



# Identification and shape restrictions in nonparametric instrumental variables estimation



Joachim Freyberger<sup>a</sup>, Joel L. Horowitz<sup>b,\*</sup>

<sup>a</sup> Department of Economics, University of Wisconsin, Madison, WI 53706, United States

<sup>b</sup> Department of Economics, Northwestern University, Evanston IL 60208, United States

## ARTICLE INFO

### Article history:

Received 23 October 2013

Received in revised form

28 April 2015

Accepted 18 June 2015

Available online 3 July 2015

### JEL classification:

C13

C14

C26

### Keywords:

Partial identification

Linear programming

Bootstrap

## ABSTRACT

This paper is concerned with inference about an unidentified linear functional,  $L(g)$ , where  $g$  satisfies  $Y = g(X) + U$ ;  $E(U|W) = 0$ . In much applied research,  $X$  and  $W$  are discrete, and  $W$  has fewer points of support than  $X$ . Consequently,  $L(g)$  is not identified nonparametrically and can have any value in  $(-\infty, \infty)$ . This paper uses shape restrictions, such as monotonicity or convexity, to achieve interval identification of  $L(g)$ . The paper shows that under shape restrictions,  $L(g)$  is contained in an interval whose bounds can be obtained by solving linear programming problems. Inference about  $L(g)$  can be carried out by using the bootstrap. An empirical application illustrates the usefulness of the method.

© 2015 Elsevier B.V. All rights reserved.

## 1. Introduction

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

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

and

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

for almost every  $w$ . Equivalently,

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

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

\* Correspondence to: Department of Economics, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208, United States. Tel.: +1 847 491 8253; fax: +1 847 491 7001.

E-mail address: [joel-horowitz@northwestern.edu](mailto:joel-horowitz@northwestern.edu) (J.L. Horowitz).

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  $\beta_0$  and  $\beta_1$  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 et al. \(2009\)](#), who treated estimation of the interval  $[\sup_{v \in \mathcal{V}} \theta^l(v), \inf_{v \in \mathcal{V}} \theta^u(v)]$ , where  $\theta^l$  and  $\theta^u$  are unknown functions and  $\mathcal{V}$  is a possibly infinite set. [Santos \(2012\)](#) treats a case in which  $g$  is partially identified but  $X$  and  $W$  are continuously distributed and there is no shape restriction.

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 extends the results of the previous sections to models with exogenous covariates. Section 7 presents concluding comments. An [Appendix](#) provides proofs that are not in the text.

It is assumed throughout this paper that  $X$  and  $W$  are discretely distributed with finitely many mass points. The use of shape restrictions with continuously distributed variables is beyond the scope of this paper.

## 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  $K < J$ . 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

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

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

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

Note that  $\text{rank}(\Pi) < J$ , because  $K < J$ . Therefore, (4) does not point identify  $\mathbf{g}$ . 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 Eq. (4) is true for some  $\mathbf{g}$ ,  $K < J$ , and  $\mathbf{c}$  is not orthogonal to the null space of  $\Pi'$ . Then any value of  $L(g)$  in  $(-\infty, \infty)$  is consistent with (1a) and (1b). ■

**Proof.** By (4), there is a vector  $\mathbf{g}_1$  in the space spanned by the rows of  $\Pi'$  satisfying  $\Pi' \mathbf{g}_1 = \mathbf{m}$ . Let  $\mathbf{g}_2$  be a vector in the null space (the orthogonal complement of the row space) of  $\Pi'$  such that  $\mathbf{c}' \mathbf{g}_2 \neq 0$ . For any real  $\gamma$ ,  $\Pi'(\mathbf{g}_1 + \gamma \mathbf{g}_2) = \mathbf{m}$  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  $\gamma$  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

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

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.** (i) The unknown function  $g$  satisfies (1a)–(1b) with  $S\mathbf{g} \leq 0$ , and (ii)  $K < J$ .

Sharp bounds on  $L(g)$  are the optimal values of the objective functions of the linear programming problems

$$\begin{aligned} & \underset{\mathbf{h}}{\text{maximize (minimize)}}: \mathbf{c}'\mathbf{h} \\ & \text{subject to: } \Pi'\mathbf{h} = \mathbf{m} \\ & \quad S\mathbf{h} \leq 0. \end{aligned} \tag{6}$$

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 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.** Let Assumption 1 hold. Then 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.<sup>1</sup> 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.<sup>2</sup>

<sup>1</sup>  $L_{\min}$  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).

<sup>2</sup> 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  $j = 1, \dots, J$ . If the range of  $g$  is bounded, then under the assumptions of Proposition 1,  $\mathbf{c}'\mathbf{g}$  can take any value in its logically permitted range.

We also assume:

**Assumption 3.** (i) There is a vector  $\mathbf{h}$  satisfying  $\Pi'\mathbf{h} - \mathbf{m} = 0$  and  $S\mathbf{h} \leq -\boldsymbol{\varepsilon}$  for some vector  $\boldsymbol{\varepsilon} > 0$ . (The inequality holds component-wise, and each component of  $\boldsymbol{\varepsilon}$  exceeds zero.) (ii) The vector  $\mathbf{c}$  is not orthogonal to the null space of  $\Pi'$ .

Assumption 3(i) ensures that problem (6) has a feasible solution with probability approaching 1 as  $n \rightarrow \infty$  when  $\Pi$  and  $\mathbf{m}$  are replaced by consistent estimators.<sup>3</sup> Assumptions 3(i) and 3(ii) imply 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. A method for testing the hypothesis that  $\mathbf{c}$  is orthogonal to the null space of  $\Pi'$  (or, equivalently, that Assumption 3(ii) does not hold) is described in Section 3.4. Assumption 3(i) 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  $\varepsilon > 0$ .

### 2.1. Relation to the local average treatment effect

This section shows that  $L_{\min}$  and  $L_{\max}$  in problem (6) are lower and upper bounds on the local average treatment effect (LATE) of Angrist and Imbens (1995). Let  $X$  be a scalar treatment indicator with support  $x_1 < \dots < x_J$  for some  $J > 2$ . Let  $W$  be a binary instrument for  $X$  with support  $\{0, 1\}$ . Let  $X^1$  and  $X^0$ , respectively, denote the treatments received by randomly selected individuals for whom  $W = 1$  and  $W = 0$ . Let  $Y_j$  denote the outcome of an individual who receives treatment  $X^j$ . Finally, let assumptions 1 and 2 of Angrist and Imbens (1995) hold with  $X^1 > X^0$ . These assumptions are stated in the Appendix of this paper. Define  $\beta_j = E(Y_j - Y_{j-1} | X^1 \geq x_j > X^0)$ . Angrist and Imbens (1995) call  $\beta_j$  a local average treatment effect. Because  $Y_j$  and  $Y_{j-1}$  are not both observed,  $\beta_j$  is not necessarily point identified. Define  $\Pi$  and  $\mathbf{m}$  as in (3)–(4). The following proposition is proved in the Appendix:

**Proposition 3.** Under the conditions of the foregoing paragraph, there is a unique vector  $\mathbf{g} = (g_1, \dots, g_J)'$  such that  $\Pi'\mathbf{g} = \mathbf{m}$  and  $g_j - g_{j-1} = \beta_j$ . ■

Choose  $\mathbf{c}$  in (6) so that  $\mathbf{c}'\mathbf{g} = g_j - g_{j-1}$ . Then  $L_{\min}$  and  $L_{\max}$  bound the LATE. Shape restrictions on  $\mathbf{g}$  correspond to restrictions on the  $\beta_j$ s. For example, if  $\mathbf{g}$  is non-increasing, then  $\beta_j \leq 0$  for all  $j = 2, \dots, J$ . If  $\mathbf{g}$  is concave, then  $\beta_{j+1} \leq \beta_j$  for all  $j = 1, \dots, J - 1$ . Proposition 3 does not imply that  $\beta_j$  is point identified, because the vector  $\mathbf{g}$  for which  $\Pi'\mathbf{g} = \mathbf{m}$  and  $g_j - g_{j-1} = \beta_j$  is not point identified under the assumptions of this paper.

### 2.2. 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  $\mathbf{z}$ .

<sup>3</sup> The feasible region of problem (6) with  $\Pi$  and  $\mathbf{m}$  replaced with consistent estimators may be empty if  $n$  is small. This problem is treated in Section 3.4.

The dimension of  $z$  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

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

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

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

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

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

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

Define the matrix

$$\tilde{A} = \begin{pmatrix} \Pi' & \mathbf{0}_{K \times M} \\ S & I_{M \times M} \end{pmatrix}.$$

Make the following assumption.

**Assumption 4.** Every set of  $K + M$  columns of the matrix  $(\tilde{A} \bar{\mathbf{m}})$  is linearly independent.

Assumption 4 implies that  $\text{rank}(\Pi) = K$  and ensures that the basic optimal solution(s) to (6) and (7) are nondegenerate (that is, no basic variables are zero) (Hadley, 1962). The assumption can be tested using methods like those described by Härdle and Hart (1992).<sup>4</sup>

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

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

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

$$Z_B = \bar{\mathbf{c}}_B' \tilde{A}_B^{-1} \bar{\mathbf{m}} \tag{8a}$$

for the maximization version of (6) and

$$\tilde{Z}_B = -\bar{\mathbf{c}}_B' \tilde{A}_B^{-1} \bar{\mathbf{m}} \tag{8b}$$

for the minimization version.

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

$$\begin{aligned} & \text{maximize: } \tilde{\mathbf{m}}' \tilde{\mathbf{q}} \quad \text{or} \quad -\tilde{\mathbf{m}}' \tilde{\mathbf{q}} \\ & \text{subject to } \tilde{A}' \tilde{\mathbf{q}} = \mathbf{c} \\ & \quad \quad \quad \tilde{\mathbf{q}} \geq 0, \end{aligned} \tag{9}$$

<sup>4</sup> A basic solution to the system of equations  $\tilde{A}x = \bar{\mathbf{m}}$  is defined as follows. Let  $B$  be a matrix formed by choosing  $K + M$  of the  $2J + M$  columns of  $\tilde{A}$ . If  $B$  is nonsingular and all of the  $2J - K$  variables not associated with these columns are set equal to zero, then the solution to the resulting system of equations is called a basic solution. The variables associated with the  $K + M$  columns are called basic variables. The remaining variables are called non-basic.

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

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

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

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

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 3.5 to form a test of Assumption 2.

### 3. Estimation of $L_{\text{max}}$ and $L_{\text{min}}$

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

#### 3.1. Consistent estimators of $L_{\text{max}}$ and $L_{\text{min}}$

$L_{\text{max}}$  and  $L_{\text{min}}$  can be estimated consistently by replacing  $\Pi$  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

$$\hat{\pi}_{jk} = n^{-1} \sum_{i=1}^n I(X_i = x_j) I(W_i = w_k);$$

$$j = 1, \dots, J; \quad k = 1, \dots, K.$$

Then  $\hat{m}_k$  and  $\hat{\pi}_{jk}$ , respectively, are strongly consistent estimators of  $m_k$  and  $\pi_{jk}$ . 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  $\hat{L}_{\text{max}}$  and  $\hat{L}_{\text{min}}$  as the optimal values of the objective functions of the linear programs

$$\begin{aligned} & \text{maximize (minimize): } \mathbf{c}'\mathbf{h} \\ & \text{subject to: } \hat{\Pi}'\mathbf{h} = \hat{\mathbf{m}} \\ & \quad \quad \quad S\mathbf{h} \leq 0. \end{aligned} \tag{10}$$

Assumptions 2 and 3 ensure that (10) has a feasible solution and a bounded optimal solution with probability approaching 1 as  $n \rightarrow \infty$ . Section 3.4 treats the possibility that (10) does not have a feasible solution if  $n$  is small. The standard form of (10) is

$$\begin{aligned} & \text{maximize: } \bar{\mathbf{c}}'z \quad \text{or} \quad -\bar{\mathbf{c}}'z \\ & \text{subject to: } \hat{A}z = \hat{\mathbf{m}} \\ & \quad \quad \quad z \geq 0, \end{aligned} \tag{11}$$

where

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

and

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

It follows from (8a) to (8b) that the objective function at each optimal basic solution and, therefore,  $L_{\text{max}}$  and  $L_{\text{min}}$  are continuous functions of the components of  $\bar{\mathbf{m}}$  and  $A$ . This result and the strong consistency of  $\hat{\Pi}$  and  $\hat{\mathbf{m}}$  for  $\Pi$  and  $\bar{\mathbf{m}}$ , respectively, imply:

**Theorem 1.** Let Assumptions 1–3 hold. As  $n \rightarrow \infty$ ,  $\hat{L}_{\text{max}} \rightarrow L_{\text{max}}$  almost surely and  $\hat{L}_{\text{min}} \rightarrow L_{\text{min}}$  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  $k = 1, \dots, K$ .

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  $\mathcal{B}_{\max}$ . 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  $\mathcal{B}_{\min}$ . The following theorem, which is proved in the Appendix, gives the asymptotic distributions of  $\hat{L}_{\max}$  and  $\hat{L}_{\min}$ .

**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  $\Sigma_{\max}$ ; (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 covariance matrices of the asymptotic distributions of  $n^{1/2}(\hat{L}_{\max} - L_{\max})$ ,  $n^{1/2}(\hat{L}_{\min} - L_{\min})$ , and  $[n^{1/2}(\hat{L}_{\max} - L_{\max}), n^{1/2}(\hat{L}_{\min} - L_{\min})]$  are algebraically complex and tedious to calculate. Section 4 presents bootstrap methods for estimating the asymptotic distributions of  $n^{1/2}(\hat{L}_{\max} - L_{\max})$ ,  $n^{1/2}(\hat{L}_{\min} - L_{\min})$ , and  $[n^{1/2}(\hat{L}_{\max} - L_{\max}), n^{1/2}(\hat{L}_{\min} - L_{\min})]$  that do not require knowledge of  $\mathcal{B}_{\max}$  or calculation of the covariance matrices.

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  $\sigma_{\max}^2$  and  $\sigma_{\min}^2$ , 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  $\rho$  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  $\rho$ . Then the following corollary to Theorem 2 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  $n^{1/2}(\hat{L}_{\min} - L_{\min})/\sigma_{\min} \rightarrow^d N(0, 1)$ . 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  $c_{\alpha}$  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 intervals 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  $\hat{L}_{\max}$ .<sup>5</sup> 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 the relevant covariance matrices are complicated and unknown, 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 the covariance matrices or  $\mathcal{B}_{\max}$  and  $\mathcal{B}_{\min}$ .

### 3.3. Confidence intervals when some suboptimal basic solutions are nearly optimal

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. As a consequence, the asymptotic approximations of Section 3.2 may be inaccurate in finite samples when there are one or more suboptimal basic solutions that are “nearly” optimal. 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  $k$ . Finally, let  $P_{\infty}(\cdot)$  denote probability with respect to the asymptotic distribution of the random variable in its argument. Thus, for example, if  $\{\tau_n\}$  is a sequence of random variables that converges in distribution,  $P_{\infty}(\tau_n \leq t) = \lim_{n \rightarrow \infty} P(\tau_n \leq t)$  and  $P_{\infty}(\tau_n > t) = \lim_{n \rightarrow \infty} P(\tau_n > t)$  at each continuity point of the limiting distribution function. We have  $n^{1/2}(L_{\max} - L_{k, \max}) > 0$  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  $\tilde{k} \in \mathcal{B}_{\max}$ ,

$$\begin{aligned} P_{\infty}[n^{1/2}(\hat{L}_{\max} - L_{\max}) > -c] &\geq P_{\infty} \left\{ \max_{k \in \mathcal{B}_{\max}} [n^{1/2}(\hat{Z}_k - L_{k, \max}) - n^{1/2}(L_{\max} - L_{k, \max})] > -c \right\} \\ &\geq P_{\infty}[n^{1/2}(\hat{Z}_{\tilde{k}} - L_{\tilde{k}, \max}) > -c]. \end{aligned}$$

Moreover

$$\begin{aligned} P_{\infty}[n^{1/2}(\hat{L}_{\max} - L_{\max}) > -c] &\geq \min_{k \in \mathcal{B}_{\max}} P_{\infty}[n^{1/2}(\hat{Z}_k - L_{k, \max}) > -c]. \end{aligned}$$

<sup>5</sup> 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.

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

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

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

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

This result holds without regard to the number of suboptimal basic solutions that may be nearly optimal.

Now define

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

$\mathcal{B}_{n,\max}$  contains all optimal basic solutions to the maximization version of (10) with probability approaching 1 as  $n \rightarrow \infty$ . Define  $c_{\alpha,\max} = \max_{k \in \mathcal{B}_{n,\max}} c_{k,\alpha,\max}$ . Then for all sufficiently large  $n$  and regardless of the number of suboptimal but nearly optimal basic solutions,

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

where  $\varepsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ . 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  $1 - \alpha$  regardless of the number of nearly optimal basic solutions.<sup>6</sup>

To obtain a 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

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

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

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

where  $\varepsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ , regardless of the number of suboptimal but nearly optimal basic solutions. 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  $1 - \alpha$ , regardless of the number or nearly optimal basic solutions.

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

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

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

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

where  $\varepsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ , regardless of the number of nearly optimal basic solutions to the maximization and minimization versions of (6). 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  $c_\alpha$  in applications.

### 3.4. Confidence intervals under near point identification

If the feasible region of problem (6) is small, as can happen if  $\mathbf{c}'\mathbf{g}$  is “nearly” point identified, then problem (10) may have no feasible solution. When this happens,  $\hat{L}_{\max}$  and  $\hat{L}_{\min}$  cannot be computed by solving (10). This section describes a method for obtaining a confidence interval for  $\mathbf{c}'\mathbf{g}$  if (10) has no feasible solution. The method requires solving a nonlinear programming problem, which is more difficult than solving (10). Moreover, for reasons explained later in this section, the method yields a conservative confidence interval. That is, it yields a confidence interval whose asymptotic coverage probability is larger than the nominal coverage probability and is longer than a confidence interval that has the nominal coverage probability. Therefore, the method of Sections 3.1 and 3.2 is preferred to the method of this section if (10) has a feasible solution. However, the method of this section is useful if (10) has no feasible solution or if there is concern that the asymptotic approximations of Sections 3.2 and 3.3 are inaccurate because the feasible region of (6) is “small” compared to the size of random sampling errors in  $\hat{L}_{\max}$  and  $\hat{L}_{\min}$ .

To describe the method, let  $\hat{\Pi}_{-1}$  and  $\Pi_{-1}$ , respectively, be the matrices that are obtained by omitting one element of  $\hat{\Pi}$  and  $\Pi$ . Define

$$\Gamma_n(\hat{\mathbf{m}}, \hat{\Pi}; \mathbf{m}, \Pi) = \begin{bmatrix} n^{1/2}(\hat{\mathbf{m}} - \mathbf{m}) \\ n^{1/2}vec(\hat{\Pi}_{-1} - \Pi_{-1}) \end{bmatrix},$$

where  $vec(\cdot)$  is the vector obtained by stacking the components of the matrix  $(\cdot)$ . The sums of the elements of  $\hat{\Pi}$  and  $\Pi$  are 1, so the covariance matrix of  $\hat{\Pi} - \Pi$  is singular. Omitting one element of  $\hat{\Pi}$  and  $\Pi$  avoids this problem. Then  $\Gamma_n(\hat{\mathbf{m}}, \hat{\Pi}_{-1}; \mathbf{m}, \Pi_{-1}) \rightarrow^d N(0, \Omega)$  for some covariance matrix,  $\Omega$ . Assume that  $\Omega$  is non-singular. Let  $\hat{\Omega}$  be a consistent estimator of  $\Omega$  (e.g., the estimator obtained from sample moments). Let  $d = \dim(\Gamma_n)$ . Then

$$\Gamma_n(\hat{\mathbf{m}}, \hat{\Pi}_{-1}; \mathbf{m}, \Pi_{-1})' \hat{\Omega}^{-1} \Gamma_n(\hat{\mathbf{m}}, \hat{\Pi}_{-1}; \mathbf{m}, \Pi_{-1}) \rightarrow^d \chi_d^2.$$

Let  $c(\alpha, d)$  denote the  $1 - \alpha$  quantile of the  $\chi_d^2$  distribution. Define  $\hat{F}_{\max}$  and  $\hat{F}_{\min}$  as the optimal values of the objective functions of the nonlinear programming problems

$$\begin{aligned} &\text{maximize (minimize): } \mathbf{c}'\mathbf{h} \\ &\text{subject to: } \begin{aligned} &\mathbf{h}, \mathbf{m}, \Pi \\ &\Pi' \mathbf{h} = \mathbf{m} \\ &\mathbf{S}\mathbf{h} \leq 0 \\ &\Gamma_n(\hat{\mathbf{m}}, \hat{\Pi}_{-1}; \mathbf{m}, \Pi_{-1})' \hat{\Omega}^{-1} \Gamma_n(\hat{\mathbf{m}}, \hat{\Pi}_{-1}; \mathbf{m}, \Pi_{-1}) \\ &\leq c(\alpha, d) \\ &\sum_{j=1}^J \sum_{k=1}^K \Pi_{jk} = 1 \\ &0 \leq \Pi \leq 1, \end{aligned} \end{aligned} \quad (12)$$

<sup>6</sup> At the cost of greater notational complexity, this argument can be formalized by considering a sequence of  $\mathbf{c}$  values for which  $\mathcal{B}_{\max}$  is constant but there are suboptimal basic solutions  $k$  for which  $L_{k,\max} - L_{\max} \rightarrow 0$  and  $n^{1/2}(\log n)^{-1}(L_{k,\max} - L_{k,\max}) \rightarrow \infty$  as  $n \rightarrow \infty$ .

where the inequalities in the second and fifth constraints hold component by component. The third constraint holds with probability  $1 - \alpha$ . The other constraints are deterministic. Therefore,

$$\lim_{n \rightarrow \infty} P(\hat{F}_{\min} \leq L_{\min} \leq L_{\max} \leq \hat{F}_{\max}) \geq 1 - \alpha.$$

Thus,  $[\hat{F}_{\min}, \hat{F}_{\max}]$  is a confidence interval for  $\mathbf{c}'\mathbf{g}$  whose asymptotic coverage probability is at least  $1 - \alpha$  and does not require (10) to have a feasible solution.

The coverage probability of the confidence interval  $[\hat{F}_{\min}, \hat{F}_{\max}]$  exceeds  $1 - \alpha$  except in special cases, because a confidence region for  $(\mathbf{m}, \Pi)'$  is usually larger than a confidence region for  $\mathbf{c}_B \bar{A}_B^{-1} \mathbf{m}$ , which (8a)–(8b) show is the parameter that determines  $L_{\min}$  and  $L_{\max}$ . It is possible that the feasible region for (12) is empty. This constitutes rejection at an asymptotic level not exceeding  $\alpha$  of the hypothesis that there is a vector  $\mathbf{g}$  satisfying (4) and (5).

### 3.5. Testing Assumptions 2 and 3(ii)

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).<sup>7</sup> 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  $\hat{\tilde{A}}'_k$  as the matrix that is obtained by replacing the components of  $\Pi$  with the corresponding components of  $\hat{\Pi}$  in  $\tilde{A}'_k$ . Then an application of the delta method yields

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

Eq. (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 et al. (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  $k = 1, \dots, k_{\max}$ .

Now, consider Assumption 3(ii). We outline a test of the hypothesis,  $H_0$ , that  $\mathbf{c}$  is orthogonal to the null space of  $\Pi'$  or,

equivalently, that Assumption 3(ii) does not hold. Under  $H_0$ ,  $\mathbf{c}$  is a linear function of columns of  $\Pi$ . Therefore,

$$S(\Pi) \equiv \min_{\mathbf{a}} (\mathbf{c} - \Pi \mathbf{a})' (\mathbf{c} - \Pi \mathbf{a}) = 0. \quad (14)$$

The matrix  $\Pi' \Pi$  is non-singular under Assumption 4, so the analytic solution to (14) is

$$\mathbf{a} = (\Pi' \Pi)^{-1} \Pi' \mathbf{c}.$$

Therefore,  $H_0$  is equivalent to

$$\Pi (\Pi' \Pi)^{-1} \Pi' \mathbf{c} - \mathbf{c} = 0. \quad (15)$$

A Wald statistic for testing (15) can be obtained by replacing  $\Pi$  with  $\hat{\Pi}$  and applying the delta method to the resulting version of (15).

## 4. Bootstrap estimation of the asymptotic distributions of $\hat{L}_{\max}$ and $\hat{L}_{\min}$

This section presents 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  $c_{\alpha, \min}$ ,  $c_{\alpha, \max}$ , and  $c_{\alpha}$ . The first procedure yields confidence regions for  $[L_{\min}, L_{\max}]$  and  $L(\mathbf{g})$  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. Compute the bootstrap versions of  $\hat{m}_k$  and  $\hat{\pi}_{jk}$ , which are  $m_k^*$  and  $\pi_{jk}^*$ . Define  $\Pi^*$  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  $\hat{A}'_k$  and  $\mathbf{m}^*$  by replacing the estimation sample with the bootstrap sample in  $\hat{A}'_k$  and  $\hat{\mathbf{m}}$ .

<sup>7</sup> We assume that  $2K + M \geq J$  as happens, for example, if  $\mathbf{g}$  is assumed to be monotone, convex, or both.

(ii) Define problem (B10) as problem (10) with  $\Pi^*$  and  $\mathbf{m}^*$  in place of  $\hat{\Pi}$  and  $\hat{\mathbf{m}}$ . Solve (B10).<sup>8</sup> 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

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

and

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

(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

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

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

Estimate  $c_\alpha$  by the quantity  $c_\alpha^*$  that satisfies

$$P^*(-\max_{k \in \hat{\mathcal{B}}_{\min}} \Delta_{2k}^* \leq c_\alpha^*, -c_\alpha^* \leq \max_{k \in \hat{\mathcal{B}}_{\max}} \Delta_{1k}^*) = 1 - \alpha.$$

Asymptotically,  $n^{1/2}(\bar{c}'_k \hat{A}_k^{-1} \hat{\mathbf{m}} - \bar{c}'_k \bar{A}_k^{-1} \bar{\mathbf{m}})$  ( $k \in \mathcal{B}_{\max} \cup \mathcal{B}_{\min}$ ) is a linear function of sample moments. Therefore, the bootstrap distributions of  $\Delta_{1k}^*$  and  $\Delta_{2k}^*$  consistently estimate the asymptotic distributions of  $\pm n^{1/2}(\bar{c}'_k \hat{A}_k^{-1} \hat{\mathbf{m}} - \bar{c}'_k \bar{A}_k^{-1} \bar{\mathbf{m}})$  uniformly over  $(-\infty, \infty)$  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 approaching 1 as  $n \rightarrow \infty$ , 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 approaching 1 as  $n \rightarrow \infty$ . 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  $(\max_{k \in \hat{\mathcal{B}}_{\max}} \Delta_{1k}^*, -\max_{k \in \hat{\mathcal{B}}_{\min}} \Delta_{2k}^*)$  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, uniformly over  $(-\infty, \infty)$  for the first two distributions and  $(-\infty, \infty)^2$  for the third.<sup>9</sup>

Let  $P^*$  denote the probability measure induced by bootstrap sampling. Then we have the following theorem, which is proved formally in the Appendix.

<sup>8</sup> The method of Section 3.4 can be used if (B10) has no feasible solution for a significant number of bootstrap samples or if  $|\hat{L}_{\max} - \hat{L}_{\min}| \leq c_n$ . If (B10) has no feasible solution for only a few bootstrap samples, then an asymptotically correct confidence interval can be obtained by disregarding those samples or replacing them with bootstrap samples for which (B10) has a feasible solution.

<sup>9</sup> 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  $\Delta_{1k}^*$ ,  $\Delta_{2k}^*$ , and the asymptotic form of  $n^{1/2}(\bar{c}'_k \hat{A}_k^{-1} \hat{\mathbf{m}} - \bar{c}'_k \bar{A}_k^{-1} \bar{\mathbf{m}})$  all have means of zero. The bootstrap is also consistent if  $n^{1/2}(\hat{Z}_k - Z_k)$  is deterministically 0 because  $Z_k$  ( $\hat{Z}_k$ ) is a smooth function of population (sample) moments.

**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)

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

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

and

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

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.<sup>10</sup>

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 (19) and (20) to compute the bootstrap versions of  $\hat{m}_k$  and  $\hat{\pi}_{jk}$ , which are  $m_k^*$  and  $\pi_{jk}^*$ . Define  $\Pi^*$  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

<sup>10</sup> 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.



replacing the estimation sample with the bootstrap sample in  $\hat{A}_k$  and  $\hat{m}$ .

(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

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

and

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

(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

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

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

and

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

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 approaching 1 as  $n \rightarrow \infty$ . Therefore, the second bootstrap procedure estimates the asymptotic distribution of  $n^{1/2}(\hat{L}_{\max} - L_{\max})$  consistently uniformly over  $(-\infty, \infty)$  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  $n^{1/2}(\hat{L}_{\min} - L_{\min})$  consistently uniformly over  $(-\infty, \infty)$ , 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)

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

(ii)

$$P(L_{\min} \geq \hat{L}_{\min} - c_{\alpha, \min}^*) \geq 1 - \alpha + o(1)$$

(iii)

$$P^*(\hat{L}_{\min} - c_{\alpha}^* \leq L_{\min}, L_{\max} \leq \hat{L}_{\max} + c_{\alpha}^*) \geq 1 - \alpha + o(1). \blacksquare$$

**Proof.** Only part (i) is proved. The proof of parts (ii) and (iii) are similar. With probability approaching 1 as  $n \rightarrow \infty$ ,  $\hat{k}_{\max} \in \mathcal{B}_{\max}$ , so

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

and

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

Therefore, by Theorem 3(i)

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

The bootstrap can also be used to obtain the 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  $\Delta_{1k}^*$ . 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  $c_{\alpha}$  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

$$\Pi' = \begin{pmatrix} 0.20 & 0.10 & 0.06 & 0.05 & 0.03 & 0.03 \\ 0.15 & 0.12 & 0.07 & 0.08 & 0.06 & 0.05 \end{pmatrix}. \quad (16)$$

In experiments with  $J = 4$ ,  $X \in \{2, 3, 4, 5\}$ , and  $\Pi'$  is obtained from (16) by

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

In experiments with  $J = 6$ ,  $g = (23, 17, 13, 11, 9, 8)'$ . Thus,  $g(x)$  is decreasing and convex. We also require  $g(1) - g(J) \leq 52$ . In experiments with  $J = 4$ ,  $g = (23, 17, 13, 11)'$ . 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  $\Pi'$  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  $c_n = cn^{-1/2} \log n$ , where  $c = 1$  or  $c = 0$ . Bootstrap procedure 1 with  $c = 0$  is equivalent to bootstrap procedure 2. The results of experiments using bootstrap procedure 1 with larger values of  $c$  were similar to those with  $c = 1$ .

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

**Table 1**  
Results of Monte Carlo experiments assuming only monotonicity.

$L(g)$	$c$	$J$	Empirical coverage probability with $n = 1000$	Empirical coverage probability with $n = 5000$
$g(3) - g(2)$	0	4	0.962	0.963
	1		0.9620	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.

$L(g)$	$c$	$J$	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

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  $\bar{c}'_k \hat{A}_k^{-1} \hat{m}$ .

**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 obtained confidence intervals for  $L(g)$  by using the first bootstrap procedure of Section 4.1 with  $c_n =$

$n^{-1/2} \log n$ . The results are not sensitive to varying  $c_n$  between 0 and  $10n^{-1/2} \log n$ .

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

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

where  $\beta_0$  and  $\beta_1$  are constants. The binary instrument  $W$  point identifies  $\beta_0$  and  $\beta_1$ . 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  $\beta_1$  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).<sup>11</sup> 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. Models with exogenous covariates**

This section extends model (1a)–(1b) and (2) to include an exogenous, possibly multidimensional covariate,  $Z$ . The extended model is

$$Y = g(X, Z) + U \tag{17a}$$

and

$$E(U|W = w, Z = z) = 0 \tag{17b}$$

for almost every  $w$  and  $z$ . Equivalently,

$$E[Y - g(X, Z)|W = w, Z = z] = 0. \tag{18}$$

<sup>11</sup> Angrist's and Evans's (1998) data are available at <http://economics.mit.edu/faculty/angrist/data1/data/angev98>.

**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	$[-22.7, -7.1]$

The object of interest is the linear functional  $L_z(g) = \sum_{j=1}^J c_j g(x_j, z)$ , where  $x_j$  is a mass point of the discretely distributed random variable  $X$ ,  $z$  is a point in the support of the possibly continuously distributed random variable  $Z$ , and the  $c_j$ 's are constants. As before,  $W$  is discretely distributed.

If  $Z$  is a discretely distributed random variable, then the results for model (1a)–(1b) hold for model (17a)–(17b) after conditioning on  $Z$ . The remainder of this section provides an informal outline of a method for estimating bounds on  $L_z(g)$  in (17a)–(17b) under a shape restriction on  $g(\cdot, z)$  at a fixed  $z$ .

Let  $z \in \text{supp}(Z)$  be given, and let  $\{w_k : k = 1, \dots, K\}$  be the set of mass points of  $W$  for which  $P(W = w_k | Z = z) > 0$ . Define

$$\pi_{jk}(z) = \frac{\partial P(X = x_j, W = w_k, Z \leq z)}{\partial z}.$$

Assume that this derivative exists. Also define

$$\pi_k(z) = \sum_{j=1}^J \pi_{jk}(z)$$

and

$$m_k(z) = E(Y | W = w_k, Z = z) \pi_k(z).$$

Let  $\mathbf{m}(z) = [m_1(z), \dots, m_K(z)]'$ , and let  $\Pi(z)$  be the  $J \times K$  matrix whose  $(j, k)$  component is  $\pi_{jk}(z)$ . Assume that  $J > K$ . Define  $\mathbf{g}(z) = [g(x_1, z), \dots, g(x_J, z)]'$ . Then at  $Z = z$ , (17a)–(17b) and (18) are equivalent to

$$\Pi'(z)\mathbf{g}(z) = \mathbf{m}(z). \tag{19}$$

Moreover,  $L_z(g) = \mathbf{c}'\mathbf{g}(z)$ .

Proposition 1 shows that any value of  $L_z(g)$  in  $(-\infty, \infty)$  is consistent with (19) except in special cases. To achieve partial identification of  $L_z(g)$ , impose the shape restriction  $S\mathbf{g}(z) \leq 0$ . Under this restriction, identified upper and lower bounds on  $L_z(g)$  are the optimal values of the objective functions of the linear programming problems

$$\begin{aligned} &\text{maximize (minimize): } \mathbf{c}'\mathbf{h} \\ &\text{subject to: } \Pi'(z)\mathbf{h} = \mathbf{m}(z) \\ &\quad S\mathbf{h} \leq 0. \end{aligned} \tag{20}$$

Denote the upper and lower bounds by  $L_{z,\max}$  and  $L_{z,\min}$ , respectively. These bounds do not change if the constraints in (20) are assumed to hold at every  $z \in \text{supp}(Z)$ . It may be possible to tighten the bounds by imposing shape restrictions on  $g(x, z)$  considered as a function of  $z$  or by imposing bounds on partial derivatives of  $g(x, z)$  with respect to  $z$ , but we do not investigate these possibilities here.

The results of Section 2 apply to (19) after replacing  $\Pi$  and  $\mathbf{m}$ , respectively, with  $\Pi(z)$  and  $\mathbf{m}(z)$ . To estimate  $L_{z,\max}$  and  $L_{z,\min}$ , we need nonparametric estimators of  $\Pi(z)$  and  $\mathbf{m}(z)$ . We use kernel

estimators here, though other nonparametric estimators can be used. Let  $d = \text{dim}(Z)$ , and let  $s_n$  denote the bandwidth. The kernel estimators of  $\pi_{jk}(z)$  and  $m_k(z)$  are

$$\hat{\pi}_{jk}(z) = \frac{1}{ns_n^d} \sum_{i=1}^n I(X_i = x_j) I(W_i = w_k) K\left(\frac{Z_i - z}{s_n}\right)$$

and

$$\hat{m}_k(z) = \frac{1}{ns_n^d} \sum_{i=1}^n Y_i I(W_i = w_k) K\left(\frac{Z_i - z}{s_n}\right),$$

where  $K$  is the kernel function.  $K$  can be a kernel of any order, but higher-order kernels often give poor finite-sample estimates. Accordingly, we assume here that  $K$  is a second-order kernel. Assume that the second derivatives of  $\pi_{jk}(z)$  and  $m_k(z)$  are bounded and  $s_n \propto n^{-r}$  for some  $r > 1/(4+d)$ . Then  $(ns_n^d)^{1/2}[\hat{\pi}_{jk}(z) - \pi_{jk}(z)]$  and  $(ns_n^d)^{1/2}[\hat{m}_k(z) - m_k(z)]$  are asymptotically normally distributed with means of zero and finite variances (Li and Racine, 2006 among many other references). Moreover, the bootstrap consistently estimates the asymptotic distributions of  $(ns_n^d)^{1/2}[\hat{\pi}_{jk}(z) - \pi_{jk}(z)]$  and  $(ns_n^d)^{1/2}[\hat{m}_k(z) - m_k(z)]$  (Hall, 1992).

Now let  $\hat{\Pi}(z)$  be the  $J \times K$  matrix whose  $(j, k)$  element is  $\hat{\pi}_{jk}(z)$ . Let  $\hat{\mathbf{m}}(z)$  be the  $K \times 1$  vector whose  $k$ 'th component is  $\hat{m}_k(z)$ .  $L_{z,\max}$  and  $L_{z,\min}$  can be estimated consistently by solving the linear programming problems

$$\begin{aligned} &\text{maximize (minimize): } \mathbf{c}'\mathbf{h} \\ &\text{subject to: } \hat{\Pi}'(z)\mathbf{h} = \hat{\mathbf{m}}(z) \\ &\quad S\mathbf{h} \leq 0. \end{aligned} \tag{21}$$

The asymptotic distributional results of Sections 3.2–3.3 and the results of Section 4.1 on using the bootstrap to form confidence intervals for  $L_{z,\max}$ ,  $L_{z,\min}$ ,  $[L_{z,\min}, L_{z,\max}]$  and  $L_z(g)$  apply to (21) after replacing  $\Pi$  and  $\mathbf{m}$  with  $\Pi(z)$  and  $\mathbf{m}(z)$ ,  $\hat{\Pi}$  and  $\hat{\mathbf{m}}$  with  $\hat{\Pi}(z)$  and  $\hat{\mathbf{m}}(z)$ , and  $n^{1/2}$  with  $(ns_n^d)^{1/2}$ . As usually happens in nonparametric estimation, there is a curse of dimensionality as  $d$  increases. Investigation of the possibility of overcoming this problem through the use of semiparametric models is left to future research.

## 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.

**Acknowledgments**

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

**Appendix. Proofs of Theorems 2 and 3 and Proposition 3**

**Proof of Theorem 2.** First, consider  $\hat{L}_{\max}$ . 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  $\Pi$  and  $m$  with probability approaching 1 as  $n \rightarrow \infty$ .<sup>12</sup> Therefore, the probability that a basic solution is optimal in (10) but not (6) approaches zero as  $n \rightarrow \infty$ .

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,

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

Moreover, with probability approaching 1 as  $n \rightarrow \infty$ ,

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

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) and, with probability approaching 1 as  $n \rightarrow \infty$ , (10),

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

An application of the delta method yields

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

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  $\hat{m}$  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  $E(\hat{m}) = \bar{m}$ . Therefore, it follows from the Lindeberg–Levy and Cramér–Wold theorems that

<sup>12</sup> The components of  $\hat{m}$  and  $\hat{\Pi}$  are sample averages. Therefore, it follows from the law of the iterated logarithm that with probability approaching 1 as  $n \rightarrow \infty$ , the components of  $\hat{m} - \bar{m}$  and  $\hat{\Pi} - \Pi$  are within arbitrarily small-neighborhoods of zero.

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.<sup>13</sup> 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  $\tilde{Z}_k = -\bar{c}'_k \bar{A}_k^{-1} \bar{m}$  and  $\hat{Z}_k = -\bar{c}'_k \hat{A}_k^{-1} \hat{m}$ . Then arguments like those made for  $L_{\max}$  show that

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

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 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  $\Sigma_{\min}$ .

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  $[n^{1/2}(\hat{L}_{\max} - L_{\max}), n^{1/2}(\hat{L}_{\min} - L_{\min})]$ . □

**Proof of Theorem 3.** We prove part (i) and part (iv) for  $L_{\max}$ . The proofs of the remaining parts are similar.

(i) From (A.2),  $n^{1/2}(\hat{Z}_k - Z_k)$  is asymptotically a linear function of centered, scaled sample moments. Asymptotically,  $\Delta_{1k}^*$  is the same linear function of centered, scaled bootstrap sample moments. It follows from Theorem 1 of (Mammen, 1992) and the Cramér–Wold device that  $\Delta_{1k}^*$  consistently estimates the asymptotic distribution of  $n^{1/2}(\hat{Z}_k - Z_k)$  and that

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

This result and (A.1) imply that

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

Part (i) now follows from the observation that  $\hat{\mathcal{B}}_{\max} = \mathcal{B}_{\max}$  with probability approaching 1 as  $n \rightarrow \infty$ .

<sup>13</sup> If  $J$  and  $K$  increase as  $n \rightarrow \infty$ , then under mild regularity conditions,  $\hat{m}$  and  $\hat{A}_k$  are consistent for  $m$  and  $\bar{A}_k$ , respectively, and normalized, centered versions of  $\hat{m}$  and  $\hat{A}_k$  are asymptotically multivariate normally distributed. The rates of convergence of  $\hat{m}$  and  $\hat{A}_k$  are slower than  $n^{-1/2}$  and depend on the rates at which  $J$  and  $K$  increase. The conclusions of Theorem 2 hold if  $J$  and  $K$  increase slowly as  $n$  increases. The upper bounds on the allowed rates of increase depend on the details of the objective function. Derivation of the bounds is beyond the scope of this paper.

(iv) Define  $\Delta^* = \max_{k \in \hat{\mathcal{B}}_{\max}} \Delta_{1k}^*$ . It follows from (i) that

$$\sup_{-\infty < z < \infty} |P^*(\Delta^* \leq z) - P[n^{1/2}(\hat{L}_{\max} - L_{\max}) \leq z]| \rightarrow^p 0.$$

But  $P^*(\Delta^* \leq c_{\alpha, \max}^*) = 1 - \alpha$  by construction. Therefore,  $P[n^{1/2}(\hat{L}_{\max} - L_{\max}) \leq c_{\alpha, \max}^*] \rightarrow^p 0$ .  $\square$

The assumptions of Angrist and Imbens (1995). Assumptions 1 and 2 of Angrist and Imbens (1995) in the notation of Section 2.1 and with exogenous covariates are:

**Assumption 1 (Independence).** The random variables  $X$  and  $Y$  are jointly independent of  $W$ .

**Assumption 2 (Monotonicity).** With probability 1, either  $X_1 - X_0 \geq 0$  or  $X_1 - X_0 \leq 0$  for each person.

**Proof of Proposition 3.** Let  $g \in \mathbb{R}^J$  be any vector such that  $g_j - g_{j-1} = E(Y_j - Y_{j-1} | X^1 \geq x_j > X^0)$ . It follows from Theorem 1 of Angrist and Imbens (1995) that

$$\begin{aligned} E(Y|W = 1) - E(Y|W = 0) &= \sum_{j=2}^J E(Y_j - Y_{j-1} | X^1 \geq x_j > X^0) P(X^1 \geq x_j > X^0) \\ &= \sum_{j=2}^J (g_j - g_{j-1}) P(X^1 \geq x_j > X^0). \end{aligned}$$

Since  $P(X^1 > x_j \geq X^0) = P(X^1 \geq x_1 > X^0) = 0$ , we can write

$$\begin{aligned} &\sum_{j=2}^J (g_j - g_{j-1}) P(X^1 \geq x_j > X^0) \\ &= \sum_{j=2}^J g_j P(X^1 \geq x_j > X^0) - \sum_{j=2}^J g_{j-1} P(X^1 \geq x_j > X^0) \\ &= \sum_{j=1}^J g_j P(X^1 \geq x_j > X^0) - \sum_{j=1}^{J-1} g_j P(X^1 \geq x_{j+1} > X^0) \\ &= \sum_{j=1}^J g_j [P(X^1 \geq x_j > X^0) - P(X^1 > x_j \geq X^0)], \end{aligned}$$

where the last line follows from  $x_{j+1} > x_j$  and  $X^1, X^0 \in \{x_1, \dots, x_J\}$ .

Assumption 2 (monotonicity) of Angrist and Imbens (1995) implies that

$$\begin{aligned} &P(X^1 \geq x_j > X^0) - P(X^1 = x_j) \\ &= P(X^1 \geq x_j > X^0) - P(X^1 = x_j, x_j \geq X^0) \\ &= P(X^1 \geq x_j > X^0) - P(X^1 = x_j, x_j = X^0) \\ &\quad - P(X^1 = x_j, x_j > X^0) \\ &= P(X^1 > x_j > X^0) - P(X^1 = x_j, x_j = X^0). \end{aligned}$$

Similarly,

$$\begin{aligned} &P(X^1 > x_j \geq X^0) - P(X^0 = x_j) \\ &= P(X^1 > x_j > X^0) - P(X^1 = x_j, x_j = X^0). \end{aligned}$$

Therefore,  $P(X^1 \geq x_j > X^0) - P(X^1 > x_j \geq X^0) = P(X^1 = x_j) - P(X^0 = x_j)$ . It follows from Assumption 1 (independence) of Angrist and Imbens (1995) that

$$\begin{aligned} &P(X^1 = x_j) - P(X^0 = x_j) \\ &= P(X = x_j | W = 1) - P(X = x_j | W = 0). \end{aligned}$$

Therefore,

$$\begin{aligned} &E(Y|W = 1) - E(Y|W = 0) \\ &= \sum_{j=1}^J g_j [P(X = x_j | W = 1) - P(X = x_j | W = 0)] \end{aligned}$$

and

$$\begin{aligned} &E(Y|W = 1) - \sum_{j=1}^J g_j P(X = x_j | W = 1) \\ &= E(Y|W = 0) - \sum_{j=1}^J g_j P(X = x_j | W = 0). \end{aligned}$$

Now choose  $\{g_j^* : j = 1, \dots, J\}$  so that

$$\begin{aligned} &E(Y|W = 1) - \sum_{j=1}^J g_j^* P(X = x_j | W = 1) \\ &= E(Y|W = 0) - \sum_{j=1}^J g_j^* P(X = x_j | W = 0) = 0. \end{aligned}$$

This can always be done by adding a constant to each  $g_j$ . Then  $g_j^* - g_{j-1}^* = E(Y_j - Y_{j-1} | X^1 \geq x_j > X^0)$  and

$$E(Y|W = k) P(W = k) - \sum_{j=1}^J g_j^* P(X = x_j | W = k) P(W = k) = 0$$

for  $k = 1, 2$ . Therefore,  $\Pi g = m$ .  $\square$

## References

- Angrist, J.D., Evans, W.N., 1998. Children and their parents' labor supply: evidence from exogenous variation in family size. *Amer. Econ. Rev.* 88, 450–477.
- Angrist, J.D., Imbens, G.W., 1995. Two-state least squares estimation of average causal effects in models with variable treatment intensity. *J. Amer. Statist. Assoc.* 90, 431–442.
- Angrist, J.D., Krueger, A.B., 1991. Does compulsory school attendance affect schooling and earnings? *Quart. J. Econ.* 106, 979–1014.
- Bronars, S.G., Grogger, J., 1994. The economic consequences of unwed motherhood: using twins as a natural experiment. *Amer. Econ. Rev.* 84, 1141–1156.
- Card, D., 1995. Using geographic variation in college proximity to estimate returns to schooling. In: Christofides, L.N., Grant, E.K., Swidinsky, R. (Eds.), *Aspects of Labor Market Behaviour: Essays in Honour of John Vanderkamp*. University of Toronto Press, Toronto.
- Chernozhukov, V., Lee, S., Rosen, A.M., 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.
- Hall, P., 1992. *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, New York.
- Härdle, W., Hart, J.D., 1992. A bootstrap test for positive definiteness of income effect matrices. *Econometric Theory* 8, 276–290.
- Imbens, G.W., Manski, C.F., 2004. Confidence intervals for partially identified parameters. *Econometrica* 72, 1845–1857.
- Lehmann, E.L., Romano, J.P., 2005. *Testing Statistical Hypotheses*, third ed. Springer, New York.
- Li, Q., Racine, J.S., 2006. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, Princeton.
- Lochner, L., Moretti, E., 2004. The effect of education on crime: evidence from prison inmates, arrests, and self reports. *Amer. Econ. Rev.* 94, 155–189.
- Mammen, E., 1992. When Does Bootstrap Work? In: *Asymptotic Results and Simulations*. Springer, New York.
- Manski, C.F., Pepper, J.V., 2000. Monotone instrumental variables: With an application to returns to schooling. *Econometrica* 68, 997–1010.
- Manski, C.F., Pepper, J.V., 2009. More on monotone instrumental variables. *Econom. J.* 119, S200–S216.
- Moran, J.R., Simon, K.I., 2006. Income and the use of prescription drugs by the elderly: evidence from the notch cohorts. *J. Hum. Res.* 41, 411–432.
- Romano, J.P., Shaikh, A.M., Wolf, M., 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.