

Imprecise Identification from Incomplete Data

Joel L. Horowitz

Department of Economics,
University of Iowa
joel-horowitz@uiowa.edu

Charles F. Manski

Department of Economics and Institute for Policy
Research, Northwestern University
cfmanski@northwestern.edu

Abstract

An incomplete data problem arises when sample realizations are not fully observable: some realizations may be entirely or partially missing; some variables may be interval-measured. Whatever the specific form of the incomplete data problem, the generic consequence is imprecise identification of the population distribution generating the data. This paper describes completed and ongoing research showing how incomplete data problems lead to imprecise identification of regressions and of parameters solving extremum problems.

Keywords. Identification regions, interval data, missing data, nonparametric regression

1 Introduction

Inference from incomplete data is a commonplace problem of empirical research. Consider a survey researcher who draws a random sample of persons from a population of interest and queries each person about some variables w taking values in some space W . The objective is to infer the population probability distribution $P(w)$. A problem of incomplete data arises if the researcher does not fully observe the realizations of w .

Incomplete data problems take various forms in practice. Some realizations of w may be entirely missing, as when a person selected for participation in a survey refuses to be interviewed. Other realizations of w may be partly missing, as when a person agrees to be interviewed but then responds to only a subset of the questions posed. Or w may be measured only within intervals, as when survey respondents are asked to provide categorical responses to questions.

Whatever the specific form of the incomplete data problem, the generic consequence is imprecise identification of the population distribution generating the data. Incomplete observation of realizations of w generally enables one to learn that $P(w)$ is a member of a set of feasible distributions, say Ψ , all of which are consistent with the

available empirical evidence. Let $\tau[P(w)]$ denote a parameter of interest, perhaps the mean or median of w . The *identification region* for this parameter is $[\tau(\Psi), \Psi \in \Psi]$.

This paper describes completed and ongoing research on imprecise identification from incomplete data. Our perspective is deliberately conservative – we focus on “worst case” scenarios in which the researcher has no prior information on the distribution of interest or on the process preventing complete observation of sample realizations. Our “worst-case” analysis contrasts with the “best-case” approach that dominates the literature on inference from incomplete data. Consider in particular the problem of missing data. A common practice is to assume that data are missing completely at random (MCAR) and to perform analyses using only observations with complete data. Conventional methods for imputing missing data assume that missingness is random conditional on specified covariates. On occasion, a model of non-random missing data may be asserted. Either way, the identification problem is solved and efficiency of estimation becomes the central matter of concern to statisticians. We have emphasized in Manski (1995), Horowitz and Manski (1998, 2000) and elsewhere that it is not sufficient for empirical researchers to know the inferences that can be made if specified assumptions hold. It is also important to be able to characterize the inferences that may be made without imposing these assumptions. An especially appealing feature of conservative analysis is that it enables establishment of a domain of consensus among researchers who may hold disparate beliefs about what assumptions are appropriate.

Section 2 describes completed research on the nonparametric identification of regressions when outcome or covariate data are missing. Sections 3 through 5 describe ongoing research on the analysis of extremum problems from incomplete data. Section 3 presents the inferential problem in abstraction and poses central open questions. The complexity of the general problem makes it prudent to examine important special cases and to develop analytical approaches that provide at least partial results. In

this vein, Section 4 discusses the inferential problem when the space W is finite; this restriction somewhat simplifies matters. Section 5 introduces a set of minmax methods for computing relatively simple *outer identification regions*. These are regions which enclose the (sharp) identification regions that fully express what can be learned about the parameter of interest.

To keep attention focused on the core problem of identification created by incomplete data, this paper does not dwell on the routine problem of statistical induction from samples to populations. Much of the discussion below supposes that the researcher knows the values of population features which are identified by the sampling process. In practice, one may generally estimate such features consistently by their sample analogs.

The body of research described in this paper constitutes part of a growing modern literature on imprecise identification of probability distributions. A paper presented at ISIPTA 1 described another part of this literature – research on imprecise identification of distributions of treatment response (Manski, 2001). A non-technical exposition of a variety of findings is given in Manski (1995).

2 Regression with Missing Outcome or Covariate Data

In this section, $W \equiv Y \times X$, where Y is a bounded real outcome space and X is a covariate space. Each member j of a population J has a value $w_j \equiv (y_j, x_j) \in Y \times X$. The range of Y is normalized to be the unit interval $[0, 1]$. Let A be any measurable subset of X such that A is on the support of $P(x)$. The objective is to learn about the conditional expectation $E(y|x \in A)$. A random sample is drawn, but some data on (y, x) are missing.¹

Empirical researchers have to contend with general patterns of missing data. It is instructive, however, to consider three polar cases: missing outcomes (Section 2.1), jointly missing outcomes and covariates (Section 2.2), and missing covariates (Section 2.3). After examining these polar cases in detail, we briefly consider more general patterns of missing data (Section 2.4). See Horowitz and Manski (1998, 2000) for empirical applications of the findings described below.

¹ Boundedness of the outcome space Y is necessary if worst-case inference on $E(y|x \in A)$ in the presence of missing data is to yield informative conclusions. However, boundedness of Y is not necessary for inference on conditional probabilities of the form $P(y \in B|x \in A)$, where $B \subset Y$. The reason is that $P(y \in B|x \in A)$ is the conditional expectation of the indicator function $1[y \in B]$, which is bounded.

2.1 Missing Outcome Data

Manski (1989) analyzed the case in which x is always observed but data on y may be missing. Let $z = 1$ if (y, x) are observed, $z = 0$ if only x is observed. Use the law of iterated expectations to write

$$(1) \quad E(y|x \in A) = E(y|x \in A, z = 1) \cdot P(z = 1|x \in A) \\ + E(y|x \in A, z = 0) \cdot P(z = 0|x \in A).$$

The quantities $E(y|x \in A, z = 1)$ and $P(z|x \in A)$ are identified by the sampling process, but $E(y|x \in A, z = 0)$ is not. The last quantity must lie in the interval $[0, 1]$. Hence we obtain the identification region

$$(2) \quad E(y|x \in A) \in \\ [E(y|x \in A, z = 1) \cdot P(z = 1|x \in A), \\ E(y|x \in A, z = 1) \cdot P(z = 1|x \in A) + P(z = 0|x \in A)].$$

Observe that the width of this interval increases from zero to one as the response probability $P(z = 1|x \in A)$ falls from one to zero

2.2 Jointly Missing Outcome and Covariate Data

Horowitz and Manski (1998) analyzed the case in which some realizations (y, x) are entirely missing, the remainder being fully observed. Let $z = 1$ if (y, x) are observed, $z = 0$ otherwise. Use the law of iterated expectations (1) and Bayes Theorem to write

$$(3) \quad E(y|x \in A) = \\ E(y|x \in A, z = 1) \frac{\pi(A, 1)P(z = 1)}{\pi(A, 1)P(z = 1) + \pi(A, 0)P(z = 0)} \\ + E(y|x \in A, z = 0) \frac{\pi(A, 0)P(z = 0)}{\pi(A, 1)P(z = 1) + \pi(A, 0)P(z = 0)},$$

where $\pi(A, j) \equiv P(x \in A|z = j)$. The quantities $E(y|x \in A, z = 1)$, $\pi(A, 1)$, and $P(z)$ are identified by the sampling process but $E(y|x \in A, z = 0)$ and $\pi(A, 0)$ are not. The identification region for $E(y|x \in A)$ is obtained by evaluating (3) over all values $E(y|x \in A, z = 0) \in [0, 1]$ and $\pi(A, 0) \in [0, 1]$. The result is

$$(4) \quad E(y|x \in A) \in \\ [E(y|x \in A, z = 1) \cdot P_e(z = 1|x \in A), \\ E(y|x \in A, z = 1) \cdot P_e(z = 1|x \in A) + P_e(z = 0|x \in A)],$$

where

$$P_c(z = 1 | x \in A) \equiv \frac{\pi(A, 1) P(z = 1)}{\pi(A, 1) P(z = 1) + P(z = 0)}$$

is the *effective response probability*.

Region (4) has the same form as (2), except that the effective response probability $P_c(z = 1 | x \in A)$ replaces the unknown $P(z = 1 | x \in A)$. Observe that the width of the identification region increases from zero to one as either $\pi(A, 1)$ or $P(z = 1)$ falls from one to zero.

2.3 Missing Covariate Data

Horowitz and Manski (1998) analyzed the case in which y is always observed but data on x may be missing; now $z = 1$ if (y, x) are observed and $z = 0$ if only y is observed. This case is more complex than those discussed thus far.

To determine the identification region, reconsider the analysis of jointly missing outcome and covariate data in Section 2.2. There the available data constrained the right side of equation (3) by identifying $E(y | x \in A, z = 1)$, $\pi(A, 1)$, and $P(z)$. If only covariate data are missing, the data also identify the distribution $P(y | z = 0)$. Knowledge of this distribution jointly constrains $E(y | x \in A, z = 0)$ and $\pi(A, 0)$ through the equation

$$(5) \quad P(y | z = 0) = P(y | x \in A, z = 0) \cdot \pi(A, 0) \\ + P(y | x \in \bar{A}, z = 0) \cdot \pi(\bar{A}, 0),$$

where \bar{A} denotes the complement of A .

To determine the implications of (5), first let $p \in [0, 1]$ and suppose that $\pi(A, 0) = p$. Let Ψ denote the set of all distributions on Y . Then the values for the distribution $P(y | x \in A, z = 0)$ that are consistent with (5) are

$$(6) \quad \Psi(p) \equiv \Psi \cap \{[P(y | z = 0) - (1 - p)\psi]/p : \psi \in \Psi\}.$$

The implied set of feasible values for $E(y | x \in A, z = 0)$ is

$$(7) \quad E(y | x \in A, z = 0) \in [g_0(p), g_1(p)],$$

where $g_0(p) \equiv \inf [\int y d\psi, \psi \in \Psi(p)]$ and $g_1(p) \equiv \sup [\int y d\psi, \psi \in \Psi(p)]$. It can be shown that $g_0(p)$ and $g_1(p)$ are the means of two truncated versions of $P(y | z = 0)$, specifically the distributions formed from the left and right tails containing mass p (see Horowitz and Manski, 1995).

Combining (3) and (7) yields

$$(8) \quad E(y | x \in A) \in \\ \{E(y | x \in A, z = 1) \frac{pP(z = 1)}{pP(z = 1) + (1 - p)P(z = 0)} \\ + g_0(p) \frac{pP(z = 0)}{pP(z = 1) + (1 - p)P(z = 0)}, \\ E(y | x \in A, z = 1) \frac{pP(z = 1)}{pP(z = 1) + (1 - p)P(z = 0)} \\ + g_1(p) \frac{pP(z = 0)}{pP(z = 1) + (1 - p)P(z = 0)}\}.$$

If it were known that $\pi(A, 0) = p$, the right side of (8) would give the identification region for $E(y | x \in A)$. However, the available data place no constraint on $\pi(A, 0)$. Hence the identification region for $E(y | x \in A)$ is the union over $p \in [0, 1]$ of the intervals on the right side of (8).

In general, this region does not have a simple form comparable to those reported in Sections 2.1 and 2.2. However, one special case yields an exceedingly simple and surprising result. Suppose that $P(y | z = 0)$ is found to be degenerate, with all mass at $E(y | x \in A, z = 1)$. Then $g_0(p) = g_1(p) = E(y | x \in A, z = 1)$ for all $p \in (0, 1]$. Hence (8) reduces to $E(y | x \in A) = E(y | x \in A, z = 1)$ for all $p \in [0, 1]$. Thus $E(y | x \in A)$ may be precisely identified, even if x is never observed.

2.4 General Patterns of Missing Data

Succinct characterization of the identification region for $E(y | x \in A)$ is elusive in general settings where some sample realizations may have missing outcome data, others have missing covariate data, and still others have jointly missing outcomes and covariates. The literature to date contains two sets of findings.

Horowitz and Manski (2000) show that the identification region has a tractable closed form expression if y is a binary outcome variable and if covariate data, when missing, are entirely missing. (Thus, if x is a vector, it is presumed that each realization of x is either fully observed or that all components of x are missing.) This article also gives a closed-form expression for the feasible values of contrasts of the form $E(y | x \in A) - E(y | x \in B)$, where A and B are any two disjoint subsets of X . Analysis of such contrasts is subtle because a missing covariate realization cannot simultaneously lie in the sets A and B . Hence the identification region for $E(y | x \in A) - E(y | x \in B)$ is a proper subset of the region formed by considering all

feasible values of $E(y|x \in A)$ and all feasible values of $E(y|x \in B)$.

Zaffalon (2001) supposes that the set $Y \times X$ is finite and permits an arbitrary pattern of missing data. This extends the setting of Horowitz and Manski (2000) in two respects; y need not be binary and realizations of the x vector may be partly observed. The price paid for this generality is that a closed-form expression for the identification region does not emerge. However, computation of a sample analog estimate of the identification region is tractable. Zaffalon shows that estimates of the smallest and largest feasible values of $E(y|x \in A)$ can be obtained by solving fractional linear programming problems.

3 Extremum Problems with Incomplete Data

A very large part of formal empirical research uses sample data to estimate the value of a finite-dimensional parameter that minimizes the expectation of a random function. Leading special cases include estimation of best predictors and maximum likelihood estimation.

Let B be a finite-dimensional real parameter space. Let $h(\cdot, \cdot): W \times B \rightarrow \mathbb{R}^1$ be a specified function. Assume that $E[h(w, \cdot)]$ exists and has a minimum at some $b \in B$. Suppose that a random sample of realizations of w are drawn. The objective is to infer b .²

Empirical researchers routinely report estimates based only on those sample realizations that are completely observed. This practice is justified if the same population probability distribution generates the realizations that are completely and incompletely observed; however, it usually is not justified otherwise. Let $z = 1$ if a realization is completely observed and $z = 0$ otherwise. The standard practice is to estimate $b_1 \equiv \operatorname{argmin}_{c \in B} E[h(w, c) | z = 1]$ rather than the parameter of interest, namely $b \equiv \operatorname{argmin}_{c \in B} E[h(w, c)]$. Obviously, $b_1 = b$ if $P(w|z = 1) = P(w)$. However, b_1 usually differs from b otherwise.

Empirical researchers need to understand what can be learned about the parameter of interest when the researcher has no prior information on the distribution generating the realizations that are incompletely observed. This motivates our ongoing research that seeks to characterize and find tractable ways to estimate the identification region for b .

3.1 The Identification Region and its Sample Analog

To begin, we need an appropriately general description of

² The nonparametric regression problem of Section 2 is a special case. Let $W = Y \times X$, $B = \mathbb{R}^1$, and $A \subset X$. Let $h[(y, x), c] = 1[x \in a](y - c)^2$. Then $b = E(y|x \in a)$.

an incomplete data problem, one that embraces both missing data problems and interval measurement problems. This is accomplished by supposing that precise realizations of w may not be observable, but sets containing w are observable. Thus, let each member j of the population J be characterized by a value $w_j \in \omega_j \subset W$. A random sample of size N is drawn from the population and the researcher observes the set-valued realizations $(\omega_i, i = 1, \dots, N)$. When a realization of ω contains a single value, the researcher has complete data on w . When ω contains multiple values, ω constitutes *incomplete data* on w .

Supposing that (w, ω) is measurable, let $P(\omega)$ denote the population distribution of the sets ω and let $P(w|\omega)$ be the distribution of w among persons who have observable characteristics ω . Consider the expectation function

$$(9) E[h(w, \cdot)] \equiv \int_{\omega} [\int_w h(w, \cdot) dP(w|\omega)] dP(\omega).$$

The sampling process identifies the distribution $P(\omega)$ but reveals nothing about the conditional distributions $P(w|\omega)$. Let $\Psi(\omega)$ denote the set of all distributions with support ω . Then the identification region for b is

$$(10) B_1 \equiv \left\{ \operatorname{argmin}_{c \in B} \int [\int h(w, c) d\psi(w|\omega)] dP(\omega), \right. \\ \left. \psi(w|\omega) \in \Psi(\omega), \omega \in \Omega \right\},$$

where Ω is the collection of all measurable subsets of W . The natural estimate of B_1 is its sample analog

$$(11) B_{1N} \equiv \left\{ \operatorname{argmin}_{c \in B} \frac{1}{N} \sum_{i=1}^N \int h(w, c) d\psi(w|\omega_i), \right. \\ \left. \psi(w|\omega_i) \in \Psi(\omega_i), i = 1, \dots, N. \right.$$

3.2 Central Questions

The above provide a notationally compact description of a very broad range of incomplete data problems. However, the abstraction of the description works against constructive analysis. It seems inevitable that to make progress, attention must be focused on suitably circumscribed classes of functions $h(\cdot, \cdot)$ and/or suitably restricted forms of the sets ω that define the incomplete data problem. This done, the central questions that we would like to address include these:

Characterization of B_1 : Is the identification region B_1 a proper subset of B , and hence informative? When B_1 is informative, what is its geometry?

Computation of B_{1N} : Does the estimate B_{1N} have a tractable closed-form? If not, is numerical computation feasible?

Statistical Inference: What is the sampling behavior of B_{1N} as an estimate of B_1 ?

Prospects for Tractable Partial Analysis: If it is intractable to characterize B_1 and compute B_{1N} , are there

tractable approaches that provide useful partial results?

Sections 4 and 5 report progress in addressing these questions. Also see Manski and Tamer (2001) for analysis of inference on regressions in the presence of interval data.

4 Inference when W is Finite

If W is finite, then each subset ω is finite and $\Psi(\omega)$ is the collection of multinomial distributions placing all mass on ω . Hence $\Psi(w|\omega) \in \Psi(\omega)$ is a point on the $|W|$ -dimensional simplex. This simplifies matters somewhat.

Let $\theta(\cdot): B \rightarrow \mathbb{R}^1$ be a specified function mapping the parameter space into the real line. The identification region for $\theta(b)$ is $[\theta(c), c \in B_0]$. Suppose that one wants to learn $\inf[\theta(c), c \in B_0]$. The sample analog is $\inf[\theta(c), c \in B_{0N}]$. If W is finite, $\inf[\theta(c), c \in B_{0N}]$ can be determined by solving a two-stage, finite-dimensional extremum problem whose second stage is a nonlinear programming problem. Solution of this problem is tractable in some cases of empirical interest.

Consider, for example, the best linear predictor (BLP) under square loss of a real outcome y given a d -dimensional covariate vector x . The BLP has the form $x'b$, where $b \equiv [E(xx')]^{-1}E(xy)$ is the solution to the familiar least squares extremum problem. A common problem in empirical research is to learn the BLP when some observations of y and/or x are missing or otherwise incomplete. In ongoing work we have found that, provided d is not too large, it is tractable to compute the smallest and largest values of $x'b$ that are consistent with the available data.

5 Minmax Outer Identification Regions

This section develops simple methods for partial inference on b when data are incomplete. The objective is to develop methods that may be implemented routinely with at most minor enhancements to standard statistical software. The methods described here achieve their simplicity in two ways. First, we treat incompletely observed realizations as entirely missing. Second, we use the data to determine *minmax outer identification regions* for b ; that is, regions which enclose B_1 . Thus, the methods examined here achieve simplicity by exploiting only part of the information available to the researcher.

5.1 Three Minmax Regions

Let $z = 1$ if w is completely observed and $z = 0$ otherwise. For $c \in B$, let $S_1(c) \equiv E[h(w, c)|z = 1]$ and $S_0(c) \equiv E[h(w, c)|z = 0]$. Then

$$(12) E[h(w, c)] = S_1(c) \cdot P(z = 1) + S_0(c) \cdot P(z = 0).$$

Let $h_L(c) \equiv \inf_{w \in W} h(w, c)$. Then

$$(13) E[h(w, c)] \geq S_1(c) \cdot P(z = 1) + h_L(c) \cdot P(z = 0).$$

Let $b_1 \in \operatorname{argmin}_{c \in B} S_1(c)$ and $h_U(b_1) \equiv \sup_{w \in W} h(w, b_1)$. Then

$$(14) E[h(w, b_1)] \leq S_1(b_1) \cdot P(z = 1) + h_U(b_1) \cdot P(z = 0).$$

Hence c cannot minimize $E[h(w, \cdot)]$ if

$$(15) S_1(c) \cdot P(z = 1) + h_L(c) \cdot P(z = 0) > S_1(b_1) \cdot P(z = 1) + h_U(b_1) \cdot P(z = 0).$$

This yields the *minmax outer identification region*

$$(16) B_m \equiv \{c \in B: S_1(c) - S_1(b_1) \leq [h_U(b_1) - h_L(c)] \cdot P(z = 0) / P(z = 1)\}.$$

Region B_m is *potentially informative* if $h(\cdot, c)$ is bounded on W for each $c \in B$. It is not necessary that $h(\cdot, \cdot)$ be uniformly bounded, nor that W be a finite set. When B_m is potentially informative, the size of the region depends on the specifics of the problem. However we can conclude that, in general, $B_m \rightarrow b$ as $P(z = 0) \rightarrow 0$.

Region B_m is a subset of an outer identification region reported in Manski (1994). There attention was restricted to cases in which $h(\cdot, \cdot)$ is uniformly bounded in a finite interval $[K_L, K_U]$. Applying the same reasoning as in (12) through (16) yields the region

$$(17) B_{m0} \equiv \{c \in B: S_1(c) - S_1(b_1) \leq (K_U - K_L) \cdot P(z = 0) / P(z = 1)\}.$$

Yet another outer region may be obtained if $h(\cdot, \cdot)$ is uniformly bounded from below but not necessarily from above. This is

$$(18) B_{m1} \equiv \{c \in B: S_1(c) - S_1(b_1) \leq [h_U(b_1) - K_L] \cdot P(z = 0) / P(z = 1)\}.$$

Clearly $B_m \subset B_{m1} \subset B_{m0}$. However, the three regions have the reverse ranking computationally, with B_{m0} being easiest to compute, followed by B_{m1} .

5.2. Minmax Estimation Regions

Now consider the sample analogs of the regions derived above. Let $N(1)$ denote the sub-sample of observations with $z = 1$ and let $N_1 \equiv |N(1)|$. Let N_0 be the number of observations with missing data and $N \equiv N_1 + N_0$. Then the sample analog of B_m is

$$(19) B_{mN} \equiv \{c \in B: S_{1N}(c) - S_{1N}(b_1) \leq [h_U(b_{1N}) - h_L(c)] \cdot N_0 / N_1\},$$

where $S_{1N}(c) \equiv (1/N_1) \sum_{i \in N(1)} h(w_i, c)$ and where $b_{1N} \in \operatorname{argmin}_{c \in B} S_{1N}(c)$. Sample analogs of B_{m0} and B_{m1} may be formed analogously.

The empirical requirements for computation of minmax estimation regions are access to the sub-sample $N(1)$ of complete data and knowledge of the number N_0 of observations with incomplete data. Computation of B_{mN} requires solution of these extremum problems:

- (i) $\min_{c \in B} S_{1N}(c)$
- (ii) $\sup_{w \in W} h(w, b_{1N})$
- (iii) $\inf_{w \in W} h(w, c), c \in B.$

Problem (i) is standard, it being the extremum problem solved by researchers who assume that $P(w|z=1) = P(w)$ and who consequently choose to discard incomplete data. The complexity of problems (ii) and (iii) depends on the form of $h(\cdot, \cdot)$ and W . Problem (ii) needs to be solved only once, at b_{1N} , and so will rarely pose a difficulty in applications. Problem (iii) must be solved at each $c \in B$, and so may more often be an issue. When solution of (iii) is problematic, an empirical researcher may prefer to compute the region B_{m1N} , which requires only solution of problems (i) and (ii).

5.3 Best Linear Prediction Under Square Loss

Computation of B_{m1N} is particularly simple in the familiar problem of best linear prediction under square loss. Let $W = Y \times X$ and $h(w, c) = (y - x'c)^2$, with Y a bounded subset of R^1 , X a rectangular bounded subset of R^K , and $B \subset R^K$. In problems of this form,

$$(20) S_1(c) - S_1(b_1) = (c - b_1)'E(xx'|z=1)(c - b_1).$$

The solution to problem (ii) occurs at an extreme point of $Y \times X$, namely

$$(21) h_U(b_1) = \max \{ [y_U - \sum_j b_{1j} \cdot (x_{Lj} \cdot 1[b_{1j} > 0] + x_{Uj} \cdot 1[b_{1j} < 0])]^2, [y_L - \sum_j b_{1j} \cdot (x_{Lj} \cdot 1[b_{1j} < 0] + x_{Uj} \cdot 1[b_{1j} > 0])]^2 \}.$$

where $y_L \equiv \inf(Y)$, $y_U \equiv \sup(Y)$, $x_L \equiv \inf(X)$, and $x_U \equiv \sup(X)$. A uniform lower bound on $(y - x'c)^2$ is provided by the value $K_L = 0$. Hence the minmax region B_{m1} is an ellipse centered on b_1 , namely

$$(22) B_{m1} = \{c \in R^K: (c - b_1)'E(xx'|z=1)(c - b_1) \leq \frac{1}{2} h_U(b_1) \cdot P(z=0)/P(z=1)\}.$$

Acknowledgments

The research of Joel Horowitz was supported in part by U. S. National Science Foundation grant SBR-9910925. The research of Charles Manski was supported in part by U. S. National Science Foundation grant SES-0001436.

References

- [1] Horowitz, J. and C. Manski (1995), "Identification and Robustness with Contaminated and Corrupted Data," Econometrica, 63, 281-302.
- [2] Horowitz, J. and C. Manski (1998), "Censoring of Outcomes and Regressors due to Survey Nonresponse: Identification and Estimation Using Weights and Imputations," Journal of Econometrics, 84, 37-58.
- [3] Horowitz, J., and C. Manski (2000), "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data," Journal of the American Statistical Association, 95, 77-84.
- [4] Manski, C. (1989), "Anatomy of the Selection Problem," Journal of Human Resources, 24, 343-360.
- [5] Manski, C. (1994), "The Selection Problem," in C. Sims (Ed.), Advances in Econometrics, Sixth World Congress, Cambridge: Cambridge University Press.
- [6] Manski, C. (1995), Identification Problems in the Social Sciences, Cambridge, Mass.: Harvard University Press.
- [7] Manski, C. (2001), "Treatment Choice under Ambiguity Induced by Inferential Problems," Journal of Statistical Planning and Inference, forthcoming.
- [8] Manski, C. and E. Tamer (2001), "Inference on Regressions with Interval Data on a Regressor or Outcome," Econometrica, forthcoming.
- [9] Zaffalon, M. (2001), "Exact Credal Treatment of Missing Data," Journal of Statistical Planning and Inference, forthcoming.