. Thus, for any value of β in this region,
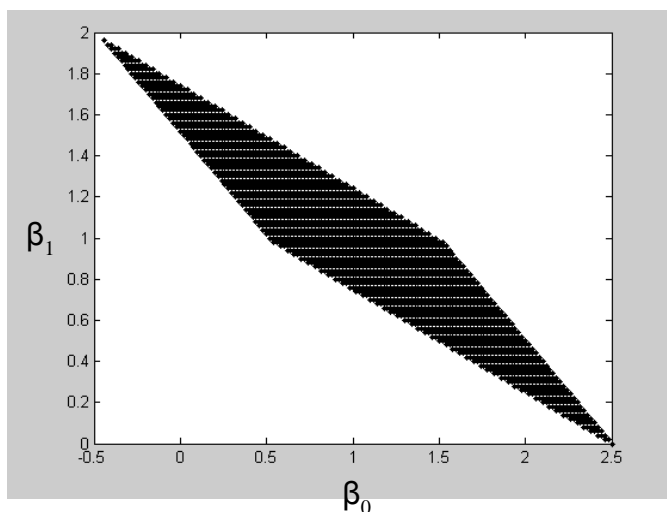
Figure 1: The identified set for (β_0, β_1) in model 2.

there is some distribution of x and u consistent with the maintained assumptions that yields the observed distribution of (y_0, y_1, x). Even though $\beta = (1, 1)$ in the simulations performed, any other value of β in this set could be used to obtain the simulated distribution of observables. Although the goal of my confidence regions is a pre-specified coverage level for the true β, the region must cover any fixed β in this set with at least the pre-specified probability, since they are all consistent with the distribution of observables and a priori knowledge. As discussed in the introduction, this is inherently a different goal than building a confidence region for the entire identified set.

With the simulated data in hand, the following procedure was used to evaluate the empirical coverage probability of nominal $1 - \alpha$ confidence regions for β constructed by computing the cutoff value for $n\hat{Q}_n(\beta)$ as described in section 4.2:

1. Specify the number of simulations to draw, R (in this case 5000), and the sample size for each simulation, n.

2. Perform the following procedure R times.

   (a) Draw a random sample of (y_0, y_1, x) of size n from the population.

   (b) For each $\theta \in \Theta^*$ compute $\hat{Q}_n(\theta)$.

   (c) If $n\hat{Q}_n(\theta) > c_\alpha$, reject the null hypothesis that $\beta \in B^*$, where c_α is the unique value that satisfies
   $$\frac{1}{2}\Pr\left\{\chi_1^2 \geq c_\alpha\right\} + \frac{1}{4}\Pr\left\{\chi_2^2 \geq c_\alpha\right\} = \alpha.$$

24

Nominal Coverage:	0.75	0.85	0.95	0.99
Actual Coverage for $n =$				
100	0.7544	0.8492	0.9460	0.9868
500	0.7430	0.8436	0.9490	0.9884
1000	0.7452	0.8438	0.9458	0.9870
2000	0.7516	0.8494	0.9452	0.9900

Table 5: Coverage Probabilities for confidence regions based on pointwise testing with 5000 repetitions.

This corresponds to weights for a 2×2 diagonal variance covariance matrix given by equation (12) from corollary 3.

3. For each $\beta \in B^*$, compute the fraction of simulations for which H_0 was not rejected, denoted $C_\alpha(\beta)$. Because any $\beta \in B^*$ can generate the observed distribution of observables, the coverage probability for β is $P(\beta \in CR_{1-\alpha}) \geq \inf_{\beta \in \beta^*} C_\alpha(\beta) \equiv \hat{C}_{1-\alpha}$, where $CR_{1-\alpha}$ is the $1 - \alpha$ level confidence region for β. $\hat{C}_{1-\alpha}$ is the observed probability with which $CR_{1-\alpha}$ was guaranteed to contain the true β in these simulations.

If the confidence sets are asymptotically valid, it should be that $\hat{C}_{1-\alpha} \xrightarrow{p} c$ as $N \to \infty$ for some $c \geq 1 - \alpha$. Table 5 shows empirical coverage probabilities obtained from the above procedure for various pre-specified values of n and α, and for a value of $R = 5000$. In all cases, the observed coverage probability was very close to the nominal level, with the greatest difference being 0.007. Interestingly, it appears that as n increased, the observed coverage probability approached the nominal confidence level from below.

Finally, Figures 3 and 4 show two examples of confidence regions for β taken to be the set of $\beta \in B$ that are not rejected by the pointwise testing procedure at the 0.95 level for $n = 100$ and $n = 1000$. The figures illustrate how, as one would expect, the size of the confidence set shrinks as n increases for a fixed coverage probability.

6   Conclusion

The confidence sets of this paper are guaranteed to provide a pre-specified level of asymptotic coverage for a parameter of interest in models that consist of a finite number of moment inequalities. Many models in this class have appeared in the literature, and these models comprise a large subset of models with partially identified parameters. The method for constructing confidence sets is conservative, but easy to implement, as no bootstrapping, subsampling, or simulation is required. Despite their conservative nature, the confidence sets performed well in the Monte Carlo experiments conducted.

Figure 2: A representative 0.95 coverage region for (β_0, β_1) based on a sample of 100 observations. The black region denotes the population identified set, while the encompassing blue/shaded region is the 0.95 confidence set for the parameter $\beta = (\beta_0, \beta_1)$.
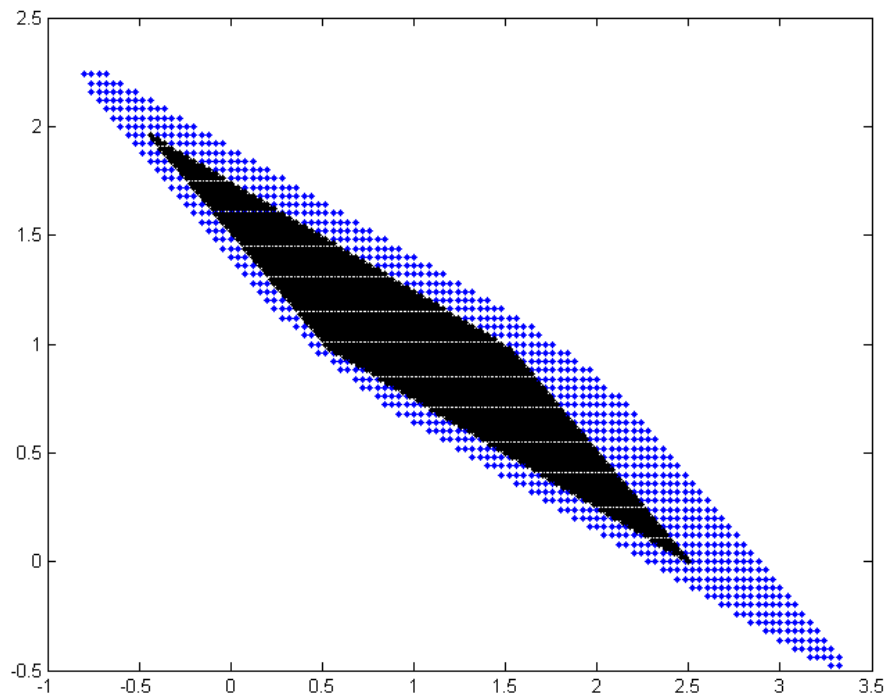
Figure 3: A representative 0.95 coverage region for (β_0, β_1) based on a sample of 1000 observations. The black region denotes the population identified set, while the encompassing blue/shaded region is the 0.95 confidence set for the parameter $\beta = (\beta_0, \beta_1)$.
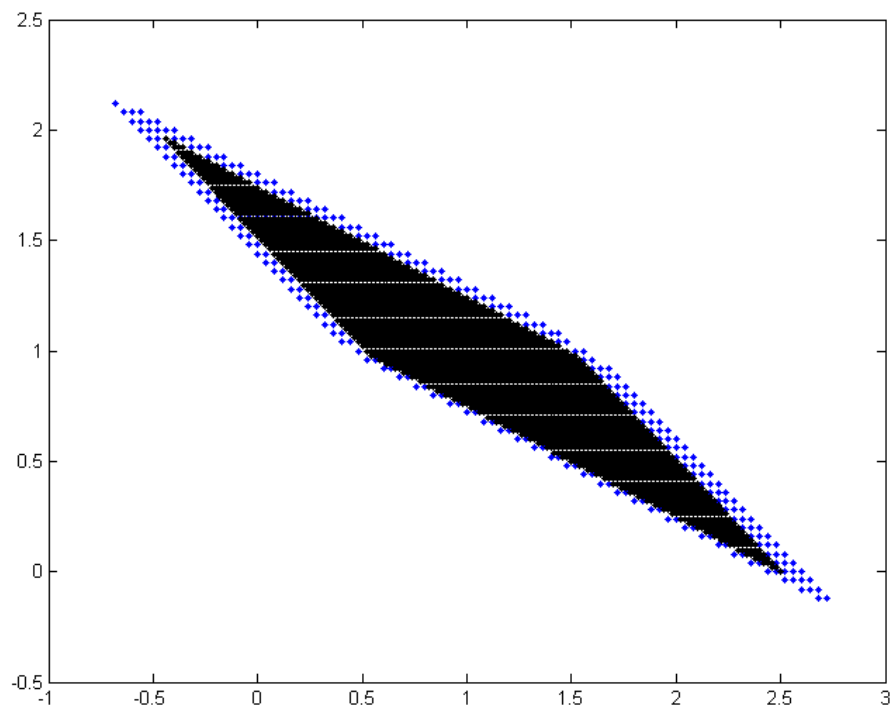
The findings of this paper have naturally lead to some avenues for further research. First, the cutoff values for the test statistic $n\hat{Q}_n(\theta)$ are computed by making use of an upper bound on the feasible number of moments that bind at θ. This provides a worst case for the values of the weights function of the asymptotic chi-bar-square distribution of $n\hat{Q}_n(\theta)$. If the true weights for the asymptotic distribution of $n\hat{Q}_n(\theta)$ can be consistently estimated, then a smaller cutoff value for $n\hat{Q}_n(\theta)$ could possibly be estimated for any size test. If this could be done, the conservative nature of the confidence sets could potentially be alleviated. However, such an approach would likely not be without computational cost, since a different cutoff would need to be computed for each value of θ.

Furthermore, this paper focuses on building confidence sets for just the parameter of interest θ_0. There have been two other types of confidence sets that have appeared in the literature on partially identified parameters, as discussed in section 2. Which type is appropriate depends on the context and the researcher's goal in any particular application. It would be of interest to determine whether the testing procedure of this paper could be modified to construct confidence sets with uniform asymptotic coverage over the identified set Θ^*, or confidence sets for Θ^* itself.

Appendix A: The Boundary of $\mathbb{E}[m(y, x, \theta)]$ in $\mathbb{R}^J_+$ and the Boundary of Θ^*

An implication of Proposition 1 is that the asymptotic distribution of $n\hat{Q}_n(\theta)$ is degenerate when $\mathbb{E}[m(y, x, \theta)] > 0$, converging to zero in probability. Put another way, $n\hat{Q}_n(\theta)$ only has a non-degenerate limiting distribution when $\mathbb{E}[m(y, x, \theta)]$ lies on the boundary of $\mathbb{R}^J_+$, the nonnegative orthant in J dimensional Euclidean space. This section examines the relationship between the boundary of $\mathbb{E}[m(y, x, \theta)]$ in $\mathbb{R}^J_+$ and the boundary of the identified set Θ^*. Toward this end, let

$$D\Theta^* \equiv \{\theta \in \Theta^* : \mathbb{E}[m_j(y, x, \theta)] = 0 \text{ for at least one } j \in \{1, ..., J\}\}$$

be the set of $\theta \in \Theta^*$ such that $\mathbb{E}[m(y, x, \theta)]$ lies on the boundary of $\mathbb{R}^J_+$. Let

$$\partial\Theta^* \equiv \left\{\theta \in \Theta^* : \text{for every open neighborhood of } \theta, N_\theta \subseteq \mathbb{R}^k, N_\theta \nsubseteq \Theta^*\right\}$$

be the boundary of Θ^* in Θ. In order to characterize the relationship between these two sets, I consider the implications of the following two assumptions.

Assumption 6 *(A6: continuity)* $\mathbb{E}[m(y, x, \theta)]$ *is continuous in* θ.

Assumption 7 *(A7: monotonicity)* $\forall j = 1, ..., J$, $\mathbb{E}[m_j(y, x, \theta)]$ *is strictly monotone in at least one component of* θ.

28

First, it is easy to see that if $\mathbb{E}\left[m\left(y,x,\theta\right)\right]$ is not continuous in θ, $\partial\Theta^*$ need not be contained in $D\Theta^*$. This is because if $\mathbb{E}\left[m\left(y,x,\cdot\right)\right]$ has jump discontinuities, it is possible that $\mathbb{E}\left[m\left(y,x,\theta\right)\right] > 0$ but that there exists an arbitrarily small ϵ in $\mathbb{R}^k$ such that $\mathbb{E}\left[m\left(y,x,\theta+\epsilon\right)\right] < 0$, i.e. $\mathbb{E}\left[m\left(y,x,\theta\right)\right]$ "jumps" from the interior of $\mathbb{R}_+^J$ to the exterior of $\mathbb{R}_+^J$ at θ. Proposition 4 shows that the contrapositive is in fact true; if $\mathbb{E}\left[m\left(y,x,\theta\right)\right]$ is continuous in θ, then $\partial\Theta^* \subseteq D\Theta^*$. In turn, this implies that if assumption (A5) holds, the asymptotic distribution of $n\hat{Q}_n\left(\theta\right)$ is degenerate at 0 on the interior of Θ^*. Proposition 3 proceeds to show that when combined with continuity, the monotonicity requirement of assumption (A6) is sufficient to conclude that $\partial\Theta^*$ and $D\Theta^*$ are equal. In the absence of monotonicity, Continuity alone is not enough for for the two sets to be equivalent.

Proposition 4 *Let assumptions (A1)-(A3) as well as (A6) hold. Then $\partial\Theta^* \subseteq D\Theta^*$.*

Proposition 5 *Let (A1)-(A3),(A6), and (A7) hold. Then $\partial\Theta^* = D\Theta^*$.*

So far the analysis has centered around the boundary of Θ^*, which is the boundary of the null hypothesis in (2). The hypothesis test can be recast however as

$$
\begin{aligned}
H_0 &: \quad Q\left(\theta\right) = 0 \\
H_1 &: \quad Q\left(\theta\right) > 0.
\end{aligned}
$$

Because $Q\left(\theta\right) = 0$ if and only if $\mathbb{E}\left[m\left(y,x,\theta\right)\right] \geq 0$, and $Q\left(\theta\right)$ is nonnegative, this is exactly the same null and alternative. Written this way, the hypothesis test has the property that $Q\left(\theta\right)$ is on the boundary of the maintained hypothesis $Q\left(\theta\right) \geq 0$. In related work, Andrews (2001) studies the problem of hypothesis testing when a parameter is on the boundary of the maintained hypothesis.

Appendix B: Proofs

As a preliminary step to proposition 1, I first prove the following Lemma.

6.1 Lemma

Consider the minimization problem

$$
QP = \min\left(\mathbf{x}-\mathbf{t}\right)' V^{-1}\left(\mathbf{x}-\mathbf{t}\right) \text{ s.t. } \mathbf{t}_1 \geq 0, \tag{19}
$$

where $x, t \in \mathbb{R}^J$, and $x_1, t_1 \in \mathbb{R}^b$, $b \leq J$, s.t. $t = \left(\mathbf{t}_1', \mathbf{t}_2'\right)'$ and $x = \left(\mathbf{x}_1', \mathbf{x}_2'\right)'$. Let V_{11} be the $b \times b$ leading submatrix of V so that

$$
V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}.
$$

Then
$$QP = \min \left(\mathbf{x}_1 - \mathbf{t}_1 \right)' V_{11}^{-1} \left(\mathbf{x}_1 - \mathbf{t}_1 \right) \text{ s.t. } \mathbf{t}_1 \geq 0. \tag{20}$$

Proof. Let $\Lambda \equiv V^{-1}$ and partition Λ so that

$$\Lambda = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix},$$

where Λ_{11} is $b \times b$ and Λ_{22} is $J - b. \times J - b$. Let t^* be the value of t that solves QP, so that

$$QP = \left(\mathbf{x} - \mathbf{t}^* \right)' \Lambda \left(\mathbf{x} - \mathbf{t}^* \right).$$

The Kuhn-Tucker conditions for (19) are

(i) For $j = 1, ..., b$, Either $t_j^* = 0$ and $\left[-\Lambda \left(\mathbf{x} - \mathbf{t}^* \right) \right]_j \geq 0$, or $t_j^* > 0$ and $\left[-\Lambda \left(\mathbf{x} - \mathbf{t}^* \right) \right]_j = 0$.

(ii) For $j = b + 1, ..., J$, $\left[-\Lambda \left(\mathbf{x} - \mathbf{t}^* \right) \right]_j = 0$.

By conditions (i) and (ii),

$$-\Lambda_{11} \left(\mathbf{x}_1 - \mathbf{t}_1^* \right) - \Lambda_{12} \left(\mathbf{x}_2 - \mathbf{t}_2^* \right) \geq 0, \tag{21}$$
$$-\Lambda_{21} \left(\mathbf{x}_1 - \mathbf{t}_1^* \right) - \Lambda_{22} \left(\mathbf{x}_2 - \mathbf{t}_2^* \right) = 0. \tag{22}$$

Solving for $\left(\mathbf{x}_2 - \mathbf{t}_2^* \right)$, the latter condition is

$$\left(\mathbf{x}_2 - \mathbf{t}_2^* \right) = -\Lambda_{22}^{-1} \Lambda_{21} \left(\mathbf{x}_1 - \mathbf{t}_1^* \right). \tag{23}$$

Now

$$\begin{aligned} QP &= \left(\mathbf{x} - \mathbf{t}^* \right)' \Lambda \left(\mathbf{x} - \mathbf{t}^* \right) \\ &= \left(\mathbf{x}_1 - \mathbf{t}_1^* \right)' \Lambda_{11} \left(\mathbf{x}_1 - \mathbf{t}_1^* \right) + \left(\mathbf{x}_1 - \mathbf{t}_1^* \right)' \Lambda_{12} \left(\mathbf{x}_2 - \mathbf{t}_2^* \right) + \left(\mathbf{x}_2 - \mathbf{t}_2^* \right) \left[\Lambda_{21} \left(\mathbf{x}_1 - \mathbf{t}_1^* \right) + \Lambda_{22} \left(\mathbf{x}_2 - \mathbf{t}_2^* \right) \right] \\ &= \left(\mathbf{x}_1 - \mathbf{t}_1^* \right)' \Lambda_{11} \left(\mathbf{x}_1 - \mathbf{t}_1^* \right) + \left(\mathbf{x}_1 - \mathbf{t}_1^* \right)' \Lambda_{12} \left(\mathbf{x}_2 - \mathbf{t}_2^* \right), \end{aligned}$$

by (22). Now using (23) it follows that

$$\begin{aligned} QP &= \left(\mathbf{x}_1 - \mathbf{t}_1^* \right)' \Lambda_{11} \left(\mathbf{x}_1 - \mathbf{t}_1^* \right) - \left(\mathbf{x}_1 - \mathbf{t}_1^* \right)' \Lambda_{12} \left[\Lambda_{22}^{-1} \Lambda_{21} \left(\mathbf{x}_1 - \mathbf{t}_1^* \right) \right] \\ &= \left(\mathbf{x}_1 - \mathbf{t}_1^* \right)' \left[\Lambda_{11} - \Lambda_{12} \Lambda_{22}^{-1} \Lambda_{21} \right] \left(\mathbf{x}_1 - \mathbf{t}_1^* \right) \\ &= \left(\mathbf{x}_1 - \mathbf{t}_1^* \right)' V_{11}^{-1} \left(\mathbf{x}_1 - \mathbf{t}_1^* \right), \end{aligned}$$

where the last equality follows by the partition inverse result.[19]   All that remains is to show that t_1^* minimizes (20): $\min \left(\mathbf{x}_1 - \mathbf{t}_1\right)' V_{11}^{-1} \left(\mathbf{x}_1 - \mathbf{t}_1\right)$ s.t. $t_1 \geq 0$, but this follows from the Kuhn-Tucker minimization condition (i) as shown below:

The Kuhn-Tucker conditions for t_1^* that solves (20) are for $j = 1, ..., b$,

$$\text{either } \mathbf{t}_j^* = 0 \text{ and } \left[-V_{11}^{-1}\left(\mathbf{x}_1 - \mathbf{t}_1^*\right)\right]_j \geq 0, \text{ or } \mathbf{t}_j^* > 0 \text{ and } \left[-V_{11}^{-1}\left(\mathbf{x}_1 - \mathbf{t}_1^*\right)\right]_j = 0.$$

$\Longleftrightarrow$

$$\begin{aligned}
\text{either } \mathbf{t}_j^* &= 0 \text{ and } \left\{-\left[\Lambda_{11} - \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21}\right]\left(\mathbf{x}_1 - \mathbf{t}_1^*\right)\right\}_j \geq 0, \\
\text{or } \mathbf{t}_j^* &> 0 \text{ and } \left\{-\left[\Lambda_{11} - \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21}\right]\left(\mathbf{x}_1 - \mathbf{t}_1^*\right)\right\}_j = 0.
\end{aligned}$$

$\Leftrightarrow$

$$\begin{aligned}
\text{either } \mathbf{t}_j^* &= 0 \text{ and } \left\{-\left[\Lambda_{11}\left(\mathbf{x}_1 - \mathbf{t}_1^*\right) - \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21}\left(\mathbf{x}_1 - \mathbf{t}_1^*\right)\right]\right\}_j \geq 0, \\
\text{or } \mathbf{t}_j^* &> 0 \text{ and } \left\{-\left[\Lambda_{11}\left(\mathbf{x}_1 - \mathbf{t}_1^*\right) - \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21}\left(\mathbf{x}_1 - \mathbf{t}_1^*\right)\right]\right\}_j = 0
\end{aligned}$$

$\Leftrightarrow$

$$\begin{aligned}
\text{either } \mathbf{t}_j^* &= 0 \text{ and } \left\{-\left[\Lambda_{11}\left(\mathbf{x}_1 - \mathbf{t}_1^*\right) + \Lambda_{12}\left(\mathbf{x}_2 - \mathbf{t}_2^*\right)\right]\right\}_j \geq 0, \\
\text{or } \mathbf{t}_j^* &> 0 \text{ and } \left\{-\left[\Lambda_{11}\left(\mathbf{x}_1 - \mathbf{t}_1^*\right) + \Lambda_{12}\left(\mathbf{x}_2 - \mathbf{t}_2^*\right)\right]\right\}_j = 0
\end{aligned}$$

by (23), but this is exactly condition (i) from the Kuhn-Tucker conditions for the initial program (19). ∎

With Lemma 6.1 in hand, I now prove Proposition 1.

6.2   Proposition 1

Proof. Let

$$v_n \equiv \sqrt{n}\left(\hat{E}_n\left[m\left(y, x, \theta\right)\right] - \mathbb{E}\left[m\left(y, x, \theta\right)\right]\right),$$

and

$$v_n^* \equiv \sqrt{n}\left(\hat{E}_n\left[m^*\left(y, x, \theta\right)\right] - \mathbb{E}\left[m^*\left(y, x, \theta\right)\right]\right).$$

---

[19]If $V = \Lambda^{-1}$ then $V_{11} = \left(\Lambda_{11} - \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21}\right)^{-1}$.

Then

$$
\begin{aligned}
n\hat{Q}_n\left(\theta_0\right) &= \min_{t \geq 0} n \cdot \left[\hat{E}_n\left[m\left(y,x,\theta\right)\right] - t\right]' \hat{V}_\theta^{-1} \left[\hat{E}_n\left[m\left(y,x,\theta\right)\right] - t\right] \\
&= \min_{t \geq 0} \left[v_n + \sqrt{n}\left(\mathbb{E}\left[m\left(y,x,\theta\right)\right] - t\right)\right]' \hat{V}_\theta^{-1} \left[v_n + \sqrt{n}\left(\mathbb{E}\left[m\left(y,x,\theta\right)\right] - t\right)\right] \\
&= \min_{t \geq 0} \left[v_n + \sqrt{n}\mathbb{E}\left[m\left(y,x,\theta\right)\right] - t\right]' \hat{V}_\theta^{-1} \left[v_n + \sqrt{n}\mathbb{E}\left[m\left(y,x,\theta\right)\right] - t\right] \\
&= \min_s \ \left[v_n\left(\theta\right) - s\right]' \hat{V}_\theta^{-1} \left[v_n\left(\theta\right) - s\right] \text{ subject to } \ s = t - \sqrt{n}\mathbb{E}\left[m\left(y,x,\theta\right)\right], \ t \geq 0 \\
&= \min_s \ \left[v_n\left(\theta\right) - s\right]' \hat{V}_\theta^{-1} \left[v_n\left(\theta\right) - s\right] : \ \ s \geq -\sqrt{n}\mathbb{E}\left[m\left(y,x,\theta\right)\right].
\end{aligned}
$$

Partition s such that $s = \left(s_b', s_c'\right)'$, so that s_b are the first b elements of s, corresponding to those inequalities that bind, and s_c the remainder. Furthermore, let $\tilde{m}\left(y,x,\theta\right) = \left(m_{b+1}\left(y,x,\theta\right),...,m_J\left(y,x,\theta\right)\right)'$. Then because $E\left[m_j\left(y,x,\theta\right)\right] = 0$ for $j \leq b$,

$$
n\hat{Q}_n\left(\theta\right) = \min_s \ \left[v_n\left(\theta\right) - s\right]' \hat{V}_\theta^{-1} \left[v_n\left(\theta\right) - s\right] : \ \ s_b \geq 0, \ s_c \geq -\sqrt{n}\mathbb{E}\left[\tilde{m}\left(y,x,\theta\right)\right].
$$

Because $\sqrt{n}E\left[\tilde{m}\left(y,x,\theta\right)\right] \to \infty$ as $n \to \infty$, and $\hat{V}_\theta \xrightarrow{p} V_\theta$, it follows by a Slutsky Theorem that

$$
n\hat{Q}_n\left(\theta\right) \xrightarrow{p} \min_s \ \left[v_n - s\right]' V_\theta^{-1} \left[v_n - s\right] : \ \ s_b \in \mathbb{R}_+^b, \ s_c \in \mathbb{R}^{J-b},
$$

and by Lemma 6.1,

$$
\min_s \ \left[v_n - s\right]' V_\theta^{-1} \left[v_n - s\right] \text{ s.t. } s_b \in \mathbb{R}_+^b, \ s_c \in \mathbb{R}^{J-b} = \min_{s \in \mathbb{R}_+^b} \ \left[v^* - s\right]' V_\theta^{*-1} \left[v^* - s\right],
$$

where $v^* \sim N\left(0, V_\theta^*\right)$ by (9) which holds under (A1) and (A4). Thus

$$
n\hat{Q}_n\left(\theta\right) \xrightarrow{p} \min_{s \in \mathbb{R}_+^{b(\theta)}} \ \left[v^* - s\right]' V_\theta^{*-1} \left[v^* - s\right].
$$

The statistic $\min_{s \in \mathbb{R}_+^b} \left[v^* - s\right]' V_\theta^{*-1} \left[v^* - s\right]$ measures the distance of the normal random variable v^* from the nonnegative orthant. By Wolak (1991)

$$
\Pr\left\{\min_{s \in \mathbb{R}_+^b} \left[v^* - s\right]' V_\theta^{*-1} \left[v^* - s\right] \geq c\right\} = \sum_{j=0}^{b(\theta)} w\left(b\left(\theta\right), b\left(\theta\right) - j, V_\theta^*\right) \Pr\left\{\chi_j^2 \geq c\right\}.
$$

∎

6.2.1 Corollary 2

Proof.

$$\sup_{\theta \in \Theta^*} \lim_{n \to \infty} \Pr\left\{n\hat{Q}_n(\theta) \geq c\right\} = \sup_{\theta \in \Theta^*}\left(\sum_{j=0}^{b(\theta)} w\left(b(\theta), b(\theta) - j, V_\theta^*\right)\Pr\left\{\chi_j^2 \geq c\right\}\right)$$

$$\leq \frac{1}{2}\Pr\left\{\chi_{b^*}^2 \geq c\right\} + \frac{1}{2}\Pr\left\{\chi_{b^*-1}^2 \geq c\right\},$$

where the equality of the first line follows from Proposition 1. The rest of the proof follows from Sen and Silvapulle (2004, pp. 80-82), but I repeat the argument here for clarity. The inequality follows because

$$0 \leq w\left(b, j, V_\theta^*\right) \leq 1/2,$$

$$\sum_{j=0}^{b} w\left(b, j, V_\theta^*\right) = 1,$$

and

$$\Pr\left\{\chi_j^2 \geq c\right\} \text{ is increasing in } j \text{ for all } c.$$

■

6.2.2 Corollary 3

Proof. The first part, (12), follows from Wolak (1987) who derives the result for $V^* = \sigma^2 I$, and from Sen and Silvapulle (2004, Proposition 3.6.1 (11)). The latter result is that the weights function only depends on the variance through its associated correlation matrix. If V^* is diagonal, the correlation matrix is the identify matrix, so that $w(b, j, V^*) = w(b, j, I_b)$. The second part, (12), follows from the fact that $\sum_{j=0}^{b} 2^{-b}\binom{b}{j}\Pr\left\{\chi_j^2 \geq c\right\}$ is monotonically increasing in b, so that

$$\sup_{\theta \in \Theta^*} \lim_{n \to \infty} \Pr\left\{n\hat{Q}_n(\theta) \geq c\right\} = \sup_{\theta \in \Theta^*}\sum_{j=0}^{b(\theta)} w\left(b(\theta), b(\theta) - j, V_\theta^*\right)\Pr\left\{\chi_j^2 \geq c\right\}$$

$$= \sup_{\theta \in \Theta^*}\sum_{j=0}^{b(\theta)} 2^{-b(\theta)}\binom{b(\theta)}{j}\Pr\left\{\chi_j^2 \geq c\right\}$$

$$\leq \sum_{j=0}^{b^*} 2^{-b^*}\binom{b^*}{j}\Pr\left\{\chi_j^2 \geq c\right\}.$$

■

6.3 Proposition 2

Proof. $\bar{C}_n$ is implicitly defined by equation (16):

$$\Phi\left(\bar{C}_n + \sqrt{n}\frac{\hat{\theta}_U - \hat{\theta}_L}{\max\left(\hat{\sigma}_u, \hat{\sigma}_l\right)}\right) - \Phi\left(-\bar{C}_n\right) = 1 - \alpha,$$

where Φ is the cumulative distribution function for the standard normal distribution. Since this function is continuous and monotone increasing, it follows that the left hand side of this equation is also continuous and monotone increasing in $\bar{C}_n$, so that $\bar{C}_n$ is uniquely defined as a function of the sample and α. Furthermore, $\bar{C}_n$ is a continuous function of $\sqrt{n}\frac{\hat{\theta}_U - \hat{\theta}_L}{\max(\hat{\sigma}_u, \hat{\sigma}_l)}$. Because for any finite $\bar{C}_n$, $\Phi\left(\bar{C}_n + \sqrt{n}\frac{\hat{\theta}_U - \hat{\theta}_L}{\max(\hat{\sigma}_u, \hat{\sigma}_l)}\right) \to 1$ in probability as $n \to \infty$ for $p < 1$, continuity of $\bar{C}_n$ as a function of $\sqrt{n}\frac{\hat{\theta}_U - \hat{\theta}_L}{\max(\hat{\sigma}_u, \hat{\sigma}_l)}$ along with a Slutsky Theorem implies that $\Phi\left(-\bar{C}_n\right) \xrightarrow{p} \alpha$, and the claims of the proposition follow directly. ∎

6.4 Proposition 3

Proof. By definition, the $1 - \alpha$ confidence set CI^{MI} for θ_0 is given by

$$CI^{MI} = \left\{\theta \in \Theta : nQ_n\left(\theta\right) \leq c_\alpha\right\},$$

where

$$\frac{1}{2}\Pr\left\{\chi_1^2 \geq c_\alpha\right\} = \alpha.$$

Simplifying yields

$$\sqrt{c_\alpha} = \Phi^{-1}\left(1 - \alpha\right),$$

and

$$nQ_n\left(\theta\right) = \frac{n}{\hat{\sigma}_l\hat{\sigma}_u - \hat{\sigma}_{lu}^2}\min_{t_l, t_u \geq 0}\left\{\begin{array}{c}\hat{\sigma}_u \cdot \left(\hat{\Delta}_l\left(\theta\right) - t_l\right)^2 + \hat{\sigma}_l \cdot \left(\hat{\Delta}_u\left(\theta\right) - t_u\right)^2 \\ -2\hat{\sigma}_{lu} \cdot \left(\hat{\Delta}_l\left(\theta\right) - t_l\right)\left(\hat{\Delta}_u\left(\theta\right) - t_u\right)\end{array}\right\}. \tag{24}$$

Clearly, $\left[\hat{\theta}_L, \hat{\theta}_U\right] \subseteq CI^{MI}$, since for any θ on this interval, $Q_n\left(\theta\right) = 0$. The cases of interest are thus those where $\theta \leq \theta_L$ $(\hat{\Delta}_l\left(\theta\right) < 0)$ and $\theta \geq \theta_U$ $(\hat{\Delta}_u\left(\theta\right) < 0)$, which are mutually exclusive. To prove the result, I consider each of these two cases separately, and use the Kuhn-Tucker conditions for (24), which are a special case of the problem analyzed by Kudo (1963). These conditions are that for $j = l, u$ either

$$\left\{t_j = 0 \text{ and } \frac{\partial Q_n}{\partial t_j}\left(\theta\right) \geq 0\right\} \text{ or } \left\{t_j \geq 0 \text{ and } \frac{\partial Q_n}{\partial t_j}\left(\theta\right) = 0\right\}, \tag{KT}$$

where

$$\frac{\partial Q_n}{\partial t_l}(\theta) = -2\hat{\sigma}_u^2\left(\hat{\Delta}_l(\theta) - t_l\right) + 2\hat{\sigma}_{lu}\left(\hat{\Delta}_u(\theta) - t_u\right),$$

and

$$\frac{\partial Q_n}{\partial t_u}(\theta) = -2\hat{\sigma}_l^2\left(\hat{\Delta}_u(\theta) - t_u\right) + 2\hat{\sigma}_{lu}\left(\hat{\Delta}_l(\theta) - t_l\right).$$

∎

Case 1 ($\Delta_l < 0$) *In this case the Kuhn-Tucker conditions yield that the minimizing values of (t_l, t_u) are $t_l^*(\theta) = 0$ and $t_u^*(\theta) = \hat{\Delta}_u(\theta) - \left(\hat{\sigma}_{lu}/\hat{\sigma}_l^2\right)\cdot\hat{\Delta}_l(\theta)$. That $t_u^*(\theta) \geq 0$ can be verified by using the conditions that $\hat{\Delta}_u(\theta) > 0$ and $\hat{\Delta}_u(\theta) > -\hat{\Delta}_l(\theta)$, which are implications of $\hat{\Delta}_l(\theta) < 0$ and $\theta_U > \theta_L$. Plugging in $t_l^*(\theta)$ and $t_u^*(\theta)$ and simplifying yields*

$$nQ_n(\theta) = \frac{n}{\hat{\sigma}_l^2}\cdot\left(\theta - \hat{\theta}_L\right)^2,$$

so that $\theta \in CI^{MI}$ iff

$$\frac{n}{\hat{\sigma}_l^2}\cdot\left(\theta - \hat{\theta}_L\right)^2 \leq c_\alpha.$$

Since the case where $\theta - \theta_L < 0$ is being considered, this holds iff

$$-\sqrt{c_\alpha} \leq \frac{\sqrt{n}}{\hat{\sigma}_l}\cdot\left(\theta - \hat{\theta}_L\right),$$

or equivalently,

$$L_n^{MI} \leq \theta,$$

where

$$L_n^{MI} \equiv \hat{\theta}_L - \frac{\hat{\sigma}_l}{\sqrt{n}}\Phi^{-1}(1 - \alpha).$$

Case 2 ($\Delta_u < 0$) *The Kuhn-Tucker conditions yield that the minimizing values of (t_l, t_u) are $t_u^*(\theta) = 0$ and $t_l^*(\theta) = \hat{\Delta}_l(\theta) - \left(\hat{\sigma}_{lu}/\hat{\sigma}_u^2\right)\cdot\hat{\Delta}_u(\theta)$. That $t_l^*(\theta) \geq 0$ can be verified by using the conditions that $\hat{\Delta}_l(\theta) > 0$ and $\hat{\Delta}_l(\theta) > -\hat{\Delta}_u(\theta)$, which are implications of $\hat{\Delta}_l(\theta) < 0$ and $\theta_U > \theta_L$. Plugging in $t_l^*(\theta)$ and $t_u^*(\theta)$ and simplifying yields*

$$nQ_n(\theta) = \frac{n}{\hat{\sigma}_u^2}\cdot(\theta_U - \theta),$$

so that $\theta \in CI^{MI}$ iff

$$\frac{n}{\hat{\sigma}_u^2}\cdot(\theta_U - \theta)^2 \leq c_\alpha.$$

Since the case where $\theta_U - \theta < 0$ is being considered, this holds iff

$$-\sqrt{c_\alpha} \leq \frac{\sqrt{n}}{\hat{\sigma}_u} \cdot (\theta_U - \theta),$$

or equivalently,

$$\theta \leq U_n^{MI},$$

where

$$U_n^{MI} \equiv \theta_U + \frac{\hat{\sigma}_u}{\sqrt{n}} \Phi^{-1}(1 - \alpha).$$

6.5 Proposition 4

Proof. Let $\theta \in \partial \Theta^*$, but suppose that $\theta \notin D\Theta^*$ for contradiction.

$\theta \in \partial \Theta^* \Rightarrow \theta \in \Theta^*$, which implies that $\mathbb{E}[m(y, x, \theta)] > 0$. Therefore, there exists an open neighborhood of $\mathbb{E}[m(y, x, \theta)]$ contained in $\mathbb{R}_+^J$, say N. Let N_θ be the inverse image of N, i.e.

$$N_\theta \equiv \{t \in \Theta : \mathbb{E}[m(y, x, t)] \subseteq N\}.$$

Because N is an open subset of $\mathbb{R}_+^J$, $\mathbb{E}[m(y, x, t)] \subseteq N \Rightarrow \mathbb{E}[m(y, x, t)] > 0$. By the continuity of $\mathbb{E}[m(y, x, \theta)]$ under (A5) N_θ is an open neighborhood of θ, and $N_\theta \subseteq \Theta^*$ since $\mathbb{E}[m(y, x, t)] > 0$ for all $t \in N$. Therefore, there exists an open neighborhood of θ that is contained in Θ^*, contradicting the supposition that $\theta \in \partial \Theta^*$. ∎

6.6 Proposition 5

Proof. Proposition 4 shows $\partial \Theta^* \subseteq D\Theta^*$, so all that is needed is to show $\partial \Theta^* \supseteq D\Theta^*$. Let $\theta \in D\Theta^*$ so that $\mathbb{E}[m_j(y, x, \theta)] = 0$ for some j. By (A6), $\mathbb{E}[m_j(y, x, \theta)]$ is monotone in some component of θ, say $\theta_{k(j)}$. Let $\epsilon > 0$, and let $\mathbf{v}(\epsilon)$ be a k-vector with $k(j)$ component ϵ and all other components zero. By the strict monotonicity of $\mathbb{E}[m_j(y, x, \theta)]$ in $\theta_{k(j)}$, $\forall \epsilon \in \mathbb{R}^k$, either $\mathbb{E}[m_j(y, x, \theta + \mathbf{v}(\epsilon))] < 0$ or $\mathbb{E}[m_j(y, x, \theta - \mathbf{v}(\epsilon))] < 0$, so that $\theta \in \partial \Theta^*$. ∎

References

ANDREWS, D. K., S. T. BERRY, AND P. JIA (2004): "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Location," Working Paper, Yale University.

ANDREWS, D. W. K. (2001): "Testing When a Parameter is on the Boundary of the Maintained Hypothesis," *Econometrica*, 69(3), 683–734.

BALKE, A., AND J. PEARL (1997): "Bounds on Treatment Effects from Studies with Imperfect Compliance," *Journal of the American Statistical Association*, 92(439), 1171–1176.

BERESTEANU, A., AND F. MOLINARI (2006): "Asymptotic Properties for a Class of Partially Identified Models," Working paper, Cornell University.

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2004): "Parameter Set Inference in a Class of Econometric Models," Working Paper, Princeton University.

FRECHET, M. (1951): "Sur Les Tableaux de Correlation Donte les Marges sont Donnees," *Annals de l'Universite de Lyon A*, (14), 53–77.

FRISCH, R. (1934): *Statistical Confluence Analysis By Means of Complete Regression Systems.* University Institute for Economics, Oslo, Norway.

GOURIEROUX, C., A. HOLLY, AND A. MONFORT (1982): "Likelihood Ratio Test, Wald Test, and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters," *Econometrica*, 50(1), 63–80.

HO, K. (2005): "Insurer-Provider Networks in the Medical Care Market," Working paper, Harvard University.

HOROWITZ, J. L., AND C. F. MANSKI (2000): "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data," *Journal of the American Statistical Association*, 95(449), 77–84.

HOTZ, V. J., C. H. MULLIN, AND S. G. SANDERS (1997): "Bounding Causal Effects Using Data From a Contaminated Natural Experiment: Analysing the Effects of Teenage Child Bearing," *Review of Economic Studies*, 64(4), 575–603.

HU, L. (2002): "Estimation of a Censored Dynamic Panel Data Model," *Econometrica*, 70, 2499–2517.

IMBENS, G., AND C. F. MANSKI (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857.

ISHII, J. (2005): "Interconnection Pricing, Compatibility, and Investment in Network Industries: An Empirical Study of ATM Surcharging in the Retail Banking Industry," Working paper, Harvard University.

KLEPPER, S., AND E. E. LEAMER (1984): "Consistent Sets of Estimates for Regressions with Errors in All Variables," *Econometrica*, 52(1), 163–184.

KODDE, D. A., AND F. C. PALM (1986): "Wald Criteria for Jointly Testing Equality and Inequality Restrictions," *Econometrica*, 54(5), 1243–1248.

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

LEHMANN, E. L. (1986): *Testing Statistical Hypotheses*. John Wiley and Sons, New York.

MANSKI, C. F. (1989): "Anatomy of the Selection Problem," *The Journal of Human Resources*, 24(3), 343–360.

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

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

MANSKI, C. F., AND J. V. PEPPER (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68(4), 997–1010.

MANSKI, C. F., AND E. TAMER (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70(2), 519–546.

MOLINARI, F. (2005): "Missing Treatments," Working paper, Cornell University.

MOON, H. R., AND F. SCHORFHEIDE (2006): "Boosting Your Instruments: Estimation with Overidentifying Inequality Moment Conditions," Working paper, University of Pennsylvania.

PAKES, A., J. PORTER, K. HO, AND J. ISHII (2006): "The Method of Moments with Inequality Constraints," Working paper, Harvard University.

PERLMAN, M. D. (1969): "One-Sided Testing Problem in Multivariate Analysis," *The Annals of Mathematical Statistics*, 40(2), 549–567.

SEN, P. K., AND M. J. SILVAPULLE (2004): *Constrained Statistical Inference: Inequality, Order, and Shape Restrictions*. Wiley-Interscience, New York.

SHAIKH, A. M. (2005): "Inference for Partially Identified Econometric Models," Working paper, Stanford University.

WOLAK, F. A. (1987): "An Exact Test for Multiple Inequality and Equality Constraints in the Linear Regression Model," *Journal of the American Statistical Association*, 92(399), 782–793.

——— (1991): "The Local Nature of Hypothesis Testing Involving Inequality Constraint in Nonlinear Models," *Econometrica*, 59(4), 981–995.