

Freyberger, Joachim; Horowitz, Joel

Working Paper

Identification and shape restrictions in nonparametric instrumental variables estimation

cemmap working paper, No. CWP15/12

Provided in Cooperation with:

The Institute for Fiscal Studies (IFS), London

Suggested Citation: Freyberger, Joachim; Horowitz, Joel (2012) : Identification and shape restrictions in nonparametric instrumental variables estimation, cemmap working paper, No. CWP15/12, Centre for Microdata Methods and Practice (cemmap), London, <https://doi.org/10.1920/wp.cem.2012.1512>

This Version is available at:

<https://hdl.handle.net/10419/64720>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Identification and shape restrictions in nonparametric instrumental variables estimation

Joachim Freyberger
Joel Horowitz

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP15/12

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

June 2012

Abstract

This paper is concerned with inference about an unidentified linear functional, $L(g)$, where the function g satisfies the relation $Y = g(X) + U$; $E(U | W) = 0$. 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 $L(g)$.

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

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

and

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

for almost every w . Equivalently,

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

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 $\{Y_i, X_i, W_i : i = 1, \dots, n\}$ 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.

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 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 $\{x_j : j = 1, \dots, J\}$ and $\{w_k : k = 1, \dots, K\}$. In this paper, it is assumed that $K < J$. Order the support points so that $x_1 < x_2 < \dots < x_J$ and $w_1 < w_2 < \dots < w_K$. 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 (1b) is equivalent to

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

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

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

Note that $\text{rank}(\Pi) < J$, because $K < J$. Therefore, (4) does not point identify g . Write the linear functional $L(g)$ as $L(g) = \mathbf{c}'\mathbf{g}$, where $\mathbf{c} = (c_1, \dots, c_J)'$ 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 \mathbf{c} is not orthogonal to 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 \mathbf{g}_1 be a vector in the space spanned by the rows of Π' that satisfies $\Pi'\mathbf{g}_1 = \mathbf{m}$. Let \mathbf{g}_2 be a vector in the orthogonal complement of the row space of Π' such that $\mathbf{c}'\mathbf{g}_2 \neq 0$. For any real γ , $\Pi'(\mathbf{g}_1 + \gamma\mathbf{g}_2) = \mathbf{m}$ and $L(\mathbf{g}_1 + \gamma\mathbf{g}_2) = \mathbf{c}'\mathbf{g}_1 + \gamma\mathbf{c}'\mathbf{g}_2$. Then $L(\mathbf{g}_1 + \gamma\mathbf{g}_2)$ is consistent with (1a)-(1b), and by choosing γ appropriately, $L(\mathbf{g}_1 + \gamma\mathbf{g}_2)$ can be made to have any value in $(-\infty, \infty)$. ■

We now impose the linear shape restriction

$$(5) \quad S\mathbf{g} \leq 0,$$

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

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

We assume that g satisfies the shape restriction.

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

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

$$(6) \quad \underset{\mathbf{h}}{\text{maximize}} \quad (\underset{\mathbf{h}}{\text{minimize}}): \quad \mathbf{c}'\mathbf{h}$$

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

$$S\mathbf{h} \leq 0.$$

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 \mathbf{g}_{\max} and \mathbf{g}_{\min} be feasible solutions of (6) such that $\mathbf{c}'\mathbf{g}_{\max} = L_{\max}$ and $\mathbf{c}'\mathbf{g}_{\min} = L_{\min}$. Then $d = \mathbf{c}'[(1 - \lambda)\mathbf{g}_{\max} + \lambda\mathbf{g}_{\min}]$. The feasible region of a linear programming problem is convex, so $(1 - \lambda)\mathbf{g}_{\max} + \lambda\mathbf{g}_{\min}$ 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 3.3. 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 3.3, 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 if \hat{L}_{\max} and \hat{L}_{\min} are finite.²

We also assume:

Assumption 3: There is a vector \mathbf{h} satisfying $\Pi'\mathbf{h} - \mathbf{m} = 0$ and $S\mathbf{h} < 0$ (strict inequality).

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

Assumption 3 is not testable. To see why; let $(S\mathbf{h})_k$ denote the k 'th component of $S\mathbf{h}$. Regardless of the sample size, no test can discriminate with high probability between $(S\mathbf{h})_k = 0$ and $(S\mathbf{h})_k = -\varepsilon$ for a sufficiently small $\varepsilon > 0$. However, the hypothesis that there is a vector \mathbf{h} satisfying $\Pi'\mathbf{h} - \mathbf{m} = 0$ and $S\mathbf{h} \leq -\varepsilon$ for some $M \times 1$ vector $\varepsilon > 0$ is testable. A method for carrying out the test is outlined in Section 3.3.

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 \mathbf{h} as the difference between its positive and negative parts. Denote the resulting vector of variables of optimization by \mathbf{z} . The dimension of \mathbf{z} is $2J + M$. There are J variables for the positive parts of the components of \mathbf{h} , J variables for the negative parts of the components of \mathbf{h} , and M slack variables for the inequality constraints. The $(2J + M) \times 1$ vector of objective function coefficients is $\bar{\mathbf{c}} = (\mathbf{c}', -\mathbf{c}', \mathbf{0}_{1 \times M})'$, where $\mathbf{0}_{1 \times M}$ is a $1 \times M$ vector of zeros. The corresponding constraint matrix has dimension $(K + M) \times (2J + M)$ and is

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

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

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

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

$$(7) \quad \begin{aligned} & \underset{\mathbf{z}}{\text{maximize}}: \bar{\mathbf{c}}'\mathbf{z} \quad \text{or} \quad -\bar{\mathbf{c}}'\mathbf{z} \\ & \text{subject to:} \quad \bar{\mathbf{A}}\mathbf{z} = \bar{\mathbf{m}} \\ & \quad \quad \quad \mathbf{z} \geq \mathbf{0}. \end{aligned}$$

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

Make the following assumption.

Assumption 4: Every set of $K + M$ columns of the augmented matrix $(\bar{\mathbf{A}}, \bar{\mathbf{m}})$ are linearly independent.

Assumption 4 ensures that the basic optimal solution(s) to (6) and (7) are nondegenerate. Under assumption 4, the linear independence condition holds with probability approaching 1 as $n \rightarrow \infty$ for the estimate of $(\bar{\mathbf{A}}, \bar{\mathbf{m}})$ described in Section 3. Therefore, asymptotically a test of assumption 4 is not needed.

Let z_{opt} be an optimal solution to either version of (7). Let $z_{B,opt}$ 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

$$z_{B,opt} = \bar{A}_B^{-1} \bar{m}$$

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

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

for the maximization version of (6) and

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

for the minimization version.

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

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

subject to

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

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

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

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

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

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

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 \mathbf{m} in (6) with consistent estimators. To this end, define

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

and

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

Then \hat{m}_k and $\hat{\pi}_{jk}$, respectively, are strongly consistent estimators of m_k and π_{jk} . Define

$\hat{\mathbf{m}} = (\hat{m}_1, \dots, \hat{m}_K)'$. 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

$$(10) \quad \underset{\mathbf{h}}{\text{maximize (minimize)}}: \mathbf{c}'\mathbf{h}$$

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

$$S\mathbf{h} \leq 0.$$

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

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

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

$$\mathbf{z} \geq 0,$$

where

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

and

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

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

Theorem 1: Let assumptions 1-3 hold. As $n \rightarrow \infty$, $\hat{L}_{\max} \rightarrow L_{\max}$ almost surely and $\hat{L}_{\min} \rightarrow L_{\min}$ 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

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

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 $k = 1, 2, \dots$ 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{A}_k and \bar{c}_k , respectively, be the versions of \hat{A}_B and \bar{c}_B associated with the k 'th basic solution of (6) or (10). Then,

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

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 ,

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

An application of the delta method yields

$$\begin{aligned}
n^{1/2}(\hat{Z}_k - Z_k) &= n^{1/2} \bar{\mathbf{c}}_k' (\hat{A}_k^{-1} \hat{\bar{\mathbf{m}}} - \bar{A}_k^{-1} \bar{\mathbf{m}}) \\
&= \bar{\mathbf{c}}_k' \bar{A}_k^{-1} [n^{1/2}(\hat{\bar{\mathbf{m}}} - \bar{\mathbf{m}}) - n^{1/2}(\hat{A}_k - \bar{A}_k) \bar{A}_k^{-1} \bar{\mathbf{m}}] + o_p(1)
\end{aligned}$$

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{A}_k and $\hat{\bar{\mathbf{m}}}$ are sample moments or constants, depending on the basic solution, and not all are constants. In addition $E(\hat{A}_k) = \bar{A}_k$ and $E(\hat{\bar{\mathbf{m}}}) = \bar{\mathbf{m}}$. 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 Σ_{\max} . In general, Σ_{\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 Σ_{\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 $\tilde{Z}_k = -\bar{\mathbf{c}}_k' \bar{A}_k^{-1} \bar{\mathbf{m}}$ and $\hat{\tilde{Z}}_k = -\bar{\mathbf{c}}_k' \hat{A}_k^{-1} \hat{\bar{\mathbf{m}}}$. Then arguments like those made for L_{\max} show that

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

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 maximum 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 Σ_{\min} . Like Σ_{\max} , Σ_{\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 Σ_{\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 $(\max Z_{\max}, \max Z_{\min})$. 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.

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

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

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

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

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

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

and

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

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 Σ_{\max} and Σ_{\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 Σ_{\max} , Σ_{\min} , \mathcal{B}_{\max} , and \mathcal{B}_{\min} .

3.3 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 $k = 1, \dots, k_{\max} \equiv \binom{2K + M}{J}$ index basic solutions to (9).⁵ A basic solution is $\tilde{\mathbf{q}} = -(\tilde{A}_k')^{-1}\mathbf{c}$ for the dual of the maximization version of (6) or $\tilde{\mathbf{q}} = (\tilde{A}_k')^{-1}\mathbf{c}$ 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 $-(\tilde{A}_k')^{-1}\mathbf{c} \geq 0$ for some k for the maximization version of (6) and $(\tilde{A}_k')^{-1}\mathbf{c} \geq 0$ for some k for the minimization version. Therefore, testing boundedness of L_{\max} (L_{\min}) is equivalent to testing the hypothesis $H_0: -(\tilde{A}_k')^{-1}\mathbf{c} \geq 0$ ($(\tilde{A}_k')^{-1}\mathbf{c} \geq 0$) for some k .

To test either hypothesis, define \hat{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

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

Equation (13) shows that the hypothesis $H_0: -(\tilde{A}_k')^{-1}\mathbf{c} \geq 0$ ($(\tilde{A}_k')^{-1}\mathbf{c} \geq 0$) is asymptotically equivalent to a one-sided hypothesis about a vector of population means. Testing $H_0: -(\tilde{A}_k')^{-1}\mathbf{c} \geq 0$ ($(\tilde{A}_k')^{-1}\mathbf{c} \geq 0$) 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 $H_0: -(\tilde{A}_k')^{-1}\mathbf{c} \geq 0$ is rejected for at least one component of $(\tilde{A}_k')^{-1}\mathbf{c}$ for each $k = 1, \dots, k_{\max}$. The hypothesis of boundedness of L_{\min} is rejected if $H_0: (\tilde{A}_k')^{-1}\mathbf{c} \geq 0$ is rejected for at least one component of $(\tilde{A}_k')^{-1}\mathbf{c}$ for each $k = 1, \dots, k_{\max}$.

We now consider assumption 3. As was discussed in Section 2, assumption 3 cannot be tested, but the hypothesis, H_0 , that there is a vector \mathbf{h} satisfying $\Pi'\mathbf{h} - \mathbf{m} = 0$ and $S\mathbf{h} \leq -\boldsymbol{\varepsilon}$ for some $M \times 1$ vector $\boldsymbol{\varepsilon} > 0$ can be tested. A test can be carried out by solving the quadratic programming problem

$$(14) \quad \underset{\mathbf{h}}{\text{minimize}}: \hat{Q}(\mathbf{h}) \equiv \|\hat{\Pi}'\mathbf{h} - \hat{\mathbf{m}}\|^2$$

subject to

$$S\mathbf{h} \leq -\boldsymbol{\varepsilon},$$

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 , $\lim_{n \rightarrow \infty} P(\hat{Q}_{opt} = 0) = 1$. Therefore, the result $\hat{Q}_{opt} = 0$ is consistent with H_0 .

A large value of \hat{Q}_{opt} is inconsistent with H_0 . To obtain an asymptotic critical value for \hat{Q}_{opt} , observe that as a consequence of the envelope theorem, \hat{Q}_{opt} is a differentiable function of the components of $\hat{\Pi}$ and \hat{m} . Therefore, an application of the delta method shows that under H_0 , \hat{Q}_{opt} is asymptotically normally distributed with a mean of 0. The analytic expression for the variance of the asymptotic distribution is lengthy and tedious to evaluate because it involves a large number of derivatives of \hat{Q}_{opt} with respect to components of $\hat{\Pi}$ and \hat{m} . An asymptotically valid critical value can be obtained more easily by using the bootstrap. The bootstrap procedure is;

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

These are

$$(15) \quad m_k^* = n^{-1} \sum_{i=1}^n Y_i^* I(W_i^* = w_k)$$

and

$$(16) \quad \pi_{jk}^* = n^{-1} \sum_{i=1}^n I(X_i^* = x_j) I(W_i^* = w_k).$$

Define Π^* and m^* , respectively, as the matrix and vector that are obtained by replacing the estimation sample with the bootstrap sample in $\hat{\Pi}$ and \hat{m} .

(ii) Solve problem (14) with Π^* and m^* in place of $\hat{\Pi}$ and \hat{m} . Denote the resulting optimal value of the objective function by Q_{opt}^* .

(iii) Estimate the asymptotic distribution of \hat{Q}_{opt} by the empirical distribution of $Q_{opt}^* - \hat{Q}_{opt}$ 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 $Q_{opt}^* - \hat{Q}_{opt}$.

The bootstrap consistently estimates the critical value of \hat{Q}_{opt} because asymptotically, \hat{Q}_{opt} is a smooth function of sample moments whose asymptotic distributions are estimated consistently by the bootstrap.

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 Σ_{\max} , Σ_{\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 $\{c_n : n=1,2,\dots\}$ be a sequence of positive constants such that $c_n \rightarrow 0$ and $c_n[n/(\log \log n)]^{1/2} \rightarrow \infty$ as $n \rightarrow \infty$. Let P^* denote the probability measure induced by bootstrap sampling.

The first bootstrap procedure is as follows.

(i) Generate a bootstrap sample $\{Y_i^*, X_i^*, W_i^* : i=1,\dots,n\}$ by sampling the estimation data $\{Y_i, X_i, W_i : i=1,\dots,n\}$ randomly with replacement. Use (15) and (16) to compute the bootstrap versions of \hat{m}_k and $\hat{\pi}_{jk}$, which are m_k^* and π_{jk}^* . Define Π^* and \mathbf{m}^* , respectively, as the matrix and vector that are obtained by replacing the estimation sample with the bootstrap sample in $\hat{\Pi}$ and $\hat{\mathbf{m}}$. For any basic solution k to problem (6), define \bar{A}_k^* and $\bar{\mathbf{m}}^*$ by replacing the estimation sample with the bootstrap sample in \hat{A}_k and $\hat{\mathbf{m}}$.

(ii) Define problem (B10) as problem (10) with Π^* and \mathbf{m}^* in place of $\hat{\Pi}$ and $\hat{\mathbf{m}}$. 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

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

and

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

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

(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

$$P^*(\min_{k \in \hat{\mathcal{B}}_{\min}} \Delta_k^* \leq c_{\alpha,\min}^*) = 1 - \alpha$$

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

Asymptotically, $n^{1/2}(\bar{\mathbf{c}}_k' \hat{\mathbf{A}}_k^{-1} \hat{\mathbf{m}} - \bar{\mathbf{c}}_k' \bar{\mathbf{A}}_k^{-1} \bar{\mathbf{m}})$ ($k \in \mathcal{B}_{\max} \cup \mathcal{B}_{\min}$) is a linear function of sample moments. Therefore, the bootstrap distributions of Δ_{1k}^* and Δ_{2k}^* uniformly consistently estimate the asymptotic distributions of $\pm n^{1/2}(\bar{\mathbf{c}}_k' \hat{\mathbf{A}}_k^{-1} \hat{\mathbf{m}} - \bar{\mathbf{c}}_k' \bar{\mathbf{A}}_k^{-1} \bar{\mathbf{m}})$ 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 $n \rightarrow \infty$, 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 $|\hat{L}_{k,\max} - \hat{L}_{\max}| \leq c_n$ and only basic solutions $k \in \mathcal{B}_{\min}$ satisfy $|\hat{L}_{k,\min} - \hat{L}_{\min}| \leq c_n$. 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.⁶ It further follows that 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 $n \rightarrow \infty$. Under the first bootstrap procedure,

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

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

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

$$(iv) |c_\alpha^* - c_\alpha| \rightarrow^p 0. \blacksquare$$

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 $|\hat{L}_{k,\max} - \hat{L}_{\max}| \leq c_n$ or $|\hat{L}_{k,\min} - \hat{L}_{\min}| \leq c_n$. However, when n is large, basic solutions for which $|\hat{L}_{k,\max} - \hat{L}_{\max}| \leq c_n$ or $|\hat{L}_{k,\min} - \hat{L}_{\min}| \leq c_n$ 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 $|\hat{L}_{k,\max} - \hat{L}_{\max}| \leq c_n$ or $|\hat{L}_{k,\min} - \hat{L}_{\min}| \leq c_n$ 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.⁷

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 $\{Y_i^*, X_i^*, W_i^* : i=1, \dots, n\}$ by sampling the estimation data $\{Y_i, X_i, W_i : i=1, \dots, n\}$ randomly with replacement. Use (15) and (16) to compute the bootstrap versions of \hat{m}_k and $\hat{\pi}_{jk}$, which are m_k^* and π_{jk}^* . Define Π^* and \mathbf{m}^* , respectively, as the matrix and vector that

are obtained by replacing the estimation sample with the bootstrap sample in $\hat{\Pi}$ and $\hat{\mathbf{m}}$. For any basic solution k to problem (6), define \bar{A}_k^* and $\bar{\mathbf{m}}^*$ by replacing the estimation sample with the bootstrap sample in \hat{A}_k and $\hat{\mathbf{m}}$.

(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

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

and

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

(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

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

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

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 $n \rightarrow \infty$. 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)

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

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

$$(17) \quad \Pi' = \begin{pmatrix} 0.20 & 0.10 & 0.06 & 0.05 & 0.03 & 0.03 \\ 0.15 & 0.12 & 0.07 & 0.08 & 0.06 & 0.05 \end{pmatrix}.$$

In experiments with $J = 4$, $X \in \{2, 3, 4, 5\}$, and Π' is obtained from (17) by

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

In experiments with $J = 6$, $\mathbf{g} = (23, 17, 13, 11, 9, 8)'$. Thus, $g(x)$ is decreasing and convex. We also require $g(1) - g(J) \leq 52$. In experiments with $J = 4$, $\mathbf{g} = (23, 17, 13, 11)'$. 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 $U = XZ^2 - E(X|W)$ 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 $\vec{c}_k \hat{A}_k^{-1} \hat{\mathbf{m}}$.

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 $\{Y_i, X_i, W_i : i = 1, \dots, n\}$, 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).⁸ 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. 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.

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	$[-22.7, -7.1]$

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., 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.
- Hadley, G. (1962). *Linear Programming*. Reading, MA: Addison-Wesley Publishing Company.
- 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*, 3rd 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.
- 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 $j=1,\dots,J$ 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 $j=1,\dots,J$.

³ The feasible region of problem (6) with Π and \mathbf{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 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 $n^{1/2}(\bar{\mathbf{c}}_k' \hat{A}_k^{-1} \hat{\mathbf{m}} - \bar{\mathbf{c}}_k' \bar{A}_k^{-1} \bar{\mathbf{m}})$ all have means of zero.

⁷ 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.

⁸ Angrist's and Evans's (1998) data are available at <http://economics.mit.edu/faculty/angrist/data1/data/angev98>.