IDENTIFICATION AND ESTIMATION
WITH CONTAMINATED AND PARTIALLY VERIFIED DATA

BY JEFF DOMINITZ* AND ROBERT P. SHERMAN† 1

*Carnegie Mellon University  †California Institute of Technology

June 5, 2001

Abstract

A sample is said to be contaminated if it is drawn from a mixture of a distribution of interest, $F$, and another distribution, $G$. Without parametric assumptions on $F$ and $G$, characteristics of $F$ such as moments and quantiles are not identified. However, given a lower bound on the probability of being drawn from $F$, Horowitz and Manski (1995) show that finite bounds on these characteristics are identified and easily estimated. In some applications, more information is available: a subset of the sample is known to be drawn from $F$. Often these verified observations are not a random sample from $F$ and so point estimation based only on verified observations is biased. The Horowitz and Manski procedure applies but is inefficient because it makes no use of the extra information on verification. This paper adapts the method of Horowitz and Manski to develop tighter bounds using the extra information. In addition, the method is more direct and easier to generalize to other models than an alternative method proposed by Lambert and Tierney (1997). The estimator is shown to be consistent and is applied to measurements of organic pollutant concentrations.

1. Introduction

Suppose that with probability $\rho > 0$ observations are randomly drawn from a distribution

1We thank Diane Lambert for providing the organic pollutant data analyzed in this paper. We also thank Arie Kapteyn, Chuck Manski, and the seminar participants at Northwestern University for helpful comments and suggestions.
of interest $F$, and with probability $1 - \rho$ they are drawn from another distribution $G$. Let $H$ denote the resulting mixture distribution, and let $X$ denote a random variable with distribution $H$. Let $(X_1, \ldots, X_n)$ denote a sample of independent observations from $H$. We say that this sample is contaminated.

Horowitz and Manski (1995) (hereafter referred to as HM) show that various characteristics of $F$, such as the mean, are not identified under this model. However, they show that under minimal assumptions on the underlying distributions, finite bounds on these characteristics are identified. For example, should $\rho$ be known to exceed some positive threshold probability $\tau$, bounds can be obtained on the mean of $F$. The lower bound is the mean of $X$ truncated from above at $Q(\tau)$, where $Q$ is the quantile function corresponding to $H$. The upper bound is the mean of $X$ truncated from below at $Q(1 - \tau)$. Similar bounds can be obtained for any parameter of the distribution that respects stochastic dominance, e.g., the quantiles of $F$ (see HM, p.290). HM show that these bounds are sharp, in the sense that they exhaust the information about the parameters that is available from the sampling process and the maintained assumptions. Nonparametric analog estimates of these bounds for means and quantiles are easily found.

In some applications, more is known than simply a lower bound on $\rho$. For example, consider the Love Canal data analyzed by Lambert and Tierney (1997). Measurements of low concentrations of organic pollutants are hypothesized to come from the mixture model described above. If the wrong pollutant is isolated by the measuring instruments, spurious measurements result. However, in addition to observations from the mixture distribution, each measurement undergoes an imperfect verification test. If a measurement passes the test, it is known to be drawn from $F$. If it fails the test, it could be a draw from either $F$ or $G$. Often, unverified measurements are thrown out, inducing upward biases in estimates of mean concentrations since lower
concentrations are harder to verify. The HM bounds can be applied using all the measurements with the proportion $v$ of verified observations playing the role of $\tau$. However, these bounds will not be sharp, because they do not exploit the information on verified measurements.

As another example, consider the recent controversy over release of the 1997 California public school test scores. In 1997, 4.1 million California public school students in grades 2 through 11 took the Stanford 9 standardized tests in English. Of these, over 1 million students were classified as having limited English skills, though they were deemed proficient enough in the language to take the test. Some school officials support using only the scores of those students who are classified as fluent in English to assess the educational standing of individual schools, districts, and the state as a whole. Others, including then governor Pete Wilson, want all the scores to be used. It seems natural to model test scores for a particular school or region with a contamination model in which some of the students classified as having limited English skills do not provide valid scores, whereas others in this group do provide valid scores. Under this model, the average score of students classified as fluent in English will likely overestimate the average valid test score, whereas the average score of all students will likely underestimate this parameter. In short, the average valid test score is not an identifiable parameter in this model. However, finite bounds on this parameter are identified and easily estimated using the method of HM. As with the Love Canal data, these bounds will not be sharp since they do not exploit the fact that the scores of students who are classified as fluent in English can reasonably be assumed to be valid. In other words, scores of students fluent in English can be viewed as verified draws from the distribution of interest.

Empirical researchers rarely employ interval estimators in practice. Rather, when faced with situations like those described above, point estimation is the norm. Under the assumption
of exogenous verification, the probability of verifying an actual draw from $F$ is independent of its value. In this case, the mean of the verified observations is an unbiased estimate of the population mean of interest. However, under endogenous verification, where the verification probability varies arbitrarily with the value of a draw from $F$ (as is likely in the two examples described above), Lambert and Tierney (1997) have shown that only interval estimates may be obtained without parametric restrictions on the distributions $F$ and $G$ and on the verification probability as a function of the observed value.

Like HM, Lambert and Tierney (1997) derive a nonparametric maximum likelihood estimator for upper and lower endpoints of this interval. However, the Lambert and Tierney estimator exploits more fully the information on verification. Their approach involves kernel density estimation of bounds on the verification probability conditional on the value of $X$. These bounds are then used to place nonparametric restrictions on the distribution of interest $F$. Once this distribution has been restricted, bounds on the mean can be calculated. This approach is quite appealing and can be used to generate bounds for other population parameters of interest.

However, there are some difficulties with the Lambert and Tierney approach. For example, with the Love Canal data it is natural to assume that the verification probability increases with the value of a draw from $F$. Under this monotonicity assumption, if no unverified value is exactly equal to the smallest verified value (as must happen, for example, with continuously distributed data) then, without an adhoc adjustment, their procedure does not provide an estimate of a lower bound for the mean of $F$. In addition, their procedure requires that upper and lower bounds on the verification probability be estimated at each observed value of $X$. This requirement cannot be met for observations that are censored below some threshold value, as is the case with the Love Canal data. Also, their procedure for estimating the mean of $F$, for
example, seems somewhat indirect, being analogous to estimating the population mean from an uncontaminated random sample by first obtaining nonparametric density estimates and then integrating over the empirical distribution.

This paper adapts the HM approach to develop a direct estimator of bounds on the mean of $F$. The method can be used to develop bounds on other characteristics of $F$ such as arbitrary moments, probabilities, and quantiles, and exploits all the information in the sample on verification, whether exogenous or endogenous. Further, it is both easy and natural to incorporate not only monotonicity in verification probabilities but also censoring into the estimation procedure. Finally, the procedure can be generalized in a straightforward manner to incorporate the effects of covariates on the response variable.

The rest of the paper is organized as follows. The model of the data-generating process is presented in Section 2. Bounds on the mean are derived in Section 3. We also indicate how to bound other characteristics of $F$. Nonparametric analog estimators of the bounds are presented in Section 4 and shown to be consistent in Section 5. We also prove that these bounds must be contained within the HM bounds. In Section 6, we apply the estimation method to the environmental pollutant concentration data analyzed by Lambert and Tierney and compare our bounds to theirs as well as to the HM bounds. Section 7 summarizes and discusses directions for future work.

2. Model of the Data-Generating Process

Recall that $F$ denotes the unknown distribution of interest, $G$ a spurious distribution, and $\rho > 0$ the probability of a draw from $F$. Let $Y_1$ denote a random variable with distribution $F$, $Y_0$ a random variable with distribution $G$, and $Z$ a 0-1 random variable with $\mathbb{P}(Z = 1) = \rho$. 

5
Z indicates whether or not an observation is drawn from \( F \). We do not observe \( Z \). Rather, we observe \( Y = Y_1 Z + Y_0 (1 - Z) \). In addition, we observe \( V \), a partial verification indicator. If \( V = 1 \) then \( Z = 1 \) implying that \( Y \) is drawn from \( F \). If \( V = 0 \), then \( Z \) could equal 1 or 0, implying that \( Y \) could be drawn from \( F \) or \( G \). Let \((Y, V, Z)\) denote a generic draw from this model.

We are interested in learning about characteristics of the distribution of \( F \), the conditional distribution of \( Y \) given \( Z = 1 \). For expositional purposes, we focus on developing bounds on \( E[Y \mid Z = 1] \), the mean of \( F \). However, as we will show, the technique can also be used to bound other characteristics of \( F \) such as moments, probabilities, and quantiles.

Write \( v \) for \( P(V = 1) \). The following conditions summarize what we assume about contamination and verification:

- **C1.** \( 0 < v < 1 \).
- **C2.** \( P(Z = 1 \mid V = 1) = 1 \).

Condition C1 says that we have partial verification. C2 says that if an observation is verified, then it must be from \( F \).

Lambert and Tierney (1997) find that point identification of \( E[Y \mid Z = 1] \) is not feasible without parametric assumptions on \( F, G \), and the conditional verification probability \( P(V = 1 \mid Z = 1, Y) \). They also consider estimation under the following monotonicity assumption:

\[
P(V = 1 \mid Z = 1, Y = y_1) \geq P(V = 1 \mid Z = 1, Y = y_0) \quad y_1 > y_0.
\]

This assumption is plausible in the context of the Love Canal data that they analyze.

Conditions C1 and C2 allow for either endogenous or exogenous verification. Allowing endogenous verification is less restrictive than condition (1), but this latter condition will be
considered as well. Point identification of $E[Y|Z = 1]$ is not feasible in either case. Note also that condition (1) implies the following ordered outcomes condition:

$$E[Y|Z = 1, V = 0] \leq E[Y|Z = 1, V = 1].$$

3. Identification

In this section, we first show how to identify bounds on the mean of $F$ using conditions C1 and C2 from Section 2 and the method of HM which does not exploit partial verification. We then adapt the HM method to develop tighter bounds that exploit all the information on verification.

Recall that $Y$ has distribution $H$, a mixture of $F$ and $G$ with mixing parameter $\rho$, the unknown probability of a draw from $F$. Our objective is to identify finite bounds on the mean of $F$, $E[Y|Z = 1]$. By condition C2, all verified data are drawn from $F$. Thus $v = P(V = 1)$ is a lower bound on $\rho$. This lower bound on $\rho$ is all that is needed to construct the HM bounds. Write $Q$ for the quantile function corresponding to $H$. By Proposition 4 in HM we obtain the following bounds:

$$E[Y|Y \leq Q(v)] \leq E[Y|Z = 1] \leq E[Y|Y > Q(1 - v)].$$

Similar bounds can be obtained for moments and quantiles of $F$.

We now develop tighter bounds on $E[Y|Z = 1]$ using a related approach. By the law of total probability, $E[Y|Z = 1]$ equals

$$E[Y|Z = 1, V = 1]P(V = 1|Z = 1) + E[Y|Z = 1, V = 0]P(V = 0|Z = 1).$$

By C2, the first conditional expectation in (3) is identified by the data, because all verified data
are drawn from $F$. Therefore,

$$\mathbb{E}[Y|Z = 1, V = 1] = \mathbb{E}[Y|V = 1].$$  \hspace{1cm} (4)$$

Note also that the two conditional probabilities in (3) are bounded based on the associated unconditional probabilities. For example,

$$v = \mathbb{P}(V = 1) \leq \mathbb{P}(V = 1|Z = 1) \leq 1.$$  \hspace{1cm} (5)$$

Next, we obtain bounds on the quantity $\mathbb{E}[Y|Z = 1, V = 0]$ in (3). By the argument given at the beginning of this section, we can develop HM-type bounds for this conditional mean if we can get a lower bound on $\mathbb{P}(Z = 1|V = 0)$, the probability of a draw from $F$ within the sample of unverified data. The next result gives an expression for $\mathbb{P}(Z = 1|V = 0)$ in terms of the unknown quantity $\mathbb{P}(V = 1|Z = 1)$. Write $p^*$ for $\mathbb{P}(V = 1|Z = 1)$ and $\pi(p^*)$ for $\mathbb{P}(Z = 1|V = 0)$. Note that by (5), $p^* \in [v, 1]$.

**Lemma 1:** If C1 and C2 hold, then $\pi(p^*) = [(1 - p^*)v]/[p^*(1 - v)]$.

**Proof:** By C1, $v < 1$. By Bayes’ Rule,

$$\pi(p^*) = \mathbb{P}(V = 0|Z = 1)\mathbb{P}(Z = 1)/\mathbb{P}(V = 0)$$

$$= (1 - p^*)\mathbb{P}(Z = 1)/(1 - v).$$

By the law of total probability and C2,

$$\mathbb{P}(Z = 1) = \mathbb{P}(V = 1) + \mathbb{P}(Z = 1|V = 0)\mathbb{P}(V = 0)$$

$$= v + \pi(p^*)(1 - v).$$

Substitute this expression for $\mathbb{P}(Z = 1)$ into the above expression for $\pi(p^*)$ and solve for $\pi(p^*)$ to obtain the result. \hfill \Box
Now consider the conditional distribution of \( Y \) given \( V = 0 \). Let \( H_0 \) denote the cdf of this distribution and \( Q_0 \) the corresponding quantile function. Since \( \pi(p^*) \) equals \( P(Z = 1|V = 0) \), it is also a lower bound on this probability. Substitute \( H_0 \) for \( H \), \( Q_0 \) for \( Q \), and \( \pi(p^*) \) for \( v \) in the argument given at the start of this section to obtain the following bounds:

\[
\mathbb{E}[Y|Y \leq Q_0(\pi(p^*)), V = 0] \leq \mathbb{E}[Y|Z = 1, V = 0] \leq \mathbb{E}[Y|Y > Q_0(1 - \pi(p^*)), V = 0].
\]

(6)

For each \( p \in (0, 1] \), define \( \pi(p) = [(1 - p)v]/[p(1 - v)] \). Next, define the population lower and upper bound functions

\[
L(p) = p\mathbb{E}[Y|V = 1] + (1 - p)\mathbb{E}[Y|Y \leq Q_0(\pi(p)), V = 0]
\]

\[
U(p) = p\mathbb{E}[Y|V = 1] + (1 - p)\mathbb{E}[Y|Y > Q_0(1 - \pi(p)), V = 0].
\]

Since \( \pi(v) = 1 \) and \( \pi(1) = 0 \), we see that \( L(v) = U(v) \) and \( L(1) = U(1) \). It follows immediately from the definitions of \( L(p) \) and \( U(p) \) that \( L(p) \leq U(p) \) for each \( p \in [v, 1] \). We show in the appendix that under mild conditions, \( L(p) \) is strictly convex on \([v, 1]\) and \( U(p) \) is strictly concave on \([v, 1]\). Thus, under these conditions, for each \( p \in (v, 1) \), \( L(p) < U(p) \).

Deduce from (3), (4), and (6) that \( L(p^*) \) and \( U(p^*) \) are lower and upper bounds for the mean \( \mathbb{E}[Y|Z = 1] \). Since we do not know \( p^* \), these bounds are infeasible. However, since \( p^* \in [v, 1] \), we can still obtain bounds (albeit wider ones) on \( \mathbb{E}[Y|Z = 1] \) by taking the infimum of \( L(p) \) and the supremum of \( U(p) \) over \( p \in [v, 1] \). That is, the population lower and upper bounds, denoted \( \lambda \) and \( u \), are given by

\[
\lambda = \inf_{p \in [v, 1]} L(p)
\]

\[
u = \sup_{p \in [v, 1]} U(p).
\]
These parameters bound the mean of $F$ since

$$\lambda \leq L(p^*) \leq \mathbb{E}[Y|Z = 1] \leq U(p^*) \leq u.$$  

Bounds on other characteristics of $F$ can be developed similarly. For example, let $f(\cdot)$ denote a known, real-valued function of a real variable. Take expectations of $f(Y)$ rather than $Y$ in the preceding arguments to obtain bounds on $\mathbb{E}[f(Y)|Z = 1]$. For example, take $f(Y) = Y^k$, $k \geq 1$, to obtain bounds on any positive integer moment of $F$. Combine bounds on the first two moments of $F$ to get bounds on the variance of $F$. As another example, fix $t \in \mathbb{R}$ and take $f(Y) = \{Y \leq t\}$ to obtain bounds on the probability $F(t) = \mathbb{P}(Y \leq t|Z = 1)$. Bounds on the cdf can then be inverted to obtain bounds on the quantiles of $F$.

4. Estimation

Let $(Y_1, V_1, Z_1), \ldots, (Y_n, V_n, Z_n)$ denote independent draws from the model described in Section 2. Define $n_1 = \sum_{i=1}^n V_i$, the number of verified $Y_i$'s, and $n_0 = n - n_1$, the number of unverified $Y_i$'s. Define $\hat{v} = n_1/n$. For each $p \in (0, 1]$, define

$$\hat{\pi}(p) = [(1 - p)\hat{v}] / [p(1 - \hat{v})].$$  

(7)

For convenience, write $V_1, \ldots, V_{n_1}$ for verified observations and $U_1, \ldots, U_{n_0}$ for unverified observations. Let $\mathcal{V}$ have the same distribution as $V_i$ and $\mathcal{U}$ the same distribution as $U_i$. Write $H_0$ for the cdf of $\mathcal{U}$ and $\hat{H}_0$ for the corresponding empirical cdf. Also, write $Q_0$ for the quantile function for $\mathcal{U}$ and $\hat{Q}_0$ for the corresponding empirical quantile function.

By definition of the empirical quantile function $\hat{Q}_0$, $\hat{\pi}(p) = n_0^{-1} \sum_{i=1}^{n_0} \{U_i \leq \hat{Q}_0(\hat{\pi}(p))\}$ when $\hat{Q}_0(\hat{\pi}(p))$ is equal to one of the $U_i$'s. Define the sample lower and upper bound functions

$$\hat{L}(p) = p \frac{1}{n_1} \sum_{i=1}^{n_1} V_i + (1 - p) \frac{1}{n_0} \sum_{i=1}^{n_0} U_i \{U_i \leq \hat{Q}_0(\hat{\pi}(p))\} / \hat{\pi}(p)$$

10
\[
\hat{U}(p) = p \frac{1}{n_1} \sum_{i=1}^{n_1} V_i + (1 - p) \frac{1}{n_0} \sum_{i=1}^{n_0} \mathcal{U}_i \{ \mathcal{U}_i > \hat{Q}_0(1 - \hat{\pi}(p)) \} / \hat{\pi}(p).
\]

Note that \(\hat{\pi}(\hat{v}) = 1\) and \(\hat{\pi}(1) = 0\). Thus, \(\hat{L}(\hat{v}) = \hat{U}(\hat{v})\) and \(\hat{L}(1) = \hat{U}(1)\). Moreover, it follows immediately from the definitions of \(\hat{L}(p)\) and \(\hat{U}(p)\) that \(\hat{L}(p) \leq \hat{U}(p)\) for all \(p \in [\hat{v}, 1]\).

Finally, define the extreme value estimators

\[
\hat{\lambda} = \inf_{p \in [\hat{v}, 1]} \hat{L}(p)
\]
\[
\hat{a} = \sup_{p \in [\hat{v}, 1]} \hat{U}(p).
\]

Next, consider estimation with covariates. Let \((Y_1, V_1, X_1, Z_1), \ldots, (Y_n, V_n, X_n, Z_n)\) denote an iid sample of draws from the model defined previously where \(X\) is a \(d\)-dimensional covariate, \(d \geq 1\). 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 \(n(x) = \sum_{i=1}^{n} \{ |X_i - x| \leq a_n \}\), \(n_1(x) = \sum_{i=1}^{n} V_i \{ |X_i - x| \leq a_n \}\), and \(n_0(x) = n(x) - n_1(x)\). Define \(v(x) = \mathbb{P}(V = 1 \mid X = x)\) and \(\hat{v}(x) = n_1(x) / n(x)\). Define \(\pi(p \mid x) = [(1 - p)v(x)]/[p(1 - v(x))]\) and \(\hat{\pi}(p \mid x) = [(1 - p)\hat{v}(x)]/[p(1 - \hat{v}(x))]\). For convenience, write \(\mathcal{V}_i(x), i = 1, \ldots, n_1(x)\), for verified observations within \(a_n\) of \(x\). Write \(\mathcal{U}_i(x), i = 1, \ldots, n_0(x)\), for unverified observations within \(a_n\) of \(x\). Let \(\mathcal{V}(x)\) have the same distribution as \(\mathcal{V}_i\) given \(X = x\). Let \(\mathcal{U}(x)\) have the same distribution as \(\mathcal{U}_i\) given \(X = x\). Let \(H_0(u \mid x) = \mathbb{P}(\mathcal{U} \leq u \mid X = x)\) and let \(\hat{H}_0(u \mid x) = [n_0(x)]^{-1} \sum_{i=1}^{n_0(x)} \{ \mathcal{U}_i(x) \leq u \}\). Let \(h_0(u \mid x) = \frac{\partial}{\partial u} H_0(u \mid x)\). Define \(Q_0(\cdot \mid x)\) to be the conditional quantile function of \(\mathcal{U}\) given \(X = x\). Let \(\hat{Q}_0(\cdot \mid x)\) denote the conditional quantile function of the \(\mathcal{U}_i\)’s corresponding to the \(X_i\)’s within \(a_n\) of \(x\).

By definition of \(\hat{Q}_0(\cdot \mid x)\), \(\hat{\pi}(p \mid x) = [n_0(x)]^{-1} \sum_{i=1}^{n_0(x)} \{ \mathcal{U}_i(x) \leq \hat{Q}_0(\hat{\pi}(p \mid x)) \}\) when \(\hat{Q}_0(\cdot \mid x)\)
is equal to one of the $\mathcal{U}_i(x)$'s. Define the sample lower and upper bound functions

$$
\hat{L}(p \mid x) = p \frac{1}{n_1(x)} \sum_{i=1}^{n_1(x)} V_i(x) + (1 - p) \frac{1}{n_0(x)} \sum_{i=1}^{n_0(x)} \mathcal{U}_i(x) \{ \mathcal{U}_i(x) \leq \hat{Q}_0(\hat{\pi}(p \mid x) \mid x) \} / \hat{\pi}(p \mid x)
$$

$$
\hat{U}(p \mid x) = p \frac{1}{n_1(x)} \sum_{i=1}^{n_1(x)} V_i(x) + (1 - p) \frac{1}{n_0(x)} \sum_{i=1}^{n_0(x)} \mathcal{U}_i(x) \{ \mathcal{U}_i(x) > \hat{Q}_0(1 - \hat{\pi}(p \mid x) \mid x) \} / \hat{\pi}(p \mid x).
$$

Note that $\hat{\pi}(\hat{\nu}(x) \mid x) = 1$ and $\hat{\pi}(1 \mid x) = 0$. Thus, $\hat{L}(\hat{\nu}(x) \mid x) = \hat{U}(\hat{\nu}(x) \mid x)$ and $\hat{L}(1 \mid x) = \hat{U}(1 \mid x)$. Moreover, it follows immediately from the definitions of $\hat{L}(p \mid x)$ and $\hat{U}(p \mid x)$ that $\hat{L}(p \mid x) \leq \hat{U}(p \mid x)$ for all $p \in [\hat{\nu}(x), 1]$.

Finally, define the extreme value estimators

$$
\hat{\lambda}(x) = \inf_{p \in [\hat{\nu}(x), 1]} \hat{L}(p \mid x)
$$

$$
\hat{u}(x) = \sup_{p \in [\hat{\nu}(x), 1]} \hat{U}(p \mid x).
$$

We close this section by considering the assumption of ordered outcomes made by Lambert and Tierney (1997) as well as the case of censored outcomes.

Consider first the assumption of ordered outcomes. Recall that the monotonicity condition (1) implies the ordered outcomes condition (2), that $\mathbb{E}[Y \mid Z = 1, V = 0] \leq \mathbb{E}[Y \mid Z = 1, V = 1]$. By C2, $\mathbb{E}[Y \mid Z = 1, V = 1] = \mathbb{E}[Y \mid V = 1]$. It follows from (6) and the definition of $U(p)$ that under condition (2), we may define

$$
u = \mathbb{E}[Y \mid V = 1].
$$

Note that under (2), $\mathbb{E}[Y \mid V = 1] \leq \sup_{p \in [\nu, 1]} U(p)$ with strict inequality except in the degenerate cases when $p^* = \nu$ or $p^* = 1$, or, in the sample, when the maximum of the unverified observations is less than or equal to the mean of the verified observations. In other words, except in very special cases, imposing the ordered outcomes condition results in a strictly smaller
upper bound. Also, note that $E[Y|V = 1]$ can be estimated with the sample average of the verified observations.

Next consider the case of censoring. Suppose that values of $Y$ are not always directly observed. In particular, suppose the data are subject to a known lower censoring threshold $\xi_i$ that may vary across observations. This type of censoring will not complicate the estimation procedure. In particular, the upper bound $u$ can be derived by simply assigning each censored observation of $Y$ the censoring threshold value $\xi_i$. One can similarly find a lower bound $\lambda$ if the support of $F$ is known to be bounded below by some finite value.

For example, the Love Canal data analyzed in Section 6 are generally censored from below at the value 0.2, and the support of $F$ is bounded below at 0.0. Therefore, $\hat{u}$ can be computed as prescribed above after assigning the value 0.2 to each censored observation. Similarly, $\hat{\lambda}$ can be computed after assigning the value 0.0 to each censored observation.

SECTION 5. PROPERTIES OF THE BOUNDS

In this section, we prove that the sample HM bounds on the mean of $F$ always contain the sample bounds defined in the last section. The argument also works for other characteristics of $F$, like moments and cdf values. In addition, we establish consistency of the sample bounds developed here.

Start with the HM bounds. Write $\hat{Q}$ for the empirical quantile function of the sample of $Y_i$’s. Write $\hat{\lambda}_{HM} = n^{-1} \sum_{i=1}^{n} Y_i \{ Y_i \leq \hat{Q}(\hat{v}) \}$ and $\hat{u}_{HM} = n^{-1} \sum_{i=1}^{n} Y_i \{ Y_i > \hat{Q}(1 - \hat{v}) \}$ for the HM lower and upper bound estimators for the mean of $F$.

**Theorem 2:** For all $n \geq 1$, $\hat{\lambda}_{HM} \leq \hat{\lambda} \leq \hat{u} \leq \hat{u}_{HM}$.

**Proof.** By definition, $\hat{L}(p) \leq \hat{U}(p)$ for all $p \in [\hat{v}, 1]$, and so $\hat{\lambda} \leq \hat{u}$.
We now show that $\hat{\lambda}_{HM} \leq \hat{\lambda}$. The proof that $\hat{u} \leq \hat{u}_{HM}$ can be established similarly.

Note that $\hat{\lambda}_{HM}$ is an arithmetic average of the $n_1$ smallest $Y_i$’s. If, for each $p \in [\hat{v}, 1]$, $\hat{L}(p)$ is an arithmetic average of at least $n_1$ of the $Y_i$’s, then $\hat{\lambda}_{HM} \leq \hat{L}(p)$ for each $p \in [\hat{v}, 1]$, implying the result. Note that

$$\hat{L}(p) = p \frac{1}{n_1} \sum_{i=1}^{n_1} V_i + (1 - p) \frac{1}{n_0 \hat{\pi}(p)} \sum_{i=1}^{n_0} U_i \{U_i \leq \hat{Q}_0(\hat{\pi}(p))\}. $$

We see that $\hat{L}(p)$ is a convex combination of an arithmetic average of the $n_1$ verified observations and an arithmetic average of the $n_0 \hat{\pi}(p)$ unverified observations that are less than or equal to $\hat{Q}_0(\hat{\pi}(p))$. Simple algebra shows that

$$p/n_1 = (1 - p)/n_0 \hat{\pi}(p) = 1/[n_1 + n_0 \hat{\pi}(p)].$$

Deduce that $\hat{L}(p)$ is an arithmetic average of the $n_1$ verified observations and $n_0 \hat{\pi}(p)$ unverified observations. This proves the result.

We now establish consistency of the extreme value estimators. For each $p$ in $(0, 1]$, we may write the population lower and upper bound functions as follows:

$$L(p) = p \mathbb{E}V + (1 - p) \mathbb{E}[U \mid U \leq Q_0(\pi(p))]$$
$$U(p) = p \mathbb{E}V + (1 - p) \mathbb{E}[U \mid U > Q_0(1 - \pi(p))] .$$

Recall that the population extreme values are given by

$$\lambda = \inf_{p \in [v, 1]} L(p)$$
$$u = \sup_{p \in [v, 1]} U(p) .$$

Our objective is to establish rates of convergence of the extreme value estimators defined in the last section to their population counterparts. We will show that $|\hat{\lambda}(x) - \lambda(x)| = O_p(n^{-2/d+4})$
as $n \to \infty$. The proof for the upper bound is similar. Write $S(x)$ for the support of $\mathcal{U}$ given $X = x$. For convenience, adopt the convention that $\hat{\pi}(p \mid x) = 1$ for $p \in [0, \hat{v}(x))$ and $\pi(p \mid x) = 1$ for $p \in [0, v(x))$. Define $f(x)$ to be the density of $X$ at $x$. Define $N(x)$ to be an open neighborhood of $x$. We make the following assumptions.

A0. $(Y_1, V_1, X_1, Z_1), \ldots, (Y_n, V_n, X_n, Z_n)$ are iid draws from the model described in Section 2.

A1. $S(x)$ is an interval (possibly infinite) of the real line.

A2. $H_0(\cdot \mid x)$ is continuously differentiable on $S(x)$.

A3. $\mathbb{E}|\mathcal{U}(x)| < \infty$ and $\mathbb{E}|\mathcal{V}(x)| < \infty$.

A4. $f(\cdot) > 0$ on $N(x)$.

A5. $f(\cdot)$ and $v(\cdot)$ have bounded first and second partial derivatives on $N(x)$.

A6. $H_0(u \mid \cdot)$ has bounded first and second partial derivatives on $N(x)$.

A7. $h_0(u \mid \cdot)$ and $Q_0(u \mid \cdot)$ have bounded first and second partial derivatives on $N(x)$.

A8. For $d \geq 1$, $a_n = O(n^{-1/d+4})$.

We begin with the following result.

**Lemma 3:** Let $\{\epsilon_n\}$ be a sequence of nonnegative real numbers satisfying $\epsilon_n \to 0$ as $n \to \infty$. If $\sup_{p \in [0,1]} |\hat{L}(p) - L(p)| = O_p(\epsilon_n)$ as $n \to \infty$, then $|\hat{\lambda} - \lambda| = O_p(\epsilon_n)$ as $n \to \infty$.

**Proof.** There exists $p_1 \in [0,1]$ such that $\hat{\lambda} = \hat{L}(p_1)$ or $\hat{\lambda} = \lim_{p \to p_1} \hat{L}(p)$. Similarly, there exists $p_0 \in [0,1]$ such that $\lambda = L(p_0)$ or $\lambda = \lim_{p \to p_0} L(p)$. Consider the case $\hat{\lambda} = \hat{L}(p_1)$ and $\lambda = L(p_0)$. (The other 3 cases can be handled similarly.) Either $\hat{\lambda} \leq \lambda$ or $\hat{\lambda} > \lambda$. Suppose $\hat{\lambda} \leq \lambda$. By definition of $\lambda$, $\lambda \leq L(p_1)$. Thus, $\hat{\lambda} \leq \lambda \leq L(p_1)$ and so $|\hat{\lambda} - \lambda| \leq |\hat{\lambda} - L(p_1)|$. Apply the uniform law to get $|\hat{\lambda} - \lambda| \leq O_p(\epsilon_n)$ as $n \to \infty$. Suppose $\hat{\lambda} > \lambda$. By definition
of $\hat{\lambda}$, $\hat{L}(p_0) \geq \hat{\lambda}$. Thus, $\hat{L}(p_0) \geq \hat{\lambda} > \lambda$ and so $|\hat{\lambda} - \lambda| \leq |\hat{L}(p_0) - \lambda|$. Apply the uniform law to get $|\hat{\lambda} - \lambda| \leq |\hat{L}(p_0) - \lambda| = O_p(\epsilon_n)\text{ as } n \to \infty$. This proves the result.

\[\square\]

**Remark.** From the proof of Lemma 3, we see that if both $\hat{L}(p)$ and $L(p)$ attain their minimum values on $[0, 1]$, then a pointwise rate of convergence of $\hat{L}(p)$ to $L(p)$ is enough to imply a corresponding rate of convergence of $\hat{\lambda}$ to $\lambda$.

**Lemma 4:** If A0 through A3 hold, then $\sup_{p \in [0, 1]} |\hat{L}(p) - L(p)| = O_p(n^{-2/(d+4)})\text{ as } n \to \infty$.

**Proof.** See Appendix. \[\square\]

**Theorem 4:** If A0 through A3 hold, then $|\hat{\lambda} - \lambda| = O_p(n^{-2/(d+4)})\text{ as } n \to \infty$.

**Proof.** Apply Lemma 3 and Lemma 4. \[\square\]

**Remark.** It is interesting to note that assumptions A1 through A3 imply that $L(p)$ is strictly convex on $[v, 1]$ and $U(p)$ is strictly concave on $[v, 1]$.

To see this, start with $L(p)$. Apply A2 and write $h_0$ for the derivative of $H_0$ on $S$. Write $c$ for $(1 - v)/v$. By definition of $Q_0$, $\pi(p) = \mathbb{P}\{U \leq Q_0(\pi(p))\}$. Therefore,

$$L(p) = p\mathbb{E}V + (1 - p)\mathbb{E}U\{U \leq Q_0(\pi(p))\}/\pi(p)$$

$$= p\left[\mathbb{E}V + c \int_{-\infty}^{Q_0(\pi(p))} t h_0(t) dt\right].$$

By A1, $h_0$ must be positive on $S$, except possibly at the boundary points. A1 also implies that $H_0$ is strictly increasing on $S$. Thus, we may interpret $Q_0$ as the inverse function of $H_0$ and apply A3 and the chain rule to get that the second derivative of $L(p)$ equals $1/[cp^3h_0(Q_0(\pi(p)))^3]$, which is strictly positive on $(v, 1)$. Deduce that $L(p)$ is a strictly convex function on $[v, 1]$, and so has a unique minimizer in $[v, 1]$.  

16
Similar calculations show that the second derivative of $U(p)$ equals $-1/[cp^3h_0(Q_0(1-\pi(p))))]$ on $(v, 1)$. Thus, $U(p)$ is a strictly concave function on $[v, 1]$ and so is uniquely maximized on $[v, 1]$.

6. An Example: The Love Canal Study

In this section, we apply the extreme value estimators developed in the previous sections to U.S. Environmental Protection Agency data on soil pollutants in the Love Canal and in a comparison region of upstate New York. These data have been analyzed by Lambert, Peterson, and Terpenning (1991) and by Lambert and Tierney (1997).

We have data on the concentration levels of two types of pollutants, namely, alpha-BHC and 2-chloronaphthalene (2-CNAP), measured with gas chromatography-mass spectroscopy. Measurements were made by six different laboratories analyzing randomly assigned soil samples. Following Lambert and Tierney, we compare mean concentration levels in Love Canal with those in the comparison region.

Table 1, reproduced in part from Table 1 in Lambert and Tierney, summarizes the data by laboratory. It is evident that the verification probability, $\hat{v}$, varies considerably across laboratories, regions, and pollutants, ranging from 0.29 to 0.97. Note also that the data are censored, but the censoring probability ($p_{cens}$) is typically less than 0.05. The official censoring threshold is a concentration of 0.2 parts per billion (ppb), but some laboratories report measurements in the interval (0.0,0.2).

Mean concentration levels are also reported, conditional on verification status. Note that single values are reported for the verified data, whereas intervals are reported for some of the unverified data. These bounds arise from the occasional censoring of values between 0.0 and 0.2,
but censoring only occurs here for the unverified data. The width of the interval is proportional to the fraction of the unverified data that are censored. Note also that the mean of the verified data exceeds that of the unverified data for each laboratory report of alpha-BHC concentration.

In Table 2, we apply our estimators to the data. The bounds we obtain, as well as the corresponding bootstrap confidence intervals, very closely resemble those depicted in Figure 3 of Lambert and Tierney (1997). As illustrated in Figures 1A and 1B, these bounds tend to be much tighter than the HM bounds reported in Table 3. Note also that estimates of mean concentration levels based on verified data alone (reported in Table 1) tend to be in the upper end of the intervals we obtain. Recall that these point estimates are biased if verification is endogenous.

Consider, for instance, the alpha-BHC concentration estimates based on Laboratory 1 data. This lab analyzed 89 soil samples from Love Canal. None of these measurements were censored, but 13 percent were not verified. The mean of the verified data is 2.7412 ppb, and we obtain an interval estimate based on verified and unverified data of [2.3794, 2.7412] ppb. In this case, imposing the ordered outcomes restriction does not reduce the upper bound (i.e., the maximal unverified value (0.2641) is less than the mean of the verified values). Note also that the HM interval estimate of [0.2319, 2.7467] ppb is considerably wider.

Now consider the measurements of Laboratory 8, which analyzed 104 soil samples from Love Canal. Again, no censoring occurs, but 30 percent of the data were not verified. Our interval estimate is [4.4755, 6.2857] ppb. The upper bound is again equal to the mean of the verified data, so the ordered outcomes restriction does not tighten the bounds. The HM bound is [0.1850, 6.3454] ppb.

In only one case does the ordered outcomes restriction reduce the upper bound of an alpha-
BHC interval estimate. This occurs for the estimate based on Laboratory 8 data for the comparison region. Here, the interval estimate is [0.1056,0.1859] ppb, whereas the mean of the verified data is slightly below the upper bound at 0.1851. Note that this interval lies far below the interval estimate of [4.4755,6.2857] for Love Canal, indicating a significantly higher level of alpha-BHC concentration in Love Canal. Thus, these interval estimates can be very informative. Also note, by comparison of the means of verified data reported in Table 1 with the bounds reported in Table 2, that the ordered outcomes restriction typically reduces the estimated upper bound on 2-CNAP concentration.

7. Summary and Directions for Future Work

This paper is concerned with estimating characteristics of a distribution of interest, $F$, when the observed data are contaminated, but a positive proportion of the data are known to be drawn from $F$. Without parametric assumptions on the distribution from which the data are drawn, characteristics of $F$ such as moments and quantiles are not identified. However, under minimal assumptions on the underlying distributions, finite bounds on these characteristics are identified.

Nonparametric extreme value estimators of bounds on the mean of $F$ are developed by extending the method of Horowitz and Manski (1995) to exploit all the information on verification. The procedure can be used to estimate bounds on all moments, probabilities, and quantiles of $F$. Except in degenerate cases where the bounds are equal, the estimated bounds developed here are always tighter than those of Horowitz and Manski. In addition, the procedure is more direct and more naturally adaptable to monotonicity properties and censoring in the data than an alternative method proposed by Lambert and Tierney (1997). The estimated bounds are
shown to be consistent for their population counterparts.

In their concluding section, Lambert and Tierney (1997) call for improvements on their methods that can handle general forms of censoring and allow for the effect of covariates on the response variable. As mentioned above, the method presented in this paper handles general forms of known censoring in a simple, natural way. In addition, the method can be extended in a straightforward manner to incorporate information on covariates. To see this, consider estimating a lower bound on $E[Y|Z = 1, X]$ where $X$ is a vector of covariates. Given a sample $(Y_1, V_1, X_1, Z_1), \ldots, (Y_n, V_n, X_n, Z_n)$, replace $\hat{L}(p)$ and its components with locally averaged counterparts using standard kernel weights. For instance, estimate $v(X) = P(V = 1|X)$ with

$$\hat{v}(X) = \frac{\sum_{i=1}^{n} V_i K_n(X - X_i)}{\sum_{j=1}^{n} K_n(X - X_j)}$$

where $K_n(t) = K(t/a_n)$ with $K$ a symmetric probability density and $a_n$ a bandwidth converging to zero as $n \to \infty$. Similarly, construct $\hat{Q}_0(\cdot|X)$ and $\hat{L}(p|X)$ and define the estimated lower bound as

$$\hat{\lambda}(X) = \inf_{p \in \mathcal{P}(X)} \hat{L}(p|X).$$

An upper bound estimator can be defined in like fashion. We plan to address questions about bandwidth selection, asymptotic behavior and estimation of sampling distributions in future work.

**APPENDIX: PROOF OF LEMMA 4 IN SECTION 5.**

In this appendix, we prove Lemma 4 in Section 5. To do this, we first establish some preliminary lemmas. For $z \in \mathbb{R}^d$, $d \geq 1$, define $\kappa_d = \int \{||z|| \leq 1\} dz$.

**Lemma A1.** If A0 through A8 hold, then $|\hat{v}(x) - v(x)| = O_p(n^{-2/d+4})$ as $n \to \infty$. 

20
Proof. Define
\[ \hat{f}(x) = \frac{1}{\kappa_d n^{d_n}} \sum_{i=1}^{n} \frac{\{ |X_i - x| \leq a_n \}}{\kappa_d}. \]

Note that
\[ \hat{v}(x) = \frac{\sum_{i=1}^{n} V_i \{ |X_i - x| \leq a_n \}}{\sum_{i=1}^{n} \{ |X_i - x| \leq a_n \}} = \frac{1}{\kappa_d f(x)} \left[ \frac{1}{\hat{v}(x)} - 1 \right] \]

where \( \hat{v}(x) = 1 - \hat{f}(x)/f(x) \). Note that
\[ \hat{v}(x) = f(x) - \mathbb{E}\hat{f}(x) \]
\[ \mathbb{E}\hat{f}(x) = \frac{\mathbb{E}f(x) - \mathbb{E}\hat{f}(x)}{f(x)} \]

where \( \mathbb{E}\hat{f}(x) = [\kappa_d]^{-1} \int \{ |y - x| \leq a_n \} f(y)dy \). Change variables, letting \( z = (y - x)/a_n \) and apply A4 and A5 along with a 2-term Taylor expansion of \( f \) about \( x \) to see that \( \mathbb{E}\hat{f}(x) = f(x) + O(a_n^2) \). Thus, \( \frac{f(x) - \mathbb{E}\hat{f}(x)}{f(x)} = O(a_n^2) \). Apply A4 and A5 and a standard central limit theorem to see that \( \mathbb{E}(f(x) - \mathbb{E}\hat{f}(x)) = O_p(1/\sqrt{n\kappa_d n^{d_n}}) \). Apply A8 to get that \( \hat{v}(x) = O_p(n^{-2/d+4}). \)

Deduce that
\[ \frac{1}{1 - \hat{v}(x)} = 1 + O_p(n^{-2/d+4}). \]

A similar argument shows that
\[ \frac{1}{\kappa_d f(x)} = v(x) + O_p(n^{-2/d+4}). \]

Deduce from (8), (9), and (10) that \( |\hat{v}(x) - v(x)| = O_p(n^{-2/d+4}). \) 

Define \( g(x) = [1 - v(x)]f(x) \). Next, define
\[ \tilde{H}_0(u \mid x) = \frac{1}{\kappa_d n^{d_n}} \sum_{i=1}^{n} \frac{\{ Y_i \leq u \} \{ 1 - V_i \} \{ |X_i - x| \leq a_n \}}{\kappa_d g(x)} \]
\[ \mathbb{E}\tilde{H}_0(u \mid x) = \frac{1}{\kappa_d n^{d_n}} \mathbb{E}\{ Y \leq u \} \{ 1 - V \} \{ |X - x| \leq a_n \} \]
where the expectation is taken over \((Y, V, X)\).

**Lemma A2.** If A0 through A8 hold, then

\[
\sup_{u \in \mathbb{R}} |\tilde{H}_0(u \mid x) - \mathbb{E}\tilde{H}_0(u \mid x)| = O_p\left(1/\sqrt{na_n^d}\right).
\]

**Proof.** For \(y \in \mathbb{R}, v \in \{0, 1\}, \) and \(x, x_0 \in \mathbb{R}^d\), define \(F_n = \{f_n(y, v, x; u \mid x_0) : u \in \mathbb{R}\}\) where \(f_n(y, v, x; u \mid x_0)\) equals

\[
\{y \leq u\}(1-v)\{|x - x_0| \leq a_n\}/\kappa dg(x_0) - \mathbb{E}\{Y \leq u\}(1-V)\{|X - x_0| \leq a_n\}/\kappa dg(x_0).
\]

Apply Lemma 2.4 and Lemma 2.12 in Pakes and Pollard (1989) to see that \(F_n\) is Euclidean for the envelope \(F_n = 2\{|x - x_0| \leq a_n\}/\kappa dg(x_0)\). Use this fact and apply the Maximal Inequality in Sherman (1994, p.446) with \(k = p = 1\) to see that

\[
\mathbb{E}\sup_{u \in \mathbb{R}} |\tilde{P}_n f_n(\cdot, \cdot, \cdot; u \mid x_0)| = O(\sqrt{\mathbb{E}\tilde{F}_n^2} / \sqrt{n}).
\]

(11)

Apply Jensen’s inequality to get that \(\mathbb{E}\sqrt{\tilde{P}_n F_n} \leq \sqrt{\mathbb{E}\tilde{F}_n^2} = O\left(\sqrt{a_n^d}\right)\). The result follows from dividing both sides of (11) by \(a_n^d\) and applying Chebyshev’s inequality.

**Lemma A3.** If A0 through A8 hold, then \(\sup_{u \in \mathbb{R}} |\tilde{H}_0(u \mid x) - H_0(u \mid x)| = O_p(n^{-2/d+4})\).

**Proof.** Recall that \(g(x) = [1 - v(x)]f(x)\) and define

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

Recall that \(\tilde{H}_0(u \mid x) = [n_0(x)]^{-1} \sum_{i=1}^{n_0(x)} \{U_i(x) \leq u\}.\) Then

\[
\tilde{H}_0(u \mid x) = \frac{\sum_{i=1}^n \{Y_i \leq u\}(1 - V_i)\{|X_i - x| \leq a_n\}}{\sum_{i=1}^n (1 - V_i)\{|X_i - x| \leq a_n\}}
\]

\[
= \tilde{H}_0(u \mid x) \left[1 - \epsilon(x) \right]
\]

(12)

\[22\]
where \( \hat{\epsilon}(x) = 1 - \hat{g}(x)/g(x) \). Argue as in the proof of Lemma A1 to get that

\[
\frac{1}{1 - \hat{\epsilon}(x)} = 1 + O_p(n^{-2/d+4}).
\] (13)

Apply Lemma A2 to get that

\[
\sup_{u \in R} |\tilde{H}_0(u \mid x) - \mathbb{E}[\tilde{H}_0(u \mid x)]| = O_p(1/\sqrt{n\alpha_n^d}).
\] (14)

Apply A4 and A6 and argue as in Lemma A1 to get that

\[
\sup_{u \in R} |\mathbb{E}[\tilde{H}_0(u \mid x) - H_0(u \mid x)]| = O_p(a_n^2).
\] (15)

Apply (12), (13), (14), and (15), together with A8 to get the result.

Let \( \{\gamma_n\} \) denote a sequence of positive real numbers converging to zero as \( n \to \infty \). Let \( \Theta_n \) denote the product space \([0, 1] \times [0, \gamma_n] \). Define

\[
S(p, \gamma \mid x) = \mathbb{E}[Y | \{Q_0(\pi(p \mid x) \mid x) \leq Y \leq Q_0(\pi(p \mid x) + \gamma \mid x)\} \mid V = 0, X = x]
\]

\[
T_n(x) = \sup_{(p, \gamma) \in \Theta_n} S(p, \gamma \mid x).
\]

**Lemma A4.** If A1 through A3 hold, then \( T_n(x) = O(\gamma_n) \) as \( n \to \infty \).

**Proof.** Since \( S(p, \gamma \mid x) \) is increasing in \( \gamma \) for each fixed \( p \), \( T_n(x) = \sup_{p \in [0, 1]} S(p, \gamma_n \mid x) \).

Write \( f(y \mid V = 0, X = x) \) for the density of \( Y \) given \( V = 0 \) and \( X = x \). Thus,

\[
S(p, \gamma_n \mid x) = \int_{Q_0(\pi(p(x)) + \gamma_n \mid x)}^{Q_0(\pi(p(x)) \mid x)} |y| f(y \mid V = 0, X = x) dy.
\]

Note that \( f(y \mid V = 0, X = x) = h_0(u \mid x) \), the density of \( U \) given \( X = x \). By assumption A2, \( f(y \mid V = 0, X = x) \) is continuous in \( y \) and \( Q_0(t \mid x) \) is continuous in \( t \). It follows that \( S(p, \gamma_n \mid x) \) 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 \( T_n(x) = S(p_n(x), \gamma_n \mid x) \). Note that \( S(p, 0 \mid x) = 0 \) for all \( p \in [0, 1] \). Assumption A2 also implies that \( Q_0(t \mid x) \) is differentiable in \( t \), implying that \( S(p, \gamma \mid x) \) is a differentiable function of \( \gamma \) for each \( p \in [0, 1] \). Taylor expand \( S(p_n(x), \gamma_n \mid x) \) about \( \gamma = 0 \) to see that

\[
T_n(x) = \gamma_n \frac{\partial}{\partial \gamma} S(p_n(x), \gamma^* \mid x)
\]

where \( \gamma^* \in [0, \gamma_n] \). Since \( \gamma_n \to 0 \) as \( n \to \infty \), the result will follow provided \( \frac{\partial}{\partial \gamma} S(p_n(x), \gamma \mid x) \) is bounded in a neighborhood of \( \gamma = 0 \). By A1, \( S(x) = (a, b) \) for \(-\infty \leq a < b \leq \infty \). Simple calculus shows that

\[
\frac{\partial}{\partial \gamma} S(p_n(x), \gamma \mid x) = Q_0(\pi(p_n(x) \mid x) + \gamma \mid x).
\]

Thus, \( \frac{\partial}{\partial \gamma} S(p_n(x), \gamma \mid x) \) is bounded in a neighborhood of \( \gamma = 0 \) provided \( \pi(p_n(x) \mid x) \) 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 \( S(p, \gamma \mid x) \) over \( \Theta_n \), these conditions on \( p_n(x) \) must hold, otherwise either \( f(-\infty \mid V = 0, X = x) > 0 \) or \( f(\infty \mid V = 0, X = x) > 0 \), contradicting A3. This proves the result. \( \square \)

**Lemma 4.** If A0 through A8 hold, then for \( d \geq 0 \), \( \sup_{p \in [0, 1]} |\hat{L}(p \mid x) - L(p \mid x)| = O_p(n^{-2/d+4}) \) as \( n \to \infty \).

**Proof.** We 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

\[
\hat{L}_1(p \mid x) = p \frac{1}{n_1(x)} \sum_{i=1}^{n_1(x)} V_i(x) + (1 - p) \frac{1}{n_0(x)} \sum_{i=1}^{n_0(x)} U_i(x) \{ U_i(x) \leq \hat{Q}_0(\hat{\pi}(p \mid x) \mid x) \} / \pi(p \mid x)
\]

24
\begin{align*}
\hat{L}_2(p \mid x) &= p \frac{1}{n_1(x)} \sum_{i=1}^{n_1(x)} V_i(x) + (1 - p) \frac{1}{n_0(x)} \sum_{i=1}^{n_0(x)} U_i(x) \{ U_i(x) \leq Q_0(\pi(p \mid x) \mid x) \} / \pi(p \mid x).
\end{align*}

Note that

\[ |\hat{L}(p \mid x) - L(p \mid x)| \leq |\hat{L}(p \mid x) - \hat{L}_1(p \mid x)| + |\hat{L}_1(p \mid x) - \hat{L}_2(p \mid x)| + |\hat{L}_2(p \mid x) - L(p \mid x)|. \tag{16} \]

A straightforward argument using A0, A3, and Lemma A1 shows that the first term on the right-hand side of (16) has order \( O_p(n^{-2/d+4}) \) uniformly over \( p \) in \([0, 1]\). Arguments similar to those given in Lemma A1 and Lemma A2 show that the third term on the right-hand side of (16) has order \( O_p(n^{-2/d+4}) \) uniformly over \( p \) in \([0, 1]\). Consider the second term on the right-hand side of (16). Write \( \hat{\delta}(u, p \mid x) \) for \( \hat{\pi}(p \mid x) - \pi(p \mid x) + H_0(u \mid x) - \hat{H}_0(u \mid x) \). Again, Lemma A1 and simple calculus imply that \(|\hat{\pi}(p \mid x) - \pi(p \mid x)| = O_p(n^{-2/d+4}) \) uniformly over \( p \in [0, 1] \).

Deduce from this and Lemma A3 that \( |\hat{\delta}(u, p \mid x)| = O_p(n^{-2/d+4}) \) uniformly over \( u \in S(x) \) and \( p \in [0, 1] \). For notational convenience, assume, without loss of generality, that \( \hat{\delta}(u, p \mid x) \) is equal to its positive part. A similar argument works for the negative part of \( \hat{\delta}(u, p \mid x) \). Let \( c(x) = [1 - v(x)] / v(x) \). After some simple algebra, we get that \( \hat{L}_1(p \mid x) - \hat{L}_2(p \mid x) \) equals

\[
\begin{align*}
pc(x) \frac{1}{n_0(x)} \sum_{i=1}^{n_0(x)} U_i(x) \{ Q_0(\pi(p \mid x) \mid x) < U_i(x) \leq Q_0(\pi(p \mid x) + \hat{\delta}(U_i(x), p \mid x) \}
&= pc(x) \sum_{i=1}^{n} Y_i(1 - V_i) \{ Q_0(\pi(p \mid x) \mid x) < Y_i \leq Q_0(\pi(p \mid x) + \hat{\delta}(Y_i, p \mid x) \} \{ |X_i - x| \leq a_n \}
\sum_{i=1}^{n} (1 - V_i) \{ |X_i - x| \leq a_n \}
\end{align*}
\]

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

\[ |y|(1 - v) \{ Q_0(\pi(p \mid x_0) \mid x_0) < y \leq Q_0(\pi(p \mid x_0) + \gamma) \} \{ |x - x_0| \leq a_n \} / \kappa_a g(x_0). \]

Recall that \( \hat{g}(x) = [na_n^d]^{-1} \sum_{i=1}^{n} (1 - V_i) \{ |X_i - x| \leq a_n \} / \kappa_d. \) Let \( \hat{\delta}(p \mid x) = \max_{1 \leq i \leq n} \hat{\delta}(Y_i, p \mid x). \)
Since $f_n$ is increasing in $\gamma$ for fixed $y, v, x, p$, and $x_0$, we see that

$$|\hat{L}_1(p \mid x) - \hat{L}_2(p \mid x)| \leq \frac{1}{na_n^d} \sum_{i=1}^{n} f_n(Y_i, V_i, X_i, p, \hat{\delta}(p \mid x) \mid x) \left| \frac{pc(x)}{1 - \hat{\epsilon}(x)} \right|$$

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

$$\hat{F}(p, \gamma \mid x) = \frac{1}{na_n^d} \sum_{i=1}^{n} f_n(Y_i, V_i, X_i, p, \gamma \mid x).$$

Define $\Theta = S(x) \times [0, 1]$. Since $\hat{\delta}(p \mid x) \leq \sup_{(u,p) \in \Theta} \hat{\delta}(u, p \mid x) = O_p(n^{-2/d+4})$ and $\hat{F}(p, \gamma \mid x)$ is increasing in $\gamma$ 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 $n \to \infty$,

$$|\hat{L}_1(p \mid x) - \hat{L}_2(p \mid x)| \leq \hat{F}(p, \gamma_n \mid x) \left| \frac{pc(x)}{1 - \hat{\epsilon}(x)} \right|. \tag{17}$$

Argue as in the proof of Lemma A1 to see that

$$\frac{pc(x)}{1 - \hat{\epsilon}(x)} = pc(x) + O_p(n^{-2/d+4}). \tag{18}$$

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

$$\sup_{p \in [0,1]} |\hat{F}(p, \gamma_n \mid x)| \leq \sup_{p \in [0,1]} |\hat{F}(p, \gamma_n \mid x) - \mathbb{E}\hat{F}(p, \gamma_n \mid x)| + \sup_{p \in [0,1]} |\mathbb{E}\hat{F}(p, \gamma_n \mid x) - S(p, \gamma_n \mid x)| + T_n(x).$$

Argue as in the proof of Lemma A2 that the first term above has order $O_p(1/\sqrt{na_n^d})$. Apply A7 and argue as in the proof of Lemma A1 that the second term above has order $O(a_n^2)$. Finally, apply A8 and Lemma A4 to see that

$$\sup_{p \in [0,1]} |\hat{F}(p, \gamma_n \mid x)| = O_p(n^{-2/d+4}). \tag{19}$$
Apply (17), (18), and (19) to see that

\[ |\hat{L}_1(p \mid x) - \hat{L}_2(p \mid x)| = O_P(n^{-2/d+1}). \]

This proves the result. \( \Box \)

REFERENCES


