Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities

Rosen, Adam M.

Accepted Manuscript

Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities

Adam M. Rosen

PII: S0304-4076(08)00079-1
DOI: 10.1016/j.jeconom.2008.08.001
Reference: ECONOM 3049

To appear in: *Journal of Econometrics*

Received date: 15 June 2006
Revised date: 20 July 2008
Accepted date: 1 August 2008

Please cite this article as: Rosen, A.M., Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities. *Journal of Econometrics* (2008), doi:10.1016/j.jeconom.2008.08.001

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

Adam M. Rosen*†
Department of Economics, University College London,
Centre for Microdata Methods and Practice,
and Institute for Fiscal Studies

July 2008

Abstract

This paper proposes a computationally simple way to construct confidence sets for a parameter of interest in models comprised of 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. Critical values for the test are easily computed, and a Monte Carlo study demonstrates implementation and 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 parameters of interest, but nonetheless impose meaningful restrictions on the values these parameters

*This is a revised version of the first chapter of my Northwestern PhD dissertation and subsequently CEMMAP working paper CWP25/06. I thank R. Blundell, V. Chernozhukov, A. Chesher, J. Horowitz, S. Lee, C. Manski, F. Molinari, R. Porter, J. Stoye, three anonymous reviewers, an Associate Editor, the Editor A. Ronald Gallant, and especially E. Tamer for insightful comments and suggestions. I have benefited from comments from seminar participants at BU, Cornell, Duke, LSE, Mannheim, Maryland, Michigan, Northwestern, Oxford, Penn., Princeton, Queen Mary, Toulouse, UBC, UC3M, UCL, UCSD, the 2005 Conference on Incomplete Models (Northwestern), the 2006 Econometric Society NASM (Minnesota), the 2006 RTN conference (Madrid), and the 2007 ESEM (Budapest). Financial support from the Robert Eisner Memorial Fellowship, the Center for the Study of Industrial Organization, and the ESRC Centre for Microdata Methods and Practice (grant reference RES-589-28-0001) are gratefully acknowledged. The usual disclaimer applies.

†email: adam.rosen@ucl.ac.uk

may take, the parameters are said to be partially identified, see Manski (2003). 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 is a necessary precursor for statistical inference. Second, in some cases, the size of the estimated identified set is significantly larger than the imprecision of estimates due to sampling variation, as in Manski and Nagin (1998). However, in order to build confidence regions, perform hypothesis tests, or compare set estimates to point estimates derived from more restrictive models, sampling variation must be taken into account.

This paper proposes a computationally attractive way to perform inference via pointwise testing in the spirit of Anderson and Rubin (1949, 1950) in a large class of models whose application often results in partial identification: moment inequality models. These are models in which the parameter of interest, denoted θ_0, is known to satisfy a moment restriction of the form $\mathbb{E}\left[m\left(z,\theta_0\right)\right] \geq 0$, where z is an observable random vector, and m is a known, vector-valued function of the data and a possibly multivariate parameter of interest θ_0. Such restrictions are common implications of optimizing behavior and appear in many econometric models. Some examples include bounds on regression parameters when there is measurement error, studied by Frisch (1934) and Klepper and Leamer (1984), bounds on treatment effects as in Balke and Pearl (1997), Hotz, Mullin, and Sanders (1997), Manski and Pepper (2000), bounds on joint cumulative distribution functions (Frechet (1951)), and bounds on regression parameters with interval data as in Manski and Tamer (2002). Moon and Schorfheide (2005) consider models comprised of both moment inequalities and equalities where there is point identification.

This paper contributes to the literature on inference on partially identified parameters by offering a way to perform inference on a (possibly multivariate) parameter θ_0 using fixed critical values based on the asymptotic distribution of a test statistic. Previously, Imbens and Manski (2004) showed one way this can be done when θ_0 is univariate and interval-identified. Chernozhukov, Hong, and Tamer (2007) were the first to provide general methods for consistent set estimation and inference with potentially multivariate θ_0, covering cases where the identification region can be written as the minimizer of a criterion function. Research on this subject has since expanded considerably, and the reader is referred to Chernozhukov, Hong, and Tamer (2007) for an excellent overview of this literature. Recent research in this area includes Andrews, Berry, and Jia (2004), Pakes, Porter, Ho, and Ishii (2004), Galichon and Henry (2006a), Galichon and Henry (2006b), Romano and Shaikh (2006), Andrews and Guggenberger (2007), Andrews and Soares (2007), Bontemps, Magnac, and Maurin (2007), Bugni (2007), Canay (2007), Guggenberger, Hahn, and Kim (2008), Fan and Park (2007), Stoye (2007), Beresteanu and Molinari (2008), and Romano and Shaikh (2008). Examples of recent papers that employ such methods include Ciliberto and Tamer (2004), Ishii (2005), Rosen (2006), Blundell, Gosling, Ichimura, and Meghir (2007), Blundell, Browning, and Crawford (2008), Ho (2008), and Molinari (2008).

Methods for inference applicable in contexts with multivariate θ_0 have relied on subsampling, bootstrapping, or simulation for approximation of asymptotic critical values. In this paper, the test statistic used to perform inference has an asymptotic chi-bar-square distribution, and can be inverted to construct an asymptotic confidence set for the parameter of interest. Relative to inferential methods based on subsampling or bootstrapping, this has the computational advantage of not requiring resampling of one's data to obtain critical values for a test statistic over each element of the parameter space.

To motivate the confidence sets of this paper, it is useful to first consider inference when there is point-identification. When θ_0 is point-identified, one may construct a confidence set $\mathcal{C}_n$ such that in repeated sampling

$$\lim_{n\to\infty} \mathbb{P}\left\{\theta_0 \in \mathcal{C}_n\right\} = 1-\alpha, \tag{1}$$

for pre-specified level $1-\alpha$. This is the starting point taken for motivation of the confidence regions constructed in this paper. However, when θ_0 is partially identified, the standard methods for constructing such a set $\mathcal{C}_n$ do not apply without modification, as they rely on point identification as a necessary condition. In this context, there is some set of values, Θ^*, which are observationally equivalent to θ_0, called the *identified set*. In the class of models considered here, a confidence set that satisfies (1) for one value of $\theta_0 = \theta' \in \Theta^*$, may not do so for another value $\theta_0 = \theta'' \in \Theta^*$. Because any two such values θ' and θ'' are by definition observationally equivalent, no amount of sample data will allow the researcher to distinguish between any two such values.

Thus, the goal of this paper is construction of sets that satisfy

$$\inf_{\theta\in\Theta^*} \lim_{n\to\infty} \mathbb{P}\left\{\theta \in \mathcal{C}_n^{pt}\right\} = 1-\alpha, \tag{2}$$

where $\mathbb{P}$ is taken to be the measure induced by repeated sampling from the *true* population distribution. Since $\theta_0 \in \Theta^*$, i.e. the true θ_0 is necessarily a member of the identified set, such sets $\mathcal{C}_n^{pt}$ will contain θ_0 with at least probability $1-\alpha$ for n sufficiently large, i.e. $\lim_{n\to\infty} \mathbb{P}\left\{\theta_0 \in \mathcal{C}_n^{pt}\right\} \geq 1-\alpha$. To this end, I employ a pointwise testing procedure, in the vein of Anderson and Rubin (1949) and Anderson and Rubin (1950). In the face of either weak or partial identification, pointwise approaches have also been employed by, for example, Dufour (1997), Staiger and Stock (1997), Stock and Wright (2000), Hu (2002), Kleibergen (2005), and Guggenberger and Smith (2005), among others. Some recent papers have also considered sets that provide uniform asymptotic coverage in both $\mathbb{P}$ and $\theta \in \Theta^*$, see for example Imbens and Manski (2004), Fan and Park (2007), and Stoye (2007). Andrews and Guggenberger (2005) and Andrews and Guggenberger (2007) provide conditions under which the confidence sets of this paper have uniformly valid asymptotic coverage.

The procedure employed in this paper makes use of results on multivariate one-sided hypothesis testing, such as Bartholomew (1959a), Bartholomew (1959b), Kudo (1963), Perlman (1969), Gourieroux, Holly, and Monfort (1982), Kodde and Palm (1986) and Wolak (1991); see Sen and Silvapulle (2004) for a thorough compendium. Results in this literature apply in cases where the parameter of interest is point-identified. This paper extends these methods to the moment inequality setting, where there is no consistent point estimate for θ_0, by relying on the asymptotic behavior of the moment restrictions. Specifically, I construct a test statistic $\hat{Q}_n(\theta)$ that, under sufficient regularity conditions, when scaled by n and evaluated at any element θ of the identified set Θ^*, has an asymptotic distribution that is a mixture of chi-square distributions, the chi-bar-square distribution. This test statistic is then inverted to construct confidence sets for θ_0 with pre-specified asymptotic coverage. The test statistic is a function of the moments that comprise the imposed modeling restrictions on θ_0. As such, the theory needed to guarantee proper asymptotic coverage relies completely on the distribution of observables. The inferential method is relatively straightforward to implement in practice and is demonstrated with a specific example in section 5.

A drawback is that in general the cutoff value for the test statistic $\hat{Q}_n(\theta)$ differs for different values of $\theta \in \Theta^*$. That is, $n\hat{Q}_n(\theta)$ is not asymptotically pivotal because its asymptotic distribution depends on the variance of those components of $m(z,\theta)$ that have expected value zero. This problem is overcome by building confidence sets for θ_0 by using an upper bound on the number of such components. The dimension of $m(z,\theta)$, J, is clearly an upper bound, but in models with partially identified parameters there is often a smaller upper bound which can be used to achieve more accurate inference. As discussed further in section 4, in some cases use of this upper bound may lead to coverage inflation, in the sense that $\inf_{\theta\in\Theta^*}\lim_{n\to\infty}\mathbb{P}\left\{\theta \in \mathcal{C}_n^{pt}\right\}$ may exceed $1-\alpha$, though the test on which the confidence sets are based is consistent regardless. In cases where there is no obvious upper bound implied by the modeling restrictions, it is straightforward to estimate.

The paper proceeds as follows. Section 2 presents the moment inequality model. Section 3 describes the pointwise testing procedure. Section 4 then presents two ways to construct confidence sets based on the hypothesis test of section 3. Section 5 presents a simple example as illustration and investigates the performance of confidence sets via Monte Carlo simulation. Section 6 concludes and offers avenues for continued research. All proofs are in the Appendix.

2 The Model

Let $\{z_i : i = 1,...,n\}$ denote a random sample of observations of z distributed with population distribution P with support $\mathcal{Z} \subseteq \mathbb{R}^s$. Each observation z_i represents *all* information observed by the econometrician for each $i = 1,...,n$. If partial identification is a result of missing data, for example, then z_i excludes those characteristics of individual i in the population that are missing. θ, rather than θ_0, is used to denote a representative value of the parameter of interest. Θ^* denotes the set of values of $\theta \in \Theta$ that satisfy the restrictions of the model, i.e. Θ^* is the identified set for θ_0. The "true" underlying value of θ in the model is denoted θ_0, but in general θ_0 might not be point-identified by the restrictions of the model.

The focus of this paper is moment inequality models. The model is summarized by the restrictions

$$\mathbb{E}\left[m\left(z,\theta_0\right)\right]=\mathbb{E}\left[\begin{array}{c} m_1\left(z,\theta_0\right) \\ \vdots \\ m_J\left(z,\theta_0\right) \end{array}\right]\geq\begin{pmatrix}0\\ \vdots \\ 0\end{pmatrix}. \tag{3}$$

$J<\infty$ is the number of moment inequalities of the model. Formally, the model is given by the following three assumptions.

Assumption A1 (random sampling) $Z\equiv\{z_i : i=1,...,N\}$ are i.i.d. observations distributed P.
Assumption A2 (compact parameter space) θ_0 is an element of the compact space $\Theta\subseteq\mathbb{R}^k$.
Assumption A3 (moment inequalities) $\mathbb{E}\left[m\left(z,\theta_0\right)\right]\geq 0$, where $m\left(\cdot,\cdot\right):\mathbb{R}^s\times\Theta\rightarrow\mathbb{R}^J$.
These assumptions yield the following identified set for θ_0.

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

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

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 misspecified. If $\Theta^*\subseteq\Theta$ but is neither empty nor singleton, then θ_0 is 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

$$\begin{aligned} H_0 &: \mathbb{E}\left[m\left(z,\theta\right)\right]\geq 0 \\ H_1 &: \mathbb{E}\left[m\left(z,\theta\right)\right]\ngeq 0, \end{aligned} \tag{4}$$

for any *fixed candidate value* of $\theta\in\Theta$. The next two sections provide theoretical justification and a description of how to perform this test with pre-specified asymptotic size α. Once the testing procedure is established for fixed θ, a $1-\alpha$ confidence set for θ_0 is constructed by taking the set of θ that are not rejected by this hypothesis test.

The hypothesis test is based on a test statistic $\hat{Q}_n\left(\theta\right)$ such that if $n\hat{Q}_n\left(\theta\right)$ exceeds a critical value, the null hypothesis is rejected. That is $\theta\in\Theta^*$ is rejected if $n\hat{Q}_n\left(\theta\right)>C^*_\alpha$, where $\sup_{\theta\in\Theta^*}\lim_{n\rightarrow\infty}\mathbb{P}\left\{n\hat{Q}_n\left(\theta\right)>C^*_\alpha\right\}=\alpha$. This implies that the set $\mathcal{C}_n^{pt}\equiv\left\{\theta:n\hat{Q}_n\left(\theta\right)\leq C^*_\alpha\right\}$

satisfies condition (2) as

$$\begin{aligned}\inf_{\theta\in\Theta^*}\lim_{n\to\infty}\mathbb{P}\left\{\theta\in\mathcal{C}_n^{pt}\right\} &= \inf_{\theta\in\Theta^*}\lim_{n\to\infty}\mathbb{P}\left\{n\hat{Q}_n(\theta)\leq C_\alpha^*\right\}\\ &= 1-\sup_{\theta\in\Theta^*}\lim_{n\to\infty}\mathbb{P}\left\{n\hat{Q}_n(\theta)>C_\alpha^*\right\}=1-\alpha.\end{aligned}$$

This further implies that $\lim_{n\to\infty}\mathbb{P}\left\{\theta_0\in\mathcal{C}_n^{pt}\right\}\geq 1-\alpha$. While the focus of this paper is pointwise inference, Andrews and Guggenberger (2005) and Andrews and Guggenberger (2007) give sufficient conditions for such confidence sets to provide uniformly valid asymptotic coverage. The next section explains how the pointwise test is carried out and characterizes the asymptotic distribution for the statistic $n\hat{Q}_n(\theta)$ on which the test is based.

3 Asymptotic Behavior of the Test Statistic

In this section, I consider a test of the hypothesis (4) for any fixed candidate value of θ. To test this hypothesis, I construct a test statistic, $\hat{Q}_n(\theta)$ whose asymptotic distribution, when scaled by n, is chi-bar-square (a mixture of chi-square random variables) under the null hypothesis, while under the alternative hypothesis, $n\hat{Q}_n(\theta)\to\infty$. The test statistic is in general not asymptotically pivotal, but can still be used to construct confidence sets for θ_0. Depending on the variance of the binding moments over Θ^*, the confidence sets may be conservative, in the sense that condition (2) may be satisfied with weak inequality ($\geq$) rather than equality. This is not relevant for the theoretical result of this section, but is an important consideration in the actual construction and accuracy of confidence regions. A more detailed discussion is deferred to the details of implementation discussed in section 4.

In order to test whether θ is contained in the identified set implied by the restrictions (3), I employ the following statistic:

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

where $\hat{V}_\theta$ is the sample variance of $m(z,\theta)$, and where the minimization is taken over the vector t of dimension J, constrained to have all elements non-negative. The value of $\hat{Q}_n(\theta)$ is a function of the sample moment functions evaluated at θ, as well as $\hat{V}_\theta$. Given any fixed value of θ being tested, $\hat{Q}_n(\theta)$ is the solution of a quadratic minimization problem over a polyhedral cone, for which the Kuhn-Tucker conditions characterize a unique minimum, see Kudo (1963). Thus, for any fixed value of θ being tested, $\hat{Q}_n(\theta)$ is straightforward to compute using the necessary and sufficient Kuhn-Tucker conditions, which are that for each $j=1,...,J$,

$$\left[\hat{V}_\theta^{-1}\left[\hat{E}_n[m(z,\theta)]-t\right]\right]_j=0 \text{ and } t_j>0.$$

or

$$\left[\hat{V}_\theta^{-1}\left[\hat{E}_n\left[m\left(z,\theta\right)\right]-t\right]\right]_j \leq 0 \text{ and } t_j = 0.$$

Explicitly imposing these conditions substantially simplifies the computation of $\hat{Q}_n(\theta)$.

If the moment restrictions $\mathbb{E}\left[m\left(z,\theta\right)\right] \geq 0$ are true, i.e. if $\theta \in \Theta^*$, then $\hat{Q}_n(\theta)$ should be small. In this case, violations of $\hat{E}_n\left[m\left(z,\theta\right)\right] \geq 0$ are attributable to no more than sampling variation. This is because the population version of $\hat{Q}_n(\theta)$ (and the probability limit of $\hat{Q}_n(\theta)$ under sufficient regularity, see Proposition 1) is

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

where V_θ is the variance of $m(z,\theta)$. $Q(\theta)$ measures the distance of θ from Θ^*, as $Q(\theta)=0$ if and only if $\mathbb{E}\left[m\left(z,\theta\right)\right] \geq 0$, and is otherwise positive. Manski and Tamer (2002) and Chernozhukov, Hong, and Tamer (2007) derive conditions for consistency of parameter sets that minimize an objective function, and their results apply here. The focus of this paper is inference, yet in practice estimation precedes inference, so the application of these results to $\hat{Q}_n(\theta)$ is stated formally in Proposition 2.

Outside the context of estimating partially identified parameters, test statistics of similar form have been used previously in the literature on multivariate one-sided hypothesis testing, e.g. Bartholomew (1959a), Bartholomew (1959b), Kudo (1963), Perlman (1969), Gourieroux, Holly, and Monfort (1982), 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. Thus, inference is based on the estimated moment functions rather than an estimate of θ_0. 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, i.e. those that bind, at the least favorable value of θ satisfying the null hypothesis, here that $\mathbb{E}\left[m\left(z,\theta\right)\right] \geq 0$. In his model, however, there is a known function which determines the boundary of the null hypothesis, $h(\theta)$ rather than $\mathbb{E}\left[m\left(z,\theta\right)\right]$. Thus, in the setting of this paper, aside from the complication that here θ_0 is only partially identified, it is also the case that $\mathbb{E}\left[m\left(z,\theta\right)\right]$, the function that determines the boundary of the null hypothesis, is not known, but rather must be estimated. This is a notable difference because, as shown in Proposition 3, the asymptotic distribution of $n\hat{Q}_n(\theta)$ is degenerate except on the boundary of the hypothesis that $\mathbb{E}\left[m\left(z,\theta\right)\right] \geq 0$ i.e. the set of θ such that $\mathbb{E}\left[m_j\left(z,\theta\right)\right] = 0$ for at least one $j \in \{1,...,J\}$.

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

Assumption A4 (finite variance of m on Θ^*) For some $K < \infty$, for each $(i,j) \in \{1,...,J\}^2$,

$\sup_{\theta\in\Theta^*}\left|\mathbb{E}\left[m\left(z,\theta\right)m\left(z,\theta\right)'\right]_{ij}\right|<K$, i.e. each element of $\mathbb{E}\left[m\left(z,\theta\right)m\left(z,\theta\right)'\right]$ is bounded for all $\theta\in\Theta^*$. This also implies that the moments $\mathbb{E}\left[m\left(z,\theta\right)\right]$ are bounded.

Assumption 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 $\mathbb{E}\left[m\left(z,\theta\right)\right]$, while assumption (A5) guarantees that V_θ is invertible. Under (A1) and (A4), it follows that for all $\theta\in\Theta^*$,

$$\hat{E}_n\left[m\left(z,\theta\right)\right]=\frac{1}{n}\sum_{i=1}^{n}m\left(z_i,\theta\right)\overset{a.s.}{\to}\mathbb{E}\left[m\left(z,\theta\right)\right], \tag{5}$$

$$\begin{aligned}\hat{V}_n\left[m\left(z,\theta\right)\right]&=\frac{1}{n}\sum_{i=1}^{n}\Big(m\left(z,\theta\right)-\hat{E}_n\left[m\left(z,\theta\right)\right]\Big)\Big(m\left(z,\theta\right)-\hat{E}_n\left[m\left(z,\theta\right)\right]\Big)' \\ &\overset{a.s.}{\to} var\left\{m\left(z,\theta\right)\right\}\equiv V_\theta,\end{aligned} \tag{6}$$

and

$$\sqrt{n}\left\{\hat{E}_n\left[m\left(z,\theta\right)\right]-\mathbb{E}\left[m\left(z,\theta\right)\right]\right\}\overset{d}{\to}N\left(0,V_\theta\right). \tag{7}$$

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 $m\left(z,\theta\right)$ over Θ^*, and a central limit theorem for $\sqrt{n}\left\{\hat{E}_n\left[m\left(z,\theta\right)\right]-\mathbb{E}\left[m\left(z,\theta\right)\right]\right\}$ for each $\theta\in\Theta^*$. Both the assumption that the observations are i.i.d. and that the rate of convergence of $\hat{E}_n\left[m\left(z,\theta\right)\right]$ to $\mathbb{E}_n\left[m\left(z,\theta\right)\right]$ is $\sqrt{n}$ can be relaxed, as long as (5), (6), and (7) can be shown to hold at each $\theta\in\Theta^*$ for some sequence of constants $a_n\to\infty$ replacing $\sqrt{n}$. Assumption (A5) rules out singularity of the asymptotic variance of $\sqrt{n}\left\{\hat{E}_n\left[m\left(z,\theta\right)\right]-\mathbb{E}\left[m\left(z,\theta\right)\right]\right\}$. While in many cases this restriction is plausible, it is restrictive. In particular, in the context of interval identification it rules out the case where the estimators for the boundaries of the interval are perfectly correlated or when they are approximated by the same estimator. Because the goal here is construction of a confidence set $\mathcal{C}_n^{pt}$ such that $\inf_{\theta\in\Theta^*}\lim_{n\to\infty}\mathbb{P}\left\{\theta\in\mathcal{C}_n^{pt}\right\}=1-\alpha$, 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 asymptotic coverage then stronger conditions would be needed, see Andrews and Guggenberger (2007).

Before proceeding with distributional results, Proposition 1 first establishes consistency of the sample objective function, and Proposition 2 offers sufficient conditions for consistent set estimation, which typically precedes inference in applications. For these results, it is convenient to define

$$q\left(\theta,t\right)\equiv\left[\mathbb{E}\left[m\left(z,\theta\right)\right]-t\right]'V_\theta^{-1}\left[\mathbb{E}\left[m\left(z,\theta\right)\right]-t\right],$$

and

$$\hat{q}_n(\theta,t) \equiv \left[\hat{E}_n[m(z,\theta)] - t\right]' \hat{V}_\theta^{-1} \left[\hat{E}_n[m(z,\theta)] - t\right],$$

so that $Q(\theta) = \min_{t\geq 0} q(\theta,t)$ and $\hat{Q}_n(\theta) = \min_{t\geq 0} \hat{q}_n(\theta,t)$. Properties of the functions q and $\hat{q}_n$ translate directly to properties of $\hat{Q}_n$ and Q.

Proposition 1 *Let (A1)-(A5) hold. Define* $\hat{t}_n^*(\theta) \equiv \arg\min_{t\geq 0} \hat{q}_n(\theta,t)$, *and* $t^*(\theta) \equiv \arg\min_{t\geq 0} q(\theta,t)$. *Then* $t^*(\theta)$ *is unique,* $\hat{t}_n^*(\theta)$ *is unique with probability approaching one as* $n \to \infty$, *and for any* $\theta \in \Theta$, $\hat{Q}_n(\theta) \xrightarrow{p} Q(\theta)$, *and* $\hat{t}_n^*(\theta) \xrightarrow{p} t^*(\theta)$. *Furthermore,* $\hat{t}_n^*(\theta) - t^*(\theta) = O_p\left(n^{-1/2}\right)$.

Proposition 1 follows from the convexity and continuity of $q(\theta,t)$ and $\hat{q}_n(\theta,t)$ in t. These properties provide sufficient regularity to apply the results of Andrews (1999), necessary to ensure $\sqrt{n}$ convergence in probability of $\hat{t}_n^*(\theta)$ to $t^*(\theta)$ when θ is on the boundary of the null hypothesis $\mathbb{E}[m(z,\theta)] \geq 0$. If, in addition, the convergence of $\hat{Q}_n(\theta) \xrightarrow{p} Q(\theta)$ is uniform over Θ, then the results of Chernozhukov, Hong, and Tamer (2007) can be applied to formulate a consistent set estimator for Θ^*, as stated in Proposition 2.

Proposition 2 *Let (A1)-(A5) hold, and assume that* $q(\theta,t)$ *is continuous in* θ *and that* $\hat{Q}_n(\theta)$ *is stochastically equicontinuous. Then* $\hat{Q}_n(\theta) \xrightarrow{p} Q(\theta)$ *uniformly over* $\theta \in \Theta$. *In addition let* ϵ_n *be a sequence of positive constants such that* $\epsilon_n \to \infty$ *and* $\epsilon_n/n \to 0$ *as* $n \to \infty$. *Then*

$$\hat{\Theta}_n^* = \left\{\theta \in \Theta : n\hat{Q}_n(\theta) \leq \epsilon_n\right\}$$

is a consistent set estimate for Θ^* *in the Hausdorff norm.*

The next proposition provides the asymptotic distribution of $n\hat{Q}_n(\theta)$, but first some additional notation is required. For expositional convenience, I refer to the subset of the J moment inequalities such that $\mathbb{E}[m_j(z,\theta)] = 0$ as the set of binding moments. Without loss of generality, let the first $b(\theta)$ moments be the subset of binding moments at θ, so that $\mathbb{E}[m_j(z,\theta)] = 0$, $j = 1,...,b(\theta)$, and $\mathbb{E}[m_j(z,\theta)] > 0$, $j = b(\theta)+1,...,J$. Let $m^*(z,\theta) = \left(m_1(z,\theta),...,m_{b(\theta)}(z,\theta)\right)'$ denote the subvector of moments that have mean zero, and let $V_\theta^* = var(m^*(z,\theta))$. $\Pr\left\{\chi_j^2 \geq c\right\}$ denotes the probability that a chi-square random variable with j degrees of freedom 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 $n\hat{Q}_n(\theta)$ under the hypothesis that $\theta \in \Theta^*$.

Proposition 3 *Let assumptions (A1)-(A5) hold. Then for any value of* $\theta \in \Theta^*$, *for any constant* c,

$$\lim_{n\to\infty} \mathbb{P}\left\{n\hat{Q}_n(\theta) > c\right\} = \sum_{j=0}^{b(\theta)} w\left(b(\theta), b(\theta)-j, V_\theta^*\right) \Pr\left\{\chi_j^2 > c\right\}, \quad (8)$$

where $w(\cdot,\cdot,\cdot)$ *is the weights function defined by Wolak (1987) and Kudo (1963), and the* χ^2_j *random variables of the summation are independent. For those* $\theta \in \Theta^*$ *such that* $\mathbb{E}[m(z,\theta)] > 0$, $\lim_{n\to\infty} \mathbb{P}\left\{n\hat{Q}_n(\theta) > 0\right\} = 0$. *If* $\theta \notin \Theta^*$ *and each element of* $\mathbb{E}\left[m(z,\theta)\, m(z,\theta)'\right]$ *is finite, then for any constant* $c > 0$, $\lim_{n\to\infty} \mathbb{P}\left\{n\hat{Q}_n(\theta) > c\right\} = 1$.

Proposition 3 closely follows Lemma 1 of Wolak (1991). The first step to the proof shows that the limiting distribution of $n\hat{Q}_n(\theta)$ is determined only by those terms that correspond to components of $\mathbb{E}[m(z,\theta)]$ that are exactly equal to 0. Multiplication of $\hat{Q}_n(\theta)$ by n is equivalent to multiplying each of the $\left(\hat{E}_n[m(z,\theta)] - t\right)$ terms in $\hat{Q}_n(\theta)$ by $\sqrt{n}$. For those moments j where $\mathbb{E}[m_j(z,\theta)] > 0$, $\sqrt{n}\hat{E}_n[m_j(z,\theta)]$ diverges to infinity, and the Kuhn-Tucker conditions for minimization guarantee that $\hat{t}^*_n(\theta)$, will satisfy the first order condition $\left[\hat{V}_\theta^{-1}\left[\hat{E}_n[m(z,\theta)] - t\right]\right]_j = 0$ with probability going to 1 as $n \to \infty$. This implies that these moments do not contribute to $n\hat{Q}_n(\theta)$ asymptotically. On the other hand, for those moments that are equal to zero, $\sqrt{n}\hat{E}_n[m^*(z,\theta)] \xrightarrow{d} N(0, V^*_\theta)$, and these components do contribute to the asymptotic distribution of $n\hat{Q}_n(\theta)$. For any realization from the $N(0, V^*_\theta)$ distribution, any number of nonnegativity constraints up to $b(\theta)$ may bind in the solution to $n\hat{Q}_n(\theta)$, $\hat{t}^*_n(\theta)$. The number of *binding constraints* on $\hat{t}^*_n(\theta)$ generally differs from the number of *binding moment inequalities* $b(\theta)$, but the latter provides an upper bound for the former. Conditional on any number r of binding nonnegativity constraints, the limit distribution of $n\hat{Q}_n(\theta)$ is χ^2_r. Unconditionally, the weights of the chi-bar-square distribution are precisely the probabilities with which exactly r constraints bind for each $r = 0, ..., b(\theta)$. An immediate implication is that when $\mathbb{E}[m(z,\theta)] > 0$, $t^*_n(\theta) = \hat{E}_n[m(z,\theta)]$ is chosen with probability going to one, i.e. none of the constraints bind, so that $n\hat{Q}_n(\theta) \xrightarrow{p} 0$. Finally, the test is consistent against fixed alternatives, as $\theta \notin \Theta^*$ implies that $n\hat{Q}_n(\theta) \to \infty$.

The weights function $w(b(\theta), b(\theta) - j, V)$ has arisen repeatedly in research on multivariate one-sided hypothesis tests. As the limit distribution of $n\hat{Q}_n(\theta)$ conditional on r constraints binding is χ^2_r, the weights correspond to the probabilities with which each feasible number of constraints bind, or equivalently the number of components of $\hat{t}^*_n(\theta)$ that are equal to zero, so that

$$w(b(\theta), b(\theta) - j, V^*_\theta) = \lim_{n\to\infty} \mathbb{P}\left\{\hat{t}^*_n(\theta) \text{ has } j \text{ components equal to zero}\right\}.$$

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^*_θ is diagonal. More generally, closed-form expressions for the weights have not been obtained, but if V^*_θ and $b(\theta)$ were known, they could be approximated with arbitrary accuracy by means of simulation. For example, one such method outlined by Sen and Silvapulle (2004, pp. 78-80) is to simulate draws of a random variable Z from the $N(0, V^*_\theta)$ distribution and compute the frequency with which $\arg\min_{t\geq 0}(Z - t)\, V^{*-1}_\theta (Z - t)$ has j components equal to zero, each j, in place of

$w\left(b\left(\theta\right), b\left(\theta\right)-j, V_{\theta}^{*}\right)$.

If V_{θ}^{*} and $b\left(\theta\right)$ were known, then it would be straightforward using such techniques to compute the cutoff value C_{α}^{*} such that $\sum_{j=0}^{b(\theta)} w\left(b\left(\theta\right), b\left(\theta\right)-j, V_{\theta}^{*}\right) \Pr\left\{\chi_{j}^{2} > C_{\alpha}^{*}\right\} = \alpha$. Unfortunately, V_{θ}^{*} and $b\left(\theta\right)$ are not known. A seemingly intuitive solution might be to use sample analogs $\hat{V}_{\theta}^{*}$ and $\hat{b}\left(\theta\right)$ in place of these, but this doesn't work here because the CDF of the limit distribution given by (8) is discontinuous in $b\left(\theta\right)$. This problem can, however, be overcome by considering the least favorable asymptotic distribution of the test statistic over Θ^{*}. Section 4 details how this can be done by using an upper bound for $b\left(\theta\right)$ to construct a cutoff value $C_{\alpha}^{b^{*}}$ such that

$$\inf_{\theta\in\Theta^{*}} \lim_{n\to\infty} \mathbb{P}\left\{n\hat{Q}_{n}\left(\theta\right) \leq C_{\alpha}^{b^{*}}\right\} = 1-\alpha, \tag{9}$$

or, in some cases

$$\inf_{\theta\in\Theta^{*}} \lim_{n\to\infty} \mathbb{P}\left\{n\hat{Q}_{n}\left(\theta\right) \leq C_{\alpha}^{b^{*}}\right\} \geq 1-\alpha. \tag{10}$$

4 Computing Confidence Sets

This section provides two ways to compute cutoff values for $n\hat{Q}_{n}\left(\theta\right)$ and build confidence sets that cover θ_{0} with at least probability $1-\alpha$ asymptotically. Both methods have the advantage that the cutoff values are easy to compute with software that provides values of chi-square CDFs. The first method is generally applicable. The second method shows how knowledge that V_{θ}^{*} is diagonal can be used to compute a cutoff value that satisfies (9). It is also shown that in this case assumption (A5), which requires that V_{θ} is nonsingular, can be relaxed. Cases where V_{θ}^{*} is diagonal include both the mean with missing data and regression with censored outcomes such as those considered by Beresteanu and Molinari (2008), Manski and Tamer (2002), and Romano and Shaikh (2006). This is a useful special case since it occurs with moment restrictions that comprise mutually exclusive conditioning events, as in the case of i.i.d. data with discrete covariates.

Both approaches make use of an upper bound on $b\left(\theta\right)$ for $\theta\in\Theta^{*}$; an obvious upper bound is the total number of moment inequalities, J. In some settings, it may be credible to impose a smaller upper bound; more generally, I use b^{*} to denote the assumed upper bound. This may happen when the model implies both upper and lower bounds on the expectation of a function of θ, a common occurrence in models with partial identification. Such knowledge can be useful for inference. In some cases, the model may not upon inspection deliver an obvious upper bound on the number of binding moments. However, it is straightforward to estimate such a bound employing similar reasoning to that of Andrews and Soares (2007) or Chernozhukov, Hong, and Tamer (2007). For example, Chernozhukov, Hong, and Tamer (2007) Remark 4.5 motivates estimation of the number of binding moments for any θ, $b(\theta)$, by

$$\hat{b}(\theta) = \sum_{j=1}^{J} 1\left[\hat{E}_n\{m_j(z,\theta)\} \leq c\sqrt{(\log n)/n}\right],$$

for some constant $c > 0$, since $\lim_{n\to\infty} \Pr\left\{\hat{b}(\theta) = b(\theta)\right\} = 1$. One might then use $\hat{b}^* = \sup_{\theta\in\hat{\Theta}_n^*} \hat{b}(\theta)$ in place of b^* in computation of critical values $C_\alpha^{b^*}$ below, where $\hat{\Theta}_n^*$ is a consistent estimator for Θ^*.[1]

As discussed in the introduction, the goal of the procedures is construction of a confidence set $\mathcal{C}_n^{pt} = \left\{\theta : n\hat{Q}_n(\theta) \leq C_\alpha^{b^*}\right\}$ with fixed cutoff $C_\alpha^{b^*}$ that satisfies

$$\inf_{\theta\in\Theta^*} \lim_{n\to\infty} \mathbb{P}\left\{\theta \in \mathcal{C}_n^{pt}\right\} = 1 - \alpha. \tag{11}$$

If equality is replaced by $\geq$, then $\mathcal{C}_n^{pt}$ is asymptotically conservative. Whether (11) holds with equality or inequality depends on the variance of the binding moments, V_θ^* over the identified set. This is because the cutoff value is based on the variance matrix that gives the highest (most conservative) possible value of $C_\alpha^{b^*}$, see Perlman (1969). If this variance matrix is a member of $\{V_\theta^* : \theta \in \Theta^*\}$, then (11) is satisfied with equality. If the worst-case variance matrix used to compute $C_\alpha^{b^*}$ is not a feasible value for V_θ^* for $\theta \in \Theta^*$, then (11) is satisfied with weak inequality ($\geq$). However, even in this case the set is not arbitrarily large, in the sense that a test based on the conservative cutoff is consistent.

Still, in some cases an estimator for the desired critical value which is not conservative asymptotically may be preferred. Critical values with this property for the test that uses $n\hat{Q}_n(\theta)$ can be computed via simulation or the bootstrap, see e.g. Chernozhukov, Hong, and Tamer (2007) and Andrews and Soares (2007). For instance, one of the generalized moment selection (GMS) procedures of Andrews and Soares (2007) can be implemented by taking a large number of simulation draws Z^* from the $N(0, I_J)$ distribution and then computing the $1-\alpha$ quantile of

$$S_n(\theta) = \min_{t\geq 0} \left[\hat{V}_\theta^{1/2} Z^* + \varphi(\xi_n) - t\right]' \hat{V}_\theta^{-1} \left[\hat{V}_\theta^{1/2} Z^* + \varphi(\xi_n) - t\right],$$

where ξ_n is a J-vector with elements $\xi_{nj} \equiv \kappa_n^{-1} n^{1/2} \hat{E}_n[m_j(z,\theta)] / \hat{V}_{\theta,jj}^{1/2}$, $\varphi(\cdot) : \mathbb{R}_J \to \mathbb{R}_J$ such that $\varphi_j(\xi_n) = \left|\xi_{nj}\right|_+$, and where κ_n is a sequence of constants such that $\kappa_n \to \infty$ and $\kappa_n^{-1} n^{1/2} \to \infty$ as $n \to \infty$. See Andrews and Soares (2007) for details as well as other feasible simulation procedures. Such an approach requires computation of separate critical values for each θ being tested, but will not be asymptotically conservative and will have favorable asymptotic power properties. There is thus a trade-off between the computational ease of employing critical values $C_\alpha^{b^*}$ described here

[1] I thank Victor Chernozhukov and Francesca Molinari for suggesting this approach for estimation of $b(\theta)$.

and the greater asymptotic precision of critical values based on GMS. A practical approach in applications might be to first construct a confidence set $\mathcal{C}_n^{pt}$ using the easy to compute but potentially conservative critical value $C_\alpha^{b^*}$. If $\mathcal{C}_n^{pt}$ is sufficiently small for the application at hand (and in particular if it is empty), then one can stop here. However, if a more precise estimator is desired, one may then compute quantiles of $S_n(\theta)$, say $C_\alpha^*(\theta)$, and construct the confidence set $\left\{\theta \in \mathcal{C}_n^{pt} : n\hat{Q}_n(\theta) \leq C_\alpha^*(\theta)\right\}$. As $n \to \infty$, this set should be smaller than $\mathcal{C}_n^{pt}$, so that only values of $\theta \in \mathcal{C}_n^{pt}$ need to be tested, circumventing the need to compute $C_\alpha^*(\theta)$ for $\theta \notin \mathcal{C}_n^{pt}$.

4.1 General Implementation

The asymptotic distribution of $n\hat{Q}_n(\theta)$ obtained in Proposition 3 is discontinuous in $b(\theta)$ and V_θ^*. However, whatever V_θ^*, an upper bound on $b(\theta)$ can be used to construct a cutoff value that can be used to perform the hypothesis test (4). This cutoff value can then be used to build conservative, asymptotically valid confidence sets for θ_0. The following corollary provides the result.

Corollary 1 *Let (A1)-(A5) hold. Let* $\sup_{\theta\in\Theta^*} b(\theta) = b^*$. *Then for any* c,

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

This result is due to Perlman (1969)[2], and follows from the fact that the weights function satisfies the properties $0 \leq w(b(\theta), b(\theta)-j, V_\theta^*) \leq 1/2$, $\sum_{j=0}^{b} w(b(\theta), b(\theta)-j, V_\theta^*) = 1$, 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 (8). Exactly how slack the inequality is depends on the feasible values of the variance matrix V_θ^* over $\theta \in \Theta^*$. Corollary 1 provides a way to construct asymptotically valid confidence sets for θ_0 since if $C_\alpha^{b^*}$ solves

$$\frac{1}{2}\Pr\left\{\chi^2_{b^*} > C_\alpha^{b^*}\right\} + \frac{1}{2}\Pr\left\{\chi^2_{b^*-1} > C_\alpha^{b^*}\right\} = \alpha, \tag{12}$$

then $\lim_{n\to\infty}\mathbb{P}\left\{n\hat{Q}_n(\theta_0) \in \mathcal{C}_n^{pt}\right\} \geq 1-\alpha$, where $\mathcal{C}_n^{pt} = \left\{\theta \in \Theta : n\hat{Q}_n(\theta) \leq C_\alpha^{b^*}\right\}$.

4.2 Implementation when V_θ^* is diagonal

When V_θ^* is a diagonal, then $w(b(\theta), b(\theta)-j, V_\theta^*)$ only depends on $b(\theta)$ 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

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

the following corollary. This result also provides a smaller cutoff value for the hypothesis test (4) than that of Corollary 1, and thus a smaller confidence region when V_θ^* is diagonal.

Corollary 2 *Let (A1)-(A5) hold. Suppose that V_θ^* is diagonal for all $\theta \in \Theta^*$ and that $\sup_{\theta\in\Theta^*} b(\theta) = b^*$. Then*

$$w\left(b(\theta), b(\theta)-j, V_\theta^*\right) = 2^{-b(\theta)}\binom{b(\theta)}{b(\theta)-j}, \tag{13}$$

and $\forall c \in \mathbb{R}$,

$$\sup_{\theta\in\Theta^*}\lim_{n\to\infty}\mathbb{P}\left\{n\hat{Q}_n(\theta) > c\right\} = \sum_{j=0}^{b^*} 2^{-b^*}\binom{b^*}{j}\Pr\left\{\chi_j^2 > c\right\}. \tag{14}$$

Just as Corollary 1 provides a way to construct confidence sets for θ_0 so does Corollary 2 when V_θ^* is diagonal. If $C_\alpha^{b^*}$ solves

$$\sum_{j=0}^{b^*} 2^{-b^*}\binom{b^*}{j}\Pr\left\{\chi_j^2 > C_\alpha^{b^*}\right\} = \alpha, \tag{15}$$

then

$$\mathcal{C}_n^{pt} = \left\{\theta\in\Theta : n\hat{Q}_n(\theta) \le C_\alpha^{b^*}\right\}$$

satisfies (11).

In addition, when the variance of the binding moments is diagonal, a simpler test statistic, $n\tilde{Q}_n(\theta)$, can be used that is in this case asymptotically equivalent to $n\hat{Q}_n(\theta)$. Define

$$\tilde{Q}_n(\theta) \equiv \sum_{j=1}^{J} 1\left[\hat{E}_n\left\{m_j(z,\theta)\right\} < 0\right]\cdot\left[\hat{E}_n\left\{m_j(z,\theta)\right\}\right]^2/\hat{V}_{\theta,jj},$$

where $\hat{V}_{\theta,jj}$ is the j^{th} diagonal entry of $\hat{V}_\theta$, the estimated variance of $m_j(z,\theta)$. Moreover, the convergence in distribution of $n\tilde{Q}_n(\theta)$ to a chi-bar square random variable holds when V_θ is singular, as long as V_θ^* is nonsingular. The result is driven by the fact that since the binding constraints have a diagonal variance matrix, replacing off-diagonal elements of $\hat{V}_\theta$ with zero in $\hat{Q}_n(\theta)$ has no effect asymptotically. This modification of $\hat{Q}_n(\theta)$ gives $\tilde{Q}_n(\theta)$. The formal result is stated below.

Proposition 4 *Suppose that V_θ^* is diagonal and nonsingular for all $\theta\in\Theta^*$, $\sup_{\theta\in\Theta^*} b(\theta) = b^*$, and that (A1)-(A4) hold. Then $n\tilde{Q}_n(\theta)$ converges in distribution to a chi-bar square random variable and $\forall c\in\mathbb{R}$,*

$$\sup_{\theta\in\Theta^*}\lim_{n\to\infty}\mathbb{P}\left\{n\tilde{Q}_n(\theta) > c\right\} = \sum_{j=0}^{b^*} 2^{-b^*}\binom{b^*}{j}\Pr\left\{\chi_j^2 > c\right\}.$$

If $\theta\notin\Theta^*$ and each element of $\mathbb{E}\left[m(z,\theta)\,m(z,\theta)'\right]$ is finite, then for any constant $c>0$, $\lim_{n\to\infty}\mathbb{P}\left\{n\tilde{Q}(\theta) > c\right\} = 1$.

4.3 Implementation Summary

In this subsection, I briefly outline the steps required to compute a confidence set $\mathcal{C}_n^{pt}$ for θ_0 with asymptotic coverage of at least $1-\alpha$, when $\sup_{\theta\in\Theta^*} b(\theta) = b^*$ and assumptions (A1)-(A5) hold.

1. Compute the unique value of $C_\alpha^{b^*}$ such that

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

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

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

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

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

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 $n\hat{Q}_n(\theta) \le C_\alpha^{b^*}$, then $\theta \in \mathcal{C}_n^{pt}$. If $n\hat{Q}_n(\theta) > C_\alpha^{b^*}$, then $\theta \notin \mathcal{C}_n^{pt}$.

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 $C_\alpha^{b^*}$. 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 $n\hat{Q}_n(\theta)$. However, the characteristics of the confidence set will depend on the particular moment functions in any given application. If the moment functions are irregular, then it may be advantageous to employ an adaptive method for selecting grid points, such as the Metropolis-Hastings algorithm. In section 5, the confidence set can be characterized sufficiently well that use of a grid is unnecessary.

5 Monte Carlo Study

This section demonstrates the application and performance of the inferential method prescribed in the context of inference on the mean with missing data. An application to an incomplete model of oligopoly behavior with data from a cartel is given by Rosen (2006).

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

$$\begin{aligned} \theta &\geq \theta_L \equiv p \cdot \mu_1, \\ \theta &\leq \theta_U \equiv p \cdot \mu_1 + 1 - p, \end{aligned}$$

or, in the form of (3),

$$\begin{aligned} E[m_1(x,d,\theta)] &= E[\theta - xd] \geq 0, \\ E[m_2(x,d,\theta)] &= E[1 - d + xd - \theta] \geq 0. \end{aligned} \tag{16}$$

The identified set for θ_0 in this model is $\Theta^* = [\theta_L, \theta_U]$, and the variance of $m(x,d,\theta)$ is

$$V_\theta = V = var(-xd, xd - d) = \begin{pmatrix} \sigma_l^2 & \sigma_{lu} \\ \sigma_{lu} & \sigma_u^2 \end{pmatrix},$$

where $\sigma_l^2 = var(xd)$, $\sigma_u^2 = var(xd - d)$,and $\sigma_{lu} = cov(xd, d) - var(xd)$. $\hat{Q}_n(\theta)$ is given by

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

where $\hat{V}$ is the sample analog of V. Since $p < 1$, only one of $E[m_1(x,d,\theta)]$ or $E[m_2(x,d,\theta)]$ can be equal to zero. Thus, the maximum number of binding constraints is one, and V^* is a scalar. Because in this case the limit distribution of $n\hat{Q}_n(\theta)$ is a sum of only two terms, by the reasoning of Corollary 1 the weights are known exactly; each of the two terms of the summation have weight $\frac{1}{2}$. Applying this result, the cutoff value for $n\hat{Q}_n(\theta)$ needed to build a confidence set for θ_0 with at least $1-\alpha$ asymptotic coverage is the unique value of $C_\alpha^{b^*}$ that solves

$$\frac{1}{2}\Pr\left\{\chi_0^2 > C_\alpha^{b^*}\right\} + \frac{1}{2}\Pr\left\{\chi_1^2 > C_\alpha^{b^*}\right\} = \alpha.$$

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

$$\frac{1}{2}\Pr\left\{\chi_1^2 > C_\alpha^{b^*}\right\} = \alpha.$$

Algebraic manipulation of $n\hat{Q}_n(\theta)$ in this context yields a simple analytical form for the associated confidence set:

$$\mathcal{C}_n^{MI} = \left[\hat{\theta}_l - z_{1-\alpha} \cdot \hat{\sigma}_l/\sqrt{n},\ \hat{\theta}_u + z_{1-\alpha} \cdot \hat{\sigma}_u/\sqrt{n}\right].$$

where $z_{1-\alpha}$ is the $1-\alpha$ quantile of the standard normal distribution, $\hat{\sigma}_l$ and $\hat{\sigma}_u$ are sample analogs of σ_l and σ_u, $\hat{\theta}_l = \hat{E}_n[xd]$, and $\hat{\theta}_u = \hat{E}_n[1-d+xd]$. This confidence set is straightforward to compute and no grid of candidate parameter values is needed to construct it.

I simulate i.i.d. draws of (x,d) in order to compare confidence regions constructed according to the moment inequality approach to those of Imbens and Manski (2004). The two methods yield nearly identical results. Let the moment inequality confidence set of level α be denoted $\mathcal{C}_\alpha^{MI}$, for moment inequalities, and the Imbens/Manski confidence set $\mathcal{C}_\alpha^{IM}$. The sets $\mathcal{C}_\alpha^{IM}$ are constructed as described in section 4 of their paper. That is the confidence sets constructed according to their method are:

$$\mathcal{C}_n^{IM} = \left[\hat{\theta}_l - \bar{C}_n \cdot \hat{\sigma}_l/\sqrt{n}, \hat{\theta}_u + \bar{C}_n \cdot \hat{\sigma}_u/\sqrt{n}\right],$$

where $\bar{C}_n$ solves

$$\Phi\left(\bar{C}_n + \sqrt{n}\frac{\hat{\theta}_u - \hat{\theta}_l}{\max(\hat{\sigma}_u, \hat{\sigma}_l)}\right) - \Phi\left(-\bar{C}_n\right) = 1-\alpha. \tag{17}$$

Their sets have the additional property that their coverage is uniform over all $\theta \in [p \cdot \mu_1, p \cdot \mu_1 + 1 - p]$ and the population distribution P, even if p is not bounded away from 1.

I provide simulations under two different specifications for the distribution of (x,d). For the first specification, I draw x from the uniform$(0,1)$ distribution and d 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,d) are not independent of each other, so that missingness is not at random. In this case, x is distributed beta$(4,2)$ conditional on $d=0$, and beta$(2,4)$ when $d=1$. In this case, x tends to be higher when it is not observed; the conditional distribution of x given $d=0$ stochastically dominates that of x given $d=1$, with $\mathbb{E}[x|d=0] = 2/3$ and $\mathbb{E}[x|d=1] = 1/3$. The simulated sample data is then $\{(\tilde{x}_i, d_i) : i = 1, ..., n,\ \tilde{x}_i = x_i \text{ if } d_i = 1, \tilde{x}_i = \emptyset \text{ if } d_i = 0\}$. Since all values of θ_0 in the interval $[\theta_L, \theta_U]$ are observationally equivalent, a confidence set is only guaranteed to have correct coverage for θ_0 if it achieves the desired asymptotic coverage for *each* $\theta \in [\theta_L, \theta_U]$. The coverage frequencies reported here are thus the infimum of observed coverage frequencies over $\theta \in [\theta_L, \theta_U]$.

Tables 1 and 2 compare the empirical coverage of each of the two confidence sets for different choices of n, p, α when $(x,d) \sim F_1$, while tables 3 and 4 do the same for $(x,d) \sim F_2$. The number of repetitions is 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 differ most are those sets with nominal level 0.99. In this case, the coverage from the 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. The overall performance of the two approaches is comparable.

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:	$\mathcal{C}_n^{IM}$	$\mathcal{C}_n^{MI}$	$\mathcal{C}_n^{IM}$	$\mathcal{C}_n^{MI}$	$\mathcal{C}_n^{IM}$	$\mathcal{C}_n^{MI}$	$\mathcal{C}_n^{IM}$	$\mathcal{C}_n^{MI}$
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:	$\mathcal{C}_n^{IM}$	$\mathcal{C}_n^{MI}$	$\mathcal{C}_n^{IM}$	$\mathcal{C}_n^{MI}$	$\mathcal{C}_n^{IM}$	$\mathcal{C}_n^{MI}$	$\mathcal{C}_n^{IM}$	$\mathcal{C}_n^{MI}$
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|d=1 is distributed beta(2,4) and x|d=0 is distributed beta(4,2).

Target Coverage ($p = 0.7$)	0.75		0.85		0.95		0.99	
Actual Coverage for θ_0:	$\mathcal{C}_n^{IM}$	$\mathcal{C}_n^{MI}$	$\mathcal{C}_n^{IM}$	$\mathcal{C}_n^{MI}$	$\mathcal{C}_n^{IM}$	$\mathcal{C}_n^{MI}$	$\mathcal{C}_n^{IM}$	$\mathcal{C}_n^{MI}$
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|d=1 is distributed beta(2,4) and x|d=0 is distributed beta(4,2).

Target Coverage ($p = 0.9$)	0.75		0.85		0.95		0.99	
Actual Coverage for θ_0:	$\mathcal{C}_n^{IM}$	$\mathcal{C}_n^{MI}$	$\mathcal{C}_n^{IM}$	$\mathcal{C}_n^{MI}$	$\mathcal{C}_n^{IM}$	$\mathcal{C}_n^{MI}$	$\mathcal{C}_n^{IM}$	$\mathcal{C}_n^{MI}$
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

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 easy to implement, as the cutoff values used to invert the test statistic are based on an analytical asymptotic distribution and do not require bootstrapping, subsampling, simulation, or tuning parameters to compute.

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. Upper bounds that are strictly smaller than the total number of inequalities are common in settings with partial identification. This is used to provide an upper bound on the $1-\alpha$ critical value for the test of interest which is easy to compute. In some cases, the method may be asymptotically conservative, in the sense that asymptotic coverage may be greater than the nominal level. Methods that provide asymptotically exact critical values may in these cases be preferred, though these typically involve greater computation in practice and may employ tuning parameters. There is thus a trade-off between ease of implementation and precision of the asymptotic approximation employed, as discussed in section 4.

This paper focuses on building confidence sets for the parameter of interest θ_0. Some other approaches to inference have resulted in confidence sets as well as confidence collections for Θ^*; see Beresteanu and Molinari (2008) for the latter. These are each conceptually different, and 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 perform inference on Θ^*. In addition, this paper, like much of the literature to date, has focussed on inference based on a finite number of moment restrictions. It seems an important direction for future research would be to devise inferential methods that can accommodate an infinite set of unconditional restrictions asymptotically, as are implied by conditional moment inequalities with continuous conditioning variables.

Appendix: Proofs

Proposition 1

Proof. Let $\hat{q}_n(\theta,t) \equiv \left(\hat{E}[m(z,\theta)]-t\right)' \hat{V}_\theta^{-1}\left(\hat{E}[m(z,\theta)]-t\right)$, so that $\hat{Q}_n(\theta) = \min_{t\geq 0} \hat{q}_n(\theta,t)$. Similarly, let $q(\theta,t) \equiv (\mathbb{E}[m(z,\theta)]-t)' V_\theta^{-1}(\mathbb{E}[m(z,\theta)]-t)$, so that $Q(\theta) = \min_{t\geq 0} q(\theta,t)$. Fix θ. Uniqueness of $t^*(\theta)$ follows from the strict convexity of $q(\theta,t)$ in t, guaranteed by (A5), and the fact that the minimizer of a strictly convex function on a convex set is unique. Consistency of $\hat{V}_\theta$ yields that $t_n^*(\theta)$, the minimizer of $\hat{q}_n(\theta,t)$, is unique with probability going to 1 as $n\to\infty$, since $\hat{V}_\theta$ is positive definite (and therefore $\hat{q}_n(\theta, t)$ is strictly convex) with probability approaching 1 under (A5). (5), (6), (7) and a Slutsky Theorem imply that $\hat{q}_n(\theta, t) \xrightarrow{p} q(\theta, t)$ pointwise for each θ, t. Since $\hat{q}_n(\theta, t)$ is convex in t, Theorem 2.7 of Newey and McFadden (1994) implies that $\hat{q}_n(\theta, t)$ converges uniformly in $t > 0$ to $q(\theta, t)$ for fixed θ. In addition, uniform convergence holds over any compact set $[0, T]$ by the continuity of $q(\theta, t)$ in t. Therefore $\hat{q}_n(\theta, t) \xrightarrow{p} q(\theta, t)$ uniformly over $t \geq 0$, so that $\hat{Q}_n(\theta) \xrightarrow{p} Q(\theta)$, further implying convergence in probability of the minimizer over $t \geq 0$ of $\hat{q}_n(\theta, t)$ to that of $q(\theta, t)$, i.e. $\hat{t}_n^*(\theta) \xrightarrow{p} t^*(\theta)$. This establishes that assumption 1 of Andrews (1999) holds, and his assumptions 2, and 3 follow because $\hat{q}_n(\theta, t)$ is quadratic in t and by (7). Thus Andrews (1999) Theorem 1 implies that $\sqrt{n}\left(\hat{t}_n^*(\theta) - t^*(\theta)\right) = O_p(1)$. ■

Proposition 2

Proof. The first result follows from pointwise convergence of $\hat{Q}_n$ to Q and Newey and McFadden (1994), Theorem 2.8. Set consistency in the Hausdorff metric under the stated conditions follows from Chernozhukov, Hong, and Tamer (2007), Theorem 3.1. ■

As a preliminary step to proposition 3, I first prove the following lemma.

Lemma 1

Consider the minimization problem

$$QP = \min_{t \in \mathbb{R}^J} (x - t)' V^{-1} (x - t) \text{ s.t. } t_1 \geq 0, \tag{18}$$

where $x, t \in \mathbb{R}^J$, and $x_1, t_1 \in \mathbb{R}^b$, $b \leq J$, s.t. $t = (t_1', t_2')'$ and $x = (x_1', x_2')'$. QP is a quadratic program in which the first b components of the minimand t are subject to nonnegativity constraints. In the application of the lemma, b will correspond to the number of elements of $\mathbb{E}[m(z, \theta)]$ equal to zero. 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

$$QP = \min_{t_1 \in \mathbb{R}_+^b} (x_1 - t_1)' V_{11}^{-1} (x_1 - t_1). \tag{19}$$

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

$$QP = (x - t^*)' \Lambda (x - t^*) .$$

The Kuhn-Tucker conditions for (18) are

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

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

By conditions (i) and (ii),

$$-\Lambda_{11} (x_1 - t_1^*) - \Lambda_{12} (x_2 - t_2^*) \geq 0, \tag{20}$$
$$-\Lambda_{21} (x_1 - t_1^*) - \Lambda_{22} (x_2 - t_2^*) = 0. \tag{21}$$

Solving for $(x_2 - t_2^*)$, the latter condition is

$$(x_2 - t_2^*) = -\Lambda_{22}^{-1} \Lambda_{21} (x_1 - t_1^*) . \tag{22}$$

Now

$$\begin{aligned} QP &= (x - t^*)' \Lambda (x - t^*) \\ &= (x_1 - t_1^*)' \Lambda_{11} (x_1 - t_1^*) + (x_1 - t_1^*)' \Lambda_{12} (x_2 - t_2^*) + (x_2 - t_2^*)' [\Lambda_{21} (x_1 - t_1^*) + \Lambda_{22} (x_2 - t_2^*)] \\ &= (x_1 - t_1^*)' \Lambda_{11} (x_1 - t_1^*) + (x_1 - t_1^*)' \Lambda_{12} (x_2 - t_2^*) , \end{aligned}$$

by (21). Now using (22) it follows that

$$\begin{aligned} QP &= (x_1 - t_1^*)' \Lambda_{11} (x_1 - t_1^*) - (x_1 - t_1^*)' \Lambda_{12} \left[\Lambda_{22}^{-1} \Lambda_{21} (x_1 - t_1^*)\right] \\ &= (x_1 - t_1^*)' \left[\Lambda_{11} - \Lambda_{12} \Lambda_{22}^{-1} \Lambda_{21}\right] (x_1 - t_1^*) \\ &= (x_1 - t_1^*)' V_{11}^{-1} (x_1 - t_1^*) , \end{aligned}$$

where the last equality follows by the partition inverse result.[3] All that remains is to show that t_1^* minimizes (19): $\min (x_1 - t_1)' V_{11}^{-1} (x_1 - t_1)$ 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 (19) are for $j = 1, ..., b$,

$$\text{either } t_j^* = 0 \text{ and } \left[-V_{11}^{-1} (x_1 - t_1^*)\right]_j \geq 0, \text{ or } t_j^* > 0 \text{ and } \left[-V_{11}^{-1} (x_1 - t_1^*)\right]_j = 0.$$

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

$\Longleftrightarrow$

$$\begin{aligned} \text{either } t_j^* &= 0 \text{ and } \left\{-\left[\Lambda_{11} - \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21}\right](x_1 - t_1^*)\right\}_j \geq 0, \\ \text{or } t_j^* &> 0 \text{ and } \left\{-\left[\Lambda_{11} - \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21}\right](x_1 - t_1^*)\right\}_j = 0. \end{aligned}$$

By (22), this is equivalent to condition (i) from the Kuhn-Tucker conditions for the initial program (18). ■

With Lemma 1 in hand, I now prove Proposition 3.

Proposition 3

Proof. Suppose $\theta \in \Theta^*$. Let $b = b(\theta)$ be the number of components of $\mathbb{E}[m(z,\theta)]$ equal to zero, and let

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

and

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

Then

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

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

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

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

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

and by Lemma 1,

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

Under (A1) and (A4) $v_n^* \xrightarrow{d} v$ where $v \sim N(0, V_\theta^*)$. By a continuous mapping theorem it follows that

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

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

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

Now suppose that $\theta \notin \Theta^*$, so that there exists $k \in \{1, ...J\}$ such that $\mathbb{E}[m_k(z,\theta)] < 0$. Assume (A1)-(A5). Let $\mu(\theta) \equiv \mathbb{E}[m(z,\theta)]$, and let $\hat{\Lambda}_\theta = \hat{V}_\theta^{-1}$, and $\Lambda_\theta = V_\theta^{-1}$.[4] Then,

$$\begin{aligned} \mathbb{P}\left\{n\hat{Q}_n(\theta) > C_\alpha^{b^*}\right\} &= \mathbb{P}\left\{n \min_{t \geq 0} \ \left[\hat{E}_n[m(z,\theta)] - t\right]' \hat{\Lambda}_\theta \left[\hat{E}_n[m(z,\theta)] - t\right] > C_\alpha^{b^*}\right\} \\ &= \mathbb{P}\left\{\min_{s} [v_n - s]' \hat{\Lambda}_\theta [v_n - s] > C_\alpha^{b^*} : s \geq -\sqrt{n}\mu(\theta)\right\} \end{aligned} \tag{23}$$

where $v_n = \sqrt{n}\left\{\hat{E}_n[m(z,\theta)] - \mu(\theta)\right\}$. Let s_n^* be the unique value of s that solves this minimization problem, so that

$$\mathbb{P}\left\{n\hat{Q}_n(\theta) > C_\alpha^{b^*}\right\} = \mathbb{P}\left\{[v_n - s_n^*]' \hat{\Lambda}_\theta [v_n - s_n^*] > C_\alpha^{b^*}\right\}.$$

Let Γ_θ be the orthogonal matrix that diagonalizes Λ_θ, so that $\Gamma_\theta \Lambda_\theta \Gamma_\theta'$ is a diagonal matrix with diagonal entries equal to the eigenvalues of Λ_θ, i.e. $\Gamma_\theta \Lambda_\theta \Gamma_\theta' = diag(d_{\theta,1}, ..., d_{\theta,J})$, where the $d_{\theta,j}$ are the eigenvalues of Λ_θ. Since Λ_θ is positive definite, each $d_{\theta,j} > 0$. Such a matrix Γ_θ exists by Corollary 21.5.9 of Harville (1997). Then

$$\begin{aligned} [v_n - s_n^*]' \hat{\Lambda}_\theta [v_n - s_n^*] &= [v_n - s_n^*]' \Lambda_\theta [v_n - s_n^*] + o_p(1) \\ &= \sum_{j=1}^{J} \left(\left[(v_n - s_n^*)' \Gamma_\theta^{-1}\right]_j\right)^2 d_{\theta,j} + o_p(1). \end{aligned}$$

The constraint $s \geq -\sqrt{n}\mu(\theta)$ in (23), implies that the k-th component of s_n^* diverges to $-\infty$. Since $v_n = O_p(1)$, $n\hat{Q}_n(\theta)$ diverges to ∞ and $\lim_{n \to \infty} \mathbb{P}\left\{n\hat{Q}_n(\theta) > C_\alpha^{b^*}\right\} = 1$. ■

[4] If V_θ or $\hat{V}_\theta$ are singular a positive definite generalized inverse may be used, and the proof goes through unchanged. Such a generalized inverse exists by Lemma 14.4.1 of Harville (1997).

Corollary 1

Proof. This follows by Perlman (1969), Theorem 8.1. ■

Corollary 2

Proof. The first part, (13), 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 identity matrix, so that $w(b, j, V^*) = w(b, j, I_b)$. The second part, (13), 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

$$\begin{aligned}
\sup_{\theta\in\Theta^*} \lim_{n\to\infty} \mathbb{P}\left\{n\hat{Q}_n(\theta) \geq c\right\} &= \sup_{\theta\in\Theta^*} \sum_{j=0}^{b(\theta)} w\left(b(\theta), b(\theta) - j, V_\theta^*\right) \Pr\left\{\chi_j^2 \geq c\right\} \\
&= \sup_{\theta\in\Theta^*} \sum_{j=0}^{b(\theta)} 2^{-b(\theta)} \binom{b(\theta)}{j} \Pr\left\{\chi_j^2 \geq c\right\} \\
&\leq \sum_{j=0}^{b^*} 2^{-b^*} \binom{b^*}{j} \Pr\left\{\chi_j^2 \geq c\right\}.
\end{aligned}$$

■

Proposition 4

Proof. Let Λ_θ ($\tilde{\Lambda}_\theta$) be a diagonal matrix with j^{th} diagonal entry $1/V_{\theta,jj}$ ($1/\hat{V}_{\theta,jj}$), the inverse of the (estimated) variance of $m(z,\theta)$. Assume (A1)-(A4) and that V_θ^* is diagonal with all diagonal entries positive. Then

$$\begin{aligned}
n\tilde{Q}_n(\theta) &= n\sum_{j=1}^{J} 1\left[\hat{E}_n\left\{m_j(z,\theta)\right\} < 0\right] \cdot \left(\hat{E}_n\left\{m_j(z,\theta)\right\}\right)^2 / \hat{V}_{\theta,jj} \\
&= n \min_{t\geq 0} \left[\hat{E}_n\left[m(z,\theta)\right] - t\right]' \tilde{\Lambda}_\theta \left[\hat{E}_n\left[m(z,\theta)\right] - t\right].
\end{aligned}$$

The proof of Proposition 3 goes through unchanged, as $\tilde{\Lambda}_\theta \xrightarrow{p} \Lambda_\theta$, with the partition inverse result used to prove lemma 1 applied to Λ_θ. ■

References

Anderson, T., and H. Rubin (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 20, 46–63.

——— (1950): "The Asymptotic Properties of Estimates of the Parameters of a Single Equation system in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 21(4), 570–582.

Andrews, D. W. K. (1999): "Estimation When a Parameter is on a Boundary," *Econometrica*, 67(6), 1341–1383.

Andrews, D. W. 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, Cowles Foundation.

Andrews, D. W. K., and P. Guggenberger (2005): "The Limit of Finite Sample Size and A Problem with Subsampling," working paper, Cowles Foundation.

——— (2007): "Validity of Subsampling and Plug-In Asymptotic Inference for Parameters Defined by Moment Inequalities," working paper, Cowles Foundation.

Andrews, D. W. K., and G. Soares (2007): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," working paper, Cowles Foundation.

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.

Bartholomew, D. (1959a): "A Test of Homogeneity for Ordered Alternatives I," *Biometrika*, 46, 36–48.

——— (1959b): "A Test of Homogeneity for Ordered Alternatives II," *Biometrika*, 46, 328–335.

Beresteanu, A., and F. Molinari (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 76(4), 763–814.

Blundell, R., M. Browning, and I. Crawford (2008): "Best Nonparametric Bounds on Demand Responses," *Econometrica*, forthcoming.

Blundell, R., A. Gosling, H. Ichimura, and C. Meghir (2007): "Changes in the Distribution of Male and Female Wages Accounting for Unemployment Composition Using Bounds," *Econometrica*, 75(2), 323–363.

Bontemps, C., T. Magnac, and E. Maurin (2007): "Set Identified Linear Models," working paper, IDEI.

Bugni, F. (2007): "Bootstrap Inference in Partially Identified Models," working paper, Northwestern University.

Canay, I. (2007): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," working paper, University of Wisconsin.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75(5), 1243–1284.

Ciliberto, F., and E. Tamer (2004): "Market Structure and Multiple Equilibria in Airline Markets," working paper, Northwestern University.

Dufour, J.-M. (1997): "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica*, 65(6), 1365–1387.

Fan, Y., and S. Park (2007): "Confidence Sets for Some Partially Identified Parameters," working paper, Vanderbilt 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.

Galichon, A., and M. Henry (2006a): "Dilation Bootstrap: A Natural Approach to Inference in Incomplete Models," working paper, Harvard University.

——— (2006b): "Inference in Incomplete Models," working paper, Harvard University.

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.

Guggenberger, P., J. Hahn, and K. Kim (2008): "Specification Testing Under Moment Inequalities," *Economics Letters*, 99(2), 375–378.

Guggenberger, P., and R. J. Smith (2005): "Generalized Empirical Likelihood Estimators and Tests Under Partial, Weak, and Strong Identification," *Econometric Theory*, 21, 667–709.

Harville, D. A. (1997): *Matrix Algebra From a Statistician's Perspective.* Springer, New York.

Ho, K. (2008): "Insurer-Provider Networks in the Medical Care Market," *American Economic Review*, forthcoming.

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(6), 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.

Kleibergen, F. (2005): "Testing Parameters in GMM Without Assuming that They Are Identified," *Econometrica*, 74(4), 1103–1123.

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.

Manski, C. F. (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. (2008): "Missing Treatments," *Journal of Business and Economic Statistics*, forthcoming.

Moon, H. R., and F. Schorfheide (2005): "Boosting Your Instruments: Estimation with Overidentifying Inequality Moment Conditions," working paper, University of Pennsylvania.

Newey, W. K., and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *The Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. 4, pp. 193–281. North-Holland.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2004): "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.

Romano, J. P., and A. M. Shaikh (2006): "Inference for the Identified Set in Partially Identified Econometric Models," working paper, Stanford University.

——— (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Planning and Statistical Inference*, 138, 2786–2807.

Rosen, A. M. (2006): "Identification and Estimation of Firms' Marginal Cost Functions with Incomplete Knowledge of Strategic Behavior," CEMMAP working paper CWP03/07.

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

Staiger, D., and J. H. Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65(3), 557–586.

Stock, J. H., and J. H. Wright (2000): "GMM with Weak Identification," *Econometrica*, 68(5), 1055–1096.

Stoye, J. (2007): "More on Confidence Regions for Partially Identified Parameters," CEMMAP working paper CWP11/08.

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 Constraints in Nonlinear Models," *Econometrica*, 59(4), 981–995.