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

March 2007

Abstract

This paper proposes a new 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 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

*This is a revised version of the first chapter of my Northwestern PhD dissertation. I thank Richard Blundell, Andrew Chesher, Joel Horowitz, Sokbae Lee, Chuck Manski, Rob Porter, Jörg Stoye, two referees and an Associate Editor for comments and suggestions. In addition, I have benefited from comments from numerous seminar participants. I am especially grateful to Elie Tamer for continued feedback and encouragement. Financial support from the Robert Eisner Memorial Fellowship and the Center for the Study of Industrial Organization at Northwestern is gratefully acknowledged. Any and all errors are my own.

†email: adam.rosen@ucl.ac.uk

[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.

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.[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.

This paper proposes a way to perform inference 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 $\mathbb{E}\left[m\left(y, x, \theta_0\right)\right] \geq 0$, where y, x are observables 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.

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 done in the case where θ_0 is univariate and interval-identified. Methods applicable in more general contexts (i.e. when θ_0 is multivariate) have relied on subsampling, bootstrapping, or simulation for 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 a large grid of parameter values. A Monte Carlo study in section 5.2 demonstrates the performance of the inferential approach in this paper relative to inference based on subsampling. The merit of the method in this paper is demonstrated both in terms of computation time and coverage accuracy in finite samples for the particular model studied. In general, however, resampling methods have the advantage that they are consistent in a wider class of models.

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

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

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. $\mathcal{C}_n^{pt}$ is then guaranteed to contain each θ that is observationally equivalent to θ_0 with at least probability $1-\alpha$ in repeated sampling. 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 also contain θ_0 with at least probability $1-\alpha$ for n sufficiently large, i.e. $\lim_{n\to\infty} \mathbb{P}\left\{\theta_0 \in \mathcal{C}_n^{pt}\right\} \geq 1 - \alpha$.[3]

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).[4] 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.[5] The inferential method is relatively straightforward to implement in practice in many cases of interest, which is demonstrated with two specific examples 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, the test statistic $\hat{Q}_n(\theta)$ is not asymptotically pivotal because its asymptotic distribution depends on the variance of those components of $m(y,x,\theta)$ 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 $m(y,x,\theta)$, J, is clearly an upper bound, but in models with partially identified parameters there is often a smaller upper bound and that can

[3]Pointwise confidence sets that satisfy (2) are one of several types of confidence sets that have been studied in the literature on partial identification. The prior literature on construction of confidence sets of this type as well as others is discussed in the literature review of the following subsection.

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

[5]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.

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 $1-\alpha$. It is however shown in section 4 that even when this issues arises, the test on which the confidence sets are based is consistent.

Some methods for inference with partial identification focus on inference on the identified set Θ^*, rather than the partially identified parameter θ_0. Of those that are also for θ_0, many others may also admit some degree of conservatism, at least in some cases. Exceptions include Imbens and Manski (2004) and subsampling methods. However, Imbens and Manski (2004) is applicable only in settings where θ_0 is univariate and interval-identified. In this case, the inferential method of this paper is nearly identical to Imbens and Manski (2004), and this is reflected in Monte Carlo experiments in section 5.1. Subsampling, on the other hand, is applicable in a wide range of contexts, but has its own drawbacks, such as computational intensity and choice of tuning parameters. In addition, the rate of convergence of subsampling distributions in regular cases is typically slower than that achieved by analytical asymptotic approximations when they exist.[6] Thus, it is in general unclear how the accuracy of different methods for pointwise inference in the context of partial identification compares in finite samples, and to date different methods have not been systematically compared. Even in cases where methods for pointwise inference are "conservative" asymptotically, they may still perform well in finite samples relative to methods that are not conservative asymptotically, if those methods have slower convergence rates. In section 5.2 I compare inference based on the method of this paper to a subsampling procedure for the case of linear regression with interval-measured outcomes. In the particular experimental design studied, the method of this paper performs favorably. However, subsampling is applicable to a larger class of models, and can also be used to perform inference on the identified set rather than the point θ_0, if that is the researcher's goal.

The paper proceeds as follows. The remainder of the introduction reviews the literature on inference on partially identified parameters. Section 2 presents the moment inequality model. Section 3 describes the pointwise hypothesis testing procedure. Section 4 then presents two ways to construct confidence sets based on the hypothesis test of section 3. This includes a discussion of when the confidence sets are potentially conservative, as well as the result that the test is consistent. Section 5 presents two examples and investigates the performance of confidence sets in these models via Monte Carlo simulations. This includes a comparison to the non-resampling based confidence sets of Imbens and Manski (2004) for the case of the mean with missing data, and to the confidence sets of Chernozhukov, Hong, and Tamer (2004) for the case of linear regression with censored outcomes. Section 6 concludes and offers avenues for continued research. All proofs are in the Appendix.

[6] For related discussion, see Horowitz (2001) section 2.2 and Andrews and Guggenberger (2006). For general results on the asymptotic properties of subsampling, see Politis and Romano (1994), and Politis, Romano, and Wolf (1999).

1.1 Related Literature

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 identified set 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 identified set, such as those constructed by Manski and Nagin (1998) and Horowitz and Manski (2000). Also in the case where the parameter of interest is univariate and interval-identified, Imbens and Manski (2004) show how to construct asymptotic pointwise confidence intervals for θ_0, rather than for the entire identified set Θ^*. 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. They also consider sets that satisfy

$$\lim_{n\to\infty} \inf_{\theta\in\Theta^*} \mathbb{P}\left\{\theta \in \tilde{\mathcal{C}}_n^{ufrm}\right\} = 1-\alpha. \tag{3}$$

Such sets are guaranteed to contain any element of the identified set uniformly over Θ^* as $n \to \infty$. Sets $\tilde{\mathcal{C}}_n^{ufrm}$ also satisfy the pointwise statement (2) with weak inequality. These sets are weakly larger than $\mathcal{C}_n^{pt}$, but also more robust in that they give uniform coverage.

In a more general context, where θ_0 is possibly multivariate, Chernozhukov, Hong, and Tamer (2004) 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 the minimizers of an objective function. They further show how to modify their procedure to build pointwise confidence sets for the parameter of interest. Romano and Shaikh (2006a, 2006b) also employ subsampling to construct confidence sets for both Θ^* and θ_0. They derive the validity of an iterative step-down procedure for inference. While the inferential approaches of Chernozhukov, Hong, and Tamer (2004), and Romano and Shaikh (2006a, 2006b) are applicable in a very general class of models, their reliance on subsampling may in some cases be computationally intensive.

There are many additional papers that relate to the problem of inference on partially identified parameters. Pakes, Porter, Ho, and Ishii (2006) study the use of moment inequalities to perform inference on θ_0 in models with agents who make optimal, or approximately optimal decisions in models used in empirical industrial organization. They develop conservative confidence sets for model parameters, and use their technique in two applications with multiple equilibria: an investigation of how banks choose their ATM locations, and an analysis of the determination of HMO hospital networks in the United States.[7] To perform inference, Pakes, Porter, Ho, and Ishii (2006) use simulations from a multivariate normal distribution to approximate the distribution of the moments in their model. 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

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

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. 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. They show how to build confidence collections for the identified set in these models. They further show how to modify this approach to build conservative confidence sets for the parameter of interest. Blundell, Browning, and Crawford (2006) use moment inequalities implied by the strong axiom of revealed preference in conjunction with consumers' observed consumption bundles. To perform inference, they use subsampling to approximate the distribution of a minimum distance function similar to the one used here.

In this paper, I focus on models that are comprised of a finite number of moment inequalities. This class of models includes many examples from the econometrics literature, dating back at least to Frisch (1934), who derived bounds on the slope parameter of a simple linear regression model with measurement error. 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 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[8], and the case of inference on the mean of a univariate distribution with missing data, studied by Manski (1989) and Imbens and Manski (2004).

2 The Model

Let $\{(x_i, y_i) : i = 1, ..., n\}$ denote a random sample of observations of (x, y) distributed with population 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 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 characteristics of individual i in the population that are missing.[9] 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. 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.

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

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

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

$$\mathbb{E}\left[m\left(y,x,\theta_0\right)\right]=\mathbb{E}\begin{bmatrix} m_1\left(y,x,\theta_0\right) \\ \vdots \\ m_J\left(y,x,\theta_0\right)\end{bmatrix}\geq\begin{pmatrix}0\\ \vdots \\ 0\end{pmatrix}. \tag{4}$$

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

Assumption A1 (random sampling) $Z\equiv\{(x_i,y_i):i=1,...,N\}$ 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) $\mathbb{E}\left[m\left(y,x,\theta_0\right)\right]\geq 0$, where $m\left(\cdot,\cdot,\cdot\right):\mathbb{R}^p\times\mathbb{R}^s\times\Theta\to\mathbb{R}^J$.

The assumptions above yield the following identified set for θ_0.

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

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

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 rejected. If Θ^* 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

$$\begin{aligned} H_0 &: \mathbb{E}\left[m\left(y,x,\theta\right)\right]\geq 0 \\ H_1 &: \mathbb{E}\left[m\left(y,x,\theta\right)\right]\ngeq 0, \end{aligned} \tag{5}$$

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 $\sup_{\theta\in\Theta^*}\lim_{n\to\infty}\Pr\left\{n\hat{Q}_n\left(\theta\right)>C^*_\alpha\right\}=\alpha$. This implies that the set $\mathcal{C}^{pt}_n\equiv\left\{\theta:n\hat{Q}_n\left(\theta\right)\leq C^*_\alpha\right\}$

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 $\lim_{n\to\infty}\mathbb{P}\left\{\theta_0\in\mathcal{C}_n^{pt}\right\}\geq 1-\alpha$. The next section explains how this 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 (5) 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 strict inequality. 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 (4), I employ 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 $m(y,x,\theta)$, and where the minimization is taken over the vector t of dimension J. 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, $\hat{Q}_n(\theta)$ it 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.[10]

[10] The necessary and sufficient Kuhn-Tucker conditions are that for each $j=1,...,J$,

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

or

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

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. As such, the statistic $\hat{Q}_n(\theta)$ satisfies the definition of a modified minimum distance (MMD) objective function, as defined by Manski and Tamer (2002). This is because the population version of $\hat{Q}_n(\theta)$ (and the probability limit of $\hat{Q}_n(\theta)$ under sufficient regularity) 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 $m(y,x,\theta)$. $Q(\theta)$ measures the distance of θ from Θ^* because $Q(\theta) = 0$ if and only if $\mathbb{E}[m(y,x,\theta)] \geq 0$, and is otherwise positive. Manski and Tamer (2002) derive conditions for consistency of MMD estimation of identified sets, and their results apply here. The focus of this paper is inference, yet in practice estimation proceeds inference, so the application of this result 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 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 notable difference because, as I show in Proposition 3, the asymptotic distribution of $n\hat{Q}_n(\theta)$ is degenerate except on the boundary of the null hypothesis.[11] 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, so that the boundary of the set $\{\theta : h(\theta) \geq 0\}$ is at least known.

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

Imposing these conditions, simplifies computation of $\hat{Q}_n(\theta)$ significantly.

[11] Andrews (2001) considers hypothesis tests when a parameter is on the boundary of the maintained hypothesis, rather than the null. However, the hypothesis test (5) 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.

Assumption A4 (finite variance of m on Θ^*) For some large K, $\sup_{\theta\in\Theta^*} \mathbb{E}\left[m(y,x,\theta)\, m(y,x,\theta)'\right]_{ij} < K < \infty$, i.e. each element of the matrix $\mathbb{E}\left[m(y,x,\theta)\, m(y,x,\theta)'\right]$ is bounded and finite for all $\theta \in \Theta^*$. This also implies that the moments $\mathbb{E}[m(y,x,\theta)]$ 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 $\mathbb{E}[m(y,x,\theta)]$, while assumption (A5) guarantees that V_θ is invertible. Under (A1) and (A4), it follows that for all $\theta \in \Theta^*$,

$$\hat{E}_n[m(y,x,\theta)] = \frac{1}{n}\sum_{i=1}^{n} m(y_i,x_i,\theta) \overset{a.s.}{\to} \mathbb{E}[m(y,x,\theta)], \tag{6}$$

$$\begin{aligned}\hat{V}_n[m(y,x,\theta)] &= \frac{1}{n}\sum_{i=1}^{n}\Big(m(y,x,\theta)-\hat{E}_n[m(y,x,\theta)]\Big)\Big(m(y,x,\theta)-\hat{E}_n[m(y,x,\theta)]\Big)' \\ &\overset{a.s.}{\to} var\{m(y,x,\theta)\} \equiv V_\theta,\end{aligned} \tag{7}$$

and

$$\sqrt{n}\left\{\hat{E}_n[m(y,x,\theta)] - \mathbb{E}[m(y,x,\theta)]\right\} \overset{d}{\to} N(0,V_\theta). \tag{8}$$

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 $m(y,x,\theta)$ 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 $\mathcal{C}_n^{pt}$ such that $\inf_{\theta\in\Theta^*}\lim_{n\to\infty}\Pr\left\{\theta\in\mathcal{C}_n^{pt}\right\} = 1-\alpha$, 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 coverage over Θ^*, i.e. sets such that $\lim_{n\to\infty}\inf_{\theta\in\Theta^*}\Pr\{\theta\in\mathcal{C}_n\} = 1-\alpha$, then stronger conditions would be needed.

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

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

and

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

[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 (6), (7), and (8) can be shown to hold at each $\theta \in \Theta^*$ for some sequence of constants $a_n \to \infty$ replacing $\sqrt{n}$.

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. Then for any $\theta \in \Theta$, $\hat{Q}_n(\theta) \xrightarrow{p} Q(\theta)$, and $\hat{t}_n^*(\theta) \xrightarrow{p} t_\theta^*(\theta)$, where $\hat{t}_n^*(\theta) \equiv \arg\min_{t \geq 0} \hat{q}_n(\theta, t)$, and $t_\theta^*(\theta) \equiv \arg\min_{t \geq 0} q(\theta, t)$.*

Proposition 1 follows from the concavity and continuity of the $\hat{q}_n(\theta, t)$ in t. If, in addition, the convergence of $\hat{Q}_n(\theta) \xrightarrow{p} Q(\theta)$ is uniform over Θ, then the results of Manski and Tamer (2002) 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 random variables such that $\epsilon_n \xrightarrow{a.s.} 0$ and*

$$\sup_{\theta \in \Theta^*} \left| \hat{Q}_n(\theta) - Q(\theta) \right| / \epsilon_n \xrightarrow{a.s.} 0.$$

Then

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

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

The first step to deriving distributional results for $n\hat{Q}_n(\theta)$ under H_0 shows formally that only those components of $\mathbb{E}[m(y, x, \theta)]$ exactly equal to zero have a non-negligible contribution asymptotically. Before proceeding with the distributional result, I define some additional notation. For expositional convenience, I refer to the subset of the J constraints that characterize the identified set as the set of binding constraints. Without loss of generality, let the first $b(\theta)$ constraints be the subset of binding constraints at θ, so that $\mathbb{E}[m_j(y, x, \theta)] = 0$, $j = 1, ..., b(\theta)$, and $\mathbb{E}[m_j(y, x, \theta)] > 0$, $j = b(\theta) + 1, ..., J$. Let $m^*(y, x, \theta) = \left(m_1(y, x, \theta), ..., m_{b(\theta)}(y, x, \theta)\right)'$ denote the subvector of moments that have mean zero, and let $V_\theta^* = var(m^*(y, x, \theta))$. Let $b \equiv b(\theta_0)$, $V \equiv var(m(y, x, \theta_0))$, and $V^* \equiv var(m^*(y, x, \theta_0))$. Finally, $\Pr\left\{\chi_j^2 \geq c\right\}$ 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 *Under assumptions (A1)-(A5), for any value of $\theta \in \Theta^*$, for any constant c,*

$$\lim_{n \to \infty} \mathbb{P}\left\{ n\hat{Q}_n(\theta) > c \right\} = \sum_{j=0}^{b(\theta)} w\left(b(\theta), b(\theta) - j, V_\theta^*\right) \Pr\left\{\chi_j^2 > c\right\}, \tag{9}$$

where $w\left(b(\theta), b(\theta) - j, V_\theta^*\right)$ is the weights function defined by Wolak (1987) and Kudo (1963) evaluated at $w\left(b(\theta), b(\theta) - j, V_\theta^*\right)$, and the χ_j^2 random variables of the summation are independent.

Corollary 1 *Under assumptions (A1)-(A5), $\forall\theta$ such that $\mathbb{E}\left[m\left(y,x,\theta\right)\right]>0$,*

$$\lim_{n\rightarrow\infty}\mathbb{P}\left\{n\hat{Q}_{n}\left(\theta\right)>0\right\}=0.$$

Proposition 3 closely follows Lemma 1 of Wolak (1991). The first step to the proof shows that the limiting distribution of $n\hat{Q}_{n}\left(\theta\right)$ 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 $n\rightarrow\infty$. 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 conditions (A1) - (A5) required for Proposition 3 are noticeably weak. They are not as strong as the conditions required for consistent set estimation, as the result is applicable pointwise over the identified set, while consistent set estimation relies on the asymptotic behavior of $\hat{Q}_{n}\left(\theta\right)$ jointly over all $\theta\in\hat{\Theta}_{n}^{*}$. In particular, the assumptions of Proposition 3 do not require that $\hat{Q}_{n}\left(\theta\right)$ converge to its population counterpart *uniformly* over θ, but merely pointwise over $\theta\in\Theta^{*}$. Unlike inference based on the asymptotics of m-estimators, this result does not rely on a characterization of the maximizer of a sample objective function around the maximizer of a population objective function. Rather, for any fixed value of θ, $n\hat{Q}_{n}\left(\theta\right)$ is a well-behaved function of sample moments evaluated at that value of θ. If $\theta\in\Theta^{*}$ then this function of sample moments converges in distribution to a chi-bar square random variable.

The weights function $w\left(b\left(\theta\right),b\left(\theta\right)-j,V\right)$ has arisen repeatedly in research on multivariate one-sided hypothesis testing. It is the probability that a random variable $Z\sim\mathcal{N}\left(0,V_{\theta}^{*}\right)$ has exactly j positive components. That is,

$$\begin{aligned}w\left(b\left(\theta\right),b\left(\theta\right)-j,V_{\theta}^{*}\right) &= \Pr\left\{Z\text{ has exactly }b\left(\theta\right)-j\text{ positive components}\right\}\\ &= \Pr\left\{Z\text{ has exactly }j\text{ components equal to zero}\right\}.\end{aligned}$$

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 $b\leq4$, or where V_{θ}^{*} is diagonal. More generally, closed-form expressions for the weights have not been obtained, but if V_{θ}^{*} and $b\left(\theta\right)$ were known, they could be approximated with arbitrary accuracy by means of simulation.[13]

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 in this case. An intuitive solution would be to plug consistent estimates into the weights function, but this won't work here because the CDF of the limit distribution given by

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

(9) is discontinuous in $b(\theta)$. This problem can, however, be overcome by considering the least favorable distribution of test statistic over Θ^*. Section 4 details how this can be done by using an upper bound for $b(\theta)$ to construct a cutoff value $C_\alpha^{b^*}$ such that

$$\inf_{\theta\in\Theta^*}\lim_{n\to\infty}\mathbb{P}\left\{n\hat{Q}_n(\theta)\le C_\alpha^{b^*}\right\}=1-\alpha, \tag{10}$$

or, in some cases, the more conservative

$$\inf_{\theta\in\Theta^*}\lim_{n\to\infty}\mathbb{P}\left\{n\hat{Q}_n(\theta)\le C_\alpha^{b^*}\right\}\ge 1-\alpha. \tag{11}$$

4 Computing Confidence Sets

This section provides two ways to compute cutoff values for $n\hat{Q}_n(\theta)$ 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 any software package 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 (10) with equality, thus ensuring exactly correct (i.e. not conservative) asymptotic coverage for the least favorable point in the identified set. 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, which are the 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 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.

A comment is in order regarding the aforementioned conservatism of the confidence sets based on these procedures. As discussed in the introduction, one cannot distinguish between any $\theta\in\Theta^*$ and θ_0 based on one's data, so that the goal of a confidence set $\mathcal{C}_n^{pt}=\left\{\theta:n\hat{Q}_n(\theta)\le C_\alpha^{b^*}\right\}$ for fixed cutoff $C_\alpha^{b^*}$ is to achieve

$$\inf_{\theta\in\Theta^*}\lim_{n\to\infty}\mathbb{P}\left\{\theta\in\mathcal{C}_n^{pt}\right\}=1-\alpha. \tag{12}$$

If equality is replaced by $\ge$, then $\mathcal{C}_n^{pt}$ is asymptotically conservative. The conservatism of the procedures below for computing $C_\alpha^{b^*}$, and thus constructing $\mathcal{C}_n^{pt}$, 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 $V_\theta^*=V$ that gives the highest (most conservative) possible value of $C_\alpha^{b^*}$. If this variance matrix is a member of $\{V_\theta^* : \theta \in \Theta^*\}$, then (12) 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 asymptotic coverage is at least $1-\alpha$, and the confidence set is asymptotically conservative. However, even in this case the set is not arbitrarily large, in the sense that a test based on the conservative cutoff is consistent against any fixed alternative, see section 4.3.

The case in which the variance of the binding moments V_θ^* at any $\theta \in \Theta^*$ is diagonal is an important special case. Indeed, in this case the source of conservatism discussed above is not present. As long as the maximal number of binding constraints is known, one can compute the worst-case variance matrix exactly, so that (12) is satisfied with equality. V_θ^* is in fact diagonal in both the mean with missing data and regression with interval-censored outcomes cases investigated in section 5. The Monte Carlo studies of Chernozhukov, Hong, and Tamer (2004), Beresteanu and Molinari (2006), and Romano and Shaikh (2006b) also fit in this category, as they perform inference on the returns-to-schooling parameter of a linear regression model with interval-measured wages as the outcome variable. These experiments feature a measure of educational attainment as their covariate, which is discrete, as is the covariate used in the experimental design of section 5.2.

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 (5). This cutoff value can then be used to build conservative, asymptotically valid confidence sets for θ_0. The following corollary provides the result.

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

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

The proof 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 $\Pr\left\{\chi^2_j > c\right\}$ 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 (9). 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 $\theta \in \Theta^*$.

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

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

$$\frac{1}{2}\Pr\left\{\chi^2_{b^*} > C_\alpha^{b^*}\right\} + \frac{1}{2}\Pr\left\{\chi^2_{b^*-1} > C_\alpha^{b^*}\right\} = \alpha, \tag{13}$$

Then

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

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

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

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 Implementation when V_θ^* is diagonal

When V_θ^* is a diagonal, then $w\left(b(\theta), b(\theta) - j, V_\theta^*\right)$ 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 the following corollary. This result also provides a smaller cutoff value for the hypothesis test (5) and thus a smaller confidence region when V_θ^* is diagonal.

Corollary 3 *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*

$$w\left(b(\theta), b(\theta) - j, V_\theta^*\right) = 2^{-b(\theta)}\binom{b(\theta)}{b(\theta) - j}, \tag{14}$$

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

$$\sup_{\theta\in\Theta^*} \lim_{n\to\infty} \mathbb{P}\left\{n\hat{Q}_n(\theta) > c\right\} = \sum_{j=0}^{b^*} 2^{-b^*}\binom{b^*}{j}\Pr\left\{\chi_j^2 > c\right\}. \tag{15}$$

Just as Corollary 2 provides a way to construct confidence sets for θ_0 so does Corollary 3 when V_θ^* is diagonal. If $C_\alpha^{b^*}$ solves

$$\sum_{j=0}^{b^*} 2^{-b^*}\binom{b^*}{j}\Pr\left\{\chi_j^2 > C_\alpha^{b^*}\right\} = \alpha, \tag{16}$$

then

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

satisfies (12).

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 asymptotically equivalent to $n\hat{Q}_n(\theta)$. Define

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

where $\hat{V}_{\theta,jj}$ is the j^{th} diagonal entry of $\hat{V}_\theta$, the estimated variance of $m_j(y,x,\theta)$. 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 $\forall c \in \mathbb{R}$,*

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

4.3 Consistency of the tests

All of the tests on which confidence sets in this section are based are consistent against any fixed alternative. Thus, even in the general case of section 4.1, where the pointwise coverage probability may asymptotically exceed $1-\alpha$, the confidence sets are not altogether arbitrary. The idea is that if $\theta \notin \Theta^*$, then $n\hat{Q}_n(\theta)$ "blows up" as $n \to \infty$. This result is given by the following Proposition.

Proposition 5 *Let $\theta \notin \Theta^*$, so that there exists $j \in \{1, ...J\}$ such that $\mathbb{E}\left[m_j(y,x,\theta)\right] < 0$. Then $\forall c < \infty$, if (A1)-(A5) hold*

$$\lim_{n\to\infty} \mathbb{P}\left\{n\hat{Q}_n(\theta) > c\right\} = 1.$$

In addition, if (A1)-(A4) hold, and V_θ^* is diagonal and nonsingular for all $\theta \in \Theta^*$, then $\forall c < \infty$

$$\lim_{n\to\infty} \mathbb{P}\left\{n\tilde{Q}_n(\theta) > c\right\} = 1.$$

In particular, the results holds for $c = C_\alpha^{b^*}$.

4.4 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)-(A4) hold.

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

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

   - If V^* 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^* 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 employed for choosing subsample grid points by Chernozhukov, Hong, and Tamer (2004). For the Monte Carlo experiments of section 5.2, I use a uniform grid over the parameter space. In section 5.1, the confidence set can be characterized sufficiently well that use of a grid is unnecessary.

4.5 Computational Considerations

In general, building confidence sets by testing H_0 over a large grid of values may be computational intensive, particularly if the parameter space is large and high-dimensional. However, the need to do this is present in other inferential approaches as well. In addition, resampling based methods also require that one compute one's test statistic over a fine grid of values for *each* resampling of the data in order to estimate the appropriate quantile of its distribution. Thus, for cases in which the need to test a large grid of parameter values is costly, the problem is more acute if resampling methods are used.

This is best illustrated by first considering the test of a single parameter θ. To test the hypothesis that $\theta \in \Theta^*$ using the chi-bar-square approximation, one computes $n\hat{Q}_n(\theta)$ once and compares it to the appropriate critical value from the chi-bar-square distribution. This is one computation of the objective function, and one computation of the chi-bar-square quantile (under 1 second in Matlab). To test this hypothesis via subsampling, a test statistic has to be computed for *each* subsample, for each value of θ in a particular level set of $n\hat{Q}_n(\theta)$ to compute the appropriate critical value, which is then compared with $n\hat{Q}_n(\theta)$. Thus, if g is the number of elements of the level set grid (which is necessarily large for accuracy) over which the statistic is subsampled, and B_n subsamples are drawn, this requires is $g \cdot B_n$ computations of the subsample statistic $b\hat{Q}_b(\theta)$, in addition to the computation of $n\hat{Q}_n(\theta)$. The $g \cdot B_n$ subsample computations take the place of the computation of the chi-bar-square critical value.[15] To construct a confidence region using either method requires that one compute $n\hat{Q}_n(\theta)$ over a grid of values in the parameter space, which in general is different from but contains the grid used for the subsample stage. Since both methods require this, the computational difference between the two methods for constructing confidence sets is entirely due to the $g \cdot B_n$ computations of the subsampling stage.

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 the researcher, but is consistently estimated by its sample analog. The goal is inference on $\theta_0 \equiv \mathbb{E}[x]$. Let $\mu_1 = \mathbb{E}[x|z = 1]$, which is identified by the

[15] Chernozhukov, Hong, and Tamer (2004) and Romano and Shaikh (2006b) also consider iterating the subsampling stage, which effectively magnifies the number of computations involved, though may improve the precision of inference.

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 (4),

$$\begin{aligned}E\left[m_1\left(x,z,\theta\right)\right] &= E\left[\theta - xz\right] \geq 0, \\ E\left[m_2\left(x,z,\theta\right)\right] &= E\left[1 - z + xz - \theta\right] \geq 0.\end{aligned} \tag{17}$$

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 (17). 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.[16] Applying this result, the cutoff value for $n\hat{Q}_n\left(\theta\right)$ needed to build a confidence set for

[16] 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 $C_\alpha^{b^*}$ 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 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, $\hat{\theta}_l = \hat{E}_n[xz]$, and $\hat{\theta}_u = \hat{E}_n[1-z+xz]$. This confidence set is straightforward to compute and no grid of candidate parameter values is needed to construct it.

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 methods yield nearly identical results.[17] 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

$$\Phi\left(\bar{C}_n + \sqrt{n}\frac{\hat{\theta}_u - \hat{\theta}_l}{\max(\hat{\sigma}_u, \hat{\sigma}_l)}\right) - \Phi\left(-\bar{C}_n\right) = 1-\alpha. \tag{18}$$

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

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

[17] Indeed, an earlier version of the paper showed analytically that the two confidence sets are nearly identical in this setting.

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 the bounds is contained in the two confidence regions.[18] 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.
   (b) Compute $\mathcal{C}_n^{IM}$ and $\mathcal{C}_n^{MI}$, which amounts to just computing their endpoints since they are both intervals.
      i. If $\theta_L \notin \mathcal{C}_n^{IM}$ increment REJ_L^{IM}, and if $\theta_U \notin \mathcal{C}_n^{MI}$ increment REJ_U^{IM}.
      ii. If $\theta_L \notin \mathcal{C}_n^{IM}$ increment REJ_L^{MI}, and if $\theta_U \notin \mathcal{C}_n^{MI}$ 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.

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 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

[18] Because the identified set is an interval, it is sufficient to check coverage of the endpoints to find the smallest coverage level for any point in the identified set.

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

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.

5.2 Example 2: Mean Regression with Interval-Censored Outcomes

In this subsection I consider the case of a simple linear regression with interval-censored outcomes. I then perform Monte Carlo simulations to investigate the finite sample performance of the inferential method proposed, and I compare its performance to the subsampling algorithm of Chernozhukov, Hong, and Tamer (2004).[19]

[19] As the confidence sets of this paper provide coverage for any fixed point, the method for pointwise inference with subsampling from their Appendix G is used.

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:	$\mathcal{C}_n^{IM}$	$\mathcal{C}_n^{MI}$	$\mathcal{C}_n^{IM}$	$\mathcal{C}_n^{MI}$	$\mathcal{C}_n^{IM}$	$\mathcal{C}_n^{MI}$	$\mathcal{C}_n^{IM}$	$\mathcal{C}_n^{MI}$
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|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:	$\mathcal{C}_n^{IM}$	$\mathcal{C}_n^{MI}$	$\mathcal{C}_n^{IM}$	$\mathcal{C}_n^{MI}$	$\mathcal{C}_n^{IM}$	$\mathcal{C}_n^{MI}$	$\mathcal{C}_n^{IM}$	$\mathcal{C}_n^{MI}$
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

Let a random sample of size n of (y_1, y_0, x) be observed by the econometrician, where

$$y^* = \beta_0 + \beta_1 x + u.$$

The econometrician observes a random sample of observations on (y_0, y_1, x), and knows that $P\{y_0 \leq y^* \leq y_1\} = 1$, but does not observe y^*. It is further assumed that $P(y_0 = y_1) < 1$, 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.[20] Thus the conditional moment restrictions

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

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 moment functions all have finite mean and variance because of the restrictions on u.

Suppose, for example, that $\mathcal{X} = \{1, 2\}$. Then (4) is

$$\mathbb{E}[m(y_1, y_0, x, \beta)] = \begin{pmatrix} E[-y_0|x=1] + \beta_0 + \beta_1 \\ E[y_1|x=1] - \beta_0 - \beta_1 \\ E[-y_0|x=2] + \beta_0 + 2\beta_1 \\ E[y_1|x=2] - \beta_0 - 2\beta_1 \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{19}$$

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

[20] 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 $\hat{c}_j = \widehat{cov}(y_1, y_0|x=j)$. 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 a random sample of (y_0, y_1, x) is observed. The econometrician knows that $y^* \in [y_0, y_1]$, but does not observe y^*. In particular, but unknown 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.

INSERT FIGURE 1 HERE
CAPTION: The identified set for (β_0, β_1) in Example 2.

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, there is some joint 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 precisely the same 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.

Checking the coverage probability of the confidence sets in the experiments requires using a grid of points representing the identified set over which to check point-wise coverage. Since both inferential methods implemented exhibit degenerate asymptotics on the interior of the identified set, it is sufficient to check a suitably fine set of boundary points of the identified set.[21] To check coverage, I used a grid ∂B^* of 400 points, comprising the four corners of the identified set, and 99 equidistant points between each pair of adjacent corners.

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 $n\tilde{Q}_n(\beta)$ as described in section 4.2.

1. Specify the number of simulations to draw, R, 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 $\beta \in \partial B^*$ compute $\tilde{Q}_n(\beta)$.
   (c) If $n\tilde{Q}_n(\beta) > C^*_\alpha$, reject the null hypothesis that $\beta \in B^*$, where C^*_α is the unique value that satisfies
   $$\frac{1}{2}\Pr\{\chi^2_1 > C^*_\alpha\} + \frac{1}{4}\Pr\{\chi^2_2 > C^*_\alpha\} = \alpha.$$
   This corresponds to weights for a 2×2 diagonal variance covariance matrix given by equation (14) from corollary 3.
3. For each $\beta \in \partial B^*$, compute the fraction of simulations for which $n\tilde{Q}_n(\beta) \leq C^*_\alpha$, denoted $\tau^{pt}_\alpha(\beta)$. Because any $\beta \in B^*$ can generate the observed distribution of observables, the coverage probability for β is $\inf_{\beta \in \beta^*} \tau^{pt}_\alpha(\beta) \equiv \hat{\tau}^{pt}_\alpha$. $\hat{\tau}^{pt}_\alpha$ is the observed probability with which $\mathcal{C}^{pt}_n$ was guaranteed to contain the true β in these simulations.

[21] It was verified in the finite sample Monte Carlo studies that the highest rejection probabilities did indeed occur on the boundary of the identified set.

For each iteration, I also construct confidence regions via subsampling. For this, I used the square loss function used in the Monte Carlos of both Chernozhukov, Hong, and Tamer (2004), and Romano and Shaikh (2006a):

$$G_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{E}[y_0|x=x_i] - \beta_0 - x_i\beta_1\right|_+^2 + \left|-\hat{E}[y_1|x=x_i] + \beta_0 + x_i\beta_1\right|_+^2,$$

where $|z|_+^2 \equiv 1[z>0]z^2$. As discussed in the previous section, this requires that one specify a grid of values over the parameter space on which to evaluate a level set of the loss function in the subsampling stage. For these simulations, I used a uniform grid B_g that ranged from -2 to 4 for β_0 and -1 to 3 for β_1. For grid construction I experimented using increments of 0.05 and 0.01 in both dimensions, and results are reported for both grid increments. For each iteration of step 2 above, I perform the following steps.

a. Compute a starting cutoff value c_0 via subsampling according to the procedure recommended by Chernozhukov, Hong, and Tamer (2004) for the interval regression case.[22].

b. Collect the following grid of points: $\hat{B}_k = \{\beta \in B_g : nG_n(\beta) \leq k = c_0 \ln n\}$. This approximates a level set of the objective function.

c. Randomly draw B_n subsamples (without replacement) of size $b << n$, and for each $\beta \in \hat{B}_k$ compute $bG_n(\beta)$. Compute the $1-\alpha$ quantile of $bG_n(\beta)$ in the subsamples, denoted $q_{1-\alpha}(\beta)$. In these experiments $B_n = 200$ random subsamples of size $b = n/4$ were drawn for each iteration.

d. Compute the maximum over $\beta \in \hat{B}_{1-\alpha}$ of the quantiles computed in step c, $q_{1-\alpha}^* = \sup_{\beta\in\hat{B}_{1-\alpha}} q_{1-\alpha}(\beta)$. Then the pointwise confidence region for β is

$$\mathcal{C}_n^{SS} = \left\{\beta : nG_n(\beta) \leq q_{1-\alpha}^*\right\}.$$

e. At the conclusion of all simulations, check the coverage probabilities. For each $\beta \in \partial B^*$, compute the fraction of simulations for which $nG_n(\beta) \leq q_{1-\alpha}^*$, denoted $\tau_\alpha^{SS}(\beta)$. Because any $\beta \in B^*$ can generate the observed distribution of observables, the coverage probability for β is $\inf_{\beta\in\beta^*} \tau_\alpha^{SS}(\beta) \equiv \hat{\tau}_\alpha^{SS}$. $\hat{\tau}_\alpha^{SS}$ is the observed probability with which $\mathcal{C}_{1-\alpha}^{SS}$ was guaranteed to contain the true β in these simulations.

[22]Specifically, I based the initial cutoff on the auxilary model where the outcome variable is taken to be $\tilde{y} = y_0/2 + y_1/2$. For a model based on a random sample of $(\tilde{y}, x)$, β is point-identified as the minimizer of the square loss function. The value of c_0 used in step b is the $1-\alpha$ quantile of the statistic $\min_{\beta\in B_g} b \cdot G(\beta)$, where b is the subsample size.

According to theory, both $\hat{\tau}_{\alpha}^{pt}$ and $\hat{\tau}_{\alpha}^{SS}$ converge to $1-\alpha$ as $n \to \infty$. Tables 5 and 6 show empirical coverage probabilities obtained from the above procedures for various pre-specified values of n and α. The results of table 5 are based on 500 repetitions, with a grid increment of 0.05 for the subsampling implementation. Table 6 is based on 100 repetitions, and grid increments of 0.01.[23] The moment inequality method performed well in these simulations, in conjunction with the asymptotic theory. The subsampling algorithm performed less accurately, though also provided suitable approximations when the initial grid of values used for subsampling was sufficiently dense (Table 6). While the method of this paper performed favorably, the results must be taken with caution, as they are particular to the experimental design of this section. Interestingly, both methods performed worse at the .75 level than at the other levels.

A key difference between the methods that is highlighted by the simulation results is the effect of the initial grid B_g on the accuracy of subsampling inference. The sensitivity of the subsampling routine to the choice of subsampling grid is illustrated in the difference in the accuracy of subsampling in Tables 5 and 6. The accuracy of the subsampling confidence sets was better in nearly every case when the grid was denser. This seems consistent with the theory on which subsampling is based, which characterizes the procedure when subsamples are taken over *all* points in a given level set. In practice, however, a grid must be used to approximate this level set. It stands to reason that the denser the grid, the better the resulting approximation. Thus, there is a tradeoff to be made in subsampling between the accuracy of inference, and the computational cost, as the use of more grid points entails more computations in each subsampling stage, as detailed in the discussion of computational considerations in section 4. Chernozhukov, Hong, and Tamer (2004) use an adaptive grid based on the Metropolis-Hastings algorithm to approximate this level set, which may yield more accurate inference, though I am unaware of formal results to this effect. No such grid is needed to compute critical values using the asymptotic chi-bar-square approximation.

Table 7 reports the amount of time in seconds that it took to construct the confidence sets $\mathcal{C}_n^{pt}$ and $\mathcal{C}_n^{SS}$. The reports times are averages across ten consecutive experiments.[24] All experiments were constructed in Matlab on a Pentium 4, 2.8 gigahertz CPU with 1 gigabyte of RAM. Subsample times are reported using the finer grid with 0.01 increments, as this gave more accurate inference. The same grid density of 0.01 was used for the larger parameter space grid over which the sample (as opposed to *subsample*) objective functions $\tilde{Q}_n$ and G_n were both evaluated. The number of subsamples taken also had a substantial effect on computation time for subsampling, and times are reported when 200, 400, and 1000 subsamples are used, respectively. When 200 subsamples were used, subsampling took a little under three times as long as using the chi-bar-square approximation. The difference reflects the added computational time of using subsampling to obtain the test's critical value. Using more subsamples obviously increases the computation

[23]The numbers of repititions with the smaller increments is due to the increased computational time required.

[24]The variance of computation time was quite low, so that averaging over additional experiments did not have much effect.

Table 5: Coverage Probabilities for confidence regions based on the chi-bar-square approximation and subsampling based on 500 repetitions and a grid increment of 0.05.

Target Coverage:	0.75		0.85		0.95		0.99	
Actual Coverage:	$\mathcal{C}_n^{SS}$	$\mathcal{C}_n^{pt}$	$\mathcal{C}_n^{SS}$	$\mathcal{C}_n^{pt}$	$\mathcal{C}_n^{SS}$	$\mathcal{C}_n^{pt}$	$\mathcal{C}_n^{SS}$	$\mathcal{C}_n^{pt}$
n								
100	0.660	0.732	0.768	0.834	0.888	0.934	0.968	0.978
200	0.670	0.742	0.764	0.824	0.884	0.946	0.962	0.988
500	0.638	0.732	0.730	0.800	0.870	0.926	0.942	0.978
1000	0.616	0.740	0.744	0.836	0.870	0.950	0.956	0.984

Table 6: Coverage Probabilities for confidence regions based on the chi-bar-square approximation and subsampling based on 100 repetitions and a grid increment of 0.01.

Target Coverage:	0.75		0.85		0.95		0.99	
Actual Coverage:	$\mathcal{C}_n^{SS}$	$\mathcal{C}_n^{pt}$	$\mathcal{C}_n^{SS}$	$\mathcal{C}_n^{pt}$	$\mathcal{C}_n^{SS}$	$\mathcal{C}_n^{pt}$	$\mathcal{C}_n^{SS}$	$\mathcal{C}_n^{pt}$
n								
100	0.64	0.67	0.79	0.85	0.93	0.94	0.98	0.99
200	0.72	0.75	0.82	0.87	0.91	0.93	0.97	0.98
500	0.68	0.71	0.77	0.82	0.89	0.92	0.97	0.96
1000	0.68	0.73	0.74	0.84	0.89	0.93	0.95	0.98

time. Because increasing the number of subsamples beyond 200 did not seem to have an effect on accuracy, the results in Tables 5 and 6 are only reported for a subsample size of 200. The results were not sensitive to either changes in subsample size b or sample size n at the order of magnitude considered, likely due to Matlab's efficient matrix manipulation.

Figures 2 through 4 show examples of 0.95 confidence regions using both methods based on random samples of 100 observations. The confidence regions look very similar in shape and size. Though those obtained from subsampling were always slightly smaller, their nominal coverage was further from the actual coverage in the simulations, as shown in Tables 5 and 6. This is because there are some cases where confidence regions from subsampling do not contain certain points in the identified set that the method based on the chi-bar-square approximation does contain. This observation is illustrated in both figures 2 and 3. Figure 2 is an example in which all points in the

Table 7: Time in seconds for constructing nominal 0.95 confidence sets using the chi-bar-square approximation and subsampling. Subsamples drawn for the computation of test statistic quantiles are given in parentheses.

n	$\mathcal{C}_n^{pt}$	$\mathcal{C}_n^{SS}$ (200)	$\mathcal{C}_n^{SS}$ (400)	$\mathcal{C}_n^{SS}$ (1000)
100	36.99	99.85	185.92	449.19
200	36.87	99.53	187.13	447.75
500	36.91	99.62	185.94	449.12
1000	37.05	99.76	186.14	447.77

identified set were contained in $\mathcal{C}_n^{pt}$, but some points were not covered by $\mathcal{C}_n^{SS}$. Figure 3 presents a case in which both confidence regions excluded some points in the identified set, though $\mathcal{C}_n^{SS}$ was slightly smaller and thus excluded some points that $\mathcal{C}_n^{pt}$ did not. Both methods should of course exclude at least some point in the identified set 5% of the time. In figure 4, both confidence regions cover each point in the identified set.

INSERT FIGURE 2 HERE

CAPTION: Nominal 0.95 *confidence regions* $\mathcal{C}_n^{pt}$ and $\mathcal{C}_n^{SS}$ *for* β *based on a random sample of 100 observations. For this sample,* $\mathcal{C}_n^{pt}$ contained all points in the identified set, while $\mathcal{C}_n^{SS}$ did not cover some points near the northwest edge of the identified set.

INSERT FIGURE 3 HERE

CAPTION: Nominal 0.95 *confidence regions* $\mathcal{C}_n^{pt}$ and $\mathcal{C}_n^{SS}$ *for* β *based on a random sample of 100 observations. For this sample, in the southwest corner of the idenitified set there were some points in the identified set that both* $\mathcal{C}_n^{pt}$ and $\mathcal{C}_n^{SS}$ did not contain, as well as some points that were in $\mathcal{C}_n^{pt}$ but not $\mathcal{C}_n^{SS}$.

INSERT FIGURE 4 HERE

CAPTION: Nominal 0.95 *confidence regions* $\mathcal{C}_n^{pt}$ and $\mathcal{C}_n^{SS}$ *for* β *based on a random sample of 100 observations. For this sample, both confidence sets contained all points in the identified set.*

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 thus do not require resampling methods to compute.

In some cases, as discussed in section 4, the method may be asymptotically conservative, in the sense that limiting coverage may be greater than the nominal level. Even in these cases, the test on which the confidence sets are based is shown to be consistent. In the important special case of a single equation regression with censored outcomes, the limiting coverage of the confidence sets for the least favorable point in the identified set is in fact shown to be exact. In this case, the method was shown to perform well in finite samples relative to a subsampling algorithm. While those results are specific to the experimental design employed, they suggest that in some cases the method proposed here may result in confidence sets with favorable finite sample properties. However, there are cases where the subsampling method achieves valid asymptotic inference, but where the method of this paper is not applicable. It would be of interest to compare other inferential methods for partially identified parameters in other contexts as well, as many such methods have been recently proposed.

The findings of this paper have naturally lead to some additional 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 for any value of θ, 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 that is present in some instances 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 the parameter of interest θ_0. There have been many other types of confidence sets that have appeared in the literature on partially identified parameters. 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 3 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

$$\partial\Theta^* \equiv \left\{\theta \in \Theta^* : \text{for every open neighborhood of } \theta,\, N_\theta \subseteq \mathbb{R}^k,\, N_\theta \not\subseteq \Theta^*\right\}$$

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

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

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 6 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 6 *Let assumptions (A1)-(A3) as well as (A6) hold. Then* $\partial\Theta^* \subseteq D\Theta^*$.

Proposition 7 *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 (5). The hypothesis test can be recast however as

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

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 $Q\left(\theta\right) \geq 0$. Andrews (2001) studies the problem of hypothesis testing when a parameter is on the boundary of the maintained hypothesis.

Appendix B: Proofs

6.1 Proposition 1

Proof . Fix θ. Let $\hat{q}_n\left(\theta, t\right) \equiv \left(\hat{E}\left[m\left(y, x, \theta\right)\right] - t\right)' \hat{V}_\theta^{-1} \left(\hat{E}\left[m\left(y, x, \theta\right)\right] - t\right)$, so that $\hat{Q}_n\left(\theta\right) = \min_{t \geq 0} q_n\left(\theta, t\right)$. Similarly, let $q\left(\theta, t\right) \equiv \left(\mathbb{E}\left[m\left(y, x, \theta\right)\right] - t\right)' V_\theta^{-1} \left(\mathbb{E}\left[m\left(y, x, \theta\right)\right] - t\right)$, so that $Q\left(\theta\right) = \min_{t \geq 0} q\left(\theta, t\right)$. (6), (7), (8) and a Slutsky Theorem imply that $q_n\left(\theta, t\right) \xrightarrow{p} q\left(\theta, t\right)$ pointwise for each θ, t. $q_n\left(\theta, t\right)$ is concave in t, so that by Theorem 2.7 of Newey and McFadden (1994), $q_n\left(\theta, t\right)$ converges uniformly in $t > 0$ to $q\left(\theta, t\right)$ for fixed θ. In addition, uniform convergence holds over any compact set $[0, T]$ by the continuity of $q\left(\theta, t\right)$ in t. Therefore $q_n\left(\theta, t\right) \xrightarrow{p} q\left(\theta, t\right)$ uniformly over $t \geq 0$, so that $\hat{Q}_n\left(\theta\right) \xrightarrow{p} Q\left(\theta\right)$, further implying convergence in probability of the minimizer over $t \geq 0$ of $q_n\left(\theta, t\right)$ to that of $q\left(\theta, t\right)$, i.e. $\hat{t}_n^*\left(\theta\right) \xrightarrow{p} t_\theta^*\left(\theta\right)$. ■

6.2 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 Manski and Tamer (2002), Proposition 5. ■

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

6.3 Lemma 1

Consider the minimization problem

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

where $x, t \in \mathbb{R}^J$, and $x_1, t_1 \in \mathbb{R}^b$, $b \leq J$, s.t. $t = (\mathbf{t}_1', \mathbf{t}_2')'$ and $x = (\mathbf{x}_1', \mathbf{x}_2')'$. 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

$$QP = \min \left(\mathbf{x}_1-\mathbf{t}_1\right)' V_{11}^{-1} \left(\mathbf{x}_1-\mathbf{t}_1\right) \text{ s.t. } \mathbf{t}_1 \geq 0. \tag{21}$$

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 (20) 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),

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

$$-\Lambda_{21}\left(\mathbf{x}_1-\mathbf{t}_1^*\right) - \Lambda_{22}\left(\mathbf{x}_2-\mathbf{t}_2^*\right) = 0. \tag{23}$$

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

$$(\mathbf{x}_2 - \mathbf{t}_2^*) = -\Lambda_{22}^{-1}\Lambda_{21}(\mathbf{x}_1 - \mathbf{t}_1^*) \,. \tag{24}$$

Now

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

by (23). Now using (24) it follows that

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

where the last equality follows by the partition inverse result.[25] All that remains is to show that t_1^* minimizes (21): $\min (\mathbf{x}_1 - \mathbf{t}_1)' V_{11}^{-1} (\mathbf{x}_1 - \mathbf{t}_1)$ 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 (21) are for $j = 1, ..., b$,

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

$\Longleftrightarrow$

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

$\Leftrightarrow$

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

$\Leftrightarrow$

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

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

by (24), but this is exactly condition (i) from the Kuhn-Tucker conditions for the initial program (20). ■

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

6.4 Proposition 3

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

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

Then

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

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 $\tilde{m}\left(y,x,\theta\right)=\left(m_{b+1}\left(y,x,\theta\right),...,m_J\left(y,x,\theta\right)\right)'$. Then because $E\left[m_j\left(y,x,\theta\right)\right]=0$ for $j\leq b$,

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

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

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

and by Lemma 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 (8) 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 $\min_{s\in\mathbb{R}_+^b} [v^* - s]' V_\theta^{*-1} [v^* - s]$ measures the distance of the normal random variable v^* from the nonnegative orthant. By Wolak (1991)

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

■

6.4.1 Corollary 2

Proof.

$$\begin{aligned}
\sup_{\theta\in\Theta^*} \lim_{n\to\infty} \mathbb{P}\left\{n\hat{Q}_n\left(\theta\right) \geq c\right\} &= \sup_{\theta\in\Theta^*} \left(\sum_{j=0}^{b(\theta)} w\left(b\left(\theta\right), b\left(\theta\right) - j, V_\theta^*\right) \Pr\left\{\chi_j^2 \geq c\right\}\right) \\
&\leq \frac{1}{2} \Pr\left\{\chi_{b^*}^2 \geq c\right\} + \frac{1}{2} \Pr\left\{\chi_{b^*-1}^2 \geq c\right\},
\end{aligned}$$

where the equality of the first line follows from Proposition 3. 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 for any j and $b = \dim\left(V_\theta^*\right)$,

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

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

and

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

■

6.4.2 Corollary 3

Proof. The first part, (14), 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 $w\left(b, j, V^*\right) = w\left(b, j, I_b\right)$. The second part, (14), 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

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

■

6.5 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 $m(y,x,\theta)$. Assume (A1)-(A4) and that V_θ^* is diagonal with all diagonal entries positive. Then

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

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. ■

6.6 Proposition 5

Proof. Let $\theta \notin \Theta^*$, so that there exists $j \in \{1, ...J\}$ such that $\mathbb{E}\left[m_j(y,x,\theta)\right] < 0$. Assume (A1)-(A5). Let $\mu(\theta) \equiv \mathbb{E}\left[m(y,x,\theta)\right]$, and let $\hat{\Lambda}_\theta = \hat{V}_\theta^{-1}$, and $.\Lambda_\theta = V_\theta^{-1}$ Proceeding as in the proof of Proposition 3,

$$\begin{aligned}
\mathbb{P}\left\{n\hat{Q}_n(\theta) > C_\alpha^{b^*}\right\} &= \mathbb{P}\left\{n\min_{t\geq 0} \left[\hat{E}_n\left[m(y,x,\theta)\right] - t\right]' \hat{\Lambda}_\theta \left[\hat{E}_n\left[m(y,x,\theta)\right] - t\right] > C_\alpha^{b^*}\right\} \quad (25) \\
&= \mathbb{P}\left\{\min_{s} \left[v_n - s\right]' \hat{\Lambda}_\theta \left[v_n - s\right] > C_\alpha^{b^*} : s \geq -\sqrt{n}\mu(\theta)\right\}
\end{aligned}$$

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

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

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. $\Gamma_\theta\Lambda_\theta\Gamma_\theta' = diag\left(d_{\theta,1},...,d_{\theta,J}\right)$, where the $d_{\theta,j}$ are the eigenvalues of Λ_θ. Since Λ_θ is nonsingular, each $d_{\theta,j} > 0$. Such a matrix Γ_θ exists by Corollary 21.5.9 of Harville (1997). Then

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

The constraint $s \geq -\sqrt{n}\mu\left(\theta\right)$ in (25), implies that s_n^* diverges to $-\infty$. Since $v_n = O_p\left(1\right)$, $n\hat{Q}_n\left(\theta\right)$ diverges to ∞ and $\lim_{n\to\infty}\mathbb{P}\left\{n\hat{Q}_n\left(\theta\right) > C_\alpha^{b^*}\right\} = 1$. The same argument applies for $\lim_{n\to\infty}\mathbb{P}\left\{n\tilde{Q}_n\left(\theta\right) > C_\alpha^{b^*}\right\}$ under the conditions of Proposition 4, by replacing $\hat{V}_\theta^{-1}$ with a diagonal matrix with diagonal elements $1/\hat{V}_{\theta,jj}$. ■

6.7 Proposition 6

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

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

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

6.8 Proposition 7

Proof. Proposition 6 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 $\mathbb{E}\left[m_j\left(y,x,\theta\right)\right] = 0$ for some j. By (A6), $\mathbb{E}\left[m_j\left(y,x,\theta\right)\right]$ is monotone in some component of θ, say $\theta_{k(j)}$. Let $\epsilon > 0$, and let $\mathbf{v}\left(\epsilon\right)$ be a k-vector with $k\left(j\right)$ component ϵ and all other components zero. By the strict monotonicity of $\mathbb{E}\left[m_j\left(y,x,\theta\right)\right]$ in $\theta_{k(j)}$, $\forall\epsilon \in \mathbb{R}^k$, either $\mathbb{E}\left[m_j\left(y,x,\theta+\mathbf{v}\left(\epsilon\right)\right)\right] < 0$ or $\mathbb{E}\left[m_j\left(y,x,\theta-\mathbf{v}\left(\epsilon\right)\right)\right] < 0$, so that $\theta \in \partial\Theta^*$. ■

References

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

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, Yale University.

ANDREWS, D. W. K., AND P. GUGGENBERGER (2006): "The Limit of Finite Sample Size and A Problem with Subsampling," working paper, UCLA.

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 (2006): "Asymptotic Properties for a Class of Partially Identified Models," working paper, Cornell University.

BLUNDELL, R., M. BROWNING, AND I. CRAWFORD (2006): "Best Nonparametric Bounds on Demand Responses," working paper, University College London.

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2004): "Parameter Set Inference in a Class of Econometric Models," working paper, MIT.

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.

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

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

Horowitz, J. (2001): "The Bootstrap," in *The Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 5, pp. 3159–3228. Elsevier Science B.V.

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.

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.

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 (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.

Politis, D. N., and J. P. Romano (1994): "Large Sample Confidence Regions Based on Subsamples Under Minimal Assumptions," *Annals of Statistics*, 22(4), 2031–2050.

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

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

——— (2006b): "Inference for the Identified Set in Partially Identified Econometric Models," working paper, Stanford University.

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

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.

Figure 1

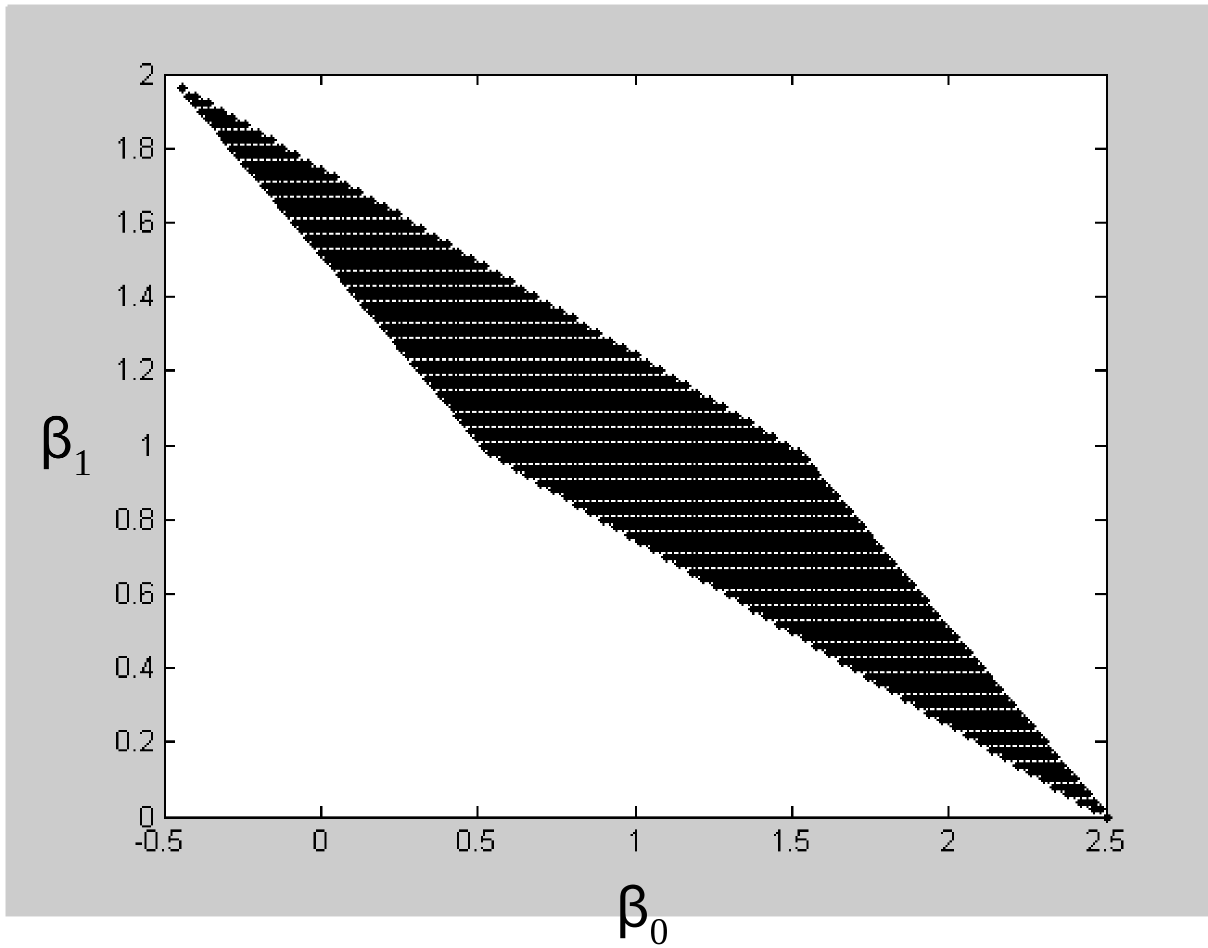
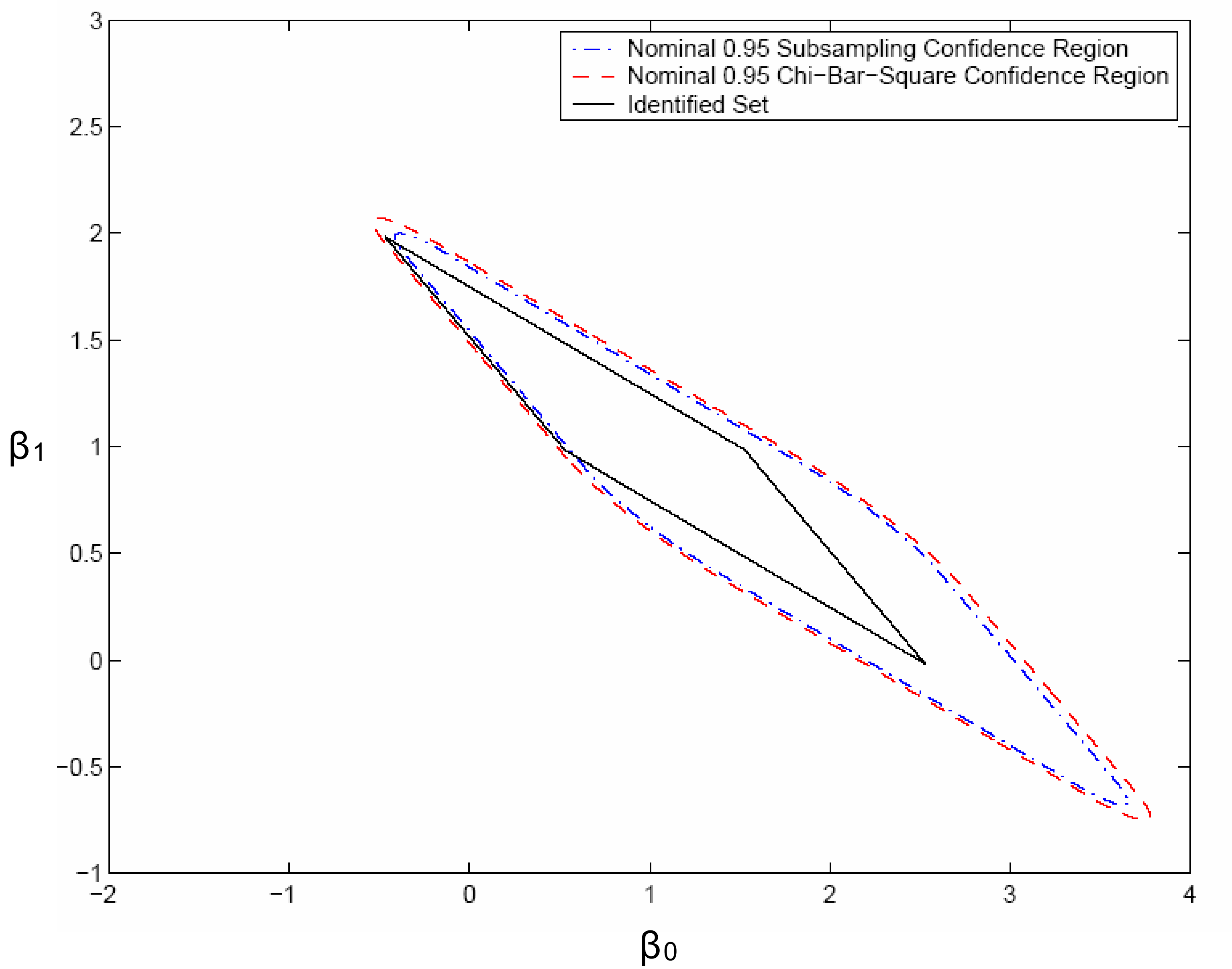


Figure 2



Nominal 0.95 Subsampling Confidence Region
Nominal 0.95 Chi−Bar−Square Confidence Region
Identified Set
3
2.5
2
1.5
1
0.5
0
−0.5
−1
−2
−1
0
1
2
3
4
β1
β0

Figure 3

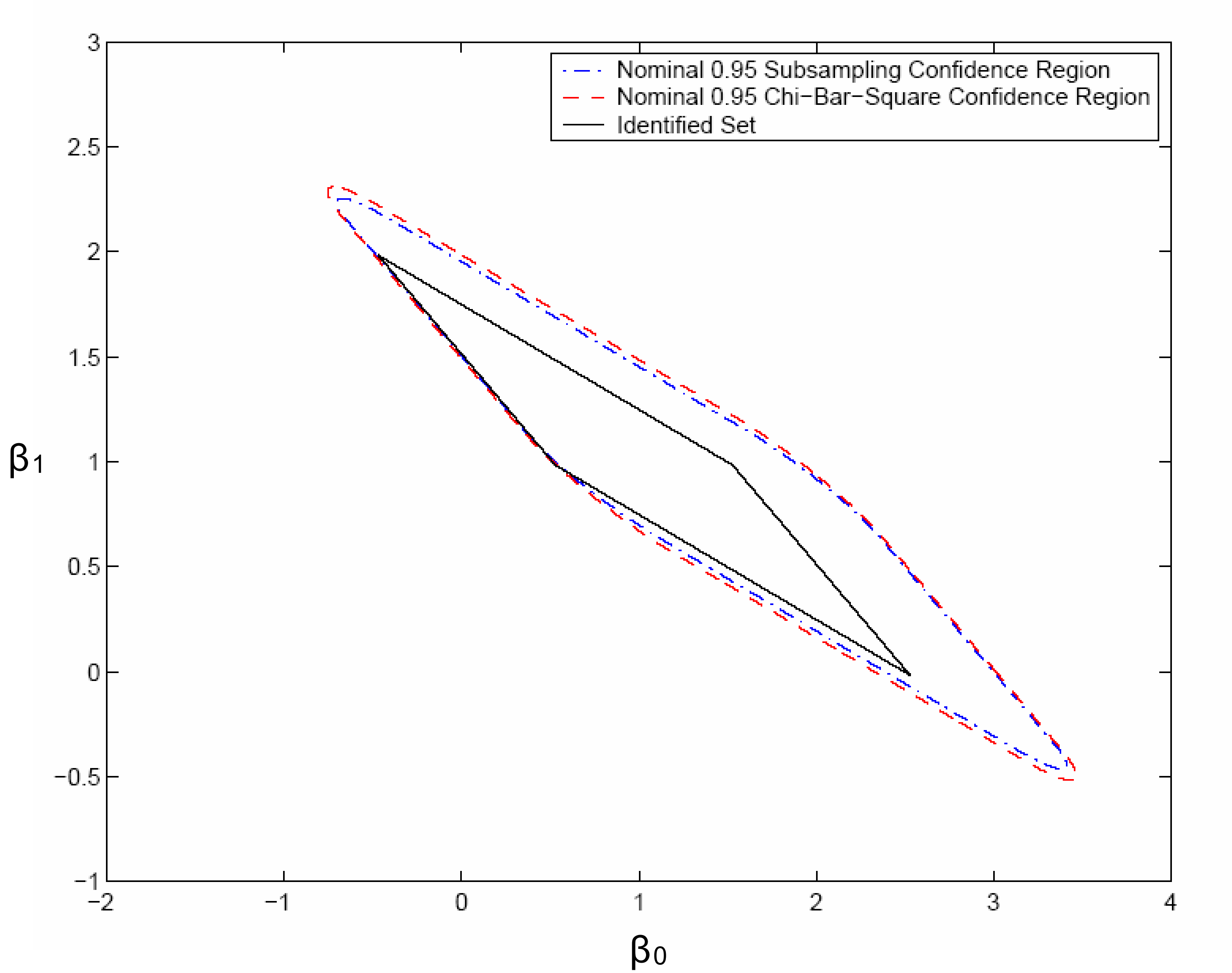
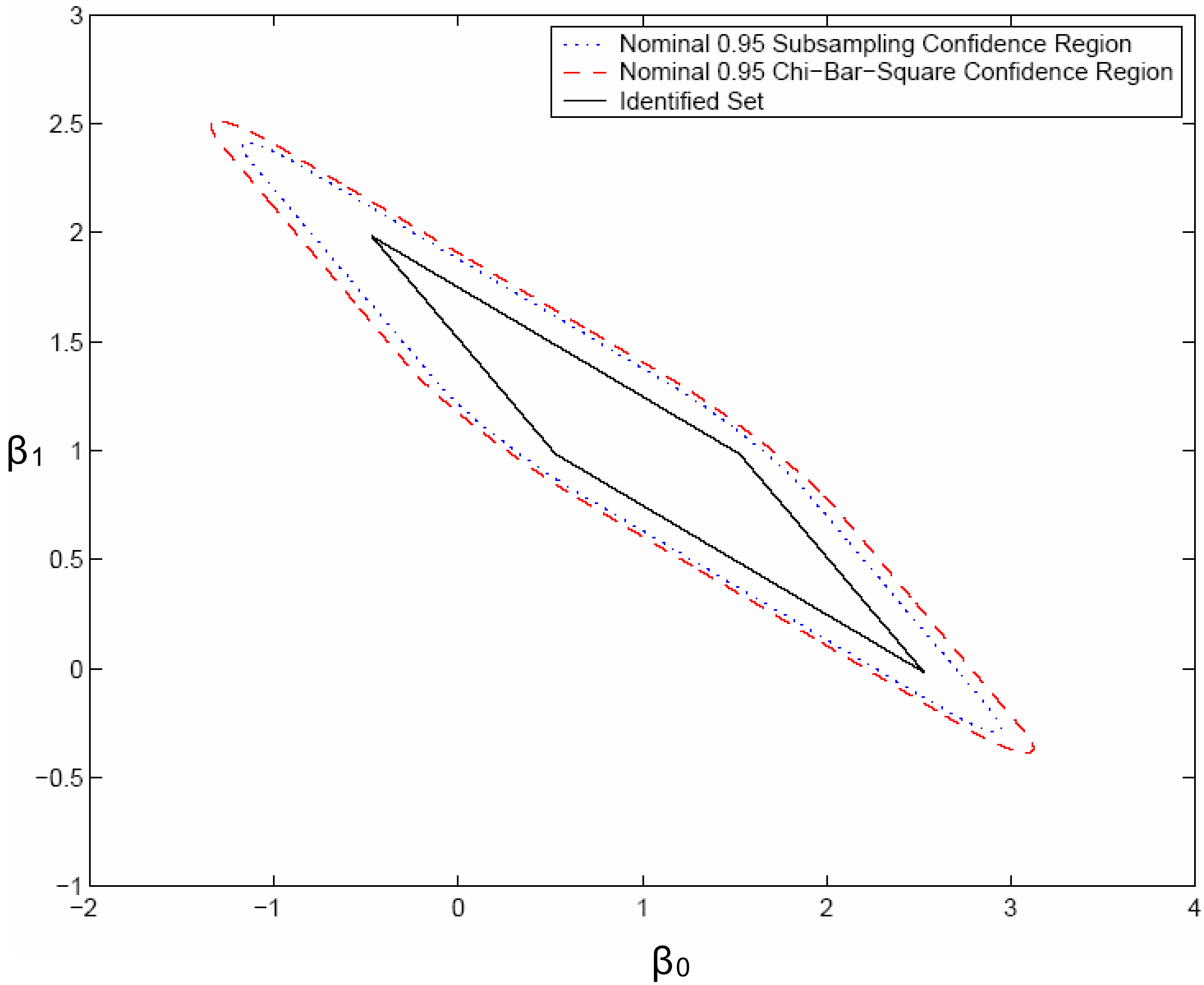

Nominal 0.95 Subsampling Confidence Region
Nominal 0.95 Chi−Bar−Square Confidence Region
Identified Set
β_1
β_0
3
2.5
2
1.5
1
0.5
0
−0.5
−1
−2
−1
0
1
2
3
4

Figure 4



Nominal 0.95 Subsampling Confidence Region
Nominal 0.95 Chi−Bar−Square Confidence Region
Identified Set
3
2.5
2
1.5
1
0.5
0
−0.5
−1
−2
−1
0
1
2
3
4
β1
β0