INCOMPLETE DRAFT

IDENTIFICATION AND SHAPE RESTRICTIONS IN NONPARAMETRIC INSTRUMENTAL VARIABLES ESTIMATION

by

Joachim Freyberger

and

Joel L. Horowitz
Department of Economics
Northwestern University
Evanston IL 60208
USA

April 2013

Abstract

This paper is concerned with inference about an unidentified linear functional, $L(g)$, where the function g satisfies the relation $Y = g(X) + U;\ E(U \mid W) = 0$. In this relation, Y is the dependent variable, X is a possibly endogenous explanatory variable, W is an instrument for X, and U is an unobserved random variable. The data are an independent random sample of (Y, X, W). In much applied research, X and W are discrete, and W has fewer points of support than X. Consequently, neither g nor $L(g)$ is nonparametrically identified. Indeed, $L(g)$ can have any value in $(-\infty, \infty)$. In applied research, this problem is typically overcome and point identification is achieved by assuming that g is a linear function of X. However, the assumption of linearity is arbitrary. It is untestable if W is binary, as is the case in many applications. This paper explores the use of shape restrictions, such as monotonicity or convexity, for achieving interval identification of $L(g)$. Economic theory often provides such shape restrictions. This paper shows that they restrict $L(g)$ to an interval whose upper and lower bounds can be obtained by solving linear programming problems. Inference about the identified interval and the functional $L(g)$ can be carried out by using the bootstrap. An empirical application illustrates the usefulness of shape restrictions for carrying out nonparametric inference about $L(g)$. An extension to nonseparable and quantile IV models is described.

JEL Classification: C13, C14, C26

We thank Ivan Canay, Xiaohong Chen, Sokbae Lee, Chuck Manski, and Elie Tamer for helpful comments. This research was supported in part by NSF grant SES-0817552

IDENTIFICATION AND SHAPE RESTRICTIONS IN NONPARAMETRIC INSTRUMENTAL VARIABLES ESTIMATION

1. INTRODUCTION

This paper is about estimation of the linear functional $L(g)$, where the unknown function g obeys the relation

(1a) $Y = g(X) + U$,

and

(1b) $E(U \mid W = w) = 0$

for almost every w. Equivalently,

(2) $E[Y - g(X) \mid W = w] = 0$.

In (1a), (1b), and (2), Y is the dependent variable, X is a possibly endogenous explanatory variable, W is an instrument for X, and U is an unobserved random variable. The data consist of an independent random sample $\{Y_i, X_i, W_i : i = 1,...,n\}$ from the distribution of (Y, X, W). In this paper, it is assumed that X and W are discretely distributed random variables with finitely many mass points. Discretely distributed explanatory variables and instruments occur frequently in applied research, as is discussed in the next paragraph. When X is discrete, g can be identified only at mass points of X. Linear functionals that may be of interest in this case are the value of g at a single mass point and the difference between the values of g at two different mass points.

Although model (1)-(2) is the main focus of this paper, we also present an extension of our methods to quantile IV models in which the mean-independence condition (2) is replaced by a quantile independence condition. As will be explained later, the quantile IV model includes a class of nonseparable models.

In much applied research, W has fewer mass points than X does. For example, in a study of returns to schooling, Card (1995) used a binary instrument for the endogenous variable years of schooling. Moran and Simon (2006) used a binary instrument for income in a study of the effects of the Social Security "notch" on the usage of prescription drugs by the elderly. Other studies in which an instrument has fewer mass points than the endogenous explanatory variable are Angrist and Krueger (1991), Bronars and Grogger (1994), and Lochner and Moretti (2004).

The function g is not identified nonparametrically when W has fewer mass points than X does. The linear functional $L(g)$ is unidentified except in special cases. Indeed, as will be shown in Section 2 of this paper, except in special cases, $L(g)$ can have any value in $(-\infty, \infty)$ when W has fewer points of support than X does. Thus, except in special cases, the data are uninformative about $L(g)$ in the absence of further information. In the applied research cited in the previous paragraph, this problem is dealt with by assuming that g is a linear function. The assumption of linearity enables g and $L(g)$ to be identified, but it is problematic in other respects. In particular, the assumption of linearity is not testable if W is binary. Moreover, any other two-parameter specification is observationally equivalent to linearity and untestable, though it might yield substantive conclusions that are very different from those obtained under the assumption of linearity. For example, the assumptions that $g(x) = \beta_0 + \beta_1 x^2$ or $g(x) = \beta_0 + \beta_1 \sin x$ for some constants β_0 and β_1 are observationally equivalent to $g(x) = \beta_0 + \beta_1 x$ and untestable if W is binary.

This paper explores the use of restrictions on the shape of g, such as monotonicity, convexity, or concavity, to achieve interval identification of $L(g)$ when X and W are discretely distributed and W has fewer mass points than X has. Specifically, the paper uses shape restrictions on g to establish an identified interval that contains $L(g)$. Shape restrictions are less restrictive than a parametric specification such as linearity. They are often plausible in applications and may be prescribed by economic theory. For example, demand and cost functions are monotonic, and cost functions are convex. It is shown in this paper that under shape restrictions, such as monotonicity, convexity, or concavity, that impose linear inequality restrictions on the values of $g(x)$ at points of support of X, $L(g)$ is restricted to an interval whose upper and lower bounds can be obtained by solving linear programming problems. The bounds can be estimated by solving sample-analog versions of the linear programming problems. The estimated bounds are asymptotically distributed as the maxima of multivariate normal random variables. Under certain conditions, the bounds are asymptotically normally distributed, but calculation of the analytic asymptotic distribution is difficult in general. We present a bootstrap procedure that can be used to estimate the asymptotic distribution of the estimated bounds in applications. The asymptotic distribution can be used to carry out inference about the identified interval that contains $L(g)$ and, using methods like those of Imbens and Manski (2004) and Stoye (2009), inference about the parameter $L(g)$.

Interval identification of g in (1a) has been investigated previously by Chesher (2004) and Manski and Pepper (2000, 2009). Chesher (2004) considered partial identification of g in (1a) but replaced (1b) with assumptions like those used in the control-function approach to estimating models with an endogenous explanatory variable. He gave conditions under which the difference between the values of g at two different mass points of X is contained in an identified interval. Manski and Pepper (2000, 2009) replaced (1b) with monotonicity restrictions on what they called “treatment selection” and “treatment response.” They derived an identified interval that contains the difference between the values of g at two different mass points of X under their assumptions. Neither Chesher (2004) nor Manski and

Pepper (2000, 2009) treated restrictions on the shape of g under (1a) and (1b). The approach described in this paper is non-nested with those of Chesher (2004) and Manski and Pepper (2000, 2009). The approach described here is also distinct from that of Chernozhukov, Lee, and Rosen (2009), who treated estimation of the interval $[\sup_{v \in \mathcal{V}} \theta^l(v), \inf_{v \in \mathcal{V}} \theta^u(v)]$, where θ^l and θ^u are unknown functions and $\mathcal{V}$ is a possibly infinite set.

The remainder of this paper is organized as follows. In Section 2, it is shown that except in special cases, $L(g)$ can have any value in $(-\infty, \infty)$ if the only information about g is that it satisfies (1a) and (1b). It is also shown that under shape restrictions on g that take the form of linear inequalities, $L(g)$ is contained in an identified interval whose upper and lower bounds can be obtained by solving linear programming problems. The bounds obtained by solving these problems are sharp. Section 3 shows that the identified bounds can be estimated consistently by replacing unknown population quantities in the linear programs with sample analogs. The asymptotic distributions of the identified bounds are obtained. Methods for obtaining confidence intervals and for testing certain hypotheses about the bounds are presented. Section 4 presents a bootstrap procedure for estimating the asymptotic distributions of the estimators of the bounds. Section 4 also presents the results of a Monte Carlo investigation of the performance of the bootstrap in finite samples. Section 5 presents an empirical example that illustrates the usefulness of shape restrictions for achieving interval identification of $L(g)$. Section 6 presents the extension to quantile IV models as well as models with exogenous covariates. Section 7 presents concluding comments.

2. INTERVAL IDENTIFICATION OF $L(g)$

This section begins by defining notation that will be used in the rest of the paper. Then it is shown that, except in special cases, the data are uninformative about $L(g)$ if the only restrictions on g are those of (1a) and (1b). It is also shown that when linear shape restrictions are imposed on g, $L(g)$ is contained in an identified interval whose upper and lower bounds are obtained by solving linear programming problems. Finally, some properties of the identified interval are obtained.

Denote the supports of X and W, respectively, by $\{x_j : j = 1,..., J\}$ and $\{w_k : k = 1,..., K\}$. In this paper, it is assumed that $K < J$. Order the support points so that $x_1 < x_2 < ... < x_J$ and $w_1 < w_2 < ... w_K$. Define $g_j = g(x_j)$, $\pi_{jk} = P(X = x_j, W = w_k)$, and $m_k = E(Y | W = w_k) P(W = w_k)$. Then (2) is equivalent to

(3) $m_k = \sum_{j=1}^{J} g_j \pi_{jk}; \; k = 1,...,K$.

Let $\boldsymbol{m} = (m_1,...,m_K)'$ and $\boldsymbol{g} = (g_1,...,g_J)'$. Define Π as the $J \times K$ matrix whose (j,k) element is π_{jk}. Then (3) is equivalent to

(4) $\boldsymbol{m} = \Pi' \boldsymbol{g}$.

Note that $rank(\Pi) < J$, because $K < J$. Therefore, (4) does not point identify $\boldsymbol{g}$. Write the linear functional $L(g)$ as $L(g) = \boldsymbol{c}'\boldsymbol{g}$, where $\boldsymbol{c} = (c_1,...,c_J)'$ is a vector of known constants.

The following proposition shows that except in special cases, the data are uninformative about $L(g)$ when $K < J$.

Proposition 1: Assume that $K < J$ and that $\boldsymbol{c}$ is not orthogonal to the orthogonal complement of the space spanned by the rows of Π'. Then any value of $L(g)$ in $(-\infty,\infty)$ is consistent with (1a) and (1b). ■

Proof: Let $\boldsymbol{g}_1$ be a vector in the space spanned by the rows of Π' that satisfies $\Pi'\boldsymbol{g}_1 = \boldsymbol{m}$. Let $\boldsymbol{g}_2$ be a vector in the orthogonal complement of the row space of Π' such that $\boldsymbol{c}'\boldsymbol{g}_2 \neq 0$. For any real γ, $\Pi'(\boldsymbol{g}_1 + \gamma \boldsymbol{g}_2) = \boldsymbol{m}$ and $L(\boldsymbol{g}_1 + \gamma \boldsymbol{g}_2) = \boldsymbol{c}'\boldsymbol{g}_1 + \gamma \boldsymbol{c}'\boldsymbol{g}_2$. Then $L(\boldsymbol{g}_1 + \gamma \boldsymbol{g}_2)$ is consistent with (1a)-(1b), and by choosing γ appropriately, $L(\boldsymbol{g}_1 + \gamma \boldsymbol{g}_2)$ can be made to have any value in $(-\infty,\infty)$. ■

We now impose the linear shape restriction

(5) $S\boldsymbol{g} \le 0$,

where S is an $M \times J$ matrix of known constants for some integer $M > 0$. For example, if g is monotone non-increasing, then S is the $(J-1) \times J$ matrix

$$S = \begin{pmatrix} -1 & 1 & 0 & 0 & ... & 0 & 0 \\ 0 & -1 & 1 & 0 & ... & 0 & 0 \\ & & & ... & & & \\ 0 & 0 & 0 & 0 & ... & -1 & 1 \end{pmatrix}.$$

We assume that g satisfies the shape restriction.

Assumption 1: The unknown function g satisfies (1a)-(1b) with $S\boldsymbol{g} \le 0$.

Sharp bounds on $L(g)$ are the solutions to the linear programming problems

(6) $\underset{\boldsymbol{h}}{\text{maximize}} \; (\underset{\boldsymbol{h}}{\text{minimize}}): \; \boldsymbol{c}'\boldsymbol{h}$

subject to: $\Pi'\boldsymbol{h} = \boldsymbol{m}$

$S\boldsymbol{h} \le 0$.

Let $L_{\min}$ and $L_{\max}$, respectively, denote the optimal values of the objective functions of the minimization and maximization versions of (6). It is clear that under (1a) and (1b), $L(g)$ cannot be less than $L_{\min}$ or greater than $L_{\max}$. The following proposition shows that $L(g)$ can also have any value between $L_{\min}$ and $L_{\max}$. Therefore, the interval $[L_{\min}, L_{\max}]$ is the sharp identification set for $L(g)$.

Proposition 2: The identification set of $L(g)$ is convex. In particular, it contains $\lambda L_{\min} + (1-\lambda)L_{\max}$ for any $\lambda \in [0,1]$. ■

Proof: Let $d = \lambda L_{\max} + (1-\lambda)L_{\min}$, where $0 < \lambda < 1$. Let $\boldsymbol{g}_{\max}$ and $\boldsymbol{g}_{\min}$ be feasible solutions of (6) such that $\boldsymbol{c}'\boldsymbol{g}_{\max} = L_{\max}$ and $\boldsymbol{c}'\boldsymbol{g}_{\min} = L_{\min}$. Then $d = \boldsymbol{c}'[(1-\lambda)\boldsymbol{g}_{\max} + \lambda \boldsymbol{g}_{\min}]$. The feasible region of a linear programming problem is convex, so $(1-\lambda)\boldsymbol{g}_{\max} + \lambda \boldsymbol{g}_{\min}$ is a feasible solution of (6). Therefore, d is a possible value of $L(g)$ and is in the identified set of $L(g)$. ■

The values of $L_{\min}$ and $L_{\max}$ need not be finite. Moreover, there are no simple, intuitively straightforward conditions under which $L_{\min}$ and $L_{\max}$ are finite.[1] Accordingly, we assume that:

Assumption 2: $L_{\min} > -\infty$ and $L_{\max} < \infty$.

Assumption 2 can be tested empirically. A method for doing this is outlined in Section 3.4. However, a test of assumption 2 is unlikely to be useful in applied research. To see one reason for this, let $\hat{L}_{\max}$ and $\hat{L}_{\min}$, respectively, denote the estimates of $L_{\max}$ and $L_{\min}$ that are described in Section 3.1. The hypothesis that assumption 2 holds can be rejected only if $\hat{L}_{\max} = \infty$ or $\hat{L}_{\min} = -\infty$. These estimates cannot be improved under the assumptions made in this paper, even if it is known that $L_{\min}$ and $L_{\max}$ are finite. If $\hat{L}_{\max} = \infty$ or $\hat{L}_{\min} = -\infty$, then a finite estimate of $L_{\max}$ or $L_{\min}$ can be obtained only by imposing stronger restrictions on g than are imposed in this paper. A further problem is that a test of boundedness of $L_{\max}$ or $L_{\min}$ has unavoidably low power because, as is explained in Section 3.4, it amounts to a test of multiple one-sided hypotheses about a population mean vector. Low power makes it unlikely that a false hypothesis of boundedness of $L_{\max}$ or $L_{\min}$ can be rejected even if $\hat{L}_{\max}$ and $\hat{L}_{\min}$ are infinite.[2]

We also assume:

Assumption 3: There is a vector $\boldsymbol{h}$ satisfying $\Pi'\boldsymbol{h} - \boldsymbol{m} = 0$ and $S\boldsymbol{h} \le -\boldsymbol{\varepsilon}$ for some vector $\varepsilon > 0$.

This assumption ensures that problem (6) has a feasible solution with probability approaching 1 as $n \to \infty$ when Π and $\boldsymbol{m}$ are replaced by consistent estimators.[3] It also implies that $L_{\min} \ne L_{\max}$, so

$L(g)$ is not point identified. The methods and results of this paper do not apply to settings in which $L(g)$ is point identified. A method for testing Assumption 3 is described in Section 3.4.

2.1 Further Properties of Problem (6)

This section presents properties of problem (6) that will be used later in this paper. These are well-known properties of linear programs. Their proofs are available in many references on linear programming, such as Hadley (1962).

We begin by putting problem (6) into standard LP form. In standard form, the objective function is maximized, all constraints are equalities, and all variables of optimization are non-negative. Problem (6) can be put into standard form by adding slack variables to the inequality constraints and writing each component of $\boldsymbol{h}$ as the difference between its positive and negative parts. Denote the resulting vector of variables of optimization by $\boldsymbol{z}$. The dimension of $\boldsymbol{z}$ is $2J+M$. There are J variables for the positive parts of the components of $\boldsymbol{h}$, J variables for the negative parts of the components of $\boldsymbol{h}$, and M slack variables for the inequality constraints. The $(2J+M)\times 1$ vector of objective function coefficients is $\bar{\boldsymbol{c}} = (\boldsymbol{c}', -\boldsymbol{c}', 0_{1\times M})'$, where $0_{1\times L}$ is a $1\times M$ vector of zeros. The corresponding constraint matrix has dimension $(K+M)\times(2J+M)$ and is

$$\bar{A} = \begin{pmatrix} \Pi' & -\Pi' & 0_{K\times M} \\ S & -S & I_{M\times M} \end{pmatrix},$$

where $I_{M\times M}$ is the $M\times M$ identity matrix. The vector of right-hand sides of the constraints is the $(K+M)\times 1$ vector

$$\bar{\boldsymbol{m}} = \begin{pmatrix} \boldsymbol{m} \\ 0_{M\times 1} \end{pmatrix}.$$

With this notation, the standard form of (6) is

$$(7) \qquad \underset{\boldsymbol{z}}{\text{maximize}}: \ \bar{\boldsymbol{c}}'\boldsymbol{z} \quad \text{or} \quad -\bar{\boldsymbol{c}}'\boldsymbol{z}$$

$$\text{subject to:} \qquad \bar{A}\boldsymbol{z} = \bar{\boldsymbol{m}}$$

$$\boldsymbol{z} \geq 0 .$$

Maximizing $-\bar{\boldsymbol{c}}'\boldsymbol{z}$ is equivalent to minimizing $\bar{\boldsymbol{c}}'\boldsymbol{z}$.

Make the following assumption.

<u>Assumption 4</u>: Let D be the matrix consisting of any $K+M$ columns of

$$\begin{pmatrix} \Pi' & 0_{K\times M} & \boldsymbol{m} \\ S & I_{M\times M} & 0_{M\times 1} \end{pmatrix}$$

Let $\boldsymbol{a}$ be any $(K+M)\times 1$ vector with $\|\boldsymbol{a}\|=1$. Then $\|D\boldsymbol{a}\|\geq\delta$ for some $\delta>0$.

Assumption 4 ensures that the basic optimal solution(s) to (6) and (7) are nondegenerate. Under assumption 4, the determinant condition holds with probability approaching 1 as $n\to\infty$ and $\delta/2$ in place of δ for the estimates of Π and $\boldsymbol{m}$ described in Section 3. Therefore, asymptotically a test of assumption 4 is not needed.

Let $\boldsymbol{z}_{opt}$ be an optimal solution to either version of (7). Let $\boldsymbol{z}_{B,opt}$ denote the $(K+M)\times 1$ vector of basic variables in the optimal solution. Let $\bar{A}_B$ denote the $(K+M)\times(K+M)$ matrix formed by the columns of $\bar{A}$ corresponding to basic variables. Then

$$\boldsymbol{z}_{B,opt}=\bar{A}_B^{-1}\bar{\boldsymbol{m}}$$

and, under assumption 4, $\boldsymbol{z}_{B,opt}>0$. Now let $\bar{\boldsymbol{c}}_B$ be the $(K+M)\times 1$ vector of components of $\bar{\boldsymbol{c}}$ corresponding to the components of $\boldsymbol{z}_{B,opt}$. The optimal value of the objective function corresponding to basic solution $z_{B,opt}$ is

(8a) $$Z_B=\bar{\boldsymbol{c}}_B{}'\bar{A}_B^{-1}\bar{\boldsymbol{m}}$$

for the maximization version of (6) and

(8b) $$\tilde{Z}_B=-\bar{\boldsymbol{c}}_B{}'\bar{A}_B^{-1}\bar{\boldsymbol{m}}$$

for the minimization version.

In standard form, the dual of problem (6) is

(9) $$\underset{\boldsymbol{q}}{\text{maximize}}:\ \tilde{\boldsymbol{m}}'\tilde{\boldsymbol{q}}\ \text{or}\ -\tilde{\boldsymbol{m}}'\tilde{\boldsymbol{q}}$$

subject to

$$\tilde{A}'\tilde{\boldsymbol{q}}=\boldsymbol{c}$$

$$\tilde{\boldsymbol{q}}\geq 0,$$

where $\tilde{\boldsymbol{q}}$ is a $(2K+M)\times 1$ vector,

$$\tilde{\boldsymbol{m}}'=(0_{1\times M},\boldsymbol{m}',-\boldsymbol{m}')$$

and $\tilde{A}'$ is the $J\times(2K+M)$ matrix

$$\tilde{A}'=\left(S',\quad \Pi,\quad -\Pi\right).$$

Under Assumptions 1-3, (6) and (9) both have feasible solutions. The optimal solutions of (6) and (9) are bounded, and the optimal values of the objective functions of (6) and (9) are the same. The dual problem is used in Section 3.3 to form a test of assumption 2.

3. ESTIMATION OF $L_{\max}$ AND $L_{\min}$

This section presents consistent estimators of $L_{\max}$ and $L_{\min}$. The asymptotic distributions of these estimators are presented, and methods obtaining confidence intervals are described. Tests of Assumptions 2 and 3 are outlined.

3.1 Consistent Estimators of $L_{\max}$ and $L_{\min}$

$L_{\max}$ and $L_{\min}$ can be estimated consistently by replacing Π and $\boldsymbol{m}$ in (6) with consistent estimators. To this end, define

$$\hat{m}_k = n^{-1}\sum_{i=1}^{n} Y_i I(W_i = w_k);\; k = 1,..,K$$

and

$$\hat{\pi}_{jk} = n^{-1}\sum_{j=1}^{J}\sum_{k=1}^{K} I(X_i = x_j)I(W_i = w_k);\; j = 1,...,J;\; k = 1,...,K .$$

Then $\hat{m}_k$ and $\hat{\pi}_{jk}$, respectively, are strongly consistent estimators of m_k and π_{jk}. Define $\hat{\boldsymbol{m}} = (\hat{m}_1,...,\hat{m}_K)'$. Define $\hat{\Pi}$ as the $J \times K$ matrix whose (j,k) element is $\hat{\pi}_{jk}$. Define $\hat{L}_{\max}$ and $\hat{L}_{\min}$ as the optimal values of the objective functions of the linear programs

(10) $$\underset{\boldsymbol{h}}{\text{maximize}}\;\; (\underset{\boldsymbol{h}}{\text{minimize}}):\; \boldsymbol{c}'\boldsymbol{h}$$

$$\text{subject to: } \hat{\Pi}'\boldsymbol{h} = \hat{\boldsymbol{m}}$$

$$S\boldsymbol{h} \le 0 .$$

Assumptions 2 and 3 ensure that (10) has a feasible solution and a bounded optimal solution with probability approaching 1 as $n \to \infty$. The standard form of (10) is

(11) $$\underset{\boldsymbol{z}}{\text{maximize}}:\; \bar{\boldsymbol{c}}'\boldsymbol{z} \quad \text{or} \quad -\bar{\boldsymbol{c}}'\boldsymbol{z}$$

$$\text{subject to: } \hat{\bar{A}}\boldsymbol{z} = \hat{\bar{\boldsymbol{m}}}$$

$$\boldsymbol{z} \ge 0 ,$$

where

$$\hat{\bar{A}} = \begin{pmatrix} \hat{\Pi}' & -\hat{\Pi}' & 0_{K\times M} \\ S & -S & I_{M\times M} \end{pmatrix}$$

and

$$\hat{\bar{\boldsymbol{m}}} = \begin{pmatrix} \hat{\boldsymbol{m}} \\ 0_{M\times 1} \end{pmatrix}.$$

As a consequence of 8(a)-8(b) and the strong consistency of $\hat{\Pi}$ and $\hat{\bar{\boldsymbol{m}}}$ for Π and $\bar{\boldsymbol{m}}$, respectively, we have

<u>Theorem 1</u>: Let assumptions 1-3 hold. As $n \to \infty$, $\hat{L}_{\max} \to L_{\max}$ almost surely and $\hat{L}_{\min} \to L_{\min}$ almost surely. ■

3.2 The Asymptotic Distributions of $\hat{L}_{\max}$ and $\hat{L}_{\min}$

This section obtains the asymptotic distributions of $\hat{L}_{\max}$ and $\hat{L}_{\min}$ and shows how to use these to obtain confidence regions for the identification interval $[L_{\min}, L_{\max}]$ and the linear functional $L(g)$. We assume that

<u>Assumption 5</u>: $E(Y^2 | W = w_k) < \infty$ for each $k = 1,...,K$.

We begin by deriving the asymptotic distribution of $\hat{L}_{\max}$. The derivation of the asymptotic distribution of $\hat{L}_{\min}$ is similar. Let $\mathcal{B}_{\max}$ denote the set of optimal basic solutions to the maximization version of (6). Let $\mathcal{K}_{\max}$ denote the number of basic solutions in $\mathcal{B}_{\max}$. The basic solutions are at vertices of the feasible region. Because there are only finitely many vertices, the difference between the optimal value of the objective function of (6) and the value of the objective function at any non-optimal feasible vertex is bounded away from zero. Moreover, the law of the iterated logarithm ensures that $\hat{\Pi}$ and $\hat{m}$, respectively, are in arbitrarily small neighborhoods of Π and m with probability 1 for all sufficiently large n. Therefore, for all sufficiently large n, the probability is zero that a basic solution is optimal in (10) but not (6).

Let $k = 1, 2,...$ index the basic solutions to (10). Let the random variable $\hat{Z}_k$ denote the value of the objective function corresponding to basic solution k. Let $\hat{\bar{A}}_k$ and $\bar{\boldsymbol{c}}_k$, respectively, be the versions of $\hat{\bar{A}}_B$ and $\bar{\boldsymbol{c}}_B$ associated with the k'th basic solution of (6) or (10). Then,

(12) $$\hat{Z}_k = \bar{c}_k{}' \hat{\bar{A}}_k^{-1} \hat{\bar{\boldsymbol{m}}}.$$

Moreover, with probability 1 for all sufficiently large n,

$$\hat{L}_{\max} = \max_k \hat{Z}_k ,$$

and

$$n^{1/2}(\hat{L}_{\max} - L_{\max}) = n^{1/2}\left(\max_k \hat{Z}_k - L_{\max}\right).$$

Let Z_k denote the value of the objective function of (6) at the k'th basic solution. Then $Z_k = L_{\max}$ if basic solution k is optimal. Because $\mathcal{B}_{\max}$ contains the optimal basic solution to (6) or (10) with probability 1 for all sufficiently large n,

$$n^{1/2}(\hat{L}_{\max} - L_{\max}) = n^{1/2} \max_{k \in \mathcal{B}_{\max}} (\hat{Z}_k - L_{\max}) + o_p(1)$$

$$= n^{1/2} \max_{k \in \mathcal{B}_{\max}} (\hat{Z}_k - Z_k) + o_p(1).$$

An application of the delta method yields

$$n^{1/2}(\hat{Z}_k - Z_k) = n^{1/2} \bar{\boldsymbol{c}}_k{}' (\hat{\bar{A}}_k^{-1} \hat{\bar{\boldsymbol{m}}} - \bar{A}_k^{-1} \bar{\boldsymbol{m}})$$

$$= \bar{\boldsymbol{c}}_k{}' \bar{A}_k^{-1} [n^{1/2}(\hat{\bar{\boldsymbol{m}}} - \bar{\boldsymbol{m}}) - n^{1/2}(\hat{\bar{A}}_k - \bar{A}_k) \bar{A}_k^{-1} \bar{\boldsymbol{m}}] + o_p(1)$$

for $k \in \mathcal{B}_{\max}$, where $\bar{A}_k$ is the version of $\bar{A}_B$ that is associated with basic solution k. The elements of $\hat{\bar{A}}_k$ and $\hat{\bar{\boldsymbol{m}}}$ are sample moments or constants, depending on the basic solution, and not all are constants. In addition $E(\hat{\bar{A}}_k) = \bar{A}_k$ and $E(\hat{\bar{m}}) = \bar{m}$. Therefore, it follows from the Lindeberg-Levy and Cramér-Wold theorems that the random components of $n^{1/2}(\hat{Z}_k - Z_k)$ ($k \in \mathcal{B}_{\max}$) are asymptotically multivariate normally distributed with mean 0. There may be some values of $k \in \mathcal{B}_{\max}$ for which $n^{1/2}(\hat{Z}_k - Z_k)$ is deterministically 0. This can happen, for example, if the objective function of (6) is proportional to the left-hand side of one of the shape constraints. In such cases, the entire vector $n^{1/2}(\hat{Z}_k - Z_k)$ ($k \in \mathcal{B}_{opt}$) has asymptotically a degenerate multivariate normal distribution. Thus, $n^{1/2}(\hat{L}_{\max} - L_{\max})$ is asymptotically distributed as the maximum of a random vector with a possibly degenerate multivariate normal distribution whose mean is zero. Denote the random vector by $Z_{\max}$ and its covariance matrix by $\Sigma_{\max}$. In general, $\Sigma_{\max}$ is a large matrix whose elements are algebraically complex and tedious to enumerate. Section 4 presents bootstrap methods for estimating the asymptotic distribution of $n^{1/2}(\hat{L}_{\max} - L_{\max})$ that do not require knowledge of $\Sigma_{\max}$ or $\mathcal{B}_{\max}$.

Now consider $\hat{L}_{\min}$. Let $\mathcal{B}_{\min}$ denote the set of optimal basic solutions to the minimization version of (6), and let $\mathcal{K}_{\min}$ denote the number of basic solutions in $\mathcal{B}_{\min}$. Define $\tilde{Z}_k = -\bar{c}_k{}' \bar{A}_k^{-1} \bar{\boldsymbol{m}}$ and $\hat{\tilde{Z}}_k = -\bar{c}_k{}' \hat{\bar{A}}_k^{-1} \hat{\bar{\boldsymbol{m}}}$. Then arguments like those made for $L_{\max}$ show that

$$-n^{1/2}(\hat{L}_{\min} - L_{\min}) = n^{1/2} \max_{k \in \mathcal{B}_{\min}} (\hat{\hat{Z}}_k + L_{\min}) + o_p(1)$$

$$= n^{1/2} \max_{k \in \mathcal{B}_{\min}} (\hat{\hat{Z}}_k - \tilde{Z}_k) + o_p(1).$$

The asymptotic distributional arguments made for $n^{1/2}(\hat{L}_{\max} - L_{\max})$ also apply to $n^{1/2}(\hat{L}_{\min} - L_{\min})$. Therefore, $n^{1/2}(\hat{L}_{\min} - L_{\min})$ is asymptotically distributed as the minimum of a random vector with a possibly degenerate multivariate normal distribution whose mean is zero. Denote this vector by $Z_{\min}$ and its covariance matrix by $\Sigma_{\min}$. Like $\Sigma_{\max}$, $\Sigma_{\min}$ is a large matrix whose elements are algebraically complex. Section 4 presents bootstrap methods for estimating the asymptotic distribution of $n^{1/2}(\hat{L}_{\min} - L_{\min})$ that do not require knowledge of $\Sigma_{\min}$ or $\mathcal{B}_{\min}$.

It follows from the foregoing discussion that $[n^{1/2}(\hat{L}_{\max} - L_{\max}), n^{1/2}(\hat{L}_{\min} - L_{\min})]$ is asymptotically distributed as $(\max Z_{\max}, \min Z_{\min})$. $Z_{\max}$ and $Z_{\min}$ are not independent of one another. The bootstrap procedure described in Section 4 consistently estimates the asymptotic distribution of $[n^{1/2}(\hat{L}_{\max} - L_{\max}), n^{1/2}(\hat{L}_{\min} - L_{\min})]$.

The foregoing results are summarized in the following theorem.

<u>Theorem 2</u>: Let assumptions 1-5 hold. As $n \to \infty$, (i) $n^{1/2}(\hat{L}_{\max} - L_{\max})$ converges in distribution to the maximum of a $\mathcal{K}_{\max} \times 1$ random vector $Z_{\max}$ with a possibly degenerate multivariate normal distribution, mean zero, and covariance matrix $\Sigma_{\max}$; (ii) $n^{1/2}(\hat{L}_{\min} - L_{\min})$ converges in distribution to the minimum of a $\mathcal{K}_{\min} \times 1$ random vector $Z_{\min}$ with a possibly degenerate multivariate normal distribution, mean zero, and covariance matrix $\Sigma_{\min}$; (iii) $[n^{1/2}(\hat{L}_{\max} - L_{\max}), n^{1/2}(\hat{L}_{\min} - L_{\min})]$ converges in distribution to $(\max Z_{\max}, \min Z_{\min})$. ■

The asymptotic distributions of $n^{1/2}(\hat{L}_{\min} - L_{\min})$, $n^{1/2}(\hat{L}_{\max} - L_{\max})$, and $[n^{1/2}(\hat{L}_{\max} - L_{\max}), n^{1/2}(\hat{L}_{\min} - L_{\min})]$ are simpler if the maximization and minimization versions of (6) have unique optimal solutions. Specifically, $n^{1/2}(\hat{L}_{\min} - L_{\min})$, $n^{1/2}(\hat{L}_{\max} - L_{\max})$ are asymptotically univariate normally distributed, and $[n^{1/2}(\hat{L}_{\max} - L_{\max}), n^{1/2}(\hat{L}_{\min} - L_{\min})]$ is asymptotically bivariate normally distributed. Let $\sigma^2_{\max}$ and $\sigma^2_{\min}$, respectively, denote the variances of the asymptotic distributions of $n^{1/2}(\hat{L}_{\max} - L_{\max})$ and $n^{1/2}(\hat{L}_{\min} - L_{\min})$. Let ρ denote the correlation coefficient of the asymptotic bivariate normal distribution of $[n^{1/2}(\hat{L}_{\max} - L_{\max}), (\hat{L}_{\min} - L_{\min})]$. Let $N_2(0,\rho)$ denote the bivariate normal distribution with variances of 1 and correlation coefficient ρ. Then the following corollary to Theorem 1 holds.

Corollary 1: Let assumptions 1-5 hold. If the optimal solution to the maximization version of (6) is unique, then $n^{1/2}(\hat{L}_{\max} - L_{\max})/\sigma_{\max} \to^d N(0,1)$. If the optimal solution to the minimization version of (6) is unique, then $n^{1/2}(\hat{L}_{\max} - L_{\max})/\sigma_{\min} \to^d N(0,1)$. If the optimal solutions to both versions of (6) are unique, then $[n^{1/2}(\hat{L}_{\max} - L_{\max})/\sigma_{\max}, n^{1/2}(\hat{L}_{\min} - L_{\min})/\sigma_{\min}] \to^d N_2(0,\rho)$. ■

Theorem 2 and Corollary 1 can be used to obtain asymptotic confidence intervals for $[L_{\min}, L_{\max}]$ and $L(g)$. A symmetrical asymptotic $1-\alpha$ confidence interval for $[L_{\min}, L_{\max}]$ is $[\hat{L}_{\min} - n^{-1/2}c_\alpha, \hat{L}_{\max} + n^{-1/2}c_\alpha]$, where c_α satisfies

$$\lim_{n\to\infty} P(\hat{L}_{\min} - n^{-1/2}c_\alpha \le L_{\min}, \hat{L}_{\max} + n^{-1/2}c_\alpha > L_{\max}) = 1-\alpha .$$

Equal-tailed and minimum length asymptotic confidence interval can be obtained in a similar way.

A confidence interval for $L(g)$ can be obtained by using ideas described by Imbens and Manski (2004) and Stoye (2009). In particular, as is discussed by Imbens and Manski (2004), an asymptotically valid pointwise $1-\alpha$ confidence interval for $L(g)$ can be obtained as the intersection of one-sided confidence intervals for $\hat{L}_{\min}$ and $\hat{L}_{\max}$.[4] Thus $[\hat{L}_{\min} - n^{-1/2}c_{\alpha,\min}, \hat{L}_{\max} + n^{-1/2}c_{\alpha,\max}]$ is an asymptotic $1-\alpha$ confidence interval for $L(g)$, where $c_{\alpha,\min}$ and $c_{\alpha,\max}$, respectively, satisfy

$$\lim_{n\to\infty} P[n^{1/2}(\hat{L}_{\min} - L_{\min}) \le c_{\alpha,\min}] = 1-\alpha$$

and

$$\lim_{n\to\infty} P[n^{1/2}(\hat{L}_{\max} - L_{\max}) \ge -c_{\alpha,\max}] = 1-\alpha .$$

Estimating the critical values $c_{\alpha,\min}$ and $c_{\alpha,\max}$, like estimating the asymptotic distributions of $n^{1/2}(\hat{L}_{\max} - L_{\max})$, $n^{1/2}(\hat{L}_{\min} - L_{\min})$, and $[n^{1/2}(\hat{L}_{\max} - L_{\max}), n^{1/2}(\hat{L}_{\min} - L_{\min})]$, is difficult because $\Sigma_{\max}$ and $\Sigma_{\min}$ are complicated unknown matrices, and $\mathcal{B}_{\max}$ and $\mathcal{B}_{\min}$ are unknown sets. Section 4 presents bootstrap methods for estimating $c_{\alpha,\min}$ and $c_{\alpha,\max}$ without knowledge of $\Sigma_{\max}$, $\Sigma_{\min}$, $\mathcal{B}_{\max}$, and $\mathcal{B}_{\min}$.

3.3 Uniformly Correct Confidence Intervals

The asymptotic distributional results presented in Section 3.2 do not hold uniformly over all distributions of (Y, X, W) satisfying assumptions 1-5. This is because the asymptotic distributions of $n^{1/2}(\hat{L}_{\max} - L_{\max})$ and $n^{1/2}(\hat{L}_{\min} - L_{\min})$ are discontinuous when the number of optimal basic solutions changes. This section explains how to overcome this problem for $n^{1/2}(\hat{L}_{\max} - L_{\max})$. Similar arguments apply to $n^{1/2}(\hat{L}_{\min} - L_{\min})$ and to the joint distribution of $[n^{1/2}(\hat{L}_{\min} - L_{\min}), n^{1/2}(\hat{L}_{\max} - L_{\max})]$. The results for these quantities are presented without further explanation.

We now consider $n^{1/2}(\hat{L}_{\max} - L_{\max})$. Let k be any basic solution to problem (6). Let $\hat{Z}_k$ denote the value of the objective function of the maximization version of (10) corresponding to basic solution k. Let $\mathcal{B}_{\max}$ be a set that contains every optimal basic solution to the maximization version of (6). Let $L_{k,\max}$ be the value of the objective function of the maximization version of (6) corresponding to basic solution k. Then $n^{1/2}(L_{\max} - L_{k,\max}) > 0$ for every $k \notin \mathcal{B}_{\max}$ and $n^{1/2}(L_{\max} - L_{k,\max}) = 0$ for every optimal k. Therefore, for any $c > 0$ and any $\tilde{k} \in \mathcal{B}_{\max}$,

$$P[n^{1/2}(\hat{L}_{\max} - L_{\max}) > -c] \geq P\{\max_{k \in \mathcal{B}_{\max}} [n^{1/2}(\hat{Z}_k - L_{k,\max}) - n^{1/2}(L_{\max} - L_{k,\max})] > -c\}$$

$$\geq P[n^{1/2}(\hat{Z}_{\tilde{k}} - L_{\tilde{k},\max}) > -c].$$

Moreover

$$P[n^{1/2}(\hat{L}_{\max} - L_{\max}) > -c] \geq \min_{k \in \mathcal{B}_{\max}} P[n^{1/2}(\hat{Z}_k - L_{k,\max}) > -c]$$

For any $k \in \mathcal{B}_{\max}$ and $\alpha \in (0,1)$, let $c_{k,\alpha,\max}$ satisfy

$$P[n^{1/2}(\hat{Z}_k - L_{k,\max}) > -c_{k,\alpha,\max}] = 1 - \alpha .$$

Define $\bar{c}_{\alpha,\max} = \max_{k \in \mathcal{B}_{\max}} c_{k,\alpha,\max}$. Then

$$P[n^{1/2}(\hat{L}_{\max} - L_{\max}) > -\bar{c}_{\alpha,\max}] \geq 1 - \alpha$$

uniformly over distributions of (Y, X, W) that satisfy assumptions 1-5.

Now define

$$\mathcal{B}_{n,\max} = \{k : \hat{L}_{\max} - \hat{Z}_k \leq n^{-1/2} \log n\}.$$

For all sufficiently large n, $\mathcal{B}_{n,\max}$ contains all optimal basic solutions to the maximization version of (10) with probability 1. Define $c_{\alpha,\max} = \max_{k \in \mathcal{B}_{n,\max}} c_{k,\alpha,\max}$. Then

$$P[n^{1/2}(\hat{L}_{\max} - L_{\max}) > -c_{\alpha,\max}] \geq 1 - \alpha + \varepsilon_n ,$$

uniformly over distributions of (Y, X, W) that satisfy assumptions 1-5, where $\varepsilon_n \to 0$ with probability 1 as $n \to \infty$. Moreover $(-\infty, \hat{L}_{\max} + n^{-1/2} c_{\alpha,\max}]$ is a confidence interval for $L_{\max}$ whose asymptotic coverage probability is at least $1-\alpha$ uniformly over distributions of (Y, X, W) that satisfy assumptions 1-5.

To obtain a uniform confidence interval for $L_{\min}$, let $\hat{Z}_k$ denote the value of the objective function of the minimization version of (10) corresponding to basic solution k. Let $L_{k,\min}$ be the value of the objective function of the minimization version of (6) corresponding to basic solution k. Define

$$\mathcal{B}_{n,\min} = \{k : \hat{Z}_k - \hat{L}_{\min} \leq n^{-1/2} \log n\} .$$

Define $c_{k,\alpha,\min}$ by

$$P[n^{1/2}(L_{k,\min} - \hat{Z}_k) \leq c_{k,\alpha,\min}] = 1 - \alpha .$$

Finally, define $c_{\alpha,\min} = \max_{k \in \mathcal{B}_{n,\min}} c_{k,\alpha,\min}$. Then arguments like those made for $L_{\max}$ show that

$$P[n^{1/2}(\hat{L}_{\min} - L_{\min}) \leq c_{\alpha,\min}] \geq 1 - \alpha + \varepsilon_n ,$$

uniformly over distributions of (Y, X, W) that satisfy assumptions 1-5, where $\varepsilon_n \to 0$ with probability 1 as $n \to \infty$. Moreover $[\hat{L}_{\min} - n^{-1/2} c_{\alpha,\min}, \infty)$ is a confidence interval for $L_{\min}$ whose asymptotic coverage probability is at least $1-\alpha$ uniformly over distributions of (Y, X, W) that satisfy assumptions 1-5.

To obtain a confidence region for $[L_{\min}, L_{\max}]$, for each $k_{\min} \in \mathcal{B}_{n,\min}$ and $k_{\max} \in \mathcal{B}_{n,\max}$, let $c_{\alpha,k_{\min},k_{\max}}$ satisfy

$$P[n^{1/2}(\hat{Z}_{k_{\min}} - L_{k_{\min},\min}) \leq c_{\alpha,k_{\min},k_{\max}} ;\ n^{1/2}(\hat{Z}_{k_{\max}} - L_{k_{\max},\max}) \geq -c_{\alpha,k_{\min},k_{\max}}] = 1 - \alpha ,$$

where $\hat{Z}_{k_{\min}}$ and $\hat{Z}_{k_{\max}}$, respectively, are the values of the objective functions of the minimization and maximization versions of (10) corresponding to basic solutions $k_{\min}$ and $k_{\max}$. Define

$$c_\alpha = \max_{k_{\min} \in \mathcal{B}_{n,\min}, k_{\max} \in \mathcal{B}_{n,\max}} c_{\alpha,k_{\min},k_{\max}} .$$

Then

$$P(\hat{L}_{\min} - n^{-1/2} c_\alpha \leq L_{\min}, L_{\max} \leq \hat{L}_{\max} + c_\alpha) \geq 1 - \alpha + \varepsilon_n$$

uniformly over distributions of (Y, X, W) that satisfy assumptions 1-5, where $\varepsilon_n \to 0$ with probability 1 as $n \to \infty$. Moreover, $[\hat{L}_{\min} - n^{-1/2} c_\alpha, \hat{L}_{\max} + n^{-1/2} c_\alpha]$ is an asymptotic $1-\alpha$ confidence region for $[L_{\min}, L_{\max}]$.

Section 4 outlines bootstrap methods for estimating the critical values $c_{\alpha,\max}$, $c_{\alpha,\min}$, and c_α in applications.

3.4 Testing Assumptions 2 and 3

We begin this section by outlining a test of assumption 2. A linear program has a bounded solution if and only if its dual has a feasible solution. A linear program has a basic feasible solution if it has a feasible solution. Therefore, assumption 2 can be tested by testing the hypothesis that the dual problem (9) has a basic feasible solution. Let $k = 1, ..., k_{\max} \equiv \binom{2K+M}{J}$ index basic solutions to (9).[5] A basic solution is $\tilde{\boldsymbol{q}} = -(\tilde{A}_k')^{-1}\boldsymbol{c}$ for the dual of the maximization version of (6) or $\tilde{\boldsymbol{q}} = (\tilde{A}_k')^{-1}\boldsymbol{c}$ for the dual of the minimization version, where $\tilde{A}_k'$ is the $J \times J$ matrix consisting of the columns of $\tilde{A}'$ corresponding to the k'th basic solution of (9). The dual problem has a basic feasible solution if $-(\tilde{A}_k')^{-1}\boldsymbol{c} \ge 0$ for some k for the maximization version of (6) and $(\tilde{A}_k')^{-1}\boldsymbol{c} \ge 0$ for some k for the minimization version. Therefore, testing boundedness of $L_{\max}$ ($L_{\min}$) is equivalent to testing the hypothesis $H_0: -(\tilde{A}_k')^{-1}\boldsymbol{c} \ge 0$ ($(\tilde{A}_k')^{-1}\boldsymbol{c} \ge 0$) for some k.

To test either hypothesis, define $\hat{\tilde{A}}_k'$ as the matrix that is obtained by replacing the components of Π with the corresponding components of $\hat{\Pi}$ in $\tilde{A}_k$. Then an application of the delta method yields

(13) $$(\hat{\tilde{A}}_k')^{-1}\boldsymbol{c} = (\tilde{A}_k')^{-1}\boldsymbol{c} - (\tilde{A}_k')^{-1}(\hat{\tilde{A}}_k' - \tilde{A}_k)(\tilde{A}_k')^{-1}\boldsymbol{c} + o_p(n^{-1/2}).$$

Equation (13) shows that the hypothesis $H_0: -(\tilde{A}_k')^{-1}\boldsymbol{c} \ge 0$ ($(\tilde{A}_k')^{-1}\boldsymbol{c} \ge 0$) is asymptotically equivalent to a one-sided hypothesis about a vector of population means. Testing $H_0: -(\tilde{A}_k')^{-1}\boldsymbol{c} \ge 0$ ($(\tilde{A}_k')^{-1}\boldsymbol{c} \ge 0$) for some k is asymptotically equivalent to testing a one-sided hypothesis about a vector of $Jk_{\max}$ non-independent population means. Methods for carrying out such tests and issues associated with tests of multiple hypotheses are discussed by Lehmann and Romano (2005) and Romano, Shaikh, and Wolf (2010), among others. The hypothesis of boundedness of $L_{\max}$ is rejected if $H_0: -(\tilde{A}_k')^{-1}\boldsymbol{c} \ge 0$ is rejected for at least one component of $(\tilde{A}_k')^{-1}\boldsymbol{c}$ for each $k = 1, ..., k_{\max}$. The hypothesis of boundedness of

$L_{\min}$ is rejected if $H_0: (\tilde{A}_k')^{-1}\boldsymbol{c} \geq 0$ is rejected for at least one component of $(\tilde{A}_k')^{-1}\boldsymbol{c}$ for each $k = 1,..., k_{\max}$.

We now consider assumption 3. Specifically, we describe a test of the null hypothesis, H_0, that there is a vector $\boldsymbol{g}$ satisfying $\Pi'\boldsymbol{g} - \boldsymbol{m} = 0$ and $S\boldsymbol{g} \leq -\boldsymbol{\varepsilon}$ for some $M \times 1$ vector $\boldsymbol{\varepsilon} > 0$. A test can be carried out by solving the quadratic programming problem

(14) $$\underset{\boldsymbol{g}}{\text{minimize}}: \hat{Q}(\boldsymbol{g}) \equiv \left\| \hat{\Pi}'\boldsymbol{g} - \hat{\boldsymbol{m}} \right\|^2$$

subject to

$$S\boldsymbol{g} \leq -\boldsymbol{\varepsilon},$$

where $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^K$. Let $\hat{Q}_{opt}$ denote the optimal value of the objective function in (14). Under H_0, $\hat{Q}_{opt} \to^p 0$. Therefore, the result $\hat{Q}_{opt} = 0$ is consistent with H_0.

A large value of $\hat{Q}_{opt}$ is inconsistent with H_0. An asymptotically valid critical value can be obtained by using the bootstrap. Let $j = 1, 2,...$ index the sets of constraints that may be binding in the optimal solution to (14), including the empty set, which represents the possibility that no constraints are binding. Let S_j denote the matrix consisting of the rows of S corresponding to constraints j. Set $S_j = 0$ if no constraints bind. The bootstrap procedure is:

(1) Use the estimation data $\{Y_i, X_i, W_i\}$ to solve problem (14). The optimal $\boldsymbol{g}$ is not unique. Use some procedure to choose a unique $\boldsymbol{g}$ from the optimal ones. One possibility is to set $\boldsymbol{g} = \hat{\boldsymbol{g}}_{\hat{j}}$, where

(15) $$\hat{\boldsymbol{g}}_{\hat{j}} = A_{11}^{+}\hat{\Pi}\hat{\boldsymbol{m}} - A_{12}^{+}\boldsymbol{\varepsilon},$$

$\hat{j}$ is the set of constraints that are binding in (14), A_{11}^{+} is a $J \times J$ matrix, A_{12}^{+} is a matrix with J rows and as many columns as there are binding constraints, $A_{12}^{+} = 0$ if there are no binding constraints, and

(16) $$\begin{pmatrix} A_{11}^{+} & A_{12}^{+} \\ A_{21}^{+} & A_{22}^{+} \end{pmatrix} = \begin{pmatrix} \hat{\Pi}\hat{\Pi}' & S_{\hat{j}}' \\ S_{\hat{j}} & 0 \end{pmatrix}^{+}$$

is the Moore-Penrose generalized inverse of

$$\begin{pmatrix} \hat{\Pi}\hat{\Pi}' & S_{\hat{j}}' \\ S_{\hat{j}} & 0 \end{pmatrix}.$$

Define $\hat{P}_{opt} = \hat{\Pi}'\hat{\boldsymbol{g}}_{\hat{j}} - \hat{\boldsymbol{m}}$ and $\hat{Q}_{opt} = \hat{P}_{opt}'\hat{P}_{opt}$.

(ii) Generate a bootstrap sample $\{Y_i^*, X_i^*, W_i^* : i = 1,...,n\}$ by sampling the estimation data $\{Y_i, X_i, W_i : i = 1,...,n\}$ randomly with replacement. Compute the bootstrap versions of $\hat{m}_k$ and $\hat{\pi}_{jk}$. These are

(17) $$m_k^* = n^{-1}\sum_{i=1}^{n} Y_i^* I(W_i^* = w_k)$$

and

(18) $$\pi_{jk}^* = n^{-1}\sum_{i=1}^{n} I(X_i^* = x_j) I(W_i^* = w_k).$$

Compute $P_{opt}^* = \hat{\Pi}^{*\prime}\boldsymbol{g}_{\hat{j}}^* - \boldsymbol{m}^*$ and $Q_{opt}^* = \hat{P}_{opt}^{*\prime} P_{opt}^*$, where $\boldsymbol{g}_{\hat{j}}^*$ is obtained by replacing $\hat{\Pi}$ and $\hat{\boldsymbol{m}}$ with Π^* and $\boldsymbol{m}^*$, respectively, in (15) and (16).

(iii) Estimate the asymptotic distribution of $\hat{Q}_{opt}$ by the empirical distribution of $(P_{opt}^* - \hat{P}_{opt})'(P_{opt}^* - \hat{P}_{opt})$ that is obtained by repeating steps (i) and (ii) many times (the bootstrap distribution). Estimate the asymptotic α level critical value of $\hat{Q}_{opt}$ by the $1-\alpha$ quantile of the bootstrap distribution of $(P_{opt}^* - \hat{P}_{opt})'(P_{opt}^* - \hat{P}_{opt})$. Denote this critical value by q_α^*.

If H_0 is true and the set of binding constraints in the population quadratic programming problem,

(19) $$\underset{\boldsymbol{g}}{\text{minimize:}} \ \|\Pi'\boldsymbol{g} - \boldsymbol{m}\|^2$$

subject to

$$S\boldsymbol{g} \le \varepsilon$$

is unique (say $j = j_{opt}$), then $\hat{j} = j_{opt}$ with probability 1 for all sufficiently large n. The bootstrap estimates the asymptotic distribution of $n\hat{Q}_{opt}$ consistently because $\hat{P}_{opt}$ is a smooth function of sample moments.[6] Therefore $P(\hat{Q}_{opt} > q_\alpha^*) \to \alpha$ as $n \to \infty$.

Now suppose that (19) has several optimal solutions with distinct sets of binding constraints, a circumstance that we consider unlikely because it requires a very special combination of values of Π, $\boldsymbol{m}$, and $\boldsymbol{\varepsilon}$. Let C_{opt} denote the sets of optimal binding constraints. Then $n\hat{Q}_{opt}$ is asymptotically distributed as $\min_{j\in C_{opt}} n\left\|\hat{\Pi}\hat{\boldsymbol{g}}_j - \hat{\boldsymbol{m}}\right\|^2$. But $\left\|\hat{\Pi}\hat{\boldsymbol{g}}_{\hat{j}} - \hat{\boldsymbol{m}}\right\|^2 \ge \min_{j\in C_{opt}}\left\|\hat{\Pi}\hat{\boldsymbol{g}}_j - \hat{\boldsymbol{m}}\right\|^2$. The bootstrap critical value is obtained under the incorrect assumption that $\hat{j}$ is asymptotically the only set of binding constraints in (19). Therefore, $\lim_{n\to\infty} P(\hat{Q}_{opt} > q_\alpha^*) \leq \alpha$. The probability of rejecting a correct H_0 is less than or equal to the nominal probability.

4. BOOTSTRAP ESTIMATION OF THE ASYMPTOTIC DISTRIBUTIONS OF $\hat{L}_{\max}$ *and* $\hat{L}_{\min}$

This section present two bootstrap procedures that estimate the asymptotic distributions of $n^{1/2}(\hat{L}_{\max} - L_{\max})$, $n^{1/2}(\hat{L}_{\min} - L_{\min})$, and $[n^{1/2}(\hat{L}_{\max} - L_{\max}), n^{1/2}(\hat{L}_{\min} - L_{\min})]$ without requiring knowledge of $\Sigma_{\max}$, $\Sigma_{\min}$, $\mathcal{B}_{\max}$, or $\mathcal{B}_{\min}$. The procedures also estimate the critical values $c_{\alpha,\min}$ and $c_{\alpha,\max}$. The first procedure yields confidence regions for $[L_{\min}, L_{\max}]$ and $L(g)$ with asymptotically correct coverage probabilities. That is, the asymptotic coverage probabilities of these regions equal the nominal coverage probabilities. However, this procedure has the disadvantage of requiring a user-selected tuning parameter. The procedure's finite-sample performance can be sensitive to the choice of the tuning parameter, and a poor choice can cause the true coverage probabilities to be considerably lower than the nominal ones. The second procedure does not require a user-selected tuning parameter. It yields confidence regions with asymptotically correct coverage probabilities if the optimal solutions to the maximization and minimization versions of problem (6) are unique (that is, if $\mathcal{B}_{\max}$, and $\mathcal{B}_{\min}$ each contain only one basic solution). Otherwise, the asymptotic coverage probabilities are equal to or greater than the nominal coverage probabilities. The procedures are described in Section 4.1. Section 4.2 presents the results of a Monte Carlo investigation of the numerical performance of the procedures.

4.1 The Bootstrap Procedures

This section describes the two bootstrap procedures. Both assume that the optimal solutions to the maximization and minimization versions of problem (10) are random. The procedures are not needed for deterministic optimal solutions. Let $\{c_n : n = 1, 2, ...\}$ be a sequence of positive constants such that $c_n \to 0$ and $c_n[n/(\log\log n)]^{1/2} \to \infty$ as $n \to \infty$. Let P^* denote the probability measure induced by bootstrap sampling.

The first bootstrap procedure is as follows.

(i) Generate a bootstrap sample $\{Y_i^*, X_i^*, W_i^* : i = 1, ..., n\}$ by sampling the estimation data $\{Y_i, X_i, W_i : i = 1, ..., n\}$ randomly with replacement. Use (17) and (18) to compute the bootstrap versions of $\hat{m}_k$ and $\hat{\pi}_{jk}$, which are m_k^* and π_{jk}^*. Define Π^* and $\boldsymbol{m}^*$, respectively, as the matrix and vector that are obtained by replacing the estimation sample with the bootstrap sample in $\hat{\Pi}$ and $\hat{\boldsymbol{m}}$. For any basic solution k to problem (6), define $\bar{A}_k^*$ and $\bar{\boldsymbol{m}}^*$ by replacing the estimation sample with the bootstrap sample in $\hat{\bar{A}}_k$ and $\hat{\bar{m}}$.

(ii) Define problem (B10) as problem (10) with Π^* and $\boldsymbol{m}^*$ in place of $\hat{\Pi}$ and $\hat{\boldsymbol{m}}$. Solve (B10). Let k denote the resulting optimal basic solution. Let $\hat{L}_{k,\max}$ and $\hat{L}_{k\min}$, respectively, denote the values of the objective function of the maximization and minimization versions of (10) at basic solution k. For basic solution k, define

$$\Delta_{1k}^* = n^{1/2}(\bar{\boldsymbol{c}}_k{}'\bar{A}_k^{*-1}\hat{\bar{\boldsymbol{m}}}^* - \bar{\boldsymbol{c}}_k{}'\hat{\bar{A}}_k^{-1}\hat{\bar{\boldsymbol{m}}})$$

and

$$\Delta_{2k}^* = -n^{1/2}(\bar{c}_k{}'\bar{A}_k^{*-1}\hat{\bar{\boldsymbol{m}}}^* - \bar{\boldsymbol{c}}_k{}'\hat{\bar{A}}_k^{-1}\hat{\bar{\boldsymbol{m}}}) .$$

(iii) Repeat steps (i) and (ii) many times. Define $\hat{\mathcal{B}}_{\max} = \{k : |\hat{L}_{k,\max} - \hat{L}_{\max}| \le c_n\}$ and $\hat{\mathcal{B}}_{\min} = \{k : |\hat{L}_{k,\min} - \hat{L}_{\min}| \le c_n\}$.

(iv) Estimate the distributions of $n^{1/2}(\hat{L}_{\max} - L_{\max})$, $n^{1/2}(\hat{L}_{\min} - L_{\min})$, and $[n^{1/2}(\hat{L}_{\max} - L_{\max}), n^{1/2}(\hat{L}_{\min} - L_{\min})]'$, respectively, by the empirical distributions of $\max_{k\in\hat{\mathcal{B}}_{\max}} \Delta_{1k}^*$, $-\max_{k\in\hat{\mathcal{B}}_{\min}} \Delta_{2k}^*$, and $(\max_{k\in\hat{\mathcal{B}}_{\max}} \Delta_{1k}^*, -\max_{k\in\hat{\mathcal{B}}_{\min}} \Delta_{2k}^*)$. Estimate $c_{\alpha,\min}$ and $c_{\alpha,\max}$, respectively, by $c_{\alpha,\min}^*$ and $c_{\alpha,\max}^*$, which solve

$$P^*[\min_{k\in\hat{\mathcal{B}}_{\min}} (-\Delta_{2k}^*) \le c_{\alpha,\min}^*] = 1-\alpha$$

$$P^*(\max_{k\in\hat{\mathcal{B}}_{\max}} \Delta_{1k}^* \ge -c_{\alpha,\max}^*) = 1-\alpha .$$

Asymptotically, $n^{1/2}(\bar{\boldsymbol{c}}_k{}'\hat{\bar{A}}_k^{-1}\hat{\bar{\boldsymbol{m}}} - \bar{\boldsymbol{c}}_k{}'\bar{A}_k^{-1}\bar{\boldsymbol{m}})$ $(k \in \mathcal{B}_{\max} \cup \mathcal{B}_{\min})$ is a linear function of sample moments. Therefore, the bootstrap distributions of Δ_{1k}^* and Δ_{2k}^* uniformly consistently estimate the asymptotic distributions of $\pm n^{1/2}(\bar{\boldsymbol{c}}_k{}'\hat{\bar{A}}_k^{-1}\hat{\bar{\boldsymbol{m}}} - \bar{\boldsymbol{c}}_k{}'\bar{A}_k^{-1}\bar{\boldsymbol{m}})$ for $k \in \mathcal{B}_{\max}$ and $k \in \mathcal{B}_{\min}$ (Mammen 1992). In addition, the foregoing procedure consistently estimates $\mathcal{B}_{\max}$ and $\mathcal{B}_{\min}$. Asymptotically, every basic solution that is feasible in problem (6) has a non-zero probability of being optimal in (B10). Therefore, with probability approaching 1 as $n \to \infty$, every feasible basic solution will be realized in sufficiently many bootstrap repetitions. Moreover, it follows from the law of the iterated logarithm that with probability 1 for all sufficiently large n, only basic solutions k in $\mathcal{B}_{\max}$ satisfy $|\hat{L}_{k,\max} - \hat{L}_{\max}| \le c_n$ and only basic solutions $k \in \mathcal{B}_{\min}$ satisfy $|\hat{L}_{k,\min} - \hat{L}_{\min}| \le c_n$. Therefore, $\hat{\mathcal{B}}_{\max} = \mathcal{B}_{\max}$ and $\hat{\mathcal{B}}_{\min} = \mathcal{B}_{\min}$ with probability 1 for all sufficiently large n. It follows that the bootstrap distributions of $\max_{k\in\hat{\mathcal{B}}_{\max}} \Delta_{1k}^*$, $-\max_{k\in\hat{\mathcal{B}}_{\min}} \Delta_{2k}^*$, and $(\max_{k\in\hat{\mathcal{B}}_{\max}} \Delta_{1k}^*, -\max_{k\in\hat{\mathcal{B}}_{\min}} \Delta_{2k}^*,)$ uniformly consistently estimate the asymptotic distributions of $n^{1/2}(\hat{L}_{\max} - L_{\max})$, $n^{1/2}(\hat{L}_{\min} - L_{\min})$ and $[n^{1/2}(\hat{L}_{\max} - L_{\max}), n^{1/2}(\hat{L}_{\min} - L_{\min})]'$, respectively.[7] These distributions are continuous, so c_α^* is a consistent estimator of c_α.

These results are summarized in the following theorem. Let P^* denote the probability measure induced by bootstrap sampling.

Theorem 3: Let assumptions 1-5 hold. Let $n \to \infty$. Under the first bootstrap procedure,

(i) $\sup_{-\infty<z<\infty} | P^*(\max_{k\in\hat{\mathcal{B}}_{\max}} \Delta_{1k}^* \le z) - P[n^{1/2}(\hat{L}_{\max} - L_{\max}) \le z] | \to^p 0$

(ii) $\sup_{-\infty<z<\infty} | P^*(-\max_{k\in\hat{\mathcal{B}}_{\min}} \Delta_{2k}^* \le z) - P[n^{1/2}(\hat{L}_{\min} - L_{\min}) \le z] | \to^p 0$

(iii) $\sup_{-\infty<z_1,z_2<\infty} \left| P^*\left[\begin{pmatrix} \max_{k\in\hat{\mathcal{B}}_{\max}} \Delta_{1k}^* \\ -\max_{k\in\hat{\mathcal{B}}_{\min}} \Delta_{2k}^* \end{pmatrix} \le \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}\right] - P\left[\begin{pmatrix} n^{1/2}(\hat{L}_{\max} - L_{\max}) \\ n^{1/2}(\hat{L}_{\min} - L_{\min}) \end{pmatrix} \le \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}\right] \right| \to^p 0$

(iv) $|c_\alpha^* - c_\alpha| \to^p 0$. ■

The theory of the bootstrap assumes that there are infinitely many bootstrap repetitions, but only finitely many are possible in practice. With finitely many repetitions, it is possible that the first bootstrap procedure does not find all basic solutions k for which $|\hat{L}_{k,\max} - \hat{L}_{\max}| \le c_n$ or $|\hat{L}_{k,\min} - \hat{L}_{\min}| \le c_n$. However, when n is large, basic solutions for which $|\hat{L}_{k,\max} - \hat{L}_{\max}| \le c_n$ or $|\hat{L}_{k,\min} - \hat{L}_{\min}| \le c_n$ have high probabilities, and basic solutions for which neither of these inequalities holds have low probabilities. Therefore, a large number of bootstrap repetitions is unlikely to be needed to find all basic solutions for which one of the inequalities holds. In addition, arguments like those used to prove Theorem 4 below show that if not all basic solutions satisfying $|\hat{L}_{k,\max} - \hat{L}_{\max}| \le c_n$ or $|\hat{L}_{k,\min} - \hat{L}_{\min}| \le c_n$ are found, then the resulting confidence regions have asymptotic coverage probabilities that equal or exceed their nominal coverage probabilities. The error made by not finding all basic solutions satisfying the inequalities is in the direction of overcoverage, not undercoverage.[8]

The second bootstrap procedure is as follows. Note that the optimal solution to the maximization or minimization version of (10) is unique if it is random.

(i) Generate a bootstrap sample $\{Y_i^*, X_i^*, W_i^* : i = 1,...,n\}$ by sampling the estimation data $\{Y_i, X_i, W_i : i = 1,...,n\}$ randomly with replacement. Use (17) and (18) to compute the bootstrap versions of $\hat{m}_k$ and $\hat{\pi}_{jk}$, which are m_k^* and π_{jk}^*. Define Π^* and $\boldsymbol{m}^*$, respectively, as the matrix and vector that are obtained by replacing the estimation sample with the bootstrap sample in $\hat{\Pi}$ and $\hat{\boldsymbol{m}}$. For any basic solution k to problem (6), define $\bar{A}_k^*$ and $\bar{\boldsymbol{m}}^*$ by replacing the estimation sample with the bootstrap sample in $\hat{\bar{A}}_k$ and $\hat{\bar{m}}$.

(ii) Let $\hat{k}_{\max}$ and $\hat{k}_{\min}$, respectively, denote the optimal basic solutions of the maximization and minimization versions of problem (10). Define

$$\Delta_{\hat{k}_{\max}}^* = n^{1/2}(\bar{\boldsymbol{c}}_{\hat{k}_{\max}}{}' \bar{A}_{\hat{k}_{\max}}^{*-1} \hat{\bar{\boldsymbol{m}}}^* - \bar{\boldsymbol{c}}_{\hat{k}_{\max}}{}' \hat{\bar{A}}_{\hat{k}_{\max}}^{-1} \hat{\bar{\boldsymbol{m}}})$$

and

$$\Delta_{\hat{k}_{\min}}^* = -n^{1/2}(\bar{c}_{\hat{k}_{\min}}{}' \bar{A}_{\hat{k}_{\min}}^{*-1} \hat{\bar{\boldsymbol{m}}}^* - \bar{\boldsymbol{c}}_{\hat{k}_{\min}}{}' \hat{\bar{A}}_{\hat{k}_{\min}}^{-1} \hat{\bar{\boldsymbol{m}}}) .$$

(iii) Repeat steps (i) and (ii) many times. Estimate the distributions of $n^{1/2}(\hat{L}_{\max} - L_{\max})$, $n^{1/2}(\hat{L}_{\min} - L_{\min})$, and $[n^{1/2}(\hat{L}_{\max} - L_{\max}), n^{1/2}(\hat{L}_{\min} - L_{\min})]'$, respectively, by the empirical distributions of $\Delta_{\hat{k}_{\max}}^*$, $-\Delta_{\hat{k}_{\min}}^*$, and $(\Delta_{\hat{k}_{\max}}^*, -\Delta_{\hat{k}_{\min}}^*)$. Estimate $c_{\alpha,\min}$ and $c_{\alpha,\max}$, respectively, by $c_{\alpha,\min}^*$ and $c_{\alpha,\max}^*$, which solve

$$P^*(-\Delta_{\hat{k}_{\min}}^* \le c_{\alpha,\min}^*) = 1 - \alpha$$

$$P^*(\Delta_{\hat{k}_{\max}}^* \ge -c_{\alpha,\max}^*) = 1 - \alpha .$$

If the maximization version of (6) has a unique optimal basic solution, $k_{\max,\text{opt}}$, then $\hat{k}_{\max} = k_{\max,\text{opt}}$ with probability 1 for all sufficiently large n. Therefore, the second bootstrap procedure estimates the asymptotic distribution of $n^{1/2}(\hat{L}_{\max} - L_{\max})$ uniformly consistently and $c_{\alpha,\max}^*$ is a consistent estimator of $c_{\alpha,\max}$. Similarly, if the minimization version of (6) has a unique optimal basic solution, then the second bootstrap procedure estimates the asymptotic distribution of $n^{1/2}(\hat{L}_{\min} - L_{\min})$ uniformly consistently, and $c_{\alpha,\min}^*$ is a consistent estimator of $c_{\alpha,\min}$.

If the maximization version of (6) has two or more optimal basic solutions that produce non-deterministic values of the objective function of (10), then the limiting bootstrap distribution of $n^{1/2}(\hat{L}_{\max} - L_{\max})$ depends on $\hat{k}_{\max}$ and is random. In this case, the second bootstrap procedure does not provide a consistent estimator of the distribution of $n^{1/2}(\hat{L}_{\max} - L_{\max})$ or $c_{\alpha,\max}$. Similarly, if the minimization version of (6) has two or more optimal basic solutions that produce non-deterministic values of the objective function of (10), then the second bootstrap procedure does not provide a consistent estimator of the distribution of $n^{1/2}(\hat{L}_{\min} - L_{\min})$ or $c_{\alpha,\min}$. However, the following theorem shows that the asymptotic coverage probabilities of confidence regions based on the inconsistent estimators of $c_{\alpha,\max}$ and $c_{\alpha,\min}$ equal or exceed the nominal coverage probabilities. Thus, the error made by the second bootstrap procedure is in the direction of overcoverage.

Theorem 4: Let assumptions 1-5 hold. Let $n \to \infty$. Under the second bootstrap procedure,

(i) $P(L_{\max} \leq \hat{L}_{\max} + c^*_{\alpha,\max}) \geq 1 - \alpha + o_p(1)$

(ii) $P(L_{\min} \geq \hat{L}_{\min} - c^*_{\alpha,\min}) \geq 1 - \alpha + o_p(1)$. ■

Proof: Only part (i) is proved. The proof of part (ii) is similar. With probability 1 all sufficiently large n, $\hat{k}_{\max} \in \mathcal{B}_{\max}$, so

$$\Delta^*_{\hat{k}_{\max}} \leq \max_{k \in \hat{\mathcal{B}}_{\max}} \Delta^*_{1k}$$

and

$$1 - \alpha = P^*(\Delta^*_{\hat{k}_{\max}} \geq -c^*_{\alpha,\max}) \leq P^*\left(\max_{k \in \hat{\mathcal{B}}_{\max}} \Delta^*_{1k} \geq -c^*_{\alpha,\max}\right).$$

Therefore, by Theorem 3(i)

$$1 - \alpha \geq P[n^{1/2}(\hat{L}_{\max} - L_{\max}) \leq c^*_{\alpha}) + o_p(1). \ ■$$

The bootstrap can also be used to obtain the conservative, uniformly consistent critical values described in Section 3.3. We now outline the bootstrap procedure for $n^{1/2}(\hat{L}_{\max} - L_{\max})$. For basic solution $k \in \mathcal{B}_{n,\max}$, estimate the distribution of $n^{1/2}(\hat{L}_{\max} - L_{k,\max})$ by the bootstrap distribution of Δ^*_{1k}. Estimate the critical value $c_{b,\alpha}$ by the bootstrap critical value $c^*_{b,\alpha}$, which is the solution to $P^*(\Delta^*_{1k} > -c^*_{b,\alpha}) = 1 - \alpha$. Estimate c_{α} by $\max_{k \in \mathcal{B}_{n,\max}} c^*_{b,\alpha}$. Similar bootstrap procedures apply to $n^{1/2}(\hat{L}_{\min} - L_{\min})$ and $[n^{1/2}(\hat{L}_{\min} - L_{\min}), n^{1/2}(\hat{L}_{\max} - L_{\max})]$.

4.2 Monte Carlo Experiments

This section reports the results of Monte Carlo experiments that investigate the numerical performance of the bootstrap procedure of Section 4.1. The design of the experiments mimics the empirical application presented in Section 5. The experiments investigate the finite-sample coverage probabilities of nominal 95% confidence intervals for $[L_{\min}, L_{\max}]$ and $L(g)$.

In the experiments, the support of W is $\{0,1\}$, and $J = 4$ or $J = 6$, depending on the experiment. In experiments with $J = 6$, $X \equiv \{2, 3, 4, 5, 6, 7\}$ and

$$(20) \qquad \Pi' = \begin{pmatrix} 0.20 & 0.10 & 0.06 & 0.05 & 0.03 & 0.03 \\ 0.15 & 0.12 & 0.07 & 0.08 & 0.06 & 0.05 \end{pmatrix}.$$

In experiments with $J = 4$, $X \in \{2, 3, 4, 5\}$, and Π' is obtained from (20) by

$$P(X = j, W = k \mid j \le J+1) = \frac{P(X = j, W = k)}{\sum_{\ell=2}^{5}[P(X = \ell, W = 0) + P(X = \ell, W = 1)]}.$$

In experiments with $J = 6$, $\boldsymbol{g} = (23, 17, 13, 11, 9, 8)'$. Thus, $g(x)$ is decreasing and convex. We also require $g(1) - g(J) \le 52$. In experiments with $J = 4$, $\boldsymbol{g} = (23, 17, 13, 11)'$. The functionals $L(g)$ are $g(3) - g(2)$, $g(5) - g(2)$, and $g(4)$.

The data are generated by sampling (X, W) from the distribution given by Π' with the specified value of J. Then Y is generated from $Y = g(X) + U$, where $U = XZ^2 - E(X \mid W)$ and $Z \sim N(0,1)$. There are 1000 Monte Carlo replications per experiment. The sample sizes are $n = 1000$ and $n = 5000$. We show the results of experiments using bootstrap procedure 1 with $c_n = 1$ and bootstrap procedure 2, which corresponds to $c_n = 0$. The results of experiments using bootstrap procedure 1 with larger values of c_n were similar to those with $c_n = 1$.

The results of the experiments are shown in Tables 1 and 2, which give empirical coverage probabilities of nominal 95% confidence intervals for $[L_{\min}, L_{\max}]$. The empirical coverage probabilities of nominal 95% confidence intervals for $L(g)$ are similar and are not shown. The empirical coverage probabilities are close to the nominal ones except when $J = 4$ and $L(g) = g(4)$. In this case, the variance of $\hat{\Pi}$ is large, which produces a large error in the asymptotic linear approximation to $\vec{c}_k' \hat{\hat{A}}_k^{-1} \hat{\boldsymbol{m}}$.

5. AN EMPIRICAL APPLICATION

This section presents an empirical application that illustrates the use of the methods described in Sections 2-4. The application is motivated by Angrist and Evans (1998), who investigated the effects of children on several labor-market outcomes of women.

We use the data and instrument of Angrist and Evans (1998) to estimate the relation between the number of children a woman has and the number of weeks she works in a year. The model is that of (1a)-(1b), where Y is the number of weeks a woman works in a year, X is the number of children the woman has, and W is an instrument for the possibly endogenous explanatory variable X. X can have the values 2, 3, 4, and 5. As in Angrist and Evans (1998), W is a binary random variable, with $W = 1$ if the woman's first two children have the same sex, and $W = 0$ otherwise. We investigate the reductions in hours worked when the number of children increases from 2 to 3 and from 2 to 5. In the first case, $L(g) = g(3) - g(2)$. In the second case, $L(g) = g(5) - g(2)$. The binary instrument W does not point identify $L(g)$ in either case. We estimate $L_{\min}$ and $L_{\max}$ under each of two assumptions about the shape of g. The first assumption is that g is monotone non-increasing. The second is that g is monotone non-increasing and convex. Both are reasonable assumptions about the shape of $g(x)$ in this application.

We also estimate $L(g)$ under the assumption that g is the linear function

$$g(x) = \beta_0 + \beta_1 x,$$

where β_0 and β_1 are constants. The binary instrument W point identifies β_0 and β_1. Therefore, $L(g)$ is also point identified under the assumption of linearity. With data $\{Y_i, X_i, W_i : i = 1,...,n\}$, the instrumental variables estimate of β_1 is

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(W_i - \bar{W})}{\sum_{i=1}^{n}(X_i - \bar{X})(W_i - \bar{W})},$$

where $\bar{Y} = n^{-1}\sum_{i=1}^{n} Y_i$, $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$, and $\bar{W} = n^{-1}\sum_{i=1}^{n} W_i$. The estimate of $L(g)$ is

$$L(\hat{g}) = \hat{\beta}_1 \Delta x,$$

where $\Delta x = 1$ for $L(g) = g(3) - g(2)$, and $\Delta x = 3$ for $L(g) = g(5) - g(2)$.

The data are a subset of those of Angrist and Evans (1998).[9] They are taken from the 1980 Census Public Use Micro Samples (PUMS). Our subset consists of 150,618 white women who are 21-35 years old, have 2-5 children, and whose oldest child is between 8 and 12 years old.

The estimation results are shown in Tables 3 and 4. Table 3 shows the estimated identification intervals $[\hat{L}_{\min}, \hat{L}_{\max}]$ and bootstrap 95% confidence intervals for $[L_{\min}, L_{\max}]$ and $L(g)$ under the two sets of shape assumptions. Table 4 shows point estimates and 95% confidence intervals for $L(g)$ under the assumption that g is linear. It can be seen from Table 3 that the bounds on $L(g)$ are very wide when g is required to be monotonic but is not otherwise restricted. The change in the number of weeks worked per year must be in the interval $[-52, 0]$, so the estimated upper bound of the identification interval $[L_{\min}, L_{\max}]$ is uninformative if $L(g) = g(3) - g(2)$, and the estimated lower bound is uninformative if $L(g) = g(5) - g(2)$. The estimated bounds are much narrower when g is required to be convex as well as monotonic. In particular, the 95% confidence intervals for $[L_{\min}, L_{\max}]$ and $L(g)$ under the assumption that g is monotonic and convex are only slightly wider than the 95% confidence interval for $L(g)$ under the much stronger assumption that g is linear.

6. QUANTILE IV AND EXOGENOUS COVARIATES

6.1 Quantile IV

We now consider a quantile version of model (1a)-(1b). The quantile model is

(21) $$Y = g(X) + U$$

(22) $$P(U \le 0 \mid W = w) = q$$

for all $w \in \text{supp}(W)$, where $0 < q < 1$, X is discretely distributed with mass points $\{x_j : j = 1,...,J\}$ for some integer $J < \infty$, W is discretely distributed with mass points $\{w_k : k = 1,...,K\}$ for some integer $K < \infty$, and $K < J$. Horowitz and Lee (2007) show that model (21)-(22) is equivalent to the non-separable model of Chernozukov, Imbens, and Newey (2007)

(23) $$Y = H(X, V),$$

where H is strictly increasing in its second argument and V is a continuously distributed random variable that is independent of the instrument W.

As before, let $g_j = g(x_j)$. Define $\pi_k = P(W = w_k)$. Then (21)-(22) is equivalent to

(24) $$\sum_{j=1}^{J} P(Y - g_j \le 0, X = x_j, W = w_k) = q\pi_k;\ \ k = 1,...,K.$$

Thus, model (21)-(22) is equivalent to K nonlinear equations in $J > K$ unknowns. As is shown by example in the appendix, (21)-(22) and (24) do not point identify $\boldsymbol{g} = (g_1,..., g_J)'$ or the linear functional $L(\boldsymbol{g}) = \boldsymbol{c'g}$ except, possibly, in special cases.

Under the shape restriction $S\boldsymbol{h} \le 0$, sharp upper and lower bounds on $L(g)$ are given by the optimal values of the objective functions of the nonlinear programming problems

(25) $$\underset{\boldsymbol{h}}{\text{maximize}}\ (\underset{\boldsymbol{h}}{\text{minimize}}) : \boldsymbol{c'h}$$

subject to

(26) $$\sum_{j=1}^{J} P(Y - h_j \le 0, X = x_j, W = w_k) = q\pi_k;\ \ k = 1,...,K$$

(27) $$S\boldsymbol{h} \le 0.$$

Denote the optimal values of the maximization and minimization versions of (25) by $L_{\max}$ and $L_{\min}$, respectively. Then $L(g)$ is contained in the interval $[L_{\min}, L_{\max}]$ and cannot be outside of this interval. However, the feasible region of (25) may be non-convex. Consequently, some points within $[L_{\min}, L_{\max}]$ may correspond to points $\boldsymbol{g}$ that are not in the feasible region. When this happens, $L_{\max}$ and $L_{\min}$ are sharp upper and lower bounds on $L(\boldsymbol{g})$, but $[L_{\min}, L_{\max}]$ is not the identification region for $L(\boldsymbol{g})$. The identification region is the union of disconnected subintervals of $[L_{\min}, L_{\max}]$. The example in Appendix A illustrates this situation.

We now consider inference about $L_{\min}$ and $L_{\max}$ in model (21)-(22). In applications, P and π_k in (26) are unknown. The probabilities π_k are estimated consistently by

$$\hat{\pi}_k = n^{-1}\sum_{i=1}^{n} I(W_i = w_k).$$

The most obvious estimator of P is the empirical probability function

$$\hat{P}(Y - h_j \le 0, X = x_j, W = w_k) = n^{-1}\sum_{i=1}^{n} I(Y_i \le h_j, X_i = x_j, W_i = w_k);\ k = 1,...,K. \tag{28}$$

This estimator is a step function of the h_j's, so there may be no h_j's that satisfy (26) when P and π_k are replaced with $\hat{P}$ and $\hat{\pi}_k$. One way of dealing with this problem is to smooth $\hat{P}$ so that it is continuous on $-\infty < h_j < \infty$. This can be done by using the estimator

$$\tilde{P}(Y - h_j \le 0, X = x_j, W = w_k) = n^{-1}\sum_{i=1}^{n} \bar{K}\left(\frac{h_j - Y_i}{s_n}\right) I(X_i = x_j, W_i = w_k);\ k = 1,...,K,$$

where s_n is a bandwidth parameter and $\bar{K}$ is a cumulative distribution function corresponding to a probability density function K that is bounded, symmetrical around 0, and supported on $[-1,1]$. In other words, $\bar{K}$ is the integral of a kernel function for nonparametric density estimation or mean regression. This approach has the disadvantage of requiring a user-selected tuning parameter, s_n. There is no good empirical way of selecting s_n in applications. We do not pursue this approach further in this paper.

Under assumptions QIV1 –QIV4 that are stated below, the need for a user-selected tuning parameter can be avoided by observing that (25)-(27) is equivalent to the unconstrained optimization problem

$$\underset{\boldsymbol{h}}{\text{minimize}}: \pm \boldsymbol{c}'\boldsymbol{h} + C\sum_{m=1}^{M}(S\boldsymbol{h})_m^{+} + C\sum_{k=1}^{K}\left|P_k(\boldsymbol{h}) - q\pi_k\right|, \tag{29}$$

where C is any sufficiently large finite constant, $(S\boldsymbol{h})_m$ is the m'th component of the vector $S\boldsymbol{h}$, $(S\boldsymbol{h})_m^+ = (S\boldsymbol{h})_m I[(S\boldsymbol{h})_m \geq 0]$, and $P_k(\boldsymbol{h}) = \sum_{j=1}^{J} P(Y - h_j \leq 0, X = x_j, W = w_k)$. "Equivalent" means that the values of $\boldsymbol{h}$ and $\boldsymbol{c}'\boldsymbol{h}$ obtained by solving (29) are the same as those obtained by solving (25)-(27) (Di Pillo and Grippo 1989). Now consider estimating $L_{\min}$ and $L_{\max}$ by the values of $\boldsymbol{c}'\boldsymbol{h}$ obtained by solving the following sample analog of (28):

$$\text{(30)} \qquad \underset{\boldsymbol{h}}{\text{minimize}}: \pm\boldsymbol{c}'\boldsymbol{h} + C\sum_{m=1}^{M}(S\boldsymbol{h})_m^+ + C\sum_{k=1}^{K}\left|\hat{P}_k(\boldsymbol{h}) - q\hat{\pi}_k\right|$$

where

$$\hat{P}_k(\boldsymbol{h}) = \sum_{j=1}^{J}\hat{P}(Y - h_j \leq 0, X = x_j, W = w_k)$$

and $\hat{P}$ is given by (28). Let $\hat{\boldsymbol{h}}$ be an optimal value of $\boldsymbol{h}$ in (30). Let $\hat{L}_{\min}$ ($\hat{L}_{\max}$) be the value of $\boldsymbol{c}'\hat{\boldsymbol{h}}$ obtained when the + (-) sign is used in (30). The values of $\hat{\boldsymbol{h}}$ that optimize (30) are not necessarily unique, but it is shown in Theorem 5 below that $\hat{L}_{\min}$ and $\hat{L}_{\max}$ converge almost surely to $L_{\min}$ and $L_{\max}$ as $n \to \infty$.

We use the following notation in our assumptions for quantile IV. Define

$$Q_{0,\min}(\boldsymbol{h}, C) = \boldsymbol{c}'\boldsymbol{h} + C\sum_{m=1}^{M}(S\boldsymbol{h})_m^+ + C\sum_{k=1}^{K}\left|P_k(\boldsymbol{h}) - q\pi_k\right|.$$

and

$$Q_{0,\max}(\boldsymbol{h}, C) = -\boldsymbol{c}'\boldsymbol{h} + C\sum_{m=1}^{M}(S\boldsymbol{h})_m^+ + C\sum_{k=1}^{K}\left|P_k(\boldsymbol{h}) - q\pi_k\right|.$$

Define $I_0(\boldsymbol{h}) = \{m = 1,...,M : (S\boldsymbol{h})_m = 0\}$. The assumptions are as follows.

<u>QIV1</u>: $\boldsymbol{g} \in \mathcal{D} \subset \mathbb{R}^K$, where $\mathcal{D}$ is a compact set consisting of vectors $\boldsymbol{h}$ satisfying the shape restriction $S\boldsymbol{h} \leq 0$.

<u>QIV2</u>: There are finitely many globally optimal solutions to (25)-(27).

<u>QIV3</u>: $P(Y \leq \xi, X = x_j, W = w_k)$ is a twice continuously differentiable function of ξ for all $j = 1,...,J$ and $k = 1,...,K$.

<u>QIV4</u>: Let $\boldsymbol{h}_0$ be a global minimum of $Q_{0,\min}(\boldsymbol{h}, C)$ or $Q_{0,\max}(\boldsymbol{h}, C)$. There exist no $u_m \geq 0$, $m \in I_0(\boldsymbol{h}_0)$ and v_k, $k = 1,...,K$ such that the u_m's and v_k's are not all 0 and

$$\sum_{m \in I_0(\boldsymbol{h}_0)} u_m \boldsymbol{S}_{mj} + \sum_{k=1}^{K} v_k \frac{\partial P_k(\boldsymbol{h}_0)}{\partial \boldsymbol{h}_j} = 0; \;\; j = 1,...,J \; .$$

This assumption is satisfied if the constraints that bind at $\boldsymbol{h}_0$ are not redundant in an arbitrarily small neighborhood of $\boldsymbol{h}_0$.

Now let $\boldsymbol{h}_{0,\min}$ and $\boldsymbol{h}_{0,\max}$, respectively, be optimal solutions (not necessarily unique) to the minimization and maximization versions of (25)-(27). Given any $\delta > 0$ define the sets $\mathcal{D}_{\delta,\min} = \{\boldsymbol{h}: \mathcal{D}, \|\boldsymbol{h} - \boldsymbol{h}_{0,\min}\| \le \delta\}$ and $\mathcal{D}_{\delta,\max} = \{\boldsymbol{h}: \mathcal{D}, \|\boldsymbol{h} - \boldsymbol{h}_{0,\max}\| \le \delta\}$. For $k = 1,...,K)$, define $\lambda_{k,\min}$ ($\lambda_{k,\max}$) to be the Lagrangian multipliers for the optimization problems

(31) $$\underset{\boldsymbol{h} \in \mathcal{D}_{\delta,\min}}{\text{minimize}} \; (\underset{\boldsymbol{h} \in \mathcal{D}_{\delta,\max}}{\text{maximize}}): \; \boldsymbol{c}'\boldsymbol{h}$$

subject to

$$P_k(\boldsymbol{h}) - q\pi_k = 0; \;\; k = 1,...,K \; .$$

We now have the following theorem, which is proved in Appendix B.

Theorem 5: Let assumptions QIV1-QIV4 hold. Let $\hat{\boldsymbol{h}}_{\min}$ ($\hat{\boldsymbol{h}}_{\max}$) optimize (30) with the + (-) sign in the objective function. Let $\hat{L}_{\min} = \boldsymbol{c}'\hat{\boldsymbol{h}}_{\min}$ and $\hat{L}_{\max} = \boldsymbol{c}'\hat{\boldsymbol{h}}_{\max}$. Let C be defined as in (30). There is a $C_0 < \infty$ that does not depend on n such that for all $C > C_0$, the following hold as $n \to \infty$:

a. $\hat{L}_{\min} \to L_{\min}$ and $\hat{L}_{\max} \to L_{\max}$ almost surely.

b. If the optimal solutions to the minimization and maximization versions of (25)-(27) are unique, then $n^{1/2}(\hat{L}_{\min} - L_{\min}) \to^d N(0, \sigma^2_{\min})$ and $n^{1/2}(\hat{L}_{\max} - L_{\max}) \to^d N(0, \sigma^2_{\max})$, where

$$\sigma^2_{\min(\max)} = E\left\{\sum_{k=1}^{K} \lambda_{k,\min(\max)} n^{1/2}[q\hat{\pi}_k - \hat{P}_k(\boldsymbol{h}_{0,\min(\max)})]\right\}^2$$

and $\boldsymbol{h}_{0,\min(\max)}$ are the optimal values of $\boldsymbol{h}$ in the minimization and maximization versions of (25)-(27). Moreover, $[n^{1/2}(\hat{L}_{\min} - L_{\min}), n^{1/2}(\hat{L}_{\max} - L_{\max})] \to^d N(0, V)$, where V is the 2×2 covariance matrix of

$$\sum_{k=1}^{K} \lambda_{k,\min(\max)} n^{1/2}[q\hat{\pi}_k - \hat{P}_k(\boldsymbol{h}_{0,\min(\max)})].$$

c. If there are $\ell_{\min}$ ($\ell_{\max}$) optimal solutions to the minimization (maximization) versions of (25)-(27), then $n^{1/2}(\hat{L}_{\min} - L_{\min})$ (($n^{1/2}(\hat{L}_{\max} - L_{\max})$) is asymptotically distributed as the minimum (maximum) of $\ell_{\min}$ ($\ell_{\max}$) multivariate normally distributed random variables with means of 0. ■

Bootstrap methods like those described in Section 4.1 can be used to consistently estimate the distributions of $n^{1/2}(\hat{L}_{\min} - L_{\min})$ $(n^{1/2}(\hat{L}_{\max} - L_{\max})$, and $[n^{1/2}(\hat{L}_{\min} - L_{\min}), n^{1/2}(\hat{L}_{\max} - L_{\max})]$. It is not necessary to calculate the asymptotic distributions analytically. As with model (1)-(2), the bootstrap estimates of the asymptotic distributions of these quantities can be used to obtain confidence intervals for $L_{\min}$, $L_{\max}$, $[L_{\min}, L_{\max}]$ and $L(g)$. The methods of Section 3.3 can be used to obtain uniformly correct critical values and confidence intervals for $L_{\min}$, $L_{\max}$, $[L_{\min}, L_{\max}]$ and $L(g)$.

6.2 Models with Exogenous Covariates

Coming soon.

7. CONCLUSIONS

This paper has been concerned with nonparametric estimation of the linear functional $L(g)$, where the unknown function g satisfies the moment condition $E[Y - g(X) | W] = 0$, Y is a dependent variable, X is an explanatory variable that may be endogenous, and W is an instrument for X. In many applications, X and W are discretely distributed, and W has fewer points of support than X does. In such settings, $L(g)$ is not identified and, in the absence of further restrictions, can take any value in $(-\infty, \infty)$. This paper has explored the use of restrictions on the shape of g, such as monotonicity and convexity, for achieving interval identification of $L(g)$. The paper has presented a sharp identification interval for $L(g)$, explained how the lower and upper bounds of this interval can be estimated consistently, and shown how the bootstrap can be used to obtain confidence regions for the identification interval and $L(g)$. The results of Monte Carlo experiments and an empirical application have illustrated the usefulness of this paper's methods.

This paper has concentrated on a model in which there is an endogenous explanatory variable and no exogenous covariates. The methods of this paper can accommodate discretely distributed exogenous covariates with essentially no change by conditioning on them. The extension to a model with a continuously distributed endogenous explanatory variable and instrument is also possible, though more challenging technically. Nonparametric identification in such a model is always problematic because any distribution of (Y, X, W) that identifies g is arbitrarily close to a distribution that does not identify g (Santos 2012), and the necessary condition for identification cannot be tested (Canay, Santos, and Shaikh 2012). The usefulness of shape restrictions for achieving partial identification of $L(g)$ and carrying out inference about $L(g)$ when point identification is uncertain will be explored in future research.

APPENDIX A: Example of Quantile IV Estimation

The following example shows that (21)-(22) does not necessarily point identify $L(g)$ and that $[L_{\min}, L_{\max}]$ may not be the identification region for $L(g)$ under the shape restriction $S\boldsymbol{g} \le 0$ is equivalent to $h_0 \le h_1$

<u>Example</u>: In model (21)-(22), let $X = W\varepsilon$, where, $\text{supp}(W) = \{0,1\}$, $P(W = 0) = P(W = 1) = 0.5$, $\text{supp}(\varepsilon) = \{-1, 0, 1\}$, and $P(\varepsilon = -1) = P(\varepsilon = 0) = P(\varepsilon = 1) = 1/3$. Then $J = 3$ and $K = 2$. Let $U = W\nu_1 + (1 - W)\nu_2$, where ν_1 and ν_2 have medians of 0 and are independent of ε and W. Then U is independent of X conditional on W. Let $q = 0.5$, $g_1 = g(-1) = 0$, $g_2 = g(0) = 1.5$, and $g_3 = g(1) = 2$. Let the shape restriction be that g is non-decreasing and concave. Then constraint $S\boldsymbol{h} \le 0$ is equivalent to $h_1 \le h_2$, $h_2 \le h_3$, and $h_2 - h_1 \ge h_3 - h_2$. Let $L(\boldsymbol{g}) = g_3 - g_2$. Finally, assume that $\boldsymbol{g}$ is in the compact set $\{\boldsymbol{g} : |g_j| \le 6; j = 1, 2, 3\}$.

Under these conditions, (24) is equivalent to

(A1) $$\sum_{j=1}^{J} P(U \le h_j - g_j \mid W = w_k) P(X = x_j \mid W = w_k) = 0.5; \; k = 1, ..., K$$

But $P(X = 0 \mid W = 0) = 1$,and $P(X = -1 \mid W = 1) = P(X = 0 \mid W = 1) = P(X = 1 \mid W = 1) = 1/3$. Therefore, (A1) is equivalent to

(A2) $$P(U \le h_2 - g_2 \mid W = 0) = 0.5$$

and

(A3) $$P(U \le h_1 \mid W = 1) + P(U \le h_2 - 1.5 \mid W = 1) + P(U \le h_3 - 2 \mid W = 1) = 1.5.$$

Assume that $P(U \le t \mid W = 0)$ is a strictly increasing function of t on $\text{supp}(U \mid W = 0)$. Then (A2) implies that $h_2 = g_2$, so g_2 is identified. Equation (A3) becomes

(A4) $$P(U \le h_1 \mid W = 1) + P(U \le h_3 - 2 \mid W = 1) = 1,$$

and the shape restriction becomes

(A5) $$h_1 \le 1.5, \; h_3 \ge 1.5, \text{ and } h_1 + h_3 \le 3.$$

Now let

$$P(U \le t \mid W = 1) = \begin{cases} 0 \text{ if } t < -1 \\ 0.5(1+t) \text{ if } -1 \le t \le 0 \\ 0.5(1 + t/4) \text{ if } 0 \le t \le 4 \\ 1 \text{ if } t > 4 \end{cases}$$

Problem (25)-(27) now becomes

(A6) $\underset{\boldsymbol{h}}{\text{maximize}}\ (\underset{\boldsymbol{h}}{\text{minimize}}):\ h_3 - h_2$

subject to

(A7) $P(U \le h_1 | W = 1) + P(U \le h_3 - 2 | W = 1) = 1,$

(A8) $h_1 \le 1.5$,

(A9) $h_3 \ge 1.5$,

(A10) $h_1 + h_3 \le 3$,

(A11) $h_1 \ge -6$,

and

(A12) $h_3 \le 6$.

Figure A1 shows the feasible region of problem (A6)-(A12). Solving (A6)-(A12) gives $[L_{\min}, L_{\max}] = [0.17, 4.5]$. However, as Figure A1 illustrates, there are no feasible values of $\boldsymbol{h}$ for which $1.83 < L(\boldsymbol{h}) < 4.50$. Thus, $L(g)$ is not point identified and the identification region for $L(g)$ is non-convex and, therefore, not an interval.

APPENDIX B: Proof of Theorem 5

This appendix gives the proof of Theorem 5 for the minimization versions of (26)-(27) and (30). The proof for the maximization versions is similar. Assumptions QIV1-QIV4 hold throughout this appendix.

Let $\{\hat{\boldsymbol{h}}_n : n = 1, 2, ...\}$ be a sequence of solutions to the minimization version of (30) (+ sign in the objective function). By assumption QIV1, 0 $\hat{\boldsymbol{h}}_n \in \mathcal{D}$. Therefore. $\{\hat{\boldsymbol{h}}_n\}$ has a convergent subsequence. Let $\boldsymbol{h}_0$ denote the limit of the convergent subsequence $\{\hat{\boldsymbol{h}}_{n_i} : i = 1, 2, ...\}$.

Lemma 1: Let $C \ge C_0$. Then $\boldsymbol{h}_0$ is a global optimum of (25)-(27).

Proof: Define

$$Q_n(\boldsymbol{h}, C) = \boldsymbol{c}'\boldsymbol{h} + C\sum_{m=1}^{M}(S\boldsymbol{h})_m^+ + C\sum_{k=1}^{K}\left|\hat{P}_k(\boldsymbol{h}) - q\hat{\pi}_k\right| \quad .$$

Then

$$\sup_{\boldsymbol{h}\in\mathcal{D}} |Q_n(\boldsymbol{h}, C) - Q_{0,\min}(\boldsymbol{h}, C)| \to^{a.s.} 0$$

for any C. Therefore, for any $\boldsymbol{h} \in \mathcal{D}$,

$$Q_{n_i}(\boldsymbol{h},C) \geq Q_{n_i}(\hat{\boldsymbol{h}}_{n_i},C) =^{a.s.} Q_{0,\min}(\hat{\boldsymbol{h}}_{n_i},C) + o(1) .$$

But $\boldsymbol{h}_{n_i} \to \boldsymbol{h}_0$, $Q_{0,\min}(\boldsymbol{h}_{n_i},C) \to Q_{0,\min}(\boldsymbol{h}_0,C)$, and $Q_{n_i}(\boldsymbol{h},C) \to^{a.s.} Q_{0,\min}(\boldsymbol{h},C)$. Therefore,

$$Q_{0,\min}(\boldsymbol{h}_0,C) \leq Q_{0,\min}(\boldsymbol{h},C)$$

for any $\boldsymbol{h} \in \mathcal{D}$, and $\boldsymbol{h}_0$ is a global optimum of $Q_{0,\min}$. It follows from Theorem 4 of Di Pillo and Lippi (1989) that $\boldsymbol{h}_0$ is also a global optimum of (25)-(27). Q.E.D.

Lemma 1 implies that if n is sufficiently large, then $\hat{\boldsymbol{h}}_n$ is almost surely in a neighborhood of some global optimizer of $Q_0(\boldsymbol{h},C)$ and optimal solution of (25)-(27).

Lemma 2: Let $C \geq C_0$, and let $\boldsymbol{h}_0$ be the limit of the subsequence $\{\hat{\boldsymbol{h}}_{n_i} : i = 1,2,...\}$. Then as $i \to \infty$

a. $\boldsymbol{c}'\hat{\boldsymbol{h}}_{n_i} \equiv \hat{L}_{\min,n_i} \to L_{\min}$.

b. $n_i^{1/2}(\hat{L}_{\min,n_i} - L_{\min}) \to^d N(0,\sigma^2_{\min})$, where

$$\sigma^2_{\min} = E\left\{\sum_{k=1}^{K} \lambda_{k,\min} n^{1/2}[q\hat{\pi}_k - \hat{P}_k(\boldsymbol{h}_0)]\right\}^2 .$$

Proof: Part a follows from $\hat{\boldsymbol{h}}_{n_i} \to \boldsymbol{h}_0$ is the limit of $\boldsymbol{h}_{n_i}$. We now prove part b. We have

$$Q_n(\boldsymbol{h},C) = \boldsymbol{c}'\boldsymbol{h} + C\sum_{m=1}^{M}(S\boldsymbol{h})^+_m + C\sum_{k=1}^{K}\left|\hat{P}_k(\boldsymbol{h}) - q\hat{\pi}_k\right|$$

Moreover, $\sum_{m=1}^{M}(S\boldsymbol{h})^+_m = 0$, because $\boldsymbol{h} \in \mathcal{D}$. Therefore, for $\boldsymbol{h} \in \mathcal{D}$

$$Q_n(\boldsymbol{h},C) = \boldsymbol{c}'\boldsymbol{h} + C\sum_{k=1}^{K}|\,\hat{P}_k(\boldsymbol{h}) - q\hat{\pi}_k\,|$$

$$= \boldsymbol{c}'\boldsymbol{h} + C\sum_{k=1}^{K}|[P_k(\boldsymbol{h}) - q\hat{\pi}_k] + [\hat{P}_k(\boldsymbol{h}_0) - P_k(\boldsymbol{h}_0)] ,$$

$$+\{[\hat{P}_k(\boldsymbol{h}) - P_k(\boldsymbol{h})] - [\hat{P}_k(\boldsymbol{h}_0) - P_k(\boldsymbol{h}_0)]|\,.$$

Because $\hat{\boldsymbol{h}}_{n_i} \to \boldsymbol{h}_0$, it suffices to consider only minimization of $Q_n(\boldsymbol{h},C)$ over $\boldsymbol{h} \in \mathcal{D}_\delta$ for any $\delta > 0$. Let $\boldsymbol{h} \in \mathcal{D}_\delta$. It follows from stochastic equicontinuity of the empirical process $n^{-1/2}(\hat{P}_k - P_k)$ that for any sequence $\{\varepsilon_{n_i}\}$ with $\varepsilon_{n_i} > 0$ and $\varepsilon_{n_i} \to 0$, there is a sequence $\{\delta(\varepsilon_{n_i})\}$ with $\delta(\varepsilon_{n_i}) > 0$ such that with probability arbitrarily close to 1 as $i \to \infty$

$$\sup_{\boldsymbol{h} \in \mathcal{D}_{\delta(\varepsilon_{n_i})}} |[\hat{P}_k(\boldsymbol{h}) - P_k(\boldsymbol{h})] - [\hat{P}_k(\boldsymbol{h}_0) - P_k(\boldsymbol{h}_0)]| < \varepsilon_{n_i} O_p(n_i^{-1/2}) .$$

Because $\varepsilon_{n_i} \to 0$, this is equivalent to

$$\sup_{\boldsymbol{h} \in \mathcal{D}_{\delta(\varepsilon_{n_i})}} |[\hat{P}_k(\boldsymbol{h}) - P_k(\boldsymbol{h})] - [\hat{P}_k(\boldsymbol{h}_0) - P_k(\boldsymbol{h}_0)]| = o_p(n_i^{-1/2}) .$$

Now use the triangle inequality, $|a| - |b| \le |a + b| \le |a| + |b|$, with

$$a = [P_k(\boldsymbol{h}) - q\hat{\pi}_k] + [\hat{P}_k(\boldsymbol{h}_0) - P_k(\boldsymbol{h}_0)]$$

and

$$b = [\hat{P}_k(\boldsymbol{h}) - P_k(\boldsymbol{h})] - [\hat{P}_k(\boldsymbol{h}_0) - P_k(\boldsymbol{h}_0)]$$

to obtain

$$|[P_k(\boldsymbol{h}) - q\hat{\pi}_k] + [\hat{P}_k(\boldsymbol{h}_0) - P_k(\boldsymbol{h}_0)] + \{[\hat{P}_k(\boldsymbol{h}) - P_k(\boldsymbol{h})] - [\hat{P}_k(\boldsymbol{h}_0) - P_k(\boldsymbol{h}_0)]\}|$$

$$= |[P_k(\boldsymbol{h}) - q\hat{\pi}_k] + [\hat{P}_k(\boldsymbol{h}_0) - P_k(\boldsymbol{h}_0)]| + o_p(n_i^{-1/2}).$$

It follows that

$$Q_n(\boldsymbol{h}, C) = \boldsymbol{c}'\boldsymbol{h} + C\sum_{k=1}^{K} |[P_k(\boldsymbol{h}) - q\hat{\pi}_k] + [\hat{P}_k(\boldsymbol{h}_0) - P_k(\boldsymbol{h}_0)]| + o_p(n_i^{-1/2}) .$$

But $P_k(\boldsymbol{h}_0) = q\pi_k$, so for $\boldsymbol{h} \in \mathcal{D}_\delta$

$$Q_n(\boldsymbol{h}, C) = \boldsymbol{c}'\boldsymbol{h} + C\sum_{k=1}^{K} |[P_k(\boldsymbol{h}) - q\pi_k] + [\hat{P}_k(\boldsymbol{h}_0) - q\hat{\pi}_k]| + o_p(n_i^{-1/2}) .$$

Now define

$$\tilde{Q}_n(\boldsymbol{h}, C) = \boldsymbol{c}'\boldsymbol{h} + C\sum_{k=1}^{K} |[P_k(\boldsymbol{h}) - q\pi_k] + [\hat{P}_k(\boldsymbol{h}_0) - q\hat{\pi}_k]| .$$

and

$$\tilde{\boldsymbol{h}}_{n_i} = \arg\min_{\boldsymbol{h} \in \mathcal{D}_\delta} \tilde{Q}(\boldsymbol{h}, C) .$$

It follows from Theorem 4 of Pillo and Grippo (1989) that $\tilde{L}_{\min,n_i} \equiv \boldsymbol{c}'\tilde{\boldsymbol{h}}_{n_i}$ is the optimal value of the objective function of

$$\underset{\boldsymbol{h} \in \mathcal{D}_\delta}{\text{minimize}}: \ \boldsymbol{c}'\boldsymbol{h}$$

subject to

(A13) $P_k(\boldsymbol{h}) - q\pi_k + \hat{P}_k(\boldsymbol{h}_0) - q\hat{\pi}_k = 0;\ k = 1,...,K$

or, equivalently,

$$\underset{\boldsymbol{h}\in\mathcal{D}_\delta}{\text{minimize}}: \boldsymbol{c}'\boldsymbol{h}$$

subject to

$$P_k(\boldsymbol{h}) - q\pi_k = q\hat{\pi}_k - \hat{P}_k(\boldsymbol{h}_0);\ k = 1,...,K .$$

ut $\boldsymbol{h}_0$ is the optimal solution to

(A14) $\underset{\boldsymbol{h}\in\mathcal{D}_\delta}{\text{minimize}}: \boldsymbol{c}'\boldsymbol{h}$

subject to

$$P_k(\boldsymbol{h}) - q\pi_k = 0;\ k = 1,...,K .$$

It follows from Theorem 3.4.1 of (Fiacco 1983) that

$$\tilde{L}_{\min,n_i} - L_{\min} = \sum_{k=1}^{K} \lambda_k [q\hat{\pi}_k - \hat{P}_k(\boldsymbol{h}_0)] + o[q\hat{\pi}_k - \hat{P}_k(\boldsymbol{h}_0)] = \sum_{k=1}^{K} \lambda_k [q\hat{\pi}_k - \hat{P}_k(\boldsymbol{h}_0)] + o_p(n^{-1/2}) .$$

In addition, it follows from $|Q_n(\boldsymbol{h},C) - \tilde{Q}_n(\boldsymbol{h},C)| = o_p(n_i^{-1/2})$ that

$$|\hat{L}_{\min,n_i} - \tilde{L}_{\min,n_i}| = o_p(n_i^{-1/2}) .$$

Therefore,

$$\hat{L}_{\min,n_i} - L_{\min} = \sum_{k=1}^{K} \lambda_k [q\hat{\pi}_k - \hat{P}_k(\boldsymbol{h}_0)] + o_p(n_i^{-1/2}) .$$

It follows from the Lindeberg-Levy central limit theorem that

$$\sum_{k=1}^{K} \lambda_k n^{1/2} [q\hat{\pi}_k - \hat{P}_k(\boldsymbol{h}_0)] \to^d N(0, \sigma_{\min}^2) .$$

Therefore, $n_i^{1/2}(\hat{L}_{\min,n_i} - L_{\min}) \to^d N(0, \sigma_{\min}^2)$. Q.E.D.

Proof of Theorem 5: Lemma 1 implies that every subsequential limit of $\hat{\boldsymbol{h}}_n$ is a global minimum of (25)-(27), thereby establishing part a of the theorem. If the global minimum of (25)-(27) is unique, then it follows from lemma 2 that $n^{1/2}(\hat{L}_{\min} - L_{\min}) \to^d N(0, \sigma_{\min}^2)$. This establishes part b. If the global minimum is not unique, then it follows from lemma 2 that $n^{1/2}(\hat{L}_{\min} - L_{\min})$ is distributed as the minimum of mean-zero normals. This established part c. Q.E.D.

Table 1: Results of Monte Carlo Experiments Assuming Only Monotonicity

$L(g)$	c_n	J	Empirical Coverage Probability with $n = 1000$	Empirical Coverage Probability with $n = 5000$
$g(3) - g(2)$	0	4	0.962	0.963
	1		0.962	0.963
$g(5) - g(2)$	0	4	0.941	0.938
	1		0.941	0.938
$g(4)$	0	4	0.882	0.895
	1		0.882	0.895
$g(3) - g(2)$	0	6	0.935	0.944
	1		0.935	0.944
$g(5) - g(2)$	0	6	0.965	0.970
	1		0.961	0.969
$g(4)$	0	6	0.936	0.923
	1		0.926	0.914

Table 2: Results of Monte Carlo Experiments Assuming Monotonicity and Convexity

$L(g)$	c_n	J	Empirical Coverage Probability with $n = 1000$	Empirical Coverage Probability with $n = 5000$
$g(3) - g(2)$	0	4	0.950	0.963
	1		0.950	0.963
$g(5) - g(2)$	0	4	0.941	0.938
	1		0.941	0.938
$g(4)$	0	4	0.962	0.970
	1		0.962	0.969
$g(3) - g(2)$	0	6	0.944	0.951
	1		0.944	0.951
$g(5) - g(2)$	0	6	0.965	0.970
	1		0.961	0.969
$g(4)$	0	6	0.958	0.950
	1		0.958	0.951

Table 3: Estimates of $[L_{\min}, L_{\max}]$ and $L(g)$ under Two Sets of Shape Restrictions

Shape Restriction	$L(g)$	$[\hat{L}_{\min} \hat{L}_{\max}]$	95% Conf. Int. for $[L_{\min}, L_{\max}]$	95% Conf. Int. for $L(g)$
g is monotone non-increasing	$g(3) - g(2)$	$[-6.0, 0]$	$[-8.6, 0]$	$[-8.6, 0]$
	$g(5) - g(2)$	$[-52.0, -6.0]$	$[-52.0, -3.4]$	$[-52.0, -3.4]$
g is monotone non-increasing and convex	$g(3) - g(2)$	$[-6.0, -5.0]$	$[-9.0, -2.3]$	$[-8.6, -2.8]$
	$g(5) - g(2)$	$[-14.9, -6.0]$	$[-22.0, -2.4]$	$[-21.2, -3.4]$

Table 4: Estimates of $L(g)$ under the Assumption that g Is Linear

$L(g)$	$L(\hat{g})$	95% Confidence Interval for $L(g)$
$g(3) - g(2)$	-5.0	$[-7.6, -2.4]$
$g(5) - g(2)$	-14.9	$[-22.7, -7.1]$

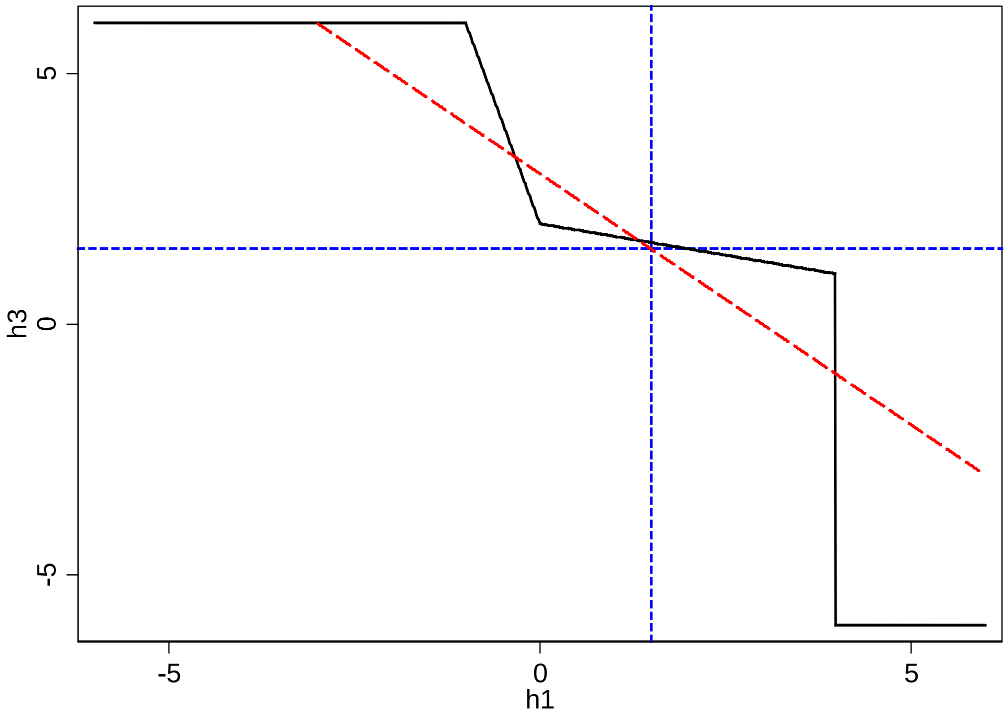

Figure A1

Feasible region for problem (A6)-(A12): Points on the solid line satisfy constraint (A7) and (A12). Points on or below the dashed line satisfy constraint (A10). Points on or to the left of the vertical dotted line and on or above the horizontal dotted line satisfy constraints (A8) and (A9). All displayed points satisfy constraint (A11). The feasible region consists of points on the solid line that are on or below the dashed line, on or to the left of the vertical dotted line, and on or above the horizontal dotted line.

REFERENCES

Angrist, J.D. and A.B. Krueger (1991). Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics*, 106, 979-1014.

Angrist, J.D. and W.N. Evans (1998). Children and their parents' labor supply: evidence from exogenous variation in family size, *American Economic Review*, 88, 450-477.

Bronars, S.G. and J. Grogger (1994). The economic consequences of unwed motherhood: using twins as a natural experiment. *American Economic Review*, 84, 1141-1156.

Canay, I.A., A. Santos, and A. Shaikh (2012). On the testability of identification in some nonparametric models with endogeneity, working paper, Department of Economics, Northwestern University, Evanston, IL, U.S.A.

Card, D. (1995). Using geographic variation in college proximity to estimate returns to schooling. In L.N. Christofides, E.K. Grant, and R. Swidinsky, editors, *Aspects of Labor Market Behaviour: Essays in Honour of John Vanderkamp*, Toronto: University of Toronto Press.

Chernozhukov, V., G.W. Imbens, and W.K. Newey (2007). Instrumental variable identification and estimation of nonseparable models via quantile conditions, *Journal of Econometrics*, 139, 4-14.

Chernozhukov, V., S. Lee, and A.M. Rosen (2009). Intersection bounds: estimation and inference, cemmap working paper CWP19/09, Centre for Microdata Methods and Practice, London.

Chesher, A. (2004). Identification in additive error models with discrete endogenous variables, working paper CWP11/04, Centre for Microdata Methods and Practice, Department of Economics, University College London.

Di Pillo, G. and L. Grippo (1989). Exact penalty functions in constrained optimization, *SIAM Journal on Control and Optimization*, 27, 1333-1360.

Fiacco, A.V. (1983). *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*, New York: Academic Press.

Hadley, G. (1962). *Linear Programming*. Reading, MA: Addison-Wesley Publishing Company.

Horowitz, J.L. and S. Lee (2007). Nonparametric instrumental variables estimation of a quantile regression model, *Econometrica*, 75, 1191-1208.

Imbens, G.W. and C.F. Manski (2004). Confidence intervals for partially identified parameters, *Econometrica*, 72, 1845-1857.

Lehmann, E.L. and J.P. Romano (2005). *Testing Statistical Hypotheses*, 3^{rd} edition. New York: Springer.

Lochner, L. and E. Moretti (2004). The effect of education on crime: evidence from prison inmates, arrests, and self reports, *American Economic Review*, 94, 155-189.

Mammen, E. (1992). *When Does Bootstrap Work? Asymptotic Results and Simulations*, New York: Springer.

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

Manski, C.F. and J.V. Pepper (2009). More on monotone instrumental variables, *Econometrics Journal*, 12, S200-S216.

Milgrom P. and I. Segal (2002). Envelope theorems for arbitrary choice sets, *Econometrica*, 70, 583-601.

Moran, J.R. and K.I. Simon (2006). Income and the use of prescription drugs by the elderly: evidence from the notch cohorts, *Journal of Human Resources*, 41, 411-432.

Romano, J.P., A.M. Shaikh, and M. Wolf (2010). Hypothesis testing in econometrics, *Annual Review of Economics*, 2, 75-104.

Santos, A. (2012). Inference in nonparametric instrumental variables with partial identification, *Econometrica*, 80, 213-275.

Stoye, J. (2009). More on confidence intervals for partially identified parameters, *Econometrica*, 77, 1299-1315.

FOOTNOTES

[1] $L_{\min}$ and $L_{\max}$, respectively, are finite if the optimal values of the objective functions of the minimization and maximization versions of (6) are finite. There are no simple conditions under which this occurs. See, for example, Hadley (1962, Sec. 3-7).

[2] In some applications, $g(x_j)$ for each $j = 1,..., J$ is contained in a finite interval by definition, so unbounded solutions to (6) cannot occur. For example, in the empirical application presented in Section 5 of this paper, $g(x_j)$ is the number of weeks a woman with x_j children works in a year and, therefore, is contained in the interval $[0, 52]$. Such restrictions can be incorporated into the framework presented here by adding constraints to (6) that require $g(x_j)$ to be in the specified interval for each $j = 1,..., J$.

[3] The feasible region of problem (6) with Π and $\boldsymbol{m}$ replaced by consistent estimators may be empty if n is small. This problem can be overcome by expanding the feasible region by an amount that is large enough to make its interior non-empty if n is small and zero if n is large.

[4] Imbens and Manski (2004) show that a confidence interval for consisting of the intersection of one-sided intervals for a partially identified parameter is not valid uniformly over a set of values of the lower and upper identification bounds that includes equality of the two ($L_{\min} = L_{\max}$ in the context of this paper). However, the possibility that $L_{\min} = L_{\max}$ is excluded by our assumption 3.

[5] We assume that $2K + M \geq J$ as happens, for example, if g is assumed to be monotone, convex, or both.

[6] In general, the bootstrap does not provide a consistent estimator of the distribution of the square of a random variable. The estimator here is consistent here because the random variable of interest and its bootstrap analog are both estimators of zero under H_0 and are centered accordingly.

[7] The bootstrap does not consistently estimate the distribution of the maximum of random variables with unknown means. The bootstrap is consistent in the case treated here because Δ_{1k}^*, Δ_{2k}^*, and the asymptotic form of $n^{1/2}(\bar{\boldsymbol{c}}_k' \hat{\bar{A}}_k^{-1} \hat{\bar{\boldsymbol{m}}} - \bar{\boldsymbol{c}}_k' \bar{A}_k^{-1} \bar{\boldsymbol{m}})$ all have means of zero.

[8] When n is small, the optimal solution in the bootstrap sample may be infeasible in the original sample. Such solutions can be excluded from $\hat{\mathcal{B}}_{\max}$ and $\hat{\mathcal{B}}_{\min}$ without affecting the asymptotic distributional results presented here.

[9] Angrist's and Evans's (1998) data are available at http://economics.mit.edu/faculty/angrist/data1/data/angev98.