SOME ASYMPTOTIC RESULTS FOR BOUNDS ESTIMATION

Robert P. Sherman *California Institute of Technology*

January 17, 2002

Abstract

This paper undertakes a nonparametric analysis of mixture models with verification. These are natural models of contaminated or corrupted data, a fraction of which is verified to be from the distribution of interest. We derive sharp bounds on population characteristics such as conditional expectations, probabilities, and quantiles. Sharp bounds under additional exclusion or monotonicity restrictions are also derived. Asymptotic properties of sample analogues of the bounds are established. Useful convexity and concavity results are proved. In addition, we show that models of survey nonresponse and treatment effects can be viewed as mixture models with verification. Finally, we apply the estimation procedures to measurements of organic pollutant concentrations in the Love Canal.

1. Introduction

Empirical researchers are often faced with a situation in which a flawed data generating mechanism produces data that are not representative of the population of interest. Often, such data can be viewed as observations from a mixture model. In such a model, each observation is selected from either a distribution of interest, say F, or another, potentially spurious, distribution. If selection is independent of the value drawn from F, the data are said to be contaminated. Otherwise, the data are said to be corrupted, a more serious deficiency. In either case, without making untestable assumptions about the data generating process, it is not possible to identify characteristics of F such as moments, probabilities, and quantiles. However, given a lower bound on the probability of selection from F, Horowitz and Manski (1995) identify sharp bounds on these characteristics under either contaminated or corrupted sampling. They also show how to nonparametrically estimate these bounds.

Sometimes there is more information than simply a lower bound on the probability of selection from F. Sometimes a subset of the sample is verified to be drawn from F. We say that data generated in this way come from a mixture model with verification. If the verified subset is a random sample from F, then characteristics of F can be consistently estimated using only verified observations. If not, as is typically the case, then characteristics of F can be bounded using the procedure of Horowitz and Manski (1995) with the probability of verification serving as a natural lower bound on the probability of selection from F. However, the Horowitz and Manski bounds are not sharp because they do not fully exploit information on verification. In this paper, we derive and nonparametrically estimate sharp bounds on characteristics of F under contaminated or corrupted sampling with verification.

Mixture models with verification apply to a wide range of interesting data problems. For example, consider data analyzed by Lambert and Tierney (1997) on organic pollution concentrations in the Love Canal. Instruments measuring concentrations sometimes isolate the wrong pollutant. An imperfect verification test positively identifies a fraction of the measurements to be on the pollutant of interest. The unverified measurements may or may not be on the pollutant of interest. As another example, consider California's Stanford-9 test scores used to evaluate individual, regional, and state educational performance. This test is administered in English to all students. Students classified as proficient in English are presumed to provide valid test scores. Their scores can be viewed as verified draws from the distribution of valid scores. Those classified as limited English speakers (roughly one-fourth of the test takers) may or may not provide valid scores.

Models of survey nonresponse and treatment effects can also be viewed as mixture models with verification. For example, consider the item nonresponse problems analyzed by Horowitz and Manski (1998). Fully observed data vectors correspond to verified draws from the distribution of interest. Vectors with missing components may or may not be from the distribution of interest. In the randomized treatment effects models analyzed by Horowitz and Manski (2000) and Molinari (2001), treatments are not always observed. Outcomes for which a given treatment is observed can be viewed as verified draws from the outcome distribution for that treatment. Outcomes for which a treatment is missing may be from the outcome distribution for either treatment.

We present the mixture model with verification in a regression setting. That is, we allow F to depend on a vector of covariates. We construct sample analogs of the population bounds and show that they converge to their population counterparts at rate $n^{-2/[d+4]}$, $d \geq 0$, where n is the sample size and d is the number of continuous covariates. We derive the limiting distribution of the estimated bounds under corruption with verification. Under contamination with verification, we introduce a parametrization that induces convexity and concavity in functions defining not only the population bounds, but also the sample bounds. Convexity and concavity in the sample functions can be exploited to significantly reduce computations when n is large, when the bounds must be calculated for many covariate values, or when it is desirable to bootstrap the distribution of the sample bounds.

In addition to identifying sharp bounds under basic contamination and corruption with verification, we also derive sharp bounds under additional exclusion and monotonicity restrictions. As we will show, these additional restrictions are plausible in certain applications, are easy to incorporate into the estimation procedure, and can considerably tighten the bounds. In addition, we show that response censoring, which occurs in the Love Canal data we analyze, can be easily incorporated into the estimation procedure. This degree of flexibility compares favorably with that of the nonparametric maximum likelihood-type estimation procedure developed by Lambert and Tierney (1997) for estimating bounds on functionals of F under contamination with verification. Not only is the latter procedure more complicated than the corresponding procedure presented here, but it does not naturally extend to allow response censoring, continuous regressors, or incorporation of a plausible monotonicity restriction.

The rest of the paper is organized as follows. In Section 2, we formally define the mixture model with verification, and present and discuss the assumptions under which we derive sharp bounds on various characteristics of F. For completeness, we also discuss sufficient conditions for point identification of characteristics of F, without making parametric assumptions about the data generating process. In Section 3, we derive sharp bounds on characteristics of F under both corruption and contamination with verification. We also incorporate exclusion and monotonicity restrictions. Section 4 defines sample analogs of the population bounds derived in Section 3 and establishes some of their asymptotic properties. We also establish convexity and concavity of the functions defining the contamination with verification bounds and discuss the consequent computational benefits. In Section 5, we show that survey nonresponse models and randomized treatment effects models are special cases of mixture models with verification. Section 6 illustrates the various estimation procedures using the Love Canal data. Section 7 summarizes and gives directions for future work. Proofs of asymptotic results are given in an appendix.

2. Mixture Models with Verification

In this section, we formally define a mixture model with verification, and state and discuss assumptions that characterize interesting submodels for which we derive sharp bounds on various functionals of the distribution of interest, such as conditional expectations, quantiles, and probabilities. For completeness, we also discuss sufficient conditions for point identification of these functionals without making parametric assumptions about the data generating process.

Let Y_1 denote an outcome of interest and X a vector of covariates on which Y_1 depends. For example, in the Love Canal data analyzed in Section 6, Y_1 denotes certain organic pollution concentrations and X denotes the lab taking the measurements. Let $P_x(Y_1)$ denote the conditional distribution of Y_1 given $X = x$. This is the distribution of interest.

What characterizes a mixture model is that the researcher may or may not observe Y_1. Rather, the researcher observes Y, a mixture of Y_1 and another random variable Y_0, where $P_x(Y_0)$, the distribution of Y_0 given $X = x$, may be quite different from $P_x(Y_1)$. Let Z denote a selection indicator. If $Z = 1$ then Y is drawn from $P_x(Y_1)$. If $Z = 0$, then Y is drawn from $P_x(Y_0)$. That is,

$$Y = Y_1 Z + Y_0(1 - Z)\,. \tag{1}$$

Finally, let V denote an observed verification indicator. If $V = 1$, then Y is verified to be drawn from $P_x(Y_1)$. If $V = 0$, then Y may be drawn from either $P_x(Y_1)$ or $P_x(Y_0)$. In the Love Canal example, an imperfect verification test positively identifies a subset of the measurements to be from $P_x(Y_1)$. The other measurements may or may not be from $P_x(Y_1)$.

In sum, in a mixture model with verification, each member of the sampling distribution is characterized by a vector (Y, X, V, Z, Y_1, Y_0), where Y and V are observed, and X and Z may or may not be fully observed. For example, in the Love Canal data, X is observed while Z is only partially observed. In the survey nonresponse models analyzed by Horowitz and Manski (1998), either X or Z may be fully or partially observed. In the treatments effects models analyzed by Horowitz and Manski (2000) and Molinari (2001), X is observed and Z is partially observed. Our objective is to make inferences about functionals of $P_x(Y_1)$ when data are generated from a mixture model with verification. These functionals include conditional expectations, probabilities, and quantiles.

Let M denote a known, real-valued function on the support of Y_1. Partition X into components X_1 and X_2. That is, $X = (X_1, X_2)$. For ease of notation, throughout we write $\mathbb{P}_x$ for probability conditional on $X = x$ and $\mathbb{E}_x$ for expectation conditional on $X = x$. For example, for any set S in the support of Y_1, $\mathbb{P}_x S = \mathbb{P}[S \mid X = x]$. Similarly, $\mathbb{E}_x Y_1 = \mathbb{E}[Y_1 \mid X = x]$ and $\mathbb{E}_x[Y \mid V = 1] = \mathbb{E}[Y \mid V = 1, X = x]$.

We shall select from the following assumptions to define interesting submodels of the basic mixture model with verification.

A1. $\mathbb{P}_x\{V = 1\} > 0$.

A2. $\mathbb{P}_x\{Z = 1 \mid V = 1\} = 1$.

A3. $\mathbb{E}_x M(Y_1) = \mathbb{E}_x[M(Y_1) \mid Z = 1]$.

A4. $\mathbb{E}_x[M(Y_1) \mid V = 1] \geq \mathbb{E}_x[M(Y_1) \mid V = 0]$.

A5. Y_1 is independent of X_1 given X_2.

A6. Z is independent of X_1 given X_2.

A7. V is independent of Y_1 given X.

Assumption A1 says that the conditional probability of verification is positive, and distinguishes mixture models with verification from the mixture models studied by Horowitz and Manski (1995), where $\mathbb{P}_x\{V = 1\}$ can be zero. In order to construct their bounds on functionals of $P_x(Y_1)$, a lower bound on $\mathbb{P}_x\{Z = 1\}$ must be available. By A2, all the verified data are from $P_x(Y_1)$. It follows that $\mathbb{P}_x\{V = 1\}$ provides a lower bound on $\mathbb{P}_x\{Z = 1\}$, making construction of the Horowitz and Manski (1995) bounds possible. However, as we will show, tighter bounds can be constructed by fully exploiting the information on verification. Assumptions A1 and A2 characterize the least restrictive corruption with verification model that we study in this paper.

Assumptions A1, A2, and A3 characterize the least restrictive contamination with verification model that we study here. Notice that the mean independence assumption, A3, follows from the following stronger independence assumption:

B3. Z is independent of Y_1 given X.

Assumption B3 says that selection is independent of the value drawn from $P_x(Y_1)$. This assumption underlies the standard mixture models in the statistics literature. We shall assume A3 instead of B3 not only because A3 is weaker than B3 and is the essential condition driving the contamination with verification bounds, but also because A3 holds for the censored regressors model considered by Horowitz and Manski (1998), whereas B3 does not.

Assumptions A4 and A5 are restrictions that can be placed on either the basic corruption with verification model or the basic contamination with verification model to tighten bounds on functionals of $P_x(Y_1)$. They can be applied either together or separately, as appropriate. Assumption A6 can only tighten bounds in a contamination with verification model.

If M is increasing, then the monotonicity restriction A4 follows from the following stronger primitive assumption:

B4. $\mathbb{P}_x\{V = 1 \mid Y_1 = y\}$ is increasing in y.

Lemma 1A in the appendix establishes that B4 implies A4 when M is increasing. Assumption B4 says that the probability of verification is increasing in the value of the draw from $P_x(Y_1)$, and can be a plausible assumption to make in certain applications. For example, in the Love Canal data, it is natural to assume that it is harder to verify small concentrations of the pollutant of interest than it is to verify large concentrations. Similarly, in the survey nonresponse context, it may be natural to assume that the probability of responding is an increasing function of the value of the outcome of interest. Assumption A4 will be used to lower upper bounds on functionals of interest. In some applications, it may be natural to assume that the inequality in A4 is reversed. This would hold, for example, if $\mathbb{P}_x\{V = 1 \mid Y_1 = y\}$ in B4 were decreasing in y. In this case, one can use the corresponding conditions to raise lower bounds on functionals of interest, as we will show.

Assumptions A5 and A6 are exclusion restrictions that can be used to tighten bounds when certain types of instrumental variables are available. A5 has been used in another context by Manski (1990, 1995) to tighten bounds. A6 is plausible for the Love Canal data provided the ability to isolate the pollutant of interest is the same across labs.

Finally, when assumption A7 is added to assumptions A1 and A2, one can point identify functionals of $P_x(Y_1)$ using only verified data. Assumption A2 says that all verified data are sampled from $P_x(Y_1)$, but does not guarantee that sampling is random. Adding A7 ensures that verified observations are a random sample from $P_x(Y_1)$. Assumptions A1, A2, and A7 are implicitly made by researchers who discard unverified data.

3. Sharp Bounds

In this section, we derive sharp bounds on various functionals of $P_x(Y_1)$ for various mixture models with verification. Section 3.1 addresses corruption with verification models. Section 3.2 addresses contamination with verification models.

3.1 Corruption with Verification Models

Consider first the basic corruption with verification model characterized by assumptions A1 and A2. Let M be a known, real-valued function on $\mathbb{R}$, and suppose the support of $M(Y_1)$ is contained in the interval (a, b), where a and b are known and satisfy $-\infty \leq a \leq b \leq \infty$. Write v_x for $\mathbb{P}_x\{V = 1\}$.

THEOREM 1. *If A1 and A2 hold, then*

$$\lambda_1(x) \leq \mathbb{E}_x M(Y_1) \leq u_1(x)$$

where

$$\begin{aligned} \lambda_1(x) &= \mathbb{E}_x[M(Y) \mid V = 1]v_x + a(1 - v_x) \\ u_1(x) &= \mathbb{E}_x[M(Y) \mid V = 1]v_x + b(1 - v_x)\,. \end{aligned}$$

Moreover, these bounds are sharp.

PROOF. By A1 and A2, $\mathbb{E}_x[M(Y_1) \mid V = 1] = \mathbb{E}_x[M(Y) \mid V = 1]$. It follows that

$$\mathbb{E}_x M(Y_1) = \mathbb{E}_x[M(Y) \mid V = 1]v_x + \mathbb{E}_x[M(Y_1) \mid V = 0](1 - v_x)\,.$$

Fix $s \in (a, b)$ and suppose that given $V = 0$ and $X = x$, the distribution of $M(Y_1)$ puts unit mass at s. This supposition is consistent with A1, A2, and the assumptions about the support of $M(Y_1)$. Moreover, under this supposition,

$$\mathbb{E}_x M(Y_1) = \mathbb{E}_x[M(Y) \mid V = 1]v_x + s(1 - v_x)\,.$$

Since $s \in (a, b)$ is arbitrary, it follows that any bound on $\mathbb{E}_x M(Y_1)$ that is consistent with the assumptions stated above must contain $[\lambda_1(x), u_1(x)]$. This proves the result. $\square$

REMARK 1. Choosing $M(s) = s^k$ in Theorem 1 produces sharp bounds on arbitrary moments of $P_x(Y_1)$. Choosing $M(s) = \{s \le t\}$ for $t \in I\!\!R$ produces sharp bounds on the conditional probabilities $I\!\!P\{Y_1 \le t \mid X = x\}$, $t \in I\!\!R$, which can be inverted to get sharp bounds on the quantiles of $P_x(Y_1)$.

Next, we consider adding the monotonicity restriction A4 to the basic corruption with verification model. This restriction can have substantial identifying power in applications where there is a known, finite lower bound on the support of $M(Y_1)$, but no known upper bound, as is the case with the Love Canal data.

THEOREM 2. *If A1, A2, and A4 hold, then*

$$\lambda_2(x) \le I\!\!E_x M(Y_1) \le u_2(x)$$

where

$$\begin{aligned} \lambda_2(x) &= \lambda_1(x) \\ u_2(x) &= I\!\!E_x[M(Y) \mid V = 1]\,. \end{aligned}$$

Moreover, these bounds are sharp.

PROOF. Apply A4 to get

$$\begin{aligned} I\!\!E_x M(Y_1) &= I\!\!E_x[M(Y_1) \mid V = 1]v_x + I\!\!E_x[M(Y_1) \mid V = 0](1 - v_x) \\ &\le I\!\!E_x[M(Y_1) \mid V = 1]\,. \end{aligned}$$

By A1 and A2, $I\!\!E_x[M(Y_1) \mid V = 1] = I\!\!E_x[M(Y) \mid V = 1]$. This yields the upper bound. Now argue as in the proof of Theorem 1, letting $\mathbb{E}_x[M(Y) \mid V = 1]$ play the role of b to get the sharpness result. □

REMARK 2. If the sense of the inequality in A4 is reversed, then $\lambda_2(x) = \mathbb{E}_x[M(Y) \mid V = 1]$ and $u_2(x) = u_1(x)$ are sharp bounds on $\mathbb{E}_x M(Y_1)$.

Next, we consider adding assumption A5 to the basic corruption with verification model. Recall that $X = (X_1, X_2)$. Let Ω_i denote the support of X_i, $i = 1, 2$. Note that when A5 holds, $\mathbb{E}_x M(Y_1) = \mathbb{E}_{x_2} M(Y_1)$, where $\mathbb{E}_{x_2}$ denotes expectation conditional on $X_2 = x_2$. In Theorem 3 below, as well as in Theorem 6 and Theorem 7 in Section 3.2, we define extreme value estimators by taking the infimum and supremum over $x_1 \in \Omega_1$ of various functions of the vector $x = (x_1, x_2)$. For ease of notation, we will write a function $f(x)$ as $f(x_1, x_2)$, rather than use the formally correct, but more cumbersome expression, $f((x_1, x_2))$.

THEOREM 3. *If A1, A2, and A5 hold, then for each $x_2 \in \Omega_2$,*

$$\lambda_3(x_2) \leq \mathbb{E}_{x_2} M(Y_1) \leq u_3(x_2)$$

where

$$\begin{aligned} \lambda_3(x_2) &= \sup_{x_1 \in \Omega_1} \lambda_1(x_1, x_2) \\ u_3(x_2) &= \inf_{x_1 \in \Omega_1} u_1(x_1, x_2) . \end{aligned}$$

Moreover, these bounds are sharp.

PROOF. Fix $x_1 \in \Omega_1$ and $x_2 \in \Omega_2$. By Theorem 1, the interval $[\lambda_1(x_1, x_2), u_1(x_1, x_2)]$ contains

$\mathbb{E}[M(Y_1) \mid X_1 = x_1, X_2 = x_2]$. By A5, for each $x_1 \in \Omega_1$,

$$\mathbb{E}[M(Y_1) \mid X_1 = x_1, X_2 = x_2] = \mathbb{E}[M(Y_1) \mid X_2 = x_2] .$$

Intersect the intervals above over $x_1 \in \Omega_1$ to get the result. □

REMARK 3. If both A4 and A5 hold, and M is an increasing function, then tighter sharp bounds can be obtained by combining the bounds in Theorem 2 and Theorem 3 in the obvious way.

3.2 CONTAMINATION WITH VERIFICATION MODELS

We now consider contamination with verification models. We begin with the basic model characterized by assumptions A1, A2, and A3, and assume that the function M is increasing. Our objective is to develop sharp bounds on $\mathbb{E}_x M(Y_1)$ under these assumptions. In what follows, we apply a certain result of Horowitz and Manski (1995) which requires that the distribution of Y given $V = 0$ and $X = x$ be continuous. Therefore, we maintain this assumption throughout this subsection. However, as we show in Section 4.3, there is no loss of generality in making this assumption, since any discrete random variable can be written as a discrete transformation of an analogous continuous random variable.

By A3, $\mathbb{E}_x M(Y_1) = \mathbb{E}_x[M(Y_1) \mid Z = 1] = \mathbb{E}_x[M(Y) \mid Z = 1]$. It follows from A1 and A2 that $\mathbb{E}_x[M(Y) \mid Z = 1, V = 1] = \mathbb{E}_x[M(Y) \mid V = 1]$. Write p_x^* for $\mathbb{P}_x\{V = 1 \mid Z = 1\}$. Then

$$\mathbb{E}_x M(Y_1) = \mathbb{E}_x[M(Y) \mid V = 1]p_x^* + \mathbb{E}_x[M(Y) \mid Z = 1, V = 0](1 - p_x^*) . \tag{2}$$

If we can find a lower bound on $\mathbb{P}_x\{Z = 1 \mid V = 0\}$, then we can apply the method of Horowitz and Manski (1995) to get sharp bounds on $\mathbb{E}_x[M(Y) \mid Z = 1, V = 0]$, which will lead to sharp bounds on $\mathbb{E}_x M(Y_1)$ in (2). To this end, let $\pi_x(p_x^*) = \mathbb{P}_x\{Z = 1 \mid V = 0\}$. Recall that $v_x = \mathbb{P}_x\{V = 1\}$. Lemma 2A in the appendix shows that $\pi_x(p_x^*) = [(1-p_x^*)v_x]/[p_x^*(1-v_x)]$. Write Q_x for the quantile function of Y given $V = 0$ and $X = x$. Then, by results in Horowitz and Manski (1995), the endpoints of the interval

$$[\mathbb{E}_x[M(Y) \mid Y \leq Q_x(\pi_x(p_x^*)), V = 0], \mathbb{E}_x[M(Y) \mid Y > Q_x(1 - \pi_x(p_x^*)), V = 0]] \qquad (3)$$

are sharp bounds on $\mathbb{E}_x[M(Y) \mid Z = 1, V = 0]$. Combining (3) with (2) produces bounds on $\mathbb{E}_x M(Y_1)$. However, these bounds are infeasible since p_x^* is unknown.

To develop feasible bounds, first note that $v_x \leq p_x^* \leq 1$. This follows from A1, A2, and Bayes' rule. For each $p \in [v_x, 1]$, define $\pi_x(p) = [(1-p)v_x]/[p(1-v_x)]$. Then define

$$\begin{aligned} L_x(p) &= p\mathbb{E}_x[M(Y) \mid V = 1] + (1-p)\mathbb{E}_x[M(Y) \mid Y \leq Q_x(\pi_x(p)), V = 0] \\ U_x(p) &= p\mathbb{E}_x[M(Y) \mid V = 1] + (1-p)\mathbb{E}_x[M(Y) \mid Y > Q_x(1 - \pi_x(p)), V = 0]\,. \end{aligned}$$

Notice that for each $p \in [v_x, 1]$, $L_x(p)$ and $U_x(p)$ are identified from the sampling process.

THEOREM 4. *If A1, A2, and A3 hold, then*

$$\lambda_4(x) \leq \mathbb{E}_x M(Y_1) \leq u_4(x)$$

where

$$\begin{aligned} \lambda_4(x) &= \inf_{p \in [v_x, 1]} L_x(p) \\ u_4(x) &= \sup_{p \in [v_x, 1]} U_x(p)\,. \end{aligned}$$

Moreover, these bounds are sharp.

PROOF. From (2), (3), and the definition of $L_x(p)$ and $U_x(p)$, we obtain the infeasible bounds $L_x(p_x^*) \leq I\!E_x M(Y_1) \leq U_x(p_x^*)$. It follows from this and the definition of $\lambda_4(x)$ and $u_4(x)$ that

$$\lambda_4(x) \leq L_x(p_x^*) \leq I\!E_x M(Y_1) \leq U_x(p_x^*) \leq u_4(x) \,.$$

That is, $\lambda_4(x)$ and $u_4(x)$ are bounds for $I\!E_x M(Y_1)$ under A1, A2, and A3. We want to prove that they are sharp bounds under these assumptions. Assume, in addition to A1, A2, and A3, that p_x^* is known. Then $L_x(p_x^*)$ and $U_x(p_x^*)$ are sharp bounds for $I\!E_x M(Y_1)$, since they are based on sharp bounds for $I\!E_x[M(Y)|Z=1, V=0]$. That is, for each s in $[L_x(p_x^*), U_x(p_x^*)]$, there is a distribution for the data vector (Y, X, V, Z, Y_1, Y_0) that is consistent with A1, A2, A3, and the assumption that p_x^* is known, such that $I\!E_x M(Y_1) = s$. By Bayes' rule and A2,

$$p_x^* = I\!P_x\{Z=1|V=1\} I\!P_x\{V=1\}/I\!P_x\{Z=1\} = I\!P_x\{V=1\}/I\!P_x\{Z=1\} = v_x/r^*$$

for $r^* \in [v_x, 1]$. Define $p(r) = v_x/r$ for $r \in [v_x, 1]$. Note that $p(r)$ takes on all the values in $[v_x, 1]$ as r ranges over $[v_x, 1]$. Thus,

$$[\lambda_4(x), u_4(x)] = \bigcup_{r \in [v_x, 1]} \{s : s \in [L_x(p(r)), U_x(p(r))]\} \,. \tag{4}$$

Now drop the assumption that p_x^* is known. Under A1, A2, and A3, $I\!P_x\{Z=1\}$ can take on any value in $[v_x, 1]$. That is, each $r \in [v_x, 1]$ is a possible value for $I\!P_x\{Z=1\}$ under A1, A2, and A3. Deduce from this and (4) that for each $s \in [\lambda_4(x), u_4(x)]$, there is a distribution for (Y, X, V, Z, Y_1, Y_0) that is consistent with our assumptions for which $I\!E_x M(Y_1) = s$. This proves sharpness. □

Next, we consider the effect of imposing the monotonicity assumption A4 on the basic contamination model.

THEOREM 5. *If A1, A2, A3, and A4 hold, then*

$$\lambda_5(x) \leq \mathbb{E}_x M(Y_1) \leq u_5(x)$$

where

$$\begin{aligned} \lambda_5(x) &= \lambda_4(x) \\ u_5(x) &= \mathbb{E}_x[M(Y) \mid V = 1]\,. \end{aligned}$$

Moreover, these bounds are sharp.

PROOF. Argue as in the proof of Theorem 2 to get the upper bound. Argue as in the proof of Theorem 4 to establish sharpness. □

REMARK 4. Note that $u_5(x) = U_x(1) \leq \sup_{p \in [v_x, 1]} U_x(p) = u_4(x)$, with strict inequality whenever $U_x(p)$ is not maximized at $p = 1$. By definition of $U_x(p)$, $u_5(x) = u_4(x)$ if and only if the upper support point of the distribution of $M(Y)$ given $V = 0$ is less than or equal to $\mathbb{E}_x[M(Y) \mid V = 1]$. In other words, except in this special case, imposing A4 will result in a strictly smaller upper bound for $\mathbb{E}_x M(Y_1)$. Also, note that $u_2(x) = u_5(x)$. That is, when A4 holds, the upper bound in the corruption with verification model is equal to the upper bound in the contamination with verification model. Finally, note that if the sense of the inequality in A4 is reversed, then $\lambda_5(x) = \mathbb{E}_x[M(Y) \mid V = 1]$ and $u_5(x) = u_4(x)$ are sharp bounds on $\mathbb{E}_x M(Y_1)$.

Next, we consider adding exclusion restrictions to the basic contamination with verification model. First, we consider the standard instrumental variable assumption, A5.

THEOREM 6. *If A1, A2, A3, and A5 hold, then for each $x_2 \in \Omega_2$,*

$$\lambda_6(x_2) \leq \mathbb{E}_{x_2} M(Y_1) \leq u_6(x_2)$$

where

$$\begin{aligned} \lambda_6(x_2) &= \sup_{x_1 \in \Omega_1} \lambda_4(x_1, x_2) \\ u_6(x_2) &= \inf_{x_1 \in \Omega_1} u_4(x_1, x_2) . \end{aligned}$$

Moreover, these bounds are sharp.

PROOF. Mimic the proof of Theorem 3. □

Next, we add assumption A6 to the basic contamination with verification model. This is a different type of instrumental variable assumption, which, to the best of our knowledge, has not been considered before.

Recall that in deriving $\lambda_4(x)$ and $u_4(x)$ in Theorem 4, we optimize over $p \in [v_x, 1]$ because $p_x^* = \mathbb{P}_x\{V = 1 \mid Z = 1\}$, though unknown, is known to lie in the interval $[v_x, 1]$. If A6 holds, we can narrow the interval over which we optimize, thus potentially tightening the bounds and reducing calculations at the same time.

Write v_x as $v(x)$. For each $x_2 \in \Omega_2$, define $v_{x_2} = \sup_{s \in \Omega_1} v(s, x_2)$ and $\rho_x = v_x / v_{x_2}$.

THEOREM 7. *If A1, A2, A3, and A6 hold, then*

$$\lambda_7(x) \le \mathbb{E}_x M(Y_1) \le u_7(x)$$

where

$$\begin{aligned} \lambda_7(x) &= \inf_{p \in [v_x, \rho_x]} L_x(p) \\ u_7(x) &= \sup_{p \in [v_x, \rho_x]} U_x(p)\,. \end{aligned}$$

Moreover, these bounds are sharp.

PROOF. Write p_x^* as $p^*(x)$. Fix $x_2 \in \Omega_2$ and $x_1, s \in \Omega_1$. By Bayes' rule and A2,

$$\begin{aligned} p^*(x_1, x_2) &= v(x_1, x_2)/\mathbb{P}\{Z = 1 \mid X_1 = x_1, X_2 = x_2\} \\ p^*(s, x_2) &= v(s, x_2)/\mathbb{P}\{Z = 1 \mid X_1 = s, X_2 = x_2\}\,. \end{aligned}$$

By A6, $\mathbb{P}\{Z = 1 \mid X_1 = x_1, X_2 = x_2\} = \mathbb{P}\{Z = 1 \mid X_1 = s, X_2 = x_2\}$. Thus,

$$p^*(x_1, x_2) = p^*(s, x_2) v(x_1, x_2)/v(s, x_2)\,.$$

Assume, without loss of generality, that $v(x_1, x_2) \le v(s, x_2)$. Since $v(x_1, x_2) \le p^*(x_1, x_2) \le 1$ and $v(s, x_2) \le p^*(s, x_2) \le 1$, it follows that

$$v(x_1, x_2) \le p^*(x_1, x_2) \le v(x_1, x_2)/v(s, x_2)\,.$$

The upper bound can be made as small as possible by taking the supremum of $v(s, x_2)$ over $s \in \Omega_1$. This establishes the stated bounds. The proof of sharpness is similar to the proof of sharpness in Theorem 4. □

REMARK 5. Suppose A6 holds. Since $v(x_1, x_2) = \mathbb{P}\{V = 1 \mid X_1 = x_1, X_2 = x_2\}$, we see that $\rho_x = 1$ for all x if and only if V is independent of X_1 given X_2. In other words, if Z is independent of X_1 given X_2, then the interval over which we optimize is reduced if and only if V depends on X_1 given X_2. If this happens, then the basic contamination with verification bounds will be tightened if and only if either one of the bounds for the basic model occurs only at a p value greater than ρ_x. We illustrate this tightening of bounds when we apply Theorem 7 in the analysis of the Love Canal data in Section 6.

Also, note that if at least two of assumptions A4, A5, or A6 hold, then tighter sharp bounds can be obtained by combining the bounds in Theorems 5, 6, and 7 in the obvious way.

4. Estimation, Asymptotics, and Computation

In this section, we develop sample analogs of the population bounds for the mixture models with verification developed in Section 3. We also establish some of their asymptotic properties. In addition, we establish convexity and concavity of the sample and population functions defining the contamination with verification bounds, and we discuss the consequent computational benefits. Section 4.1 addresses estimation and asymptotics for corruption with verification models. Section 4.2 does the same for contamination with verification models. Section 4.3 addresses convexity, concavity, and computation in contamination with verification models.

4.1 Corruption with Verification Models

Let $(Y_i, X_i, V_i, Z_i, Y_{1i}, Y_{0i})$, $i = 1, \ldots, n$, be draws from from the mixture model with verification defined in Section 2. For notational simplicity, we take X to be a d-dimensional vector, $d \geq 0$, with all continuous components. The results are the same and the conditions easily generalize to the case where X is k-dimensional with d continuous components, $k \geq d \geq 0$. Let $\{a_n\}$ denote a sequence of positive real numbers converging to zero as $n \to \infty$. For x in the support of X, define $n_x = \sum_{i=1}^n \{|X_i - x| \leq a_n\}$, $n_{1x} = \sum_{i=1}^n V_i\{|X_i - x| \leq a_n\}$, and $\hat{v}_x = n_{1x}/n_x$. Let $f(x)$ denote the density of X at x.

Refer to Theorem 1 in Section 3. Define

$$\begin{aligned}
\hat{\lambda}_1(x) &= \hat{v}_x \sum_{i=1}^n M(Y_i)V_i\{|X_i - x| \leq a_n\}/n_{1x} + a(1 - \hat{v}_x) \\
\hat{u}_1(x) &= \hat{v}_x \sum_{i=1}^n M(Y_i)V_i\{|X_i - x| \leq a_n\}/n_{1x} + b(1 - \hat{v}_x)\,.
\end{aligned}$$

Note that we are implicitly using a uniform d-dimensional kernel smoother. We do this for the sake of simplicity. Any bounded, symmetric kernel density could be used. Define $G(x) = \mathbb{E}_x[M(Y)V + a(1-V)]$. Let $\mathcal{N}_x$ denote an open, convex neighborhood of x. We make the following assumptions:

C0. $(Y_i, X_i, V_i, Z_i, Y_{1i}, Y_{0i})$, $i = 1, \ldots, n$, are iid and satisfy A1 and A2.

C1. $f(x) > 0$ on $\mathcal{N}_x$.

C2. $f(x)$ and $G(x)$ have continuous first and second partial derivatives in x on $\mathcal{N}_x$.

C3. For $d \geq 1$, $a_n = O(n^{-1/d+4})$.

Write $G_1(x)$ and $f_1(x)$ for $\frac{\partial}{\partial x}G(x)$ and $\frac{\partial}{\partial x}f(x)$. Write $G_2(x)$ and $f_2(x)$ for $\frac{\partial^2}{\partial x \partial x}G(x)$ and $\frac{\partial^2}{\partial x \partial x}f(x)$. Let $\kappa_d = \int\{|z| \leq 1\}dz$. The symbol $\Longrightarrow$ denotes convergence in distribution.

THEOREM 8. *If C0 through C3 hold, then*

$$\sqrt{na_n^d}\left[\hat{\lambda}_1(x) - \lambda_1(x)\right] \Longrightarrow N(\mu_1(x), \sigma_1^2(x))$$

where

$$\begin{aligned}
\mu_1(x) &= \int \left[G_1(x)'zz'f_1(x) + z'G_2(x)z + z'f_2(x)z\right]\{|z| \le 1\}dz/\kappa_d f(x) \\
\sigma_1^2(x) &= \mathbb{E}_x[M(Y)V + a(1-V)]^2/\kappa_d f(x).
\end{aligned}$$

PROOF. See Appendix. □

REMARK 6. The limiting distribution of $\hat{u}_1(x)$ is obtained by replacing a with b in the definition of $G(x)$ and in the statement of Theorem 8. It also follows from the proof of Theorem 8 in the Appendix that $\mu_1(x) = 0$ if $a_n \ll n^{-1/[d+4]}$.

Refer to Theorem 2 in Section 3. Define

$$\begin{aligned}
\hat{\lambda}_2(x) &= \hat{\lambda}_1(x) \\
\hat{u}_2(x) &= \sum_{i=1}^{n} M(Y_i)V_i\{|X_i - x| \le a_n\}/n_{1x}.
\end{aligned}$$

Let $G(x) = \mathbb{E}_x[M(Y) \mid V = 1]$. Define $\mathcal{N}_x$ as before.

C0′. $(Y_i, X_i, V_i, Z_i, Y_{1i}, Y_{0i})$, $i = 1, \ldots, n$, are iid and satisfy A1, A2, and A4.

C2′. $f(x)$, $G(x)$, and v_x have continuous first and second partial derivatives in x on $\mathcal{N}_x$.

Define G_1, f_1, G_2, and f_2 as before.

THEOREM 9. *If C0′, C1, C2′, and C3 hold, then*

$$\sqrt{na_n^d}\,[\hat{u}_2(x) - u_2(x)] \Longrightarrow N(\mu_2(x), \sigma_2^2(x))$$

where

$$\begin{aligned}
\mu_2(x) &= \int [G_1(x)'zz'f_1(x) + z'G_2(x)z + z'f_2(x)z]\,\{|z| \le 1\}dz/\kappa_d v_x \\
\sigma_2^2(x) &= \mathbb{E}_x[M(Y) \mid V = 1]^2/\kappa_d v_x\,.
\end{aligned}$$

PROOF. See Appendix. □

Refer to Theorem 3 in Section 3. For each $x_2 \in \Omega_2$, define

$$\begin{aligned}
\hat{\lambda}_3(x_2) &= \sup_{x_1 \in \Omega_1} \hat{\lambda}_1(x_1, x_2) \\
\hat{u}_3(x_2) &= \inf_{x_1 \in \Omega_1} \hat{u}_1(x_1, x_2)\,.
\end{aligned}$$

Rates of convergence of $\hat{\lambda}_3(x_2)$ to $\lambda_3(x_2)$ and $\hat{u}_3(x_2)$ to $u_3(x_2)$ can be established using Lemma 1 developed in Section 4.2 below. However, we leave to future work the derivation of primitive conditions implying the high level conditions of Lemma 1.

4.2 Contamination with Verification Models

We now consider contamination with verification models. As before, let $(Y_i, X_i, V_i, Z_i, Y_{1i}, Y_{0i})$, $i = 1, \ldots, n$, denote independent draws from the mixture model with verification described in Section 2. Let X, a_n, n_x, n_{1x}, $f(x)$, and $\mathcal{N}_x$ be defined as before. In addition, define $n_{0x} = n_x - n_{1x}$

and $\hat{\pi}_x(p) = [(1-p)\hat{v}_x]/[p(1-\hat{v}_x)]$. Let $\hat{Q}_x$ denote the quantile function of the Y_i's for V_i's equal to zero and X_i's within a_n of x.

Define the sample lower and upper bound functions

$$\begin{aligned}
\hat{L}_x(p) &= p\sum_{i=1}^{n} M(Y_i)V_i\{|X_i - x| \leq a_n\}/n_{1x} \\
&+ (1-p)\sum_{i=1}^{n} M(Y_i)(1-V_i)\{Y_i \leq \hat{Q}_x(\hat{\pi}_x(p))\}\{|X_i - x\} \leq a_n\}/\hat{\pi}_x(p)n_{0x} \\
\hat{U}_x(p) &= p\sum_{i=1}^{n} M(Y_i)V_i\{|X_i - x| \leq a_n\}/n_{1x} \\
&+ (1-p)\sum_{i=1}^{n} M(Y_i)(1-V_i)\{Y_i > \hat{Q}_x(1-\hat{\pi}_x(p))\}\{|X_i - x\} \leq a_n\}/\hat{\pi}_x(p)n_{0x}\,.
\end{aligned}$$

Finally, define the extreme value estimators

$$\begin{aligned}
\hat{\lambda}_4(x) &= \inf_{p\in[\hat{v}_x,1]} \hat{L}_x(p) \\
\hat{u}_4(x) &= \sup_{p\in[\hat{v}_x,1]} \hat{U}_x(p)\,.
\end{aligned}$$

Our objective is to establish rates of convergence of the extreme value estimators to their population counterparts.

We begin by proving a general result about extreme value estimators. Let Θ denote a subset of $I\!R^k$, $k \geq 1$. For $\theta \in \Theta$, let $\tau(\theta)$ and $\hat{\tau}(\theta)$ denote real-valued functions. Define

$$\begin{aligned}
\lambda &= \inf_{\theta\in\Theta} \tau(\theta)\,, \qquad \hat{\lambda} = \inf_{\theta\in\Theta} \hat{\tau}(\theta) \\
u &= \sup_{\theta\in\Theta} \tau(\theta)\,, \qquad \hat{u} = \sup_{\theta\in\Theta} \hat{\tau}(\theta)\,.
\end{aligned}$$

LEMMA 1: *Let $\{\epsilon_n\}$ be a sequence of nonnegative real numbers satisfying $\epsilon_n \to 0$ as $n \to \infty$.*

If Θ is compact, τ is continuous on Θ, and $\sup_{\theta\in\Theta}|\hat{\tau}(\theta)-\tau(\theta)| = O_p(\epsilon_n)$ *as* $n\to\infty$, *then* $|\hat{\lambda}-\lambda| = O_p(\epsilon_n)$ *and* $|\hat{u}-u| = O_p(\epsilon_n)$ *as* $n\to\infty$.

PROOF. We prove the result for the infima. The proof for the suprema is similar.

Since Θ is compact, there exists $\hat{\theta}\in\Theta$ such that $\hat{\lambda}=\hat{\tau}(\hat{\theta})$ or $\hat{\lambda}=\lim_{\theta\to\hat{\theta}}\hat{\tau}(\theta)$ for some sequence of θ values in Θ converging to $\hat{\theta}$. The compactness of Θ and the continuity of τ imply that there exists $\theta_0\in\Theta$ such that $\lambda=\tau(\theta_0)$. Consider the case $\hat{\lambda}=\lim_{\theta\to\hat{\theta}}\hat{\tau}(\theta)$. (The other case can be handled similarly.) Either $\hat{\lambda}\le\lambda$ or $\hat{\lambda}>\lambda$. Suppose $\hat{\lambda}\le\lambda$. By definition of λ, $\lambda\le\tau(\hat{\theta})$. Thus, $\hat{\lambda}\le\lambda\le\tau(\hat{\theta})$, and so

$$|\hat{\lambda}-\lambda|\le|\hat{\lambda}-\tau(\hat{\theta})| = |\lim_{\theta\to\hat{\theta}}\hat{\tau}(\theta)-\lim_{\theta\to\hat{\theta}}\tau(\theta)|\le\lim_{\theta\to\hat{\theta}}|\hat{\tau}(\theta)-\tau(\theta)|\,.$$

Apply the uniform law to get $|\hat{\lambda}-\lambda| = O_p(\epsilon_n)$ as $n\to\infty$. Suppose $\hat{\lambda}>\lambda$. By definition of $\hat{\lambda}$, $\hat{\tau}(\theta_0)\ge\hat{\lambda}$. Thus, $\hat{\tau}(\theta_0)\ge\hat{\lambda}>\lambda$, and so

$$|\hat{\lambda}-\lambda|\le|\hat{\tau}(\theta_0)-\lambda| = |\hat{\tau}(\theta_0)-\tau(\theta_0)|\,.$$

Once again, apply the uniform law to get $|\hat{\lambda}-\lambda| = O_p(\epsilon_n)$ as $n\to\infty$. This proves the result. □

We now develop primitive conditions implying the uniform strong law in Lemma 1 for the contamination with verification bounds. Write S_x for the support of Y given $V=0$ and $X=x$. For convenience, adopt the convention that $\hat{\pi}_x(p)=1$ for $p\in[0,\hat{v}_x)$ and $\pi_x(p)=1$ for $p\in[0,v_x)$. Let $H_x(y)=I\!P_x\{Y\le y \mid V=0\}$ and $h_x(y)=\frac{\partial}{\partial u}H_x(y)$. We make the following assumptions.

D0. $(Y_i,X_i,V_i,Z_i,Y_{1i},Y_{0i})$, $i=1,\ldots,n$, are iid and satisfy A1, A2, and A3.

D1. S_x is an interval (possibly infinite) of the real line.

D2. $H_x(y)$ is continuously differentiable in y on S_x.

D3. $\mathbb{E}_x[|M(Y)| \mid V=0] < \infty$ and $\mathbb{E}_x[|M(Y)| \mid V=1] < \infty$.

D4. $f(x) > 0$ on $\mathcal{N}_x$.

D5. $f(x)$ and v_x have bounded first and second partial derivatives in x on $\mathcal{N}_x$.

D6. $H_x(y)$ has bounded first and second partial derivatives in x on $\mathcal{N}_x$.

D7. $h_x(y)$ and $Q_x(t)$ have bounded first and second partial derivatives in x on $\mathcal{N}_x$.

D8. For $d \geq 1$, $a_n \propto n^{-1/[d+4]}$.

Lemma 2 below is proved in the appendix, and requires some delicate arguments to obtain rates of uniform convergence of $\hat{L}_x(p)$ and $\hat{U}_x(p)$ to their population analogs. The source of the difficulty is the fact that these sample objective functions are averages of functions of indicator functions of sets involving $\hat{Q}_x(\hat{\pi}_x(p))$, the estimated conditional quantile function. In particular, note that the proof of Lemma 2 invokes Lemma 4A, which helps establish a rate of convergence of an estimated conditional distribution function to its population counterpart. The proof of Lemma 4A requires a refined use of an existing maximal inequality bound.

LEMMA 2: *If D0 through D8 hold, then* $\sup_{p\in[0,1]} |\hat{L}_x(p) - L_x(p)| = O_p(n^{-2/[d+4]})$ *and* $\sup_{p\in[0,1]} |\hat{U}_x(p) - U_x(p)| = O_p(n^{-2/[d+4]})$ *as* $n \to \infty$.

PROOF. See Appendix. □

THEOREM 10: *If D0 through D8 hold, then* $|\hat{\lambda}_4(x) - \lambda_4(x)| = O_p(n^{-2/[d+4]})$ *as* $n \to \infty$ *and* $|\hat{u}_4(x) - u_4(x)| = O_p(n^{-2/[d+4]})$ *as* $n \to \infty$.

PROOF. Assumption D2 implies that L_x and U_x are continuous in p on $[0,1]$. Apply Lemma 1 and Lemma 2 with $\theta = p$ and $\Theta = [0,1]$. For the infima, take $\tau = L_x$, $\hat{\tau} = \hat{L}_x$, $\lambda = \lambda_4(x)$, and $\hat{\lambda} = \hat{\lambda}_4(x)$. For the suprema, take $\tau = U_x$, $\hat{\tau} = \hat{U}_x$, $u = u_4(x)$, and $\hat{u} = \hat{u}_4(x)$. □

REMARK 7. The proof of Lemma 2 in the appendix shows that the rates of convergence established in Theorem 10 are optimal in the mean squared error sense relative to the uniform kernel that is used.

Sample analogs of the upper and lower bounds in Theorem 5, Theorem 6, and Theorem 7 can be defined in the obvious way. The asymptotics developed in Theorem 9 and Theorem 10 apply immediately to sample analogs for the bounds in Theorem 5. As in the case of Theorem 3, we leave for future work the establishment of primitive conditions implying the high-level conditions of Lemma 1 for sample analogs of the bounds in Theorem 6. A generalization of Lemma 1 covering random parameter spaces is required to establish rates of convergence for sample analogs of the bounds in Theorem 7. We also leave this for future work.

4.3 Convexity, Concavity, and Computation

In this subsection, we establish convexity and concavity of the functions defining the contamination with verification bounds and discuss the consequent computational benefits. We begin with the case where Y has a continuous distribution and establish the convexity of $L_x(p)$ and the concavity of $U_x(p)$ under appropriate conditions. We then show that these results also hold if the response variable takes on a finite number of values. Next, we extend these results to the sample analogs $\hat{L}_x(p)$ and $\hat{U}_x(p)$. Finally, we discuss the consequent computational benefits.

THEOREM 11. *Suppose A1 through A3 hold. Suppose, in addition, that D1 through D3 hold and M is strictly increasing and differentiable on $\mathbb{R}$. Then $L_x(p)$ is strictly convex and $U_x(p)$ is strictly concave on $[v_x, 1]$.*

PROOF. Start with $L_x(p)$. Write c_x for $(1 - v_x)/v_x$. By definition of the quantile function Q_x,

$\pi_x(p) = I\!P_x\{Y \le Q_x(\pi_x(p)) \mid V = 0\}$. Therefore,

$$\begin{aligned} L_x(p) &= pI\!E_x[M(Y) \mid V=1] + (1-p)I\!E_x[M(Y)\{Y \le Q_x(\pi_x(p))\} \mid V=0]/\pi_x(p) \\ &= p\left[I\!E_x[M(Y) \mid V=1] + c_x \int_{-\infty}^{Q_x(\pi_x(p))} M(t)\, h_x(t)\, dt\right]. \end{aligned}$$

By D1, h_x must be positive on S_x, except possibly at the boundary points. D1 also implies that H_x is strictly increasing on S_x. Thus, we may interpret Q_x as the inverse function of H_x and apply D2 and D3 along with the chain rule to get that the second derivative of $L_x(p)$ with respect to p equals $M'(Q_x(\pi_x(p)))/[c_x p^3 h_x(Q_x(\pi_x(p)))]$, which is strictly positive on $(v_x, 1)$. Deduce that $L_x(p)$ is a strictly convex function on $[v_x, 1]$.

Similar calculations show that the second derivative of $U_x(p)$ with respect to p equals $-M'(Q_x(1-\pi_x(p)))/[c_x p^3 h_x(Q_x(1-\pi_x(p)))]$, which is strictly negative on $(v_x, 1)$. Thus, $U_x(p)$ is a strictly concave function on $[v_x, 1]$. □

Next, we show that the convexity of $L_x(p)$ and concavity of $U_x(p)$ carry over to the discrete case. To this end, suppose that given $V = 0$ and $X = x$, the response variable D takes on the value δ_i with probability $\gamma_i > 0$, $i = 1, 2, \ldots, k$, with $\sum_{i=1}^{k} \gamma_i = 1$. For convenience, take $\delta_1 < \delta_2 < \cdots < \delta_k$. Let $M_2(t)$, $t \in I\!R$ be an arbitrary increasing function and suppose that we are interested in developing sharp bounds on $I\!E_x M_2(D)$ under assumptions A1, A2, and A3. Since the distribution of D given $V = 0$ and $X = x$ is not continuous, technically, none of the results established for contamination with verification models apply. For the same reason, Theorem 11 does not apply. However, by writing D as a discrete transformation of an analogous continuous random variable, all the results just mentioned carry over to the discrete case as well.

To see this, condition on $V = 0$ and $X = x$, and for $\epsilon > 0$, define $\delta_0 = \delta_1 - \epsilon$. For each i,

$i = 1, 2, \ldots, k$, generate an observation from any continuous probability distribution on $(\delta_{i-1}, \delta_i]$. With probability γ_i, select the observation that falls in the interval $(\delta_{i-1}, \delta_i]$. Call the resulting continuously distributed random variable Y. Next, define

$$M_1(Y) = \sum_{i=1}^{k} \delta_i \{\delta_{i-1} < Y \leq \delta_i\}\,.$$

Note that $\mathbb{P}_x\{M_1(Y) = \delta_i \mid V = 0\} = \gamma_i$, $i = 1, 2, \ldots, k$. That is, given $V = 0$ and $X = x$, $M_1(Y)$ and D are identically distributed. Next, define

$$M(Y) = M_2(M_1(Y)) = \sum_{i=1}^{k} M_2(\delta_i)\{\delta_{i-1} < Y \leq \delta_i\}\,. \tag{5}$$

We see that given $V = 0$ and $X = x$, $M(Y)$ and $M_2(D)$ are identically distributed. Since the distribution of Y given $V = 0$ and $X = x$ is continuous, all the results established previously for contamination with verification models apply. In addition, we can prove the following analog of Theorem 11 for the discrete case.

THEOREM 12. *Suppose A1 through A3 hold. Suppose, in addition, that D1 through D3 hold, and $M(Y)$ is defined as in (5). Then $L_x(p)$ is a piecewise linear convex function and $U_x(p)$ is a piecewise linear concave function on $[v_x, 1]$.*

PROOF. Start with $L_x(p)$. As in the proof of Theorem 11, we have that

$$L_x(p) = p\left[\mathbb{E}_x[M_2(D) \mid V = 1] + c_x \int_{-\infty}^{Q_x(\pi_x(p))} M(t)\, h_x(t)\, dt\right].$$

Define $\tau_0 = 0$ and $\tau_i = M_2(\delta_i)$, $i = 1, 2, \ldots, k$. Define $\tau_{k+1} = \tau_k$. Straightforward, but tedious, calculations show that

$$L_x(p) = \sum_{i=k}^{1}(\beta_i p + \tau_i)\{\pi_x^{-1}(H_x(\delta_i)) \le p < \pi_x^{-1}(H_x(\delta_{i-1}))\} \tag{6}$$

where

$$\beta_i = I\!\!E_x[M_2(D) \mid V = 1] - \tau_i - c_x \sum_{j=1}^{i}(\tau_j - \tau_{j-1})H_x(\delta_{j-1})$$

and, for $p \in [0,1]$,

$$\pi_x^{-1}(p) = v_x/[v_x + p(1 - v_x)]\,.$$

Note that the summation in (6) runs from k to 1, not from 1 to k. It is straightforward to show that β_i increases as i increases, and that $L_p(x)$ is continuous on $[v_x, 1]$. Deduce that $L_p(x)$ is a piecewise linear convex function on $[v_x, 1]$.

Similar calculations show that

$$U_x(p) = \sum_{i=1}^{k}(\beta_i p + \tau_i)\{\pi_x^{-1}(1 - H_x(\delta_{i-1})) < p \le \pi_x^{-1}(1 - H_x(\delta_i))\} \tag{7}$$

where

$$\beta_i = I\!\!E_x[M_2(D) \mid V = 1] - (c_x + 1)\tau_i + c_x\tau_k - c_x \sum_{j=i}^{k}(\tau_{j+1} - \tau_j)H_x(\delta_j)\,.$$

It is straightforward to show that β_i decreases as i increases, and that $U_p(x)$ is continuous on $[v_x, 1]$. Deduce that $U_x(p)$ is a piecewise linear concave function on $[v_x, 1]$. □

When, given $V = 0$ and $X = x$, the response variable takes on a finite number of values, then $\hat{L}_x(p)$ is a piecewise linear convex function and $\hat{U}_x(p)$ is a piecewise linear concave function on $[\hat{v}_x, 1]$. This follows from the fact that $\hat{L}_x(p)$ is equal to the expression on the right-hand side of

(6) and $\hat{U}_x(p)$ is equal to the expression on the right-hand side of (7) after replacing the population quantities β_i, H_x, and π_x^{-1} with their sample analogs.

Theorem 12 also implies that $\hat{L}_x(p)$ is a piecewise linear convex function and $\hat{U}_x(p)$ is a piecewise linear concave function on $[\hat{v}_x, 1]$ when the response variable is continuously distributed given $V = 0$ and $X = x$. To see this, consider a discrete response variable that takes on each of the n distinct realizations of the continuous response variable with probability n^{-1}, and then apply Theorem 12.

The sample results stated in the last two paragraphs can be useful computationally. For example, when the response variable takes on only k values, then by Theorem 12, it is sufficient to evaluate $\hat{L}_x(p)$ at the k kink points $\hat{\pi}_x^{-1}(\hat{H}_x(\delta_i))$, $i = 1, 2, \ldots, k$, to find $\hat{\lambda}_4(x)$. Similarly, it is sufficient to evaluate $\hat{U}_x(p)$ at the k kink points $\hat{\pi}_x^{-1}(1 - \hat{H}_x(\delta_i))$, $i = 1, 2, \ldots, k$, to find $\hat{u}_4(x)$. In the important special case of a binary response where $\delta_1 = 0$, $\delta_2 = 1$, and $M_2(t) = t$, it is easy to show that $\hat{\lambda}_4(x) = \hat{L}_x(\hat{\pi}_x^{-1}(\hat{H}_x(\delta_1)))$ and $\hat{u}_4(x) = \hat{U}_x(\hat{\pi}_x^{-1}(1 - \hat{H}_x(\delta_1)))$. That is, only a single evaluation of the sample functions is needed to find the extreme value estimators. This is also the case when the response is continuous and $M(Y)$ is binary. Horowitz and Manski (2000) have found a similar result in the binary case.

By the same reasoning, when $M(Y)$ is continuously distributed given $V = 0$ and $X = x$, then it is sufficient to evaluate each sample criterion function at n kink points. However, because of convexity and concavity, a binary search over n kink points will yield an extreme value estimator after only $O(\log n)$ function evaluations. This can result in substantial savings in computation time when n is large, when the extreme value estimators need to be evaluated for many x values, or when bootstrap estimates of the distribution of the extreme value estimators are desired.

5. Survey Nonresponse and Treatment Effects

In this section, we show that survey nonresponse models analyzed by Horowitz and Manski (1998) and randomized treatment effects models analyzed by Horowitz and Manski (2000) and Molinari (2001) can be viewed as mixture models with verification. Consequently, sharp bounds on characteristics of interest in these models can be derived using the results in this paper.

We begin with the survey nonresponse models. Suppose measurements are collected on respondents of the form (Y_1, X), where Y_1 is a univariate outcome of interest and X is a vector of regressors on which Y_1 depends. The objective is to make inferences about various aspects of $P_A(Y_1)$, the conditional distribution of Y_1 given $X \in A$, where A is some subset of the support of X. The special case $A = x$, where x is a point in the support of X, has been the focus of this paper.

Horowitz and Manski (1998) consider the problem of making such inferences from survey data subject to various types of nonresponse. They consider three types: outcome censoring, regressor censoring, and joint censoring of outcomes and regressors. Outcome censoring occurs when X is always observed, but sometimes Y_1 is not observed. Regressor censoring occurs when Y_1 is always observed, but sometimes X is not observed. Finally, joint censoring of outcomes and regressors occurs when sometimes both Y_1 and X are observed, and sometimes neither are observed. For each of these polar cases, Horowitz and Manski (1998) develop sharp bounds on various functionals of $P_A(Y_1)$ without making untestable assumptions about the data generating process.

First, consider outcome censoring. Replace missing Y_1 values with any logically possible values. Define $Z = 1$ if Y_1 is observed and $Z = 0$ otherwise. Define $V = Z$. Replace $\mathbb{P}_x$ with $\mathbb{P}_A$ and $\mathbb{E}_x$ with $\mathbb{E}_A$ in all our assumptions, where $\mathbb{P}_A$ denotes probability conditional on $X \in A$ and $\mathbb{E}_A$ denotes expectation conditional on $X \in A$. Assumptions A1 and A2 hold trivially, while assumption A3 need not hold. Thus, the outcome censoring model is a special case of a basic corruption with verification model.

Next, consider regressor censoring. Assume that $\mathbb{P}\{X \in A\} > 0$ and disregard all observed (Y_1, X) pairs for which X is in the complement of A since these observations provide no information about $\mathbb{E}_A M(Y_1)$, the object of interest. View the remaining (Y_1, X) pairs as the relevant sample. For these pairs, either X is observed to be in A or X is unobserved. Define $Z = \{X \in A\}$. That is, $Z = 1$ if $X \in A$ and $Z = 0$ otherwise. Define $V = 1$ if X is observed to be in A and $V = 0$ otherwise. In this model, $Y_0 = Y_1$ and so $Y = Y_1$. To avoid degenerate cases, assume that $\mathbb{P}\{V = 1\} > 0$. Note that $\mathbb{P}\{Z = 1 \mid V = 1\} = 1$ and $\mathbb{E}_A M(Y_1) \equiv \mathbb{E}[M(Y_1) \mid Z = 1]$. That is, assumptions A1, A2, and A3 hold. Thus, the regressor censoring model is a special case of a basic contamination with verification model. Note that B3 does not make sense in this model. We also note that the bounds derived by Horowitz and Manski (1998) for this model are based on a different parametrization than the one used in this paper. Their counterparts of $L_x(p)$ and $U_x(p)$ are not respectively convex and concave. (See Section 4.3 for more discussion of this point.)

Finally, consider joint censoring of outcomes and regressors. As with regressor censoring, assume that $\mathbb{P}\{X \in A\} > 0$ and disregard all observed (Y_1, X) pairs for which X is in the complement of A. For the remaining pairs, replace missing (Y_1, X) values with any logically possible values, define $Z = \{X \in A\}$, and define $V = 1$ if (Y_1, X) is observed and $V = 0$ otherwise. For these remaining pairs, assume that $\mathbb{P}\{V = 1\} > 0$ and note that $\mathbb{P}\{Z = 1 \mid V = 1\} = 1$. That is, A1 and A2 hold. Since A3 need not hold, this model is a special case of a corruption with verification model.

Next, consider randomized treatment effects models. In these models, Y_1 is the outcome obtained from receiving treatment 1, say, and Y_0 is the outcome obtained from receiving treatment 0. The objective is to derive sharp bounds on the difference $\mathbb{E}_x M(Y_1) - \mathbb{E}_x M(Y_0)$ for various choices of M. Leading choices for the function M include the identity function $M(y) = y$ and the indicator function $M(y) = \{y \in B\}$ where B is a subset of $\mathbb{R}$.

Define $Z = 1$ if treatment 1 is received and $Z = 0$ if treatment 0 is received. In standard treatment effects models, (see, for example, Manski, 1995) Z is fully observed. However, in the applications that Horowitz and Manski (2000) and Molinari (2001) study, Z is observed for only a subset of the sample. For simplicity, assume that all outcomes and covariates are observed. Since treatments are randomized, assumption B3 holds. Focus, for the moment, on those observations for which either Z is observed to equal 1 or Z is unobserved. Since treatment assignment is random, only these observations can contain information relevant for estimating $\mathbb{E}_x M(Y_1)$. For these observations, define $V = 1$ if Z is observed to equal 1, and $V = 0$ otherwise. Since B3 implies A3 and assumptions A1 and A2 hold, sharp contamination with verification bounds on $\mathbb{E}_x M(Y_1)$ can be obtained using the results in Theorem 4. Let λ_1 and u_1 denote the lower and upper bounds on $\mathbb{E}_x M(Y_1)$. Next, let Y_1 and Y_0 change roles, and focus on those observations for which either Z is observed to equal 0 or Z is unobserved. For these observations, define $V = 1$ if Z is observed to equal 0, and $V = 0$ otherwise. As before, assumptions A1, A2, and A3 hold and so sharp contamination with verification bounds on $\mathbb{E}_x M(Y_0)$ can be obtained using the results in Theorem 4. Let λ_0 and u_0 denote the lower and upper bounds on $\mathbb{E}_x M(Y_0)$. The sharp lower bound on $\mathbb{E}_x M(Y_1) - \mathbb{E}_x M(Y_0)$ is given by $\lambda_1 - u_0$; the sharp upper bound is given by $u_1 - \lambda_0$.

6. An Example: The Love Canal Study

To illustrate the usefulness of our results on corruption and contamination with verification, we analyze data on environmental pollutants – α-BHC and 2-chloronapthalene (2-CNAP) – in the Love Canal and a comparison region. The measurements, derived from soil samples analyzed in 1988, were produced by a process that is believed to satisfy assumptions A1 and A2.

Lambert and Tierney (1997) analyze these data under the further assumption that B3 holds. Unlike the estimator proposed in Section 4.2, the nonparametric maximum likelihood-type estimators they derive require kernel smoothing even in the absence of continuous covariates. This approach does not allow them to easily handle the left-censored data that are present in the analysis or to add continuous covariates to the analysis. In addition, Lambert and Tierney motivate assumption A4 but can only incorporate it in estimation with an ad hoc adjustment unless an "unverified measurement is tied with (equal to) the smallest verified measurement" (p. 941). We shall illustrate the identifying power of assumptions A3 and A4, as well as A5 and A6, and the ease with which these assumptions and censoring can be incorporated into our estimation procedure.

As described by Lambert and Tierney, the measurement and verification process for each compound proceeded as follows. Soil samples were collected from numerous, randomly-selected locations in Love Canal and in a comparison region. Samples were then randomly assigned to labs for pollutant analysis by gas chromatography-mass spectroscopy (GC-MS). The gas chromatograph was used to attempt to isolate the pollutant of interest. The mass spectrometer was used to ionize the isolated compound and then (1) verify that the compound was the pollutant of interest and (2) measure the concentration of the most abundant ion. The verification procedure yielded reports of either verified ($V = 1$) or not verified ($V = 0$). At each lab, a positive fraction was verified, and so A1 holds.

According to Lambert and Tierney, "Chemists believe that all of the verified GC-MS measurements belong to the pollutant of interest, but unverified GC-MS measurements may or may not belong to the pollutant." If this is correct, then A2 holds. Lambert and Tierney further assume B3, namely, that isolation of the compound is independent of its concentration. This assumption is somewhat questionable, especially since they also assume B4, namely, that the probability of verification when the correct pollutant has actually been isolated is increasing in the concentration level. This underscores the need for corruption with verification bounds.

Measurements were reported in parts per billion (ppb). However, when unverified measurements were below 0.2 ppb, the protocol allowed the lab to simply report "less than .2 ppb." Unlike the approach taken by Lambert and Tierney, our bounds easily incorporate such censoring. We obtain sharp bounds on $\mathbb{E}_{labx} Y_1$ by assigning the lower limit value of 0.00 to every censored response to calculate the lower bound and the upper limit value 0.20 to every censored response to calculate the upper bound.

Table 1 presents descriptive statistics for the data produced by each of six labs, as well as for the aggregated data. Note that bounds on the sample means are reported when values are censored. Several other features of the table are noteworthy. The proportion of verified measurements varies greatly across labs, ranging from 0.29 to 0.87 for α-BHC and from 0.32 to 0.97 for 2-CNAP. The proportion censored among the unverified also varies greatly, ranging from 0.00 to 0.86 for α-BHC and from 0.00 to 0.50 for 2-CNAP. Note that not all unverified values below 0.2 are censored. For α-BHC, the verified measurements tend to be much larger than the unverified measurements, but this relationship is not clear for 2-CNAP, which has mean verified concentration levels below 0.2 ppb at each lab.

Consider now the corruption model for 2-CNAP concentration in Love Canal using data for all labs combined. Let $M(Y_1) = \{Y_1 \leq t\}$ for $t \in R$. To focus on the identifying power of the various assumptions, we shall treat the sample as the population in this example. We may then use Theorem 1 to generate the lower bound λ_1 and the upper bound u_1 for any value of t.

Suppose that we also adopt the independence assumption B3. We may then use Theorem 4 to generate the contamination bounds λ_4 and u_4. As noted in Section 4.3, we may solve for the bounds with a single evaluation of $\hat{L}_x(p)$ and $\hat{U}_x(p)$ when $M(Y_1)$ is an indicator function.

Figure 1 presents our direct calculation of these corruption and contamination bounds for a fine grid of values of t on the interval $[0, 0.3]$; that is, bounds on the cumulative distribution function (CDF) over this interval. We see that the bounds under corruption are much wider than, and of course contain, the contamination bounds. For example, the bounds on $\mathbb{P}\{Y_1 \leq 0.070\}$ are $[0.32, 0.60]$ under corruption versus $[0.37, 0.54]$ under contamination. Note also that the lower bound on $\mathbb{P}\{Y_1 \leq 0\}$ equals 0.00 under either corruption or contamination, because all measurements that may equal 0 are unverified censored values, so it may be that there are no true zero concentrations of 2-CNAP. In contrast, the corruption upper bound at $t = 0$ is $u_1 = \mathbb{P}\{V = 0\} = 0.28$. The contamination upper bound at $t = 0$ is $u_4 = 0.02$.

Note also that we may invert these bounds to obtain bounds on quantiles. For instance, the bounds on the median of Y_1 are $[0.055, 0.100]$ under corruption versus $[0.066, 0.087]$ under contamination.

Now suppose we adopt the monotonic verification assumption A4. The effect on the bounds on the CDF is straightforward. That is, for each value t, the upper bound is unchanged whereas the lower bound equals the proportion satisfying $\{Y \leq t\}$ among the verified observations. This relationship holds for both corruption and contamination. Figure 2 presents these bounds.

Further, under A4, we may obtain an informative upper bound on $\mathbb{E}_{labx} Y_1$ under corruption. We do not need to invoke A4 to obtain informative bounds on $\mathbb{E}_{labx} Y_1$ under contamination, so we begin by considering bounds on $\mathbb{E}_{labx} Y_1$ under A1, A2, and A3.

Figure 3 illustrates the calculation of contamination bounds on the mean concentration of 2-CNAP in Love Canal, based on Lab 2 measurements. Figure 4 uses Lab 6 measurements. The pictures appear different because Lab 2 has no censored observations, whereas 6.5 percent of the Lab 6 unverified observations are censored.

Consider Figure 3. Note that $U_{lab2}(v_{lab2}) = \mathbb{E}_{lab2} Y = L_{lab2}(v_{lab2})$. As p increases, $U_{lab2}(p)$

increases monotonically until $p = 0.903$, with $U_{lab2}(0.903) = u_4(lab2) = 0.068$. From there, $U_{lab2}(p)$ decreases monotonically until $U_{lab2}(1) = \mathbb{E}_{lab2}[Y|V = 1]$. Similarly, $L_{lab2}(p)$ decreases monotonically from either direction, reaching its infimum at $p = 0.871$, with $L_{lab2}(0.871) = \lambda_4(lab2) = 0.062$. Thus, we have $0.062 \leq \mathbb{E}_{lab2}Y_1 \leq 0.068$.

Now, suppose we invoke A4. Then, as depicted in Figure 3, the upper bound is reduced to $u_5(lab2) = \mathbb{E}_{lab2}[Y|V = 1] = 0.065$. This assumption has no effect on the lower bound. In contrast, invoking A4 reduces the lower bound but does not effect the upper bound in Figures 1 and 2. The effects are reversed in that case because $M(Y_1)$ is decreasing in Y_1. Figure 4 presents similar bounds for Lab 6, where the main qualitative difference arises from the censoring that separates $L_{lab6}(v_{lab6})$ and $U_{lab6}(v_{lab6})$.

Table 2 presents contamination bounds on the mean for every lab, for both α-BHC and 2-CNAP and for both Love Canal and the comparison region. The bounds are often quite wide. However, one may wish to conclude, as do Lambert and Tierney, that the mean concentration of α-BHC in Love Canal exceeds the concentration in the comparison region, under assumptions A1, A2, and A3, as well as A4 if desired. In unreported results, we find that bootstrap estimates of confidence intervals support this inference on the population means for some labs.

The importance of A3 in reaching this conclusion is illustrated by considering the entries in Table 3. Without further assumptions, informative corruption bounds on the mean are not identified. When A4 is added to A1 and A2, we still cannot obtain informative lower bounds on the mean, so the bounds will necessarily overlap across geographic regions. Under contamination, the bounds on the median concentration of α-BHC overlap slightly–[0.077,0.211] in the comparison region versus [0.198,0.297] in Love Canal. In contrast, the corruption bounds for the comparison region contain the bounds for Love Canal–[0.000,∞] versus [0.182,0.546]. When A4 is invoked, both pairs overlap slightly–[0.077,0.020] versus [0.020,0.283] (contamination) and [0.000,0.020] versus [0.182,0.283] (corruption).

It is worth commenting on the uninformative bounds on the concentration of α-BHC in the comparison region. This result arises simply because less than half of the data are verified. However, when monotonic verification is assumed, this upper bound falls to just 0.020, which is almost below the interval estimate for Love Canal. Thus, assumption A4 may be more plausible than A3 and, as is evident from comparison of the second and third rows of Table 3, yields nearly the identical conclusion about regional variation in the median concentration of α-BHC.

Finally, we shall illustrate the potential usefulness of exclusion restrictions in the contamination with verification model. Suppose that labs are equally proficient at isolating the compound of interest, so that A6 holds (i.e., isolation indicator Z is independent of lab X_1), but some labs are better at verification than others, yielding variation in V. Then the maximal verification rate may be used to reduce the upper limit of the set over which the bound functions are optimized for each lab. For Lab 1 with the maximal (2-CNAP) verification rate of 0.94, the bounds do not change, but they may change for others according to Theorem 7. For example, Figure 3 presents the tightened bounds–$\lambda_7(x)$ and $u_7(x)$–for Lab 2, which has $v_{lab2} = 0.77$ and, therefore, $\rho_{lab2} = 0.82$. Figure 4 presents the new bounds for Lab 6, where $\rho_{lab6} = 0.72$.

Consideration of A5 sheds new light on the usefulness of bounds derived here under independence assumption A3. Recall that soil samples were randomly assigned to labs, which would then attempt to isolate and verify the pollutant. Clearly, random assignment implies that the true concentration should be independent of the lab, so that A5 holds and $\mathbb{E}_{labx} Y_1$ should not vary across labs. Yet the bounds on $\mathbb{E}_{labx} Y_1$ reported in Table 2 typically do not overlap across labs, so application of Theorem 6 yields lower bounds that exceed upper bounds. Clearly, either at least one independence assumption does not hold–A3 and/or A5–or these apparent violations arise from sampling variation, which has been assumed away for illustrative purposes. Regardless, this analysis illustrates another use for such bounds, namely, specification tests for independence assumptions adopted in semiparametric and fully parametric models. We leave this topic for future research.

7. Summary and Directions for Future Work

This paper undertakes a nonparametric analysis of mixture models with verification, and extends and improves upon much of the work begun by Lambert and Tierney (1997) on verification problems. Recall that F denotes the unknown distribution of interest. Lambert and Tierney develop nonparametric maximum likelihood-type estimators of bounds on functionals of F in a contamination with verification model. Their estimation procedure requires that F be continuous, does not accommodate continuous regressors, and does not naturally handle censored responses or incorporate plausible monotonicity restrictions.

This paper develops sharp bounds on functionals of F not only for basic contamination with verification models, but also for basic corruption with verification models. In addition, for both contamination and corruption with verification models, we develop sharp bounds under plausible monotonicity and exclusion restrictions. We develop nonparametric analog estimators of these population bounds and state primitive conditions implying rates of convergence and some distributional results for these estimators. The estimation procedure allows F to be continuous, discrete, or mixed, accommodates continuous regressors, and can easily and naturally incorporate response censoring as well as monotonicity and exclusion restrictions.

We also establish convexity and concavity of lower and upper bound functions used to define extremum estimators for contamination with verification models. These results can yield substantial computational benefits. We also show that survey nonresponse and randomized treatment effects models are special cases of mixture models with verification. Finally, we illustrate the use of the estimators on data from the Love Canal.

In future work, we plan to establish the limiting distribution of the extremum estimators for the contamination with verification models. The convexity and concavity results will play a pivotal role in establishing these results. In Section 4.3 we show that when the response variable takes on k values, then each extremum estimator can be expressed as a sample objective function evaluated at one of k estimable points. Each one of these evaluations is a function of sample averages, and so a limiting normal distribution should obtain. When the response variable is continuous, the situation is more complicated. Still, because of convexity and concavity of the sample objective functions, one can write the extremum estimators as continuous functionals of optimization estimators. Thus, if the distribution of the optimization estimators can be derived, the distribution of the extremum estimators may be derivable through an application of a version of the continuous mapping theorem.

We also plan to apply these methods to econometric analysis of survey data. Missing data problems are often ignored or are overcome with strong parametric assumptions on the data-generating process. Our framework not only provides a method for producing interval estimates in the absence of these assumptions but also suggests a natural set of specification tests. For instance, mononotonicty assumptions are often implicitly invoked when estimating structural models of attrition, and these assumptions may considerably tighten the bounds on population parameters of interest without the extra assumptions of a fully parametric model. Exclusion restrictions of the form A6 may be useful whenever data are available on the behavior of the survey organization. For example, interviewers are often randomly assigned within a geographic area or respondent payments may be subject to experimental variation. In such cases, the bounds may become tight enough to reject instrumental variable assumptions that are simultaneously maintained in the econometric model.

Thus, careful application of assumptions A3 through A6 may yield significant benefits in terms of identification and may yield useful specification tests.

APPENDIX

LEMMA 1A. *Suppose M is an increasing function on the support of Y_1. If, in addition, $\mathbb{P}_x\{V=1 \mid Y_1=y\}$ is increasing in y, then*

$$\mathbb{E}_x[M(Y_1) \mid V=1] \geq \mathbb{E}_x[M(Y_1) \mid V=0].$$

PROOF. We will show that the monotonicity condition implies that for each t in the support of Y_1,

$$\mathbb{P}_x\{Y_1 \leq t \mid V=1\} \leq \mathbb{P}_x\{Y_1 \leq t\} \leq \mathbb{P}_x\{Y_1 \leq t \mid V=0\}. \tag{8}$$

The result will follow from this stochastic dominance condition and the fact that M is increasing.

By Bayes' rule,

$$\mathbb{P}_x\{Y_1 \leq t \mid V=1\} = \mathbb{P}_x\{V=1 \mid Y_1 \leq t\}\mathbb{P}_x\{Y_1 \leq t\}/\mathbb{P}_x\{V=1\}.$$

Thus, $\mathbb{P}_x\{Y_1 \leq t \mid V=1\} \leq \mathbb{P}_x\{Y_1 \leq t\}$ if and only if $\mathbb{P}_x\{V=1 \mid Y_1 \leq t\} \leq \mathbb{P}_x\{V=1\}$. But

$$\mathbb{P}_x\{V=1\} = \mathbb{P}_x\{V=1 \mid Y_1 \leq t\}\mathbb{P}_x\{Y_1 \leq t\} + \mathbb{P}_x\{V=1 \mid Y_1 > t\}\mathbb{P}_x\{Y_1 > t\}.$$

By the monotonicity condition, $\mathbb{P}_x\{V=1 \mid Y_1 \leq t\} \leq \mathbb{P}_x\{V=1 \mid Y_1 > t\}$, implying that $\mathbb{P}_x\{V=1 \mid Y_1 \leq t\} \leq \mathbb{P}_x\{V=1\}$. The proof that $\mathbb{P}_x\{Y_1 \leq t\} \leq \mathbb{P}_x\{Y_1 \leq t \mid V=0\}$ is similar.

This proves (8). □

LEMMA 2A. *If A1 and A2 hold, then* $\pi_x(p_x^*) = [(1-p_x^*)v_x]/[p_x^*(1-v_x)]$.

PROOF: Invoke A1 and assume that $v_x < 1$. (If $v_x = 1$, then by A2, $P_x(Y_1)$ is identified by the verified data.) By Bayes' Rule,

$$
\begin{aligned}
\pi_x(p_x^*) &= I\!P_x\{V=0|Z=1\}I\!P_x\{Z=1\}/I\!P_x\{V=0\} \\
&= (1-p_x^*)I\!P_x\{Z=1\}/(1-v_x)\,.
\end{aligned}
$$

By A2,

$$
\begin{aligned}
I\!P_x\{Z=1\} &= I\!P_x\{V=1\} + I\!P_x\{Z=1|V=0\}I\!P_x\{V=0\} \\
&= v_x + \pi_x(p_x^*)(1-v_x)\,.
\end{aligned}
$$

Substitute this expression for $I\!P_x\{Z=1\}$ into the above expression for $\pi_x(p_x^*)$ and solve for $\pi_x(p_x^*)$ to obtain the result. □

We now establish some preliminary results needed to prove Lemma 2 in Section 4, as well as Theorem 8 and Theorem 9 in Section 3. Recall that for $z \in I\!R^d$, $d \geq 1$, $\kappa_d = \int\{|z| \leq 1\}dz$.

LEMMA 3A. *If D0, D4, D5, and D8 hold, then* $|\hat{v}_x - v_x| = O_p(n^{-2/[d+4]})$ *as* $n \to \infty$.

PROOF. Define

$$
\hat{f}(x) = [na_n^d]^{-1}\sum_{i=1}^{n}\{|X_i - x| \leq a_n\}/\kappa_d\,.
$$

Note that

$$\begin{aligned}\hat{v}_x &= \sum_{i=1}^{n} V_i\{|X_i - x| \le a_n\} / \sum_{i=1}^{n} \{|X_i - x| \le a_n\} \\ &= \tilde{v}_x/(1-\hat{\epsilon}_x) \qquad (9)\end{aligned}$$

where

$$\begin{aligned}\hat{\epsilon}_x &= 1 - \hat{f}(x)/f(x) \\ \tilde{v}_x &= [na_n^d]^{-1} \sum_{i=1}^{n} V_i\{|X_i - x| \le a_n\}/\kappa_d f(x)\,.\end{aligned}$$

Note that

$$\hat{\epsilon}_x = [f(x) - \mathbb{E}\hat{f}(x)]/f(x) + [\mathbb{E}\hat{f}(x) - \hat{f}(x)]/f(x)$$

where $\mathbb{E}\hat{f}(x) = [\kappa_d]^{-1} \int \{|y - x| \le a_n\} f(y)dy$. Change variables, letting $z = (y-x)/a_n$ and apply D4 and D5 along with a 2-term Taylor expansion of $f(x + a_n z)$ about x to see that $\mathbb{E}\hat{f}(x) = f(x) + O(a_n^2)$. Thus, $[f(x) - \mathbb{E}\hat{f}(x)]/f(x) = O(a_n^2)$. Apply D0, D4, D5, and a standard central limit theorem to see that $[\mathbb{E}\hat{f}(x) - \hat{f}(x)]/f(x) = O_p(1/\sqrt{na_n^d})$. Apply D8 to get that $\hat{\epsilon}_x = O_p(n^{-2/[d+4]})$. Deduce that

$$1/(1-\hat{\epsilon}_x) = 1 + O_p(n^{-2/[d+4]})\,. \qquad (10)$$

A similar argument shows that

$$\tilde{v}_x = v_x + O_p(n^{-2/[d+4]})\,. \qquad (11)$$

Deduce from (9), (10), and (11) that $|\hat{v}_x - v_x| = O_p(n^{-2/[d+4]})$. □

Define $g(x) = (1 - v_x)f(x)$. Next, define

$$
\begin{aligned}
\hat{H}_x(u) &= \sum_{i=1}^{n}\{Y_i \le u\}(1-V_i)\{|X_i - x| \le a_n\} / \sum_{i=1}^{n}(1-V_i)\{|X_i - x| \le a_n\} \\
\tilde{H}_x(u) &= [na_n^d]^{-1}\sum_{i=1}^{n}\{Y_i \le u\}(1-V_i)\{|X_i - x| \le a_n\}/\kappa_d g(x)\,.
\end{aligned}
$$

LEMMA 4A. *If D0 and D4 hold, then*

$$
\sup_{u\in\boldsymbol{R}} |\tilde{H}_x(u) - I\!E\tilde{H}_x(u)| = O_p(1/\sqrt{na_n^d})\,.
$$

PROOF. For $y \in I\!R$, $v \in \{0,1\}$, and $x, x_0 \in I\!R^d$, define $\mathcal{F}_n = \{f_n(y,v,x;u \mid x_0) : u \in I\!R\}$ where $f_n(y,v,x;u \mid x_0)$ equals

$$
\{y \le u\}(1-v)\{|x - x_0| \le a_n\}/\kappa_d g(x_0) - I\!E\{Y \le u\}(1-V)\{|X - x_0| \le a_n\}/\kappa_d g(x_0)\,.
$$

Apply Lemma 2.4 and Lemma 2.12 in Pakes and Pollard (1989) to see that $\mathcal{F}_n$ is Euclidean for the envelope $F_n = 2\{|x - x_0| \le a_n\}/\kappa_d g(x_0)$. Use this fact and apply D0, D4, and the Maximal Inequality in Sherman (1994, p.446) with $k = p = 1$ to see that

$$
I\!E\sup_{u\in\boldsymbol{R}} |P_n f_n(\cdot,\cdot,\cdot;u \mid x_0)| = O(I\!E\sqrt{P_{2n}F^2}/\sqrt{n})\,. \tag{12}
$$

Apply Jensen's inequality to get that $I\!E\sqrt{P_{2n}F^2} \le \sqrt{I\!EF^2} = O(\sqrt{a_n^d})$. The result follows from dividing both sides of (12) by a_n^d and then applying Chebyshev's inequality. □

LEMMA 5A. *If D0, D4, D5, D6, and D8 hold, then* $\sup_{u\in\mathbf{R}} |\hat{H}_x(u) - H_x(u)| = O_p(n^{-2/[d+4]})$.

PROOF. Recall that $g(x) = (1 - v_x)f(x)$ and define

$$\hat{g}(x) = [na_n^d]^{-1} \sum_{i=1}^{n} (1 - V_i)\{|X_i - x| \leq a_n\}/\kappa_d \,.$$

Note that

$$\hat{H}_x(u) = \tilde{H}_x(u)/(1 - \hat{\epsilon}_x) \tag{13}$$

where $\hat{\epsilon}_x = 1 - \hat{g}(x)/g(x)$. Argue as in the proof of Lemma 3A to get that

$$1/(1 - \hat{\epsilon}_x) = 1 + O_p(n^{-2/[d+4]})\,. \tag{14}$$

Apply Lemma 4A to get that

$$\sup_{u\in\mathbf{R}} |\tilde{H}_x(u) - \mathbb{E}\tilde{H}_x(u)| = O_p(1/\sqrt{na_n^d})\,. \tag{15}$$

Apply D4 and D6 and argue as in Lemma 3A to get that

$$\sup_{u\in\mathbf{R}} |\mathbb{E}\tilde{H}_x(u) - H_x(u)| = O_p(a_n^2)\,. \tag{16}$$

Apply (13), (14), (15), and (16), together with D8 to get the result. □

Let $\{\gamma_n\}$ denote a sequence of postive real numbers converging to zero as $n \to \infty$. Let Θ_n

denote the product space $[0,1] \times [0,\gamma_n]$. Define

$$\begin{aligned} S_x(p,\gamma) &= \mathbb{E}_x[|Y|\{Q_x(\pi_x(p)) \le Y \le Q_x(\pi_x(p)+\gamma)\} \mid V=0] \\ T_n(x) &= \sup_{(p,\gamma)\in\Theta_n} S_x(p,\gamma)\,. \end{aligned}$$

LEMMA 6A. *If D1 through D3 hold, then $T_n(x) = O(\gamma_n)$ as $n \to \infty$.*

PROOF. Notes: Assumption D2 is not needed for this result when $|Y|$ is bounded, as it is in the Love Canal and test score examples. To see this, note that $S_x(p,\gamma)$ is increasing in γ for fixed p, and so $T_n(x) = \sup_{p\in[0,1]} S_x(p,\gamma_n)$. Now,

$$S_x(p,\gamma_n) = \int_{Q_x(\pi_x(p))}^{Q_x(\pi_x(p)+\gamma_n)} |y|dH_x(y)\,.$$

Let M denote a bound on $|Y|$. Then

$$\begin{aligned} S_x(p,\gamma_n) &\le M\int_{Q_x(\pi_x(p))}^{Q_x(\pi_x(p)+\gamma_n)} dH_x(y) \\ &\le M\gamma_n\,. \end{aligned}$$

That does it! Also, note that if you want to just establish a pointwise (in p) rate of convergence of $S_x(p,\gamma_n)$, then the above argument can be adapted to handle the case where Y is unbounded. Simply let M denote a bound for Y over the interval $[Q_x(\pi_x(p)), Q_x(\pi_x(p)+\gamma_n)]$ and the argument goes through as before. This is used in the Love Canal example to prove sharpness in Theorem 3.

The argument that follows proves the result when Y is unbounded and uses D2, but probably doesn't need it.

Since $S_x(p,\gamma)$ is increasing in γ for fixed p, $T_n(x) = \sup_{p\in[0,1]} S_x(p,\gamma_n)$. Note

$$S_x(p,\gamma_n) = \int_{Q_x(\pi_x(p))}^{Q_x(\pi_x(p)+\gamma_n)} |y| h_x(y) dy \, .$$

By D2, $h_x(y)$ is continuous in y and $Q_x(t)$ is continuous in t. It follows that $S_x(p,\gamma_n)$ is a continuous function of p on $[0,1]$ and so must attain its maximum value on $[0,1]$. Thus, there exists $p_n(x) \in [0,1]$ such that $T_n(x) = S_x(p_n(x),\gamma_n)$. Note that $S_x(p,0) = 0$ for all $p \in [0,1]$. Assumption D2 also implies that $Q_x(t)$ is differentiable in t, implying that $S_x(p,\gamma)$ is a differentiable function of γ for each $p \in [0,1]$. Taylor expand $S_x(p_n(x),\gamma_n)$ about $\gamma = 0$ to see that

$$T_n(x) = \gamma_n \frac{\partial}{\partial\gamma} S_x(p_n(x),\gamma^*)$$

where $\gamma^* \in [0,\gamma_n]$. Since $\gamma_n \to 0$ as $n \to \infty$, the result will follow provided $\frac{\partial}{\partial\gamma} S_x(p_n(x),\gamma)$ is bounded in a neighborhood of $\gamma = 0$. By D1, $S_x = (a,b)$ for $-\infty \le a < b \le \infty$. Simple calculus shows that

$$\frac{\partial}{\partial\gamma} S_x(p_n(x),\gamma) = Q_x(\pi_x(p_n(x)) + \gamma) \, .$$

Thus, $\frac{\partial}{\partial\gamma} S_x(p_n(x),\gamma)$ is bounded in a neighborhood of $\gamma = 0$ provided $\pi_x(p_n(x))$ is bounded away from zero and unity, or, equivalently, provided $p_n(x)$ is bounded below unity when $a = -\infty$ and above v_x when $b = \infty$. Since $(p_n(x),\gamma_n)$ maximizes $S_x(p,\gamma)$ over Θ_n, these conditions on $p_n(x)$ must hold, otherwise either $h_x(-\infty) > 0$ or $h_x(\infty) > 0$, contradicting D3. This proves the result.□

LEMMA 2. *If D0 through D8 hold, then for* $d \ge 0$, $\sup_{p\in[0,1]} |\hat{L}_x(p) - L_x(p)| = O_p(n^{-2/[d+4]})$ *and* $\sup_{p\in[0,1]} |\hat{U}_x(p) - U_x(p)| = O_p(n^{-2/[d+4]})$ *as* $n \to \infty$.

PROOF. We prove the result for the lower bounds. The proof for the upper bounds is similar.

Also, we will prove the result for the case $d \geq 1$. The proof for the case $d = 0$ will follow by taking $\kappa_0 = 1$, replacing $f(x)$ and $\{|X_i - x| \leq a_n\}$ with unity, and removing all dependence on x from all the notation and arguments in the paper.

Define

$$\begin{aligned}
\hat{L}_{1x}(p) &= p \sum_{i=1}^{n} M(Y_i) V_i \{|X_i - x| \leq a_n\}/n_{1x} \\
&+ (1-p) \sum_{i=1}^{n} M(Y_i)(1 - V_i)\{Y_i \leq \hat{Q}_x(\hat{\pi}_x(p))\}\{|X_i - x\} \leq a_n\}/\pi_x(p) n_{0x} \\
\hat{L}_{2x}(p) &= p \sum_{i=1}^{n} M(Y_i) V_i \{|X_i - x| \leq a_n\}/n_{1x} \\
&+ (1-p) \sum_{i=1}^{n} M(Y_i)(1 - V_i)\{Y_i \leq Q_x(\pi_x(p))\}\{|X_i - x\} \leq a_n\}/\pi_x(p) n_{0x} \,.
\end{aligned}$$

Note that

$$|\hat{L}_x(p) - L_x(p)| \leq |\hat{L}_x(p) - \hat{L}_{1x}(p)| + |\hat{L}_{1x}(p) - \hat{L}_{2x}(p)| + |\hat{L}_{2x}(p) - L_x(p)| \,. \tag{17}$$

A straightforward argument using D0, D3, and Lemma 3A shows that the first term on the right-hand side of (17) has order $O_p(n^{-2/[d+4]})$ uniformly over p in $[0, 1]$. Arguments similar to those given in Lemma 3A and Lemma 4A show that the third term on the right-hand side of (17) has order $O_p(n^{-2/[d+4]})$ uniformly over p in $[0, 1]$. Consider the second term on the right-hand side of (17). Write $\hat{\delta}_x(u,p)$ for $\hat{\pi}_x(p) - \pi_x(p) + H_x(u) - \hat{H}_x(u)$. Again, Lemma 3A and simple calculus imply that $|\hat{\pi}_x(p) - \pi_x(p)| = O_p(n^{-2/[d+4]})$ uniformly over $p \in [0, 1]$. Deduce from this and Lemma 5A that $|\hat{\delta}_x(u,p)| = O_p(n^{-2/[d+4]})$ uniformly over $u \in S_x$ and $p \in [0, 1]$. For notational convenience, assume, without loss of generality, that $\hat{\delta}_x(u,p)$ is equal to its positive part. A similar argument works for the negative part of $\hat{\delta}_x(u,p)$. Let $c_x = (1 - v_x)/v_x$. After some simple algebra, we get that

$$\hat{L}_{1x}(p) - \hat{L}_{2x}(p) = pc_x \frac{\sum_{i=1}^n Y_i(1-V_i)\{Q_x(\pi_x(p)) < Y_i \le Q_x(\pi_x(p) + \hat{\delta}_x(Y_i,p))\}\{|X_i - x| \le a_n\}}{\sum_{i=1}^n (1-V_i)\{|X_i - x| \le a_n\}}\,.$$

Recall that $g(x) = (1-v_x)f(x)$. For $y \in I\!R$, $v \in \{0,1\}$, $x, x_0 \in I\!R^d$, $p \in [0,1]$, and $\gamma \ge 0$, define $f_n(y,v,x,p,\gamma \mid x_0)$ to be equal to

$$|y|(1-v)\{Q_x(\pi_x(p_0)) < y \le Q_x(\pi_x(p_0)+\gamma)\}\{|x-x_0| \le a_n\}/\kappa_d g(x_0)\,.$$

Recall that $\hat{g}(x) = [na_n^d]^{-1}\sum_{i=1}^n (1-V_i)\{|X_i - x| \le a_n\}/\kappa_d$. Define $\Theta = S_x \times [0,1]$ and $\hat{\delta}_x = \sup_{(u,p)\in\Theta} \hat{\delta}_x(u,p)$. Since f_n is increasing in γ for fixed y, v, x, p, and x_0, we see that

$$|\hat{L}_{1x}(p) - \hat{L}_{2x}(p)| \le [na_n^d]^{-1}\sum_{i=1}^n f_n(Y_i, V_i, X_i, p, \hat{\delta}_x \mid x) pc_x/|1-\hat{\epsilon}_x|$$

where $\epsilon_x = 1 - \hat{g}(x)/g(x)$. Define

$$\tilde{F}_x(p,\gamma) = [na_n^d]^{-1}\sum_{i=1}^n f_n(Y_i, V_i, X_i, p, \gamma \mid x)\,.$$

Since $\hat{\delta}_x = O_p(n^{-2/[d+4]})$ and $\tilde{F}_x(p,\gamma)$ is increasing in γ for each fixed p, there exists a sequence $\{\gamma_n\}$ of positive real numbers of order $O(n^{-2/[d+4]})$ such that $wp \to 1$ as $n \to \infty$,

$$|\hat{L}_{1x}(p) - \hat{L}_{2x}(p)| \le \tilde{F}_x(p,\gamma_n)pc_x/|1-\hat{\epsilon}_x|\,. \tag{18}$$

Argue as in the proof of Lemma 3A to see that

$$pc_x/|1-\hat{\epsilon}_x| = pc_x + O_p(n^{-2/[d+4]})\,. \tag{19}$$

Recall the definitions of $S_x(p,\gamma)$ and $T_n(x)$. We have that $wp \to 1$ as $n \to \infty$,

$$\begin{aligned}
\sup_{p\in[0,1]} |\tilde{F}_x(p,\gamma_n)| \;\le\; & \sup_{p\in[0,1]} |\tilde{F}_x(p,\gamma_n) - I\!E\tilde{F}_x(p,\gamma_n)| \\
+ \; & \sup_{p\in[0,1]} |I\!E\tilde{F}_x(p,\gamma_n) - S_x(p,\gamma_n)| \\
+ \; & T_n(x)\,.
\end{aligned}$$

Argue as in the proof of Lemma 4A that the first term above has order $O_p(1/\sqrt{na_n^d})$. Apply D7 and argue as in the proof of Lemma 3A that the second term above has order $O(a_n^2)$. Finally, apply D8 and Lemma 6A to see that

$$\sup_{p\in[0,1]} |\tilde{F}_x(p,\gamma_n)| = O_p(n^{-2/[d+4]})\,. \tag{20}$$

Apply (18), (19), and (20) to see that

$$|\hat{L}_{1x}(p) - \hat{L}_{2x}(p)| = O_p(n^{-2/[d+4]})\,.$$

This proves the result. □

THEOREM 8. *If C0 through C3 hold, then*

$$\sqrt{na_n^d}\left[\hat{\lambda}_1(x) - \lambda_1(x)\right] \Longrightarrow N(\mu_1(x), \sigma_1^2(x))$$

where

$$\mu_1(x) = \int [G_1(x)'zz'f_1(x) + z'G_2(x)z + z'f_2(x)z]\,\{|z| \le 1\}dz/\kappa_d f(x)$$

$$\sigma_1^2(x) = \mathbb{E}_x[M(Y)V + a(1-V)]^2/\kappa_d f(x)\,.$$

Proof. Recall the definition of $\hat{f}(x)$ in the proof of Lemma A1 and note that

$$\hat{\lambda}_1(x) = \tilde{\lambda}_1(x)/(1-\hat{\epsilon}_x)$$

where

$$\hat{\epsilon}_x = 1 - \hat{f}(x)/f(x)$$

$$\tilde{\lambda}_1(x) = [na_n^d]^{-1}\sum_{i=1}^{n}[M(Y_i)V_i + a(1-V_i)]\{|X_i - x| \le a_n\}/\kappa_d f(x)\,.$$

Apply C0, C1, and C2, and argue as in the proof of Lemma A1 that

$$1/(1-\hat{\epsilon}_x) = 1 + O_p(n^{-2/d+4})$$

$$\tilde{\lambda}_1(x) = \lambda_1(x) + a_n^2\mu_1(x) + o_p(a_n^2)\,.$$

Apply C3 and a standard central limit theorem to get $\sqrt{na_n^d}\left[\tilde{\lambda}_1(x) - \lambda_1(x)\right] \Longrightarrow N(\mu_1(x), \sigma_1^2(x))$.
Apply Slutsky's Theorem to get the result. □

THEOREM 9. *If C0′, C1, C2′, and C3 hold, then*

$$\sqrt{na_n^d}\,[\hat{u}_2(x) - u_2(x)] \Longrightarrow N(\mu_2(x), \sigma_2^2(x))$$

where

$$\begin{aligned} \mu_2(x) &= \int [G_1(x)'zz'f_1(x) + z'G_2(x)z + z'f_2(x)z]\,\{|z| \le 1\}dz/\kappa_d v_x \\ \sigma_2^2(x) &= \mathbb{E}_x[M(Y) \mid V = 1]^2/\kappa_d v_x\,. \end{aligned}$$

PROOF. Mimic the proof of Theorem 8. □

REFERENCES

AMEMIYA, T. (1985): *Advanced Econometrics.* Harvard Univ. Press, Cambridge, Mass.

HOROWITZ, J., AND C. MANSKI (1995): "Identification and Robustness with Contaminated and Corrupted Data," *Econometrica*, 63, 281–302.

HOROWITZ, J., AND C. MANSKI (1996): "What Can Be Learned about Population Parameters when the Data are Contaminated", in *The Handbook of Statistics: Robust Inference*, Vol.15, C.R. Rao and G.S. Maddala, eds., North Holland, Amsterdam.

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.

HOROWITZ, J., AND C. MANSKI (2000): "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data", *Journal of the American Statistical Association*, 95, No. 449, 77–84.

LAMBERT, D., PETERSON, B., AND I. TERPENNING (1991): "Nondetects, Detection Limits, and the Probability of Detection," *Journal of the American Statistical Association*, 86, 266–277.

LAMBERT, D. AND L. TIERNEY (1997): "Nonparametric Maximum Likelihood Estimation from Samples with Irrelevant Data and Verification Bias," *Journal of the American Statistical Association*, 92, 937–944.

MANSKI, C. (1990): "Nonparametric Bounds on Treatment Effects", *American Economic Review Papers and Proceedings*, 80, 319–323.

MANSKI, C. (1995): *Identification Problems in the Social Sciences.* Harvard University Press, Cambridge, Mass.

MOLINARI, F. (2001): "Missing Treatments," Manuscript, Northwestern University, July 16.

PAKES, A., AND D. POLLARD (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027–1057.

SHERMAN, R. P. (1994): "Maximal Inequalities for Degenerate U-processes with Applications to Optimization Estimators," *Annals of Statistics*, 22, 439–459.