IDENTIFICATION AND ESTIMATION OF BOUNDS
ON SCHOOL PERFORMANCE MEASURES:
A NONPARAMETRIC ANALYSIS OF A MIXTURE MODEL
WITH VERIFICATION

BY JEFF DOMINITZ* AND ROBERT P. SHERMAN†

*Carnegie Mellon University †California Institute of Technology


Abstract

This paper shows how to identify and nonparametrically estimate sharp bounds on school performance measures using test scores that may be valid for some, but not all, students. In order to analyze this type of problem, we develop a mixture model with verification. This is a mixture model for data that can be partitioned into two sets, one of which (the so-called verified set) is more likely to be from the distribution of interest than the other. In the scores application, an administrative classification of a student as either English proficient or limited English proficient serves to identify these two sets. We derive sharp bounds on characteristics of distributions of interest in this model, allowing for mismeasurement in valid test scores. Sharp bounds under additional monotonicity conditions are also derived. Useful convexity and concavity results are proved, and asymptotic properties of sample analogues of the bounds are established. An analysis of performance measures of schools in a California public school district illustrates the identifying power of verification information.

†We thank William A. Bibbiano, Director of Research, Evaluation, and Testing for the Pasadena Unified School District, for providing the data analyzed in this paper and for sharing his knowledge and insights. We also thank Valentina Bali for help in obtaining the data. Finally, we thank Arie Kapteyn, Chuck Manski, and the seminar participants at Carnegie Mellon, Johns Hopkins, Northwestern, Princeton, Tilburg, UCLA, the Joint Statistical Meetings in Atlanta, August, 2001, and the Econometric Society meetings in Los Angeles, July, 2002 for helpful comments and suggestions.
1. INTRODUCTION

School administrators have long used aggregated tests scores of individual students to assess the performance of schools and school districts. With the passage of the No Child Left Behind Act of 2001 (P.L. 107-119, H.R.1), this practice is now federally mandated. According to this legislation, “School districts and schools that fail to make adequate yearly progress toward statewide proficiency goals will, over time, be subject to improvement, corrective action, and restructuring measures aimed at getting them back to meet State standards.” For instance, students who attend failing schools may transfer to other schools with free transportation provided. Failing schools are also subject to loss of accreditation or even takeover by private companies (see, for example, “Half of State’s Schools Don’t Make the Grade”, Pittsburgh Post-Gazette, 8/13/03 and http://www.cnn.com/2003/EDUCATION/08/13/sprj.sch.overview/).

Effective implementation of the No Child Left Behind Act will require proper evaluation of school performance measures. It is well understood by researchers, if not policy makers, that questions of statistical precision must be considered when evaluating performance criteria based on student test scores. See, for example, Kane and Staiger (2002) for a review and critique of performance evaluations based solely on point estimates. In this paper, we are principally concerned with the more fundamental issue of identification that arises when some test scores are likely to be invalid. In particular, we focus on concerns about the validity of scores obtained by limited English proficient students who take standardized tests administered in English. We develop a method for assessing school performance based on aggregated test scores when a possibly imperfect indicator of English proficiency is available. This method uses available information on test scores and English proficiency to obtain sharp bounds on school performance measures. These bounds may be tightened when additional monotonicity restrictions are adopted.
The validity of test scores obtained by students classified as LEP, or limited English proficient, has been the focus of much debate and litigation in California, where each year public school students in grades 2 through 11 take the Stanford 9 standardized tests in reading, mathematics, language, science, and social science, and all tests are administered in English. Of the over 4 million public school students in the state who take the tests each year, approximately one-fourth are classified by school officials as LEP. The remaining are classified as EP, or English proficient. Prior to 1998, the scores of LEP students were reported separately from the scores of EP students, and only the latter scores were used to evaluate school performance. Moreover, school comparisons were only made within school districts. Since 1998, state law has required that the scores of all students, regardless of English proficiency, be used to evaluate school performance at all levels: district, county, and state. This requirement has been the source of controversy among school and public officials (see, for example, “Judge Blocks Release of Test Scores”, Los Angeles Times, 6/28/98). School officials generally support the practice of using only the scores of EP students to assess school performance. Public officials generally want all the scores to be used.

In evaluating these scores, it is important to distinguish between being classified as EP and being truly English proficient. We say that students are truly English proficient if their English language scores equal their native language scores, where native language scores are scores they would get if they took the tests in their native languages. Certainly, native English speakers are truly English proficient in this sense. However, as we shall see, some students classified as EP may not be truly English proficient. Conversely, some students classified as LEP may be truly English proficient. We say that a student's test score is valid if that student is truly English proficient. We focus on two distributions of interest: the distribution of valid scores for truly English proficient students and the distribution of valid scores for all students. The former distribution is considered
the more appropriate distribution to study by the school officials mentioned above, whereas the latter distribution is favored by the public officials.

Empirical researchers are often faced with the type of situation described above, where a flawed data generating mechanism produces data that are not always representative of a population of interest. Often, such data can be viewed as observations from a mixture model. According to such a model, each observation is generated from either a distribution of interest, say $F$, or another, potentially spurious, distribution. Unless untestable assumptions about the data generating process hold, it is not possible to identify characteristics of $F$ such as moments, probabilities, and quantiles. However, given a lower bound on the probability of generation from $F$, Horowitz and Manski (1995) identify sharp bounds on such characteristics and show how to nonparametrically estimate these bounds.

Sometimes there is more information than simply a lower bound on the probability of generation from $F$. Sometimes data generated from a mixture model can be partitioned into two sets, and it is reasonable to assume that observations from one set are more likely to be from $F$ than observations from the other set. We call the former set the \textit{verified} set, and say that data generated in this way come from a mixture model with verification.

It is natural to model the problem of test-based assessment of school performance with a mixture model where an observation is a valid test score when a student is truly English proficient, and an invalid test score otherwise. While the observed EP indicator is not a perfect indicator of being truly English proficient, it is reasonable to assume that students classified as EP are more likely to be truly English proficient than students classified as LEP. Thus, we have a mixture model with verification, where the EP indicator acts as an imperfect verification indicator. As mentioned above, we are interested in the distribution of valid scores for truly English proficient students, as
well as the distribution of valid scores for all students. As we will show, verification information can be used to develop bounds on characteristics of both of these two distributions that are tighter than the corresponding bounds of Horowitz and Manski (1995).

We derive sharp bounds on characteristics of both distributions mentioned above. Our methods allow not only for an imperfect verification indicator, but also for measurement error in the scores. Both types of error are likely to be present in the data we analyze. In addition, we show how to tighten bounds by imposing natural monotonicity conditions.

Call the distributions of interest for the school officials and public officials distributions $A$ and $B$, respectively. We construct sample analogs of the population bounds for characteristics of these distributions, and show that they are $\sqrt{n}$-consistent and asymptotically normally distributed, where $n$ is the sample size. Extensions to allow for discrete covariates are immediate. The establishment of the limiting normal distribution for the sample bounds for characteristics of distribution $A$ depends on a parametrization that induces convexity and concavity in the functions defining these bounds. This makes short work of a problem that would otherwise be difficult to solve. Convexity and concavity in the sample functions can also be exploited to significantly reduce computations in various settings, as will be explained.

Mixture models with verification apply to a wide range of other interesting data problems. Consider self-reported data on income. Self-reports by some respondents can be verified by administrative records, as is done in the Survey of Income and Program Participation Record Check Study (U.S. Bureau of the Census, 1998, Section 6.3.4). Alternatively, some respondents may report that they consulted pay stubs when reporting income, as is done routinely in the Family Expenditure Survey conducted by the United Kingdom’s Office for National Statistics (documentation available at http://www.mimas.ac.uk/surveys/fes/). In both cases, it is reasonable to assume that the
verified data are more likely to be from the distribution of interest than the unverified data.

Some models of survey nonresponse and treatment effects can also be viewed as mixture models with verification. For example, consider the model of regressor censoring analyzed in Horowitz and Manski (1998). Fully observed data vectors can be viewed as verified draws from the distribution of interest. Vectors with missing or imputed regressors may or may not be from the distribution of interest. In the randomized treatment effects models analyzed by Molinari (2002), 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.

For quite a different application, consider data analyzed by Lambert and Tierney (1997) on organic pollution concentrations in the Love Canal in Niagra Falls, New York. 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.

Lambert and Tierney develop a nonparametric maximum likelihood approach to estimating sharp bounds on characteristics of a distribution of interest with data generated from a mixture model with verification. Their methods can be applied to bound characteristics of distribution A described above. However, they assume that the process generating observations is independent of the values generated. In the scores application, this amounts to assuming that distributions A and B are equal. This assumption is implausible; being truly English proficient is not likely to be independent of valid test scores. In addition, their methods do not easily handle censored data, incorporate natural monotonicity conditions, or generalize to allow for conditioning on continuous regressors. Nor do their methods allow for measurement error in the response or misclassification
of verified observations. The estimation methods developed in this paper cover all these situations in a natural way, as demonstrated here and in Dominitz and Sherman (2003).

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 distributions $A$ and $B$. In Section 3, we derive the sharp bounds. Section 4 shows how to tighten these bounds by imposing various monotonicity conditions. Section 5 defines sample analogs of the population bounds derived in Sections 3 and 4. We also establish convexity and concavity of the functions defining the sample bounds for characteristics of distribution $A$, and discuss the consequent computational and asymptotic benefits. In Section 6, we apply the methodology to math test scores of ninth graders from the five high schools in the Pasadena Unified School District in California and compare the verification bounds to the bounds of Horowitz and Manski (1995). Section 7 summarizes. Proofs of some theorems are given in an appendix.

2. Mixture Models with Verification

In this section, we formally define a mixture model with verification in the context of the test score data and distributions $A$ and $B$ described in the introduction. We also define the characteristics of interest of these distributions and state and discuss the basic assumptions under which we derive sharp bounds on these characteristics.

Recall that native language scores are scores students would get if they took the Stanford 9 tests in their native languages. Define $Y_1$ to be a student’s native language math score, where the score is not subject to measurement error. We call $Y_1$ a valid test score. Let $\tilde{Y}_1$ denote $Y_1$ measured with error. Note that $\tilde{Y}_1$ need not equal $Y_1$. For example, suppose there are 100 questions on the math test and each question is given unit weight. Suppose a student’s valid score is $Y_1 = 90$. If he
guesses correctly on 2 of the remaining 10 problems, then his observed score is \( \hat{Y}_1 = 92 \). Likewise, a student may make an inadvertent mistake on a solvable problem. Thus, \( \hat{Y}_1 \) may be less than \( Y_1 \).

We say students are truly English proficient in math if their English language math scores equal their native language math scores. Define \( Z = 1 \) if a student is truly English proficient and \( Z = 0 \) otherwise. Define \( Y_0 \) to be the score a student obtains on the Stanford 9 math test when \( Z = 0 \). We observe \( Y \), a mixture of \( \hat{Y}_1 \) and \( Y_0 \). That is,

\[
Y = \hat{Y}_1 Z + Y_0 (1 - Z) .
\]  

Note that each student taking the test has a \( Y_1 \) score, but only \( \hat{Y}_1 \) is observed when \( Z = 1 \).

Finally, define the verification indicator \( V = 1 \) if a student is classified as EP, and \( V = 0 \) otherwise, that is, if the student is classified as LEP. As mentioned in the introduction, it is important to realize that \( V \) is subject to the following misclassification errors: (i) \( V = 1 \) and \( Z = 0 \) and (ii) \( V = 0 \) and \( Z = 1 \). For example, during registration at a California public school, parents must fill out a home language survey indicating whether or not a language other than English is spoken at home. Only students whose parents indicate that a non-English language is spoken at home are tested to see if they merit the LEP designation. All others are classified as EP. Parents who speak a non-English language at home may fail to indicate this fact on the survey. In addition, judgement errors of type (i) can occur in testing for limited English proficiency. According to William Bibbiani, Director of Research and Testing for the Pasadena Unified School District (PUSD), it is not unreasonable to assume that up to 5% of PUSD students are designated as EP when they should be classified as LEP. Error (ii) may be more common, and can occur, for example, when children who are initially classified as LEP become truly English proficient, but, for
various reasons, are not reclassified. According to Bibbiani, at least one-third of the PUSD high
school students who are classified as LEP should be reclassified as EP. Judgement errors in testing
for limited English proficiency can also contribute to error (ii).

To summarize, in this mixture model with verification, each member of the sampling distribution
is characterized by a vector \((Y, V, Z, \tilde{Y}_1, Y_1, Y_0)\), where \(Y\) and \(V\) are observed.

Note that the school officials mentioned in the introduction consider \(P(Y_1 | Z = 1)\), the distri-
bution of valid test scores for students who are truly English proficient, to be the more appropriate
distribution to study. The public officials consider \(P(Y_1)\), the distribution of valid test scores for
all students, to be more appropriate. Let \(M\) denote a known, real-valued function on the support
of \(Y_1\). Assuming a mixture model with verification, we develop and estimate sharp bounds on
\(E[M(Y_1) | Z = 1]\) and \(EM(Y_1)\) for \(M(t) = t\) and \(M(t) = \{t \geq 50\}\). Thus, we are interested in
the mean and the probability of exceeding 50 for each of the distributions mentioned above. The
threshold score of 50 corresponds to the national median score for the Stanford 9 tests, and the
proportion of students exceeding 50 in a given test is reported in the School Accountability Report
Card issued for each school (see Appendix).

We now state and discuss the basic assumptions we make to establish sharp bounds on the
quantities \(E[M(Y_1) | Z = 1]\) and \(EM(Y_1)\).

**A1.** The data are draws from a mixture model with verification where \(P\{V = 1\} > 0\).

**A2.** \(P\{Z = 1 | V = 1\} \geq P\{Z = 1 | V = 0\}\).

**A3.** There exists a known constant \(d_0 \geq 0\) for which \(P\{Z = 1 | V = 0\} \geq d_0\).

**A4.** There exists a known constant \(d_1 > 0\) for which \(P\{Z = 1 | V = 1\} \geq d_1\).

**A5.** \(E[M(Y_1) | Z = 1] = E[M(Y) | Z = 1]\).
Assumptions A1 and A2 are plausible assumptions for the PUSD data analyzed in Section 6. Note that assumption A2 states that students classified as EP are more likely to be truly English proficient than students classified as LEP. Assumptions A3 and A4 say that there exist known lower bounds on the probability of being truly English proficient given that one is classified as LEP or EP, respectively. Note that A2, A3, and A4 together imply that $d_1 \geq d_0$. As discussed above, it may be reasonable to take $d_0 = .33$ and $d_1 = .95$ for the PUSD data. Note, however, that some school officials and public officials mentioned in the introduction may be prepared to assume that $d_1 = 1$. Assumption A5 is a measurement error assumption on scores for students who are truly English proficient. We will apply A5 with $M(t) = t$ and $M(t) = \{t > 50\}$. Thus, A5 says that, for students who are truly English proficient, (i) the expected value of their valid scores is equal to the expected value of their observed scores and (ii) the probability that their valid scores exceed 50 equals the probability that their observed scores exceed 50. Condition (i) holds, for example, under the classical measurement error model $\hat{Y}_1 = Y_1 + \epsilon$ where $E[\epsilon | Z = 1] = 0$. Condition (ii) holds exactly under the classical measurement error model if, given $Z = 1$, the distribution of $Y_1$ is uniform in a neighborhood of 50 containing the support of $\epsilon$. It holds approximately if $Y_1$ is approximately uniform in this neighborhood. This is a reasonable assumption if the support of $\epsilon$ is much smaller than the support of $Y_1$, as is likely in the scores application. Of course, if measurement error can be neglected, then one can dispense with A5.

Assumptions A1 and A2 are the main assumptions that distinguish mixture models with verification from the mixture models studied in Horowitz and Manski (1995) (hereafter, HM95), where verification information is not available. The key aspect of A1 is that the verification indicator, $V$, is observed for each member of the sample. This extra information can be used to develop bounds that are tighter than the corresponding HM95 bounds.
In order to construct estimable HM95 bounds on characteristics of either \( P(Y_1 \mid Z = 1) \) or \( P(Y_1) \), a positive lower bound on \( P(Z = 1) \) must be known or estimable from the data. It follows from assumptions A1, A3, and A4 that \( P(Z = 1) \geq d_1 P(V = 1) + d_0 P(V = 0) > 0 \). Thus, if A3 and A4 hold and \( P(V = 1) \) is known or estimable, then it is possible to construct estimable HM95 bounds. However, the verification bounds developed in this paper exploit the verification status of individual observations, whereas the HM95 bounds do not. Because of this, the verification bounds are always contained in the HM95 bounds, as we will show. When it is known that \( P(Z = 1 \mid V = 1) = P(Z = 1 \mid V = 0) \) in A2, then the two sets of bounds are equal. This makes sense, since under equality in A2, \( Z \) and \( V \) are independent, and so verification status of individual observations is not informative about the distributions of interest.

3. SHARP BOUNDS

In this section, we derive sharp bounds on \( E[M(Y_1) \mid Z = 1] \) and \( EM(Y_1) \) for \( M(t) = t \) and \( M(t) = \{t \geq 50\} \) under assumptions A1 through A5 described in the last section. We also derive the corresponding HM95 bounds and compare them to the verification bounds.

We begin by noting that the observed scores variable \( Y \) in (1) is discrete, taking integer values between 1 and 99. Define a continuous analogue \( \mathcal{Y} = Y + U \) where \( U \) is distributed uniformly on \((-1, 0]\). We introduce \( \mathcal{Y} \) for notational convenience. Bounds on all quantities of interest mentioned above can be easily expressed in terms of various quantiles of \( \mathcal{Y} \).

We now develop sharp bounds on \( E[M(Y_1) \mid Z = 1] \) under assumptions A1 through A5.

By A5, \( E[M(Y_1) \mid Z = 1] = E[M(Y) \mid Z = 1] \). Write \( p_1 \) for \( P(V = 1 \mid Z = 1) \). Then

\[
E[M(Y_1) \mid Z = 1] = E[M(Y) \mid Z = 1, V = 1]p_1 + E[M(Y) \mid Z = 1, V = 0](1 - p_1). \tag{2}
\]
Write $\delta_1$ for $\mathbb{P}\{Z = 1 \mid V = 1\}$ and $Q_i$ for the quantile function of $\mathcal{Y}$ given $V = i$, $i = 0, 1$. By Proposition 4 in HM95, the interval

$$[\mathbb{E}[M(Y) \mid \mathcal{Y} \leq Q_1(\delta_1), V = 1], \mathbb{E}[M(Y) \mid \mathcal{Y} > Q_1(1 - \delta_1), V = 1]]$$

contains $\mathbb{E}[M(Y) \mid Z = 1, V = 1]$.

Define $v_1 = \mathbb{P}\{V = 1\}$. Bayes' Rule implies that $\mathbb{P}\{Z = 1 \mid V = 0\} = [(1 - p_1)v_1\delta_1]/[p_1(1 - v_1)]$. Write $\pi(p_1, \delta_1)$ for this quantity. By Proposition 4 in HM95, the interval

$$[\mathbb{E}[M(Y) \mid \mathcal{Y} \leq Q_0(\pi(p_1, \delta_1)), V = 0], \mathbb{E}[M(Y) \mid \mathcal{Y} > Q_0(1 - \pi(p_1, \delta_1)), V = 0]]$$

contains $\mathbb{E}[M(Y) \mid Z = 1, V = 0]$. Combine (2), (3), and (4) to bound $\mathbb{E}[M(Y_1) \mid Z = 1]$. Note, however, that these bounds are infeasible since $p_1$ and $\delta_1$ are unknown.

To develop feasible bounds, note that by A4, $\delta_1 \in [d_1, 1]$. Apply Bayes' rule once again to get $p_1 = \delta_1 v_1/[(\delta_1 v_1 + \pi(p_1, \delta_1)(1 - v_1)]$. Apply A2 and A3 to get $v_1 \leq p_1 \leq \delta_1 v_1/[\delta_1 v_1 + d_0(1 - v_1)]$. For each $\delta \in [d_1, 1]$, define $\sigma(\delta) = \delta v_1/[(\delta_1 v_1 + d_0(1 - v_1)]$. For each $\delta \in [d_1, 1]$ and $p \in [v_1, \sigma(\delta)]$ define $\pi(p, \delta) = (1 - p)v_1\delta/p(1 - v_1)]$. Define the lower and upper bound functions

$$L(p, \delta) = p\mathbb{E}[M(Y) \mid \mathcal{Y} \leq Q_1(\delta), V = 1] + (1 - p)\mathbb{E}[M(Y) \mid \mathcal{Y} \leq Q_0(\pi(p, \delta)), V = 0]$$

$$U(p, \delta) = p\mathbb{E}[M(Y) \mid \mathcal{Y} > Q_1(1 - \delta), V = 1] + (1 - p)\mathbb{E}[M(Y) \mid \mathcal{Y} > Q_0(1 - \pi(p, \delta)), V = 0].$$

Note that $\sigma(\delta)$ is strictly increasing on $[d_1, 1]$. For each $p \in (\sigma(d_1), \sigma(1)]$ define the inverse function $\sigma^{-1}(p) = [p(1 - v_1)d_0]/[(1 - p)v_1]$. Define $\delta(p) = d_1\{v_1 \leq p \leq \sigma(d_1)\} + \sigma^{-1}(p)\{\sigma(d_1) < p \leq \sigma(1)\}$. Note that for each $p \in [v_1, \sigma(1)]$, the function $L(p, \delta)$ is increasing in $\delta$, and so is minimized over
\(\delta \in [\delta(p), 1]\) at \(\delta = \delta(p)\). Similarly, for each \(p \in [\nu_1, \sigma(1)]\), the function \(U(p, \delta)\) is decreasing in \(\delta\), and so is maximized over \(\delta \in [\delta(p), 1]\) at \(\delta = \delta(p)\). This leads to the following result.

**Theorem 1.** If A1 through A3 hold, then \(\lambda_1 \leq \mathbb{E}[M(Y_1) \mid Z = 1] \leq u_1\), where

\[
\lambda_1 = \inf_{p \in [\nu_1, \sigma(1)]} L(p, \delta(p))
\]

\[
u_1 = \sup_{p \in [\nu_1, \sigma(1)]} U(p, \delta(p)) .
\]

Moreover, these bounds are sharp.

**Remark 1.** Under assumptions A1, A3, and A4, \(P\{Z = 1\} \geq \nu_1 + d_0(1 - \nu_1)\). Since \(\nu_1\) is estimable from the data, this lower bound on \(P\{Z = 1\}\) is sufficient to construct estimable HM95 bounds on \(\mathbb{E}[M(Y_1) \mid Z = 1]\). Write \(Q\) for the unconditional quantile function for \(\mathcal{Y}\). By Proposition 4 in HM95, the HM95 bounds on \(\mathbb{E}[M(Y_1) \mid Z = 1]\) are the endpoints of the interval

\[
[\mathbb{E}[M(Y) \mid \mathcal{Y} \leq Q(d_1 \nu_1 + d_0(1 - \nu_1))], \mathbb{E}[M(Y) \mid \mathcal{Y} > Q(1 - d_1 \nu_1 - d_0(1 - \nu_1))] .
\]

It can be shown that this interval must contain \([\lambda_1, u_1]\). If \(P\{Z = 1 \mid V = 1\} = P\{Z = 1 \mid V = 0\}\) in A2, then \(Z\) is independent of \(V\), and so \(P\{Z = 1\} = P\{Z = 1 \mid V = 1\} \geq d_1\). In this case, the HM95 bounds on \(\mathbb{E}[M(Y_1) \mid Z = 1]\) are the endpoints of the interval

\[
[\mathbb{E}[M(Y) \mid \mathcal{Y} \leq Q(d_1)], \mathbb{E}[M(Y) \mid \mathcal{Y} > Q(1 - d_1)] .
\]

These bounds equal the verification bounds when A2 holds with equality: when \(Z\) is independent of \(V\), the verification status of each individual is uninformative about \(P(Y_1 \mid Z = 1)\).

It is also interesting to note that if it is reasonable to make the additional assumption that
$Y_1$ is independent of $Z$ (or the weaker mean independence assumption that $\mathbb{E}[M(Y_1) \mid Z = 1] = \mathbb{E}M(Y_1)$), then the bounds in Theorem 1 are sharp bounds on $\mathbb{E}M(Y_1)$. This independence assumption is among the defining properties of contaminated mixture models (see HM95, for example). As mentioned previously, for the test scores application analyzed in Section 6, this assumption is implausible: valid test scores ($Y_1$) are not likely to be independent of true English proficiency ($Z = 1$). However, if, in addition to assumptions A1 through A5, this type of independence assumption holds, then Theorem 1 provides sharp bounds on $\mathbb{E}M(Y_1)$ for all contaminated mixture model with verification applications.

Next, we develop sharp bounds on $\mathbb{E}M(Y_1)$ under assumptions A1 through A5.

Let $M$ be a known, real-valued function on $\mathcal{R}$, and suppose the support of $M(Y_1)$ is contained in the closed interval $[a, b]$, where $a$ and $b$ are known. In the scores application, when $M(t) = t$, then $[a, b] = [1, 99]$; when $M(t) = \{t \geq 50\}$, then $[a, b] = [0, 1]$.

**Theorem 2.** If A1 through A5 hold, then $\lambda_2 \leq \mathbb{E}M(Y_1) \leq u_2$, where

\[
\begin{align*}
\lambda_2 &= L(\sigma(d_1), d_1)[d_1 v_1 + d_0(1 - v_1)] + a[1 - d_1 v_1 - d_0(1 - v_1)] \\
u_2 &= U(\sigma(d_1), d_1)[d_1 v_1 + d_0(1 - v_1)] + b[1 - d_1 v_1 - d_0(1 - v_1)].
\end{align*}
\]

Moreover, these bounds are sharp.

**Remark 2.** Using the results from Remark 1, it is easy to show that under assumptions A1 through A5, the HM95 bounds on $\mathbb{E}M(Y_1)$ are given by

\[
\begin{align*}
\mathbb{E}[M(Y) \mid Y \leq Q(d_1 v_1 + d_0(1 - v_1))][d_1 v_1 + d_0(1 - v_1)] + a[1 - d_1 v_1 - d_0(1 - v_1)] \\
\mathbb{E}[M(Y) \mid Y > Q(1 - d_1 v_1 - d_0(1 - v_1))][d_1 v_1 + d_0(1 - v_1)] + b[1 - d_1 v_1 - d_0(1 - v_1)].
\end{align*}
\]

14
The interval with these endpoints must contain $[\lambda_2, u_2]$. This can be seen by considering the quantities $L(\sigma(d_1), d_1)$ and $U(\sigma(d_1), d_1)$. If $P\{Z = 1 | V = 1\} = P\{Z = 1 | V = 0\}$ in A2, then the HM95 bounds on $EM(Y_1)$ are the endpoints of the interval

$$[E[M(Y) | Y \leq Q(d_1)]d_1 + a(1 - d_1), E[M(Y) | Y > Q(1 - d_1)]d_1 + b(1 - d_1)] .$$

As before, under equality in A2, these bounds can be shown to equal the verification bounds.

4. Sharp Bounds under Monotonicity

In this section, we derive sharp bounds on $E[M(Y_1) | Z = 1]$ and $EM(Y_1)$ under additional monotonicity assumptions that can considerably tighten the bounds derived in Section 3. We also derive the corresponding HM95 bounds and compare them to the verification bounds.

We consider the following monotonicity assumptions:

A6. $E[M(Y_1) | Z = 1] \geq E[M(Y_0) | Z = 0]$.

A7. $E[M(Y) | Z = 1, V = 1] \geq E[M(Y) | Z = 1, V = 0]$.

A8. $E[M(Y_1) | Z = 0] \geq E[M(Y_0) | Z = 0]$.

A9. $E[M(Y_1) | Z = 1] \geq E[M(Y_1) | Z = 0]$.

Suppose $M(t) = t$. In this case, assumption A6 says that the average valid score of students who are truly English proficient is at least as high as the average invalid score of students who are not truly English proficient. Assumption A7 says that the average observed score of students who are truly English proficient and are classified as EP is at least as high as the average observed score of students who are truly English proficient but are classified as LEP. Adding assumption A6 can raise the lower bound, while adding A7 can lower the upper bound, on $E[M(Y_1) | Z = 1]$.
Assumption A8 says that for students who are not truly English proficient, their average valid score is at least as high as their average invalid score. Assumption A9 says that the average valid score of students who are truly English proficient is at least as high as the average valid score of students who are not truly English proficient. Adding assumption A8 can raise the lower bound, while adding A9 can lower the upper bound, on $EM(Y_1)$.

**Theorem 3.** If A1 through A7 hold, then $\lambda_3 \leq EM(Y_1) | Z = 1 \leq u_3$, where

$$\lambda_3 = EM(Y)$$

$$u_3 = \min \{ EM(Y) \mid Y > Q_1(1 - d_1), V = 1, u_1 \}.$$

If A1 through A5, A8, and A9 hold, then $\lambda_3 \leq EM(Y_1) \leq u_3$. Moreover, these bounds are sharp.

**Remark 3.** It is easy to show that under the corresponding assumptions in Theorem 3, the HM95 bounds on both $EM(Y_1) | Z = 1$ and $EM(Y_1)$ are the endpoints of the interval

$$[EM(Y), EM(Y) | Y > Q(1 - d_1 v_1 - d_0 (1 - v_1))] \,.$$

Note that the HM95 upper bounds on $EM(Y_1) | Z = 1$ and $EM(Y_1)$ under monotonicity are the same as the HM95 upper bound on $EM(Y_1) | Z = 1$ under assumptions A1 through A5 (see Remark 1 after the statement of Theorem 1 in Section 3). As before, the HM95 upper bound under monotonicity must be at least as large as $u_3$. Also, under equality in A2, both the HM95 bounds and the verification bounds are the endpoints of the interval

$$[EM(Y), EM(Y) | Y > Q(1 - d_1)] \,.$$

16
It is interesting to note that in the special case of A3 and A4 when \( d_0 = 0 \) and \( d_1 = 1 \), \( u_3 = \mathbb{E}[M(Y) \mid V = 1] \). Recall from the introduction that the lower bound, \( \lambda_3 = \mathbb{E}M(Y) \), was championed as the better measure of educational achievement by the political officials, while the upper bound, \( \mathbb{E}[M(Y) \mid V = 1] \), was preferred by the school officials.

**Remark 4.** Under certain circumstances, it may be reasonable to (i) say that a student is truly English proficient \( (Z = 1) \) if the student’s English language score is at least as high as the student’s native language score and (ii) define a valid score \( (Y_1) \) to be the greater of a student’s English language score and the student’s native language score. This would allow for the possibility that some students who are not native English speakers may, over time, acquire English language skills that exceed their native language skills. If definitions (i) and (ii) are adopted, then assumption A8 is automatically satisfied, and so \( \lambda_2 \) in Theorem 2 is equal to \( \lambda_3 \) in Theorem 3.

5. Estimation

We begin by developing sample analogs of the population bounds on \( \mathbb{E}[M(Y_1) \mid Z = 1] \) given in Theorem 1 in Section 3. We establish convexity and concavity of the sample functions defining these bounds, and discuss the consequent computational and asymptotic benefits.

Let \( (Y_i, V_i, Z_i, \bar{Y}_{i1}, Y_{i1}, Y_{i0}) \), \( i = 1, \ldots, n \), be independent draws from from the mixture model with verification defined in Section 2. Define \( n_1 = \sum_{i=1}^n V_i \), \( n_0 = n - n_1 \), and \( \hat{\nu}_1 = n_1/n \). Recall the definition of \( \mathcal{Y} \) given at the beginning of Section 3. Define \( \mathcal{Y}_i = Y_i + U_i \) where the \( U_i \)'s are independent \( U(-1, 0] \) random variables. Let \( \hat{Q}_i \) denote the empirical quantile function of \( \mathcal{Y} \) given \( V = i, i = 0, 1 \). For each \( \delta \in [d_1, 1] \), define \( \hat{\sigma}(\delta) = \delta \hat{\nu}_1/[\delta \hat{\nu}_1 + d_0(1 - \hat{\nu}_1)] \). For each \( \delta \in [d_1, 1] \) and \( p \in [\hat{\nu}_1, \hat{\sigma}(\delta)] \) define \( \hat{\pi}(p, \delta) = (1 - p)\hat{\nu}_1 \delta/[p(1 - \hat{\nu}_1)] \).

17
Define the sample lower and upper bound functions

\[
\hat{L}(p, \delta) = p \sum_{i=1}^{n} M(Y_i) V_i \left\{ Y_i \leq \hat{Q}_1(\delta) \right\}/[\delta n_1] \\
+ (1 - p) \sum_{i=1}^{n} M(Y_i) (1 - V_i) \left\{ Y_i \leq Q_0(\hat{\pi}(p, \delta)) \right\}/[\hat{\pi}(p, \delta) n_0] \\
\hat{U}(p, \delta) = p \sum_{i=1}^{n} M(Y_i) V_i \left\{ Y_i > \hat{Q}_1(1 - \delta) \right\}/[\delta n_1] \\
+ (1 - p) \sum_{i=1}^{n} M(Y_i) (1 - V_i) \left\{ Y_i > Q_0(1 - \hat{\pi}(p, \delta)) \right\}/[\hat{\pi}(p, \delta) n_0].
\]

For each \( p \in (\hat{\sigma}(d_1), \hat{\sigma}(1)) \) define the inverse function \( \hat{\sigma}^{-1}(p) = [p(1 - \hat{v}_1) d_0]/[(1 - p) \hat{v}_1] \). Define \( \hat{\delta}(p) = d_1 \{ \hat{v}_1 \leq p \leq \hat{\sigma}(d_1) \} + \hat{\sigma}^{-1}(p) \{ \hat{\sigma}(d_1) < p \leq \hat{\sigma}(1) \} \). Finally, define the extreme value estimators

\[
\hat{\lambda}_1 = \inf_{p \in [\hat{v}_1, \hat{\sigma}(1)]} \hat{L}(p, \hat{\delta}(p)), \\
\hat{u}_1 = \sup_{p \in [\hat{v}_1, \hat{\sigma}(1)]} \hat{U}(p, \hat{\delta}(p)).
\]

One can show that \( \hat{L}(p, \hat{\delta}(p)) \) is a piecewise linear convex function and \( \hat{U}(p, \hat{\delta}(p)) \) is a piecewise linear concave function for \( p \in [\hat{v}_1, \hat{\sigma}(1)] \). This result holds whether \( Y \) is discrete or continuous, and for any feasible values of \( d_0 \) and \( d_1 \). Below, we prove the result for a discrete \( Y \) taking values \( y_1 < y_2 < \cdots < y_m \). (For the scores application, we may take \( m = 99 \) and \( y_k = k, k = 1, 2, \ldots, 99 \).) Also, for ease of exposition, we prove the result for the special case \( d_0 = 0 \) and \( d_1 = 1 \), so that \( \hat{\sigma}(1) = 1 \) and \( \hat{\delta}(p) = 1 \{ \hat{v}_1 \leq p \leq 1 \} \).

**Theorem 4.** Suppose \( Y \) is a discrete random variable, taking values \( y_1 < y_2 < \cdots < y_m \). Also, suppose that \( d_0 = 0 \) and \( d_1 = 1 \). Then \( \hat{L}(p, 1) \) is a piecewise linear convex function and \( \hat{U}(p, 1) \) is a piecewise linear concave function on \([\hat{v}_1, 1]\).
PROOF. Start with $\hat{L}(p, 1)$. Write $\hat{\gamma}$ for $\sum_{i=1}^{n} M(Y_i)V_i/n_1$ and $\hat{c}$ for $(1 - \hat{v}_1)/\hat{v}_1$. We have that

$$
\hat{L}(p, 1) = p \left[ \hat{\gamma} + \hat{c} \sum_{i=1}^{n} M(Y_i)(1 - V_i)\{Y_i \leq \hat{Q}_0(\hat{\pi}(p, 1))/n_0 \} \right].
$$

Define $y_0 = -\infty$. Define $\tau_0 = 0$, $\tau_k = M(y_k)$, $k = 1, 2, \ldots, m$, and $\tau_{m+1} = \tau_m$. Finally, define $\hat{H}_0$ to be the empirical cumulative distribution function for $\mathcal{Y}$ given $V = 0$. Straightforward calculations show that

$$
\hat{L}(p, 1) = \sum_{k=m}^{1} (\hat{\beta}_k p + \tau_k)\{\hat{\pi}^{-1}(\hat{H}_0(y_k)) \leq p < \hat{\pi}^{-1}(\hat{H}_0(y_{k-1})) \}
$$

(5)

where

$$
\hat{\beta}_k = \hat{\gamma} - \tau_k - \hat{c} \sum_{j=1}^{k} (\tau_j - \tau_{j-1})\hat{H}_0(y_{j-1})
$$

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

$$
\hat{\pi}^{-1}(p) = \hat{v}_1/[\hat{v}_1 + p(1 - \hat{v}_1)].
$$

Note that the summation in (5) runs from $m$ to 1, not from 1 to $m$. It is easy to show that $\hat{\beta}_k$ is nondecreasing as $k$ decreases, and that $\hat{L}(p, 1)$ is continuous on $[\hat{v}_1, 1]$. Deduce that $\hat{L}(p, d)$ is a piecewise linear convex function on $[\hat{v}_1, 1]$.

Similar calculations show that

$$
\hat{U}(p, 1) = \sum_{k=1}^{m} (\hat{\beta}_k p + \tau_k)\{\hat{\pi}^{-1}(1 - \hat{H}_0(y_{k-1})) < p \leq \hat{\pi}^{-1}(1 - \hat{H}_0(y_k)) \}
$$

(6)

where

$$
\hat{\beta}_k = \hat{\gamma} - (\hat{c} + 1)\tau_k + \hat{c}\tau_m - \hat{c} \sum_{j=k}^{m} (\tau_{j+1} - \tau_j)\hat{H}_0(y_j).
$$

It is easy to show that $\beta_k$ is nonincreasing as $k$ increases, and that $\hat{U}(p, 1)$ is continuous on $[\hat{v}_1, 1].$
Deduce that $\hat{U}(p, 1)$ is a piecewise linear concave function on $[\hat{v}_1, 1]$. 

Theorem 4 can be useful computationally. For example, in the scores application, $M(t) = t$ or $M(t) = \{t \ge 50\}$. From the proof of Theorem 4, we see that a search to find $\hat{\lambda}_1$ when $M(t) = t$ can be limited to the potential kink point ordinates $\hat{\pi}^{-1}(\hat{H}_0(k - 1)), k = 1, 2, \ldots, 99$. Similarly, a search to find $\hat{u}_1$ can be limited to the points $\hat{\pi}^{-1}(1 - \hat{H}_0(k - 1)), k = 1, 2, \ldots, 99$. Moreover, convexity and concavity make binary searches over these points possible. When $M(t) = \{t \ge 50\}$, it is easy to show that $\hat{\lambda}_1 = \hat{L}(\hat{\pi}^{-1}(\hat{H}_0(49)), 1)$ and $\hat{u}_1 = \hat{U}(\hat{\pi}^{-1}(1 - \hat{H}_0(49)), 1)$. That is, only a single evaluation of the sample functions is needed to find the extreme value estimators.

The computational shortcuts described above can result in substantial savings in computation time when $n$ is large, when bootstrap estimates of the distribution of the extreme value estimators are desired, or when it is of interest to compute the estimators for many discrete covariate values. For example, in the scores application, it may be of interest to compute the extreme value estimators conditional on student gender, parental marital status, or level of parental income.

Theorem 4 also confers asymptotic benefits. Deriving the asymptotic distribution of an extreme value estimator is, in general, a difficult problem. However, when such an estimator is the infimum of a piecewise linear convex function or the supremum of a piecewise linear concave function, then it can be relatively straightforward to determine the limiting distribution. We illustrate this now by deriving the limiting distribution of $\hat{\lambda}_1$.

Recall the definitions of $y_k$ and $\tau_k$, $k = 0, 1, \ldots, m$. Define $H_0$ to be the cumulative distribution function for $Y$ given $V = 0$. Define $\gamma = E[M(Y_1) \mid V = 1]$ and $c = (1 - v_1)/v_1$. Using the proof technique of Theorem 4, it is straightforward to show that

$$L(p, 1) = \sum_{k=m}^{1} (\beta_k p + \tau_k)\{\pi^{-1}(H_0(y_k)) \le p < \pi^{-1}(H_0(y_{k-1}))\}$$ (7)
where
\[
\beta_k = \gamma - \tau_k - \epsilon \sum_{j=1}^{k} (\tau_j - \tau_{j-1}) H_0(y_{j-1})
\]
and, for \( p \in [0, 1] \),
\[
\pi^{-1}(p) = v_1/[v_1 + p(1 - v_1)].
\]
Likewise, it is easy to check that \( L(p, 1) \) is a piecewise linear convex function on \([v_1, 1]\). It follows
that there exists a kink point ordinate \( \pi^{-1}(H_0(y_k)) \in [v_1, 1] \) such that \( \lambda_1 = L(\pi^{-1}(H_0(y_k)), 1) \).

Extend \( \hat{L}(p, 1) \) and \( L(p, 1) \) in the obvious way so that \( \hat{L}(p, 1) \) and \( L(p, 1) \) are defined for all \( p \in [0, 1] \) and are piecewise linear convex functions on \([0, 1]\). By weak laws of large numbers, \( \hat{\gamma} \) converges in probability to \( \gamma \), \( \hat{\nu}_1 \) converges in probability to \( v_1 \), and \( \hat{H}_0(y_j) \) converges in probability to \( H_0(y_j) \), \( j = 1, 2, \ldots, m \). These facts, together with (5) and (7), imply that (i) \( \hat{\pi}^{-1}(\hat{H}_0(y_j)) \) converges in probability to \( \pi^{-1}(H_0(y_j)) \), \( j = 1, 2, \ldots, m \) and (ii) for each \( p \in [0, 1] \), \( \hat{L}(p, 1) \) converges in probability to \( L(p, 1) \). It easily follows from (i), (ii), \( \lambda_1 = L(\pi^{-1}(H_0(y_k)), 1) \), and piecewise linear convexity of \( \hat{L}(p, 1) \) and \( L(p, 1) \), that with probability tending to one as \( n \to \infty \), \( \hat{\lambda}_1 = \hat{L}(\hat{\pi}^{-1}(\hat{H}_0(y_k)), 1) \).

In sum, with probability tending to one as \( n \to \infty \), there exists a fixed point \( y_k \) in the support of \( Y \) such that \( \hat{\lambda}_1 = \hat{L}(\hat{\pi}^{-1}(\hat{H}_0(y_k)), 1) \) and \( \lambda_1 = L(\pi^{-1}(H_0(y_k)), 1) \). When this happens, it follows from (5) and (7) that
\[
\hat{\lambda}_1 - \lambda_1 = f(\hat{\theta}) - f(\theta_0)
\]
where \( \hat{\theta} = (\hat{\gamma}, \hat{\nu}_1, \hat{H}_0(y_1), \ldots, \hat{H}_0(y_k))' \), \( \theta_0 = (\gamma, v_1, H_0(y_1), \ldots, H_0(y_k))' \), and, for each parameter
vector $\theta = (\theta_1, \theta_2, \ldots, \theta_{k+2})' \in \mathbb{R}^{k+2}$,

$$f(\theta) = \left[ \theta_1 - \tau_k - \frac{1 - \theta_2}{\theta_2 d} \sum_{j=2}^{k} (\tau_j - \tau_{j-1}) \theta_{j+1} \right] \frac{\theta_2}{\theta_2 + \theta_{k+2}(1 - \theta_2)} .$$

Let $n^{-1} \sum_{i=1}^{n} \Delta_{1i}$ denote the zero-mean average $\hat{\gamma} - \gamma$. Similarly, let $n^{-1} \sum_{i=1}^{n} \Delta_{2i}$ denote $\hat{v}_1 - v_1$ and let $n^{-1} \sum_{i=1}^{n} \Delta_{2+j,i}$ denote $\hat{H}_0(y_j) - H_0(y_j)$, $j = 1, 2, \ldots, k$. Let the symbol $\Longrightarrow$ denote convergence in distribution. By the Multivariate Central Limit Theorem,

$$\sqrt{n}(\hat{\theta} - \theta_0) \Longrightarrow N(0, \Sigma)$$

where $\Sigma$ is a $(k+2) \times (k+2)$ matrix with $s$th entry equal to the probability limit of $n^{-1} \sum_{i=1}^{n} \Delta_{si} \Delta_{ti}$.

The next result follows by an application of the delta method.

**Theorem 5.** Under the assumptions of Theorem 4, $\sqrt{n}(\hat{\lambda}_1 - \lambda_1) \Longrightarrow N(0, \frac{\partial f'}{\partial \theta} \Sigma \frac{\partial f'}{\partial \theta})$.

The asymptotic variance in Theorem 5 can be estimated from the data by replacing population quantities with sample analogues. An argument similar to the one used to prove Theorem 5 can be used to show that $\sqrt{n}(\hat{u}_1 - u_1)$ is asymptotically normal. These results, in turn, can be used to develop an asymptotic confidence interval for $E[M(Y_i) \mid Z = 1]$ using Bonferroni's inequality, as is done in Section 3.3 of Horowitz and Manski (1997).

Finally, we note that sample analogs of the population bounds derived in Theorems 2 and 3 in Sections 3 and 4 can be easily constructed by replacing expectations with sample averages and population quantile functions with sample quantile functions. Standard asymptotic methods can be applied to prove $\sqrt{n}$-consistency and asymptotic normality of these estimators. Then asymptotic confidence intervals can be developed as discussed in the last paragraph.
As mentioned in the introduction, the state of California requires that all public school students in grades 2 through 11 take the Stanford 9 standardized tests in reading, mathematics, language, science, and social science. These tests are administered in English. Prior to 1998, only the scores of EP students were used to evaluate school performance, and then, only within school districts. Since then California law has required that the scores of all students, both EP and LEP, be used to evaluate educational performance at all levels: district, county, and state.

As discussed in Section 2, data of this sort can be modeled with a mixture model with verification, where the indicator of English proficiency (1 if a student is classified as EP and 0 if the student is classified as LEP) serves as a verification indicator. In this section, using the results developed in Sections 3, 4, and 5, we construct verification bounds on math scores of ninth-graders in the Pasadena Unified School District (PUSD) who took the Stanford 9 tests in the year 2000. We illustrate the identifying power of verification information by comparing the verification bounds to the bounds of Horowitz and Manki (1995).

The data we analyze is norm-referenced data. That is, a representative national sample of 9th grade students took the Stanford 9 math test and generated a distribution of test scores. The score of each 9th grade student in PUSD who took this math test in 2000 is compared to this national distribution of scores. A score of 50 corresponds to the mean and median of the national distribution of scores.²

²More specifically, each 9th grader in the national sample obtained a raw score on the math test. This is simply the number of correct answers on the test. Each correct answer was reweighted to account for differences in questions (e.g., more difficult questions received greater weight) and the new weights were added to form an adjusted score. The 1st through 99th percentiles of the adjusted scores were then computed. Each PUSD 9th grader who took the math test in 2000 generated a raw score that was adjusted and then mapped to the nearest percentile of the national distribution of adjusted scores. These percentiles were then mapped to the 1st through 99th percentiles of a normal distribution with mean 50 and standard deviation 21.06. These scores are called NCE, or normal curve equivalent, scores. (Note that an NCE score of 50 corresponds to the 50th percentile of the national distribution of adjusted
Each PUSD high school reports results from the Stanford 9 tests in its annual School Accountability Report Card. The appendix includes pages from the report cards of two such schools describing Stanford 9 math scores for the school, the district, the county, and the state. Note that one statistic reported is the percentage of students who score at or above the 50th percentile of the national distribution of adjusted scores. (For further details, see California Department of Education Score Explanations, at http://star.cde.ca.gov/star2000f/reporthelp_b.html).

Table 1 describes the math test score data for ninth graders at each of the five high schools in the school district, as well as for PUSD as a whole. The share of students who are classified as LEP ranges from 15% to 17% for schools 64, 80, 82, and 84. For school 90, LEP students constitute 35% of the student body. Overall, 18% of PUSD high school students are officially classified as LEP. The mean math score is lower for the LEP students than EP students. The difference is least pronounced at School 90 (about 4 points), whereas the difference ranges from about 9 to 14 points at the other schools.

The analysis below excludes those students who did not take the test, which is shown in Table 1 to be about 6% overall. The proportion missing varies across schools and within schools by EP status. The bounds we derive in this section can easily be revised to account for this censoring, but this would entail additional notation and revisions to the theorems in the previous sections without adding much to the substance of the analysis (other than a widening of the bounds). These results are available from the authors.

We begin by presenting estimated bounds on $E[M(Y_1)|Z = 1]$ for $M(Y_1) = Y_1$ and $M(Y_1) = \{Y_1 \geq 50\}$. When $M(Y_1) = Y_1$, the quantity of interest is the mean math score of truly English

scores.) These NCE scores were then rounded to the nearest integer. The rounded NCE scores of the PUSD students are the data we received from the Pasadena Unified School District and which we analyze in this section. These data take values in the set of integers from 1 to 99.

aSchool 90 is a continuation high school serving students with behavioral or academic problems.
proficient students. When \( M(Y_1) = \{ Y_1 \geq 50 \} \), the quantity of interest is the probability that the scores of truly English proficient students exceed the national median score. The corresponding population bounds are given in Theorem 1 in Section 3 and Theorem 3 in Section 4.

A brief word on notation. For notational simplicity, in previous sections, we have suppressed the dependence of \( \hat{\sigma}(\delta) \) on \( d_0 \) and \( \hat{\delta}(p) \), the \( \hat{\lambda}_i \)'s, and the \( \hat{u}_i \)'s on \( d_1 \) and \( d_0 \). In this section, it will be convenient to make these dependencies explicit. Thus, we will write \( \hat{\sigma}(\delta, d_0) \) for \( \hat{\sigma}(\delta, \hat{\delta}(p, d_1, d_0)) \) for \( \hat{\delta}(p) \), \( \hat{\lambda}_i(d_1, d_0) \) for \( \hat{\lambda}_i \) and \( \hat{u}_i(d_1, d_0) \) for \( \hat{u}_i \).

Figure 1 depicts the bounds for \( M(Y_1) = Y_1 \), along with the sample lower and upper bound functions \( \hat{L}(p, \hat{\delta}(p, d_1, d_0)) \) and \( \hat{U}(p, \hat{\delta}(p, d_1, d_0)) \) for \( (d_1, d_0) = (1, 0), (.95, 0) \), and \( (.95, .33) \). The bounds are based on the data for all 1543 ninth-grade math test takers in PUSD in 2000. Of these, 82\% were classified as EP. That is, \( \hat{v}_1 = .82 \).

First, consider the functions \( \hat{L}(p, \hat{\delta}(p, 1, 0)) \) and \( \hat{U}(p, \hat{\delta}(p, 1, 0)) \) in Figure 1. These functions are constructed under the assumption that all students classified as EP are truly English proficient \( (d_1 = 1) \), while possibly none of the students classified as LEP are truly English proficient \( (d_0 = 0) \). These functions are optimized over \( p \in [\hat{v}_1, \hat{\sigma}(1, 0)] = [.82, 1] \). Starting from \( p = \hat{v}_1 \), we see that \( \hat{L}(\hat{v}_1, \hat{\delta}(\hat{v}_1, 1, 0)) = \hat{U}(\hat{v}_1, \hat{\delta}(\hat{v}_1, 1, 0)) = \hat{E}[Y] \), the sample mean of the observed scores of all students. Note that \( \hat{L}(p, \hat{\delta}(p, 1, 0)) \) is minimized at \( p = .85 \) where it takes the value 46.05. That is, \( \hat{\lambda}_1(1, 0) = 46.05 \). Next, at \( p = 1 \), we see that \( \hat{L}(1, \hat{\delta}(1, 1, 0)) = \hat{U}(1, \hat{\delta}(1, 1, 0)) = \hat{E}[Y|V = 1] \), the sample mean of the observed scores of EP students. Note that \( \hat{E}[Y] \) and \( \hat{E}[Y|V = 1] \) are the sample analogues of the Theorem 3 lower and upper bounds, \( \lambda_3 \) and \( u_3 \), under monotonicity restrictions A6 and A7 when \( d_1 = 1 \) and \( d_0 = 0 \). Finally, note that the upper bound function \( \hat{U}(p, \hat{\delta}(p, 1, 0)) \) is maximized at \( p = .97 \) where it takes the value 48.89. That is, \( \hat{u}_1(1, 0) = 48.89 \).

Next, consider the functions \( \hat{L}(p, \hat{\delta}(p, .95, 0)) \) and \( \hat{U}(p, \hat{\delta}(p, .95, 0)) \) in Figure 1. These functions
illustrate changes that occur when misclassification of EP students is allowed \((d_1 = .95)\). As before, they are optimized over \(p \in [\hat{v}_1, \hat{\sigma}(1,0)] = [.82, 1]\). The most prominent change is the introduction of gaps between the functions at \(p = \hat{v}_1\) and \(p = 1\). Otherwise, the shapes of the functions are similar to those just described, except with lower infimum \((\hat{\lambda}_1(.95,0) = 44.18)\) and higher supremum \((\hat{u}_1(.95,0) = 50.43)\). Also, note that the lower bound, under monotonicity conditions A6 and A7 when \(d_1 = .95\) and \(d_0 = 0\), is still \(\hat{\lambda}_3 = \hat{E}Y\), whereas the upper bound increases to \(\hat{u}_3(.95,0) = \hat{U}(1, \hat{\delta}(1, .95, 0))\).

Next, consider the functions \(\hat{L}(p, \hat{\delta}(p, .95, .33))\) and \(\hat{U}(p, \hat{\delta}(p, .95, .33))\) in Figure 1. These functions are computed under the assumption that at least 95% of EP students are truly English proficient \((d_1 = .95)\) and at least 33% of LEP students are truly English proficient \((d_0 = .33)\); they are optimized over \(p \in [\hat{v}_1, \hat{\sigma}(1, .33)] = [.82, .93]\). Note that \(\hat{\delta}(p, d_1, d_0)\) is increasing in \(d_0\) while \(\hat{L}(p, \hat{\delta})\) is increasing in \(\hat{\delta}\) and \(\hat{U}(p, \hat{\delta})\) is decreasing in \(\hat{\delta}\). It follows that when \(d_0^* \geq d_0\), \(\hat{L}(p, \hat{\delta}(p, d_1, d_0^*)) \geq \hat{L}(p, \hat{\delta}(p, d_1, d_0))\) and \(\hat{U}(p, \hat{\delta}(p, d_1, d_0^*)) \leq \hat{U}(p, \hat{\delta}(p, d_1, d_0))\) for \(p \in [\hat{v}_1, \hat{\sigma}(1, d_0^*)]\). Deduce that when \(d_0^* \geq d_0\), \(\hat{\lambda}_1(d_1, d_0^*) \geq \hat{\lambda}_1(d_1, d_0)\) and \(\hat{u}_1(d_1, d_0^*) \leq \hat{u}_1(d_1, d_0)\). Note that for the scores data, \(\hat{\lambda}_1(.95, .33) = \hat{\lambda}_1(.95, 0)\) while \(\hat{u}_1(.95, .33) < \hat{u}_1(.95, 0)\).

Finally, note that all the sample lower bound functions in Figure 1 are piecewise linear convex functions while all the sample upper bound functions are piecewise linear concave functions on their respective domains, as guaranteed by Theorem 4 and the paragraph immediately preceding it in Section 5.

Results parallel to those presented in Figure 1 are presented in Figure 2 for the case \(M(Y_1) = \{Y_1 \geq 50\}\). As in Figure 1, the piecewise linear convexity of the sample lower bound functions and the piecewise linear concavity of the upper bound functions are evident. We also see the gaps at \(p = \hat{v}_1\) and \(p = 1\) that are introduced when \(d_1\) decreases from 1 to .95, as well as the sharp rise
in the sample lower bound function and the sharp drop in the sample upper bound function at \( \hat{\sigma}(d_1 = .95, .33) \) when \( d_0 \) increases from 0 to .33.

Tables 2 and 3 report point estimates and estimated confidence intervals for HM95 bounds, verification bounds, and verification bounds under monotonicity restrictions on \( \mathbb{E}[M(Y_1) \mid Z = 1] \) and \( \mathbb{E}M(Y_1) \) when \( M(Y_1) = Y_1 \) and \( M(Y_1) = \{Y_1 \geq 50\} \) for PUSD as well as for each school in the district. This is done for the cases \( (d_1, d_0) = (1, 0), (1, .33), \) and \( (.95, .33) \). Figures 3 through 6 graphically present point estimates for easy comparison across types of bounds and schools.

Consider first Table 2, which treats the case \( M(Y_1) = Y_1 \). We present joint 95% confidence interval estimates using the bootstrap. For each school and for PUSD as a whole, we draw 1000 bootstrap samples from the original sample, with each bootstrap sample size equal to the original sample size. Each bootstrap sample produces a bootstrap estimate of a given pair of population bounds. The estimates are denoted \((\lambda^*, u^*)\). Following Horowitz and Manski (2000), we use the empirical distribution of \((\lambda^*, u^*)\) pairs to find the smallest value \( z^* \) such that 95% of the pairs \((\lambda^* - z^*, u^* + z^*)\) contain \((\lambda, u)\), the original point estimate of the given population bounds. The joint 95% confidence interval estimate that we report is then \((\lambda - z^*, u + z^*)\).\(^4\)

Next, consider the point estimates depicted in Figures 3 and 4. The estimated bounds for PUSD are reported in the right-most panel of each figure. The PUSD bounds in Figure 3 include the bounds in Figure 1 with \((d_1, d_0) = (1, 0)\) and are reported in the sixth row of Table 2. They illustrate the findings in Sections 3 and 4 that the verification bounds on \( \mathbb{E}Y_1 \) (thin line in column

\(^4\)HM95 bounds always contain the corresponding verification bounds. However, the method of computing bootstrap confidence intervals described above can produce verification confidence intervals that contain the corresponding HM95 confidence intervals. For instance, consider the lower bound under monotonicity for PUSD, with \((d_1, d_0) = (.95, .33)\) (bottom-right corner of Table 2). HM95 and verification lower bounds are equal here. However, the verification upper bound is more variable (yet never larger) than the HM95 upper bound. This variability leads to a larger value of \( z^* \) for the verification bounds and, hence, a slightly smaller lower limit for the confidence interval (45.4 vs. 45.5). Slight anomalies of this sort occur several times in Tables 2 and 3. Also, for some cases in Table 3, estimated confidence intervals extend beyond the unit interval. When this happens, we enforce the restriction that the support of \( M(Y_1) \) is \([0, 1]\).
marked “V”) contain the bounds on $E[Y_1|Z = 1]$ (black-filled rectangle in column V), which, in turn, contain the monotonicity bounds $EY$ and $E[Y|V = 1]$ (limits of gray-filled rectangle in column marked “M”).

Further, note that the estimated HM95 bounds on $EY_1$ (thin line in column marked “HM”) contain the verification bounds on $EY_1$, while the HM95 bounds on $E[Y_1|Z = 1]$ (black-filled rectangle in column HM) contain the verification bounds on $E[Y_1|Z = 1]$. The extent to which the verification bounds are tighter than the HM95 bounds quantifies the identifying power of the verification information. For example, the estimated HM95 bounds on $EY_1$ are almost 50% wider than the corresponding verification bounds (26.0 versus 17.4 points wide), and the HM95 bounds on $E[Y_1|Z = 1]$ are nearly 4 times as wide as the verification bounds on $E[Y_1|Z = 1]$ (10.4 versus 2.8 points wide). The corresponding comparisons of 95% confidence intervals (Table 2, row 6) are 28.7 versus 19.7 and 12.5 versus 4.7, respectively.

Next, consider the relationships among these bounds and the verification bounds obtained under monotonicity restrictions, as depicted by the gray-filled rectangle in column M. Recall from Theorem 3 that, when $(d_1, d_0) = (1, 0)$, the observed mean test scores $E[Y]$ and $E[Y|V = 1]$ are the lower and upper verification bounds, respectively, for $E[Y_1|Z = 1]$ when A6 and A7 hold. They are also the lower and upper bounds for $EY_1$ when A8 and A9 hold. As noted in Remark 3, under the corresponding assumptions, the HM95 lower bounds under monotonicity are identical to the verification lower bounds, but the HM95 upper bounds under monotonicity are equal to the unrestricted HM95 upper bound on $E[Y_1|Z = 1]$. We see in Figure 3 (and row 6 of Table 2) that the estimated verification bounds under monotonicity are $[46.4, 48.5]$, whereas the estimated HM95 bounds under monotonicity are more than 2 times as wide—$[46.4, 51.1]$ and contain the value 50. The estimated confidence intervals are $[45.5, 49.5]$ and $[45.5, 52.0]$, respectively.
The qualitative relationships among the estimated bounds on mean PUSD scores reported in Figure 3 continue to hold when \((d_1, d_0) = (.95, .33)\) in Figure 4, and when we let \(M(t) = \{ t \geq 50 \}\) with \((d_1, d_0) = (1, 0)\) and \(.95, .33)\) in Figures 5 and 6, respectively. Note, however, that HM95 and verification upper bounds are closer when \((d_1, d_0) = (.95, .33)\).

Finally, we consider what can be learned by comparing scores across schools, as is done annually with the issuance of school report cards. The report cards in the Appendix focus on \(P\{Y \geq 50\}\) for each school and higher aggregations, and they also include school-level values of \(P^\prime\{Y \geq 50\} | V = 1\) and \(P\{Y \geq 50\} | V = 0\).

Consider again Figure 5 (or Table 3, top panel) and suppose that, as suggested by some public officials, we should focus on the scores of all students. The lower limits of the gray-filled rectangles correspond to the school report card reports of the percentage of students who scored at least 50 in a given school or for the district as a whole. Note that these values are all below .50. This suggests that the scores of more than half of the students at any school are below the national median score. However, when we study the verification bounds on valid scores for all students (thin line in column V), we find that only two schools—80 and 90—necessarily fall below this benchmark performance level. When we let \((d_1, d_0) = (.95, .33)\) (Figure 6), we still find that more than half of the scores of students at Schools 80 and 90 fall short of the national median. The estimated confidence interval (Table 3, bottom panel) for School 80 does contain .50, but that is not the case for School 90.

Now suppose it is the case that, as suggested by school officials, we should focus on the scores of those students who are truly English proficient. When \((d_1, d_0) = (1, 0)\), (Figure 5, and top panel of Table 3), we see that we can derive a tighter ordering of performance across schools. That is, by comparing estimated bounds on \(E\{Y \geq 50\} | Z = 1\) (black-filled rectangle in column V), we find
that Schools 64, 82, and 84 have the highest levels of performance, followed by School 80, and then School 90. When \((d_1, d_0) = (.95, .33)\) (Figure 6 and bottom panel of Table 3), we can no longer distinguish the performance of School 80 from School 82, yet we can say that Schools 64 and 84 have higher measures of performance than School 80. Note that similar conclusions may be reached when we focus on mean scores, as depicted in Figures 3 and 4.

The benefits of student-specific verification information can also be seen in these comparisons. Without student-specific verification indicators, the available information yields the HM95 bounds. In every case in Figures 3 through 6, the estimated HM95 bounds on \(E[M(Y_1)|Z = 1]\) are fully overlapping for Schools 64, 80, 82, and 84. In Figure 3, the estimated HM95 bounds on \(E[Y_1|Z = 1]\) for School 80 also overlap those for School 90.

Finally, suppose we are willing to maintain the aforementioned monotonicity assumptions. Once again, the student-specific verification information enables distinctions among schools that are not otherwise possible. For example, the HM95 bounds under monotonicity for School 80 are seen in Figure 3 to overlap with those of Schools 64 and 82, whereas the verification bounds do not. Also, in Figure 5, the HM95 bounds under monotonicity for School 64 contains the important threshold value of 0.50, whereas the verification bounds do not.

7. SUMMARY

This paper undertakes a nonparametric analysis of mixture models with verification. These are mixture models for data that can be partitioned into two sets: a verified set and an unverified set. According to these models, observations from the verified set are more likely to be from the distribution of interest than observations from the unverified set. As indicated in the introduction, these models apply to a wide range of interesting data problems.
We derive sharp bounds on characteristics of distributions of interest in these models, allowing for measurement error in the outcome of interest and misclassification in the verification indicator. Sharp bounds under additional monotonicity conditions are also derived. For a certain distribution of interest, we show that the functions defining the lower and upper bounds are respectively piecewise linear convex and piecewise linear concave. We show how these results lead to computational and asymptotic benefits. In particular, we show how convexity and concavity can be used to establish the limiting distribution of the extremum estimators.

Finally, we illustrate the identifying power of verification information in analyzing math test scores of ninth graders in a California public school district, where an indicator of English proficiency plays the role of verification indicator.

APPENDIX

PROOF OF THEOREM 1. Recall the definition of $L(p, \delta)$ and $U(p, \delta)$ given prior to the statement of Theorem 1. From (2), (3), and (4) we obtain the infeasible bounds

$$L(p_1, \delta_1) \leq E[M(Y_1) \mid Z = 1] \leq U(p_1, \delta_1).$$

It follows from this and the definition of $\lambda_1$ and $u_1$ that

$$\lambda_1 \leq L(p_1, \delta_1) \leq E[M(Y_1) \mid Z = 1] \leq U(p_1, \delta_1) \leq u_1.$$

That is, $\lambda_1$ and $u_1$ are bounds for $E[M(Y_1) \mid Z = 1]$ under A1 through A5. We want to prove that they are sharp bounds under these assumptions.

Assume, in addition to A1 through A5, that $p_1$ and $\delta_1$ are known. Then $L(p_1, \delta_1)$ and $U(p_1, \delta_1)$
are sharp bounds for $\mathbb{E}[M(Y_1) \mid Z = 1]$ since they are based on simultaneously attainable sharp
HM95 bounds for $\mathbb{E}[M(Y) \mid Z = 1, V = 1]$ and $\mathbb{E}[M(Y) \mid Z = 1, V = 0]$. The lower bounds (upper
bounds) on these expectations are simultaneously attainable since the expectations are over disjoint
subsets of the sample space.

Now drop the assumption that $p_1$ and $\delta_1$ are known. Write $\Theta$ for the compact set $[v_1, \sigma(\delta)] \times
[d_1, 1]$. Since $Y$ is a continuous random variable, both $L(p, \delta)$ and $U(p, \delta)$ are continuous on $\Theta$.
Since any pair $(p, \delta) \in \Theta$ is feasible under assumptions A1 through A5, so are the pairs that either
minimize $L(p, \delta)$ or maximize $U(p, \delta)$. This proves sharpness.

\begin{proof}

We shall prove the lower bound result. The proof of the upper bound
result is similar. By A5, $\mathbb{E}[M(Y_1) \mid Z = 1] = \mathbb{E}[M(Y) \mid Z = 1]$. It follows that

$$\mathbb{E}M(Y_1) = \mathbb{E}[M(Y) \mid Z = 1]P\{Z = 1\} + \mathbb{E}[M(Y_1) \mid Z = 0]P\{Z = 0\}.$$ \hfill (8)

Recall the following definitions: $p_1 = P\{V = 1 \mid Z = 1\}$, $\delta_1 = P\{Z = 1 \mid V = 1\}$, and $\pi(p_1, \delta_1) =$
$P\{Z = 1 \mid V = 0\}$, where $\pi(p, \delta) = (1 - p)v_1\delta / [p(1 - v_1)]$ for each $(p, \delta) \in \Theta \equiv [v_1, \sigma(\delta)] \times [d_1, 1]$
with $\sigma(\delta) = \delta v_1 / [\delta v_1 + d_0(1 - v_1)]$. Recall the definition of $L(p, \delta)$ given before the statement of
Theorem 1.

Note that $P\{Z = 1\} = \delta_1v_1 + \pi(p_1, \delta_1)(1 - v_1)$. For each $(p, \delta) \in \Theta$, write $w(p, \delta)$ for the
weight function $\delta v_1 + \pi(p, \delta)(1 - v_1)$. Deduce from (8), the argument preceding the statement of
Theorem 1, and the lower bound on $M(Y_1)$ that

$$\mathbb{E}M(Y_1) \geq L(p_1, \delta_1)w(p_1, \delta_1) + a[1 - w(p_1, \delta_1)]$$

$$\geq \inf_{(p, \delta) \in \Theta} [L(p, \delta)w(p, \delta) + a[1 - w(p, \delta)]] .$$

32
Since $L(p, \delta) \geq a$ for each $(p, \delta) \in \Theta$, the last bound is minimized when $w(p, \delta)$ is minimized over $(p, \delta) \in \Theta$. This occurs at $\delta = d_1$ and $p = \sigma(d_1)$ since $\pi(\sigma(d_1), d_1) = d_0$. This yields the stated lower bound.

To prove that the lower bound is sharp, consider a distribution for the data satisfying the following conditions: $P[Z = 1 \mid V = 1] = d_1$, $P[Z = 1 \mid V = 0] = d_0$, $M(Y_1) = a$ whenever $Z = 0$, $E[M(Y) \mid Z = 1, V = 1] = E[M(Y) \mid Y \leq Q_1(d_1), V = 1]$, $E[M(Y) \mid Z = 1, V = 0] = E[M(Y) \mid Y \leq Q_0(d_0), V = 1]$. Note that the first two conditions imply that $p_1 = \sigma(d_1)$. Thus, the lower bound $\lambda_2$ is attained for this distribution. Moreover, this distribution is consistent with A1 through A5 and the assumptions about the support of $M(Y_1)$.

**Proof of Theorem 3.** First, we show that $u_3$ is the sharp upper bound for $E[M(Y_1) \mid Z = 1]$ under assumptions A1 through A5 and A7.

Temporarily, assume only that A1, A4, A5, and A7 hold. Recall $p_1 = P[V = 1 \mid Z = 1]$. Apply A5, the law of total probability, and A7 to get

$$E[M(Y_1) \mid Z = 1] = E[M(Y) \mid Z = 1]$$
$$= E[M(Y) \mid Z = 1, V = 1]p_1 + E[M(Y) \mid Z = 1, V = 0](1 - p_1)$$
$$\leq E[M(Y) \mid Z = 1, V = 1].$$

Apply A4 and Proposition 4 in HM95 to get

$$E[M(Y) \mid Z = 1, V = 1] \leq E[M(Y) \mid Y > Q_1(1 - d_1), V = 1]. \quad (9)$$

Consider a distribution for the data for which $E[M(Y) \mid Z = 1, V = 1] = E[M(Y) \mid Z = 1, V = 0]$. 

33
and $E[M(Y) \mid Z = 1, V = 1] = E[M(Y) \mid Y > Q_1(1 - d_1), V = 1]$. The upper bound in (9) is 
attained for this feasible distribution, proving that it is sharp under assumptions A1, A4, A5, and 
A7. Now add assumptions A2 and A3 and argue as in the proof of Theorem 1 to show that $u_3$ is 
sharp under assumptions A1 through A5 and A7.

Next, apply A6, A5, and the fact that $E[M(Y_0) \mid Z = 0] = E[M(Y) \mid Z = 0]$ to get

$$
E[M(Y_1) \mid Z = 1] \geq E[M(Y_1) \mid Z = 1]P\{Z = 1\} + E[M(Y_0) \mid Z = 0]P\{Z = 0\}
= E[M(Y) \mid Z = 1]P\{Z = 1\} + E[M(Y_0) \mid Z = 0]P\{Z = 0\}
= E[M(Y) \mid Z = 1]P\{Z = 1\} + E[M(Y) \mid Z = 0]P\{Z = 0\} = EM(Y).
$$

Consider a distribution for the data for which $E[M(Y_1) \mid Z = 1] = E[M(Y_0) \mid Z = 0]$. The lower 
bound $\lambda_3$ is attained for this feasible distribution, proving that it is sharp under assumptions A1 
through A6.

Next, we show that $u_3$ is the sharp upper bound on $EM(Y_1)$ under assumptions A1 through 
A5 and A9. Temporarily, assume only that A1, A4, A5, and A9 hold. Apply the law of total 
probability and A9 to get

$$
EM(Y_1) = E[M(Y_1) \mid Z = 1]P\{Z = 1\} + E[M(Y_1) \mid Z = 0]P\{Z = 0\}
\leq E[M(Y_1) \mid Z = 1].
$$

Now use (9) to get an upper bound on $EM(Y_1)$. Consider a distribution for the data for which (i) 
$E[M(Y_1) \mid Z = 1] = E[M(Y_1) \mid Z = 0]$ (ii) $E[M(Y) \mid Z = 1, V = 1] = E[M(Y) \mid Z = 1, V = 0]$ 
and (iii) $E[M(Y) \mid Z = 1, V = 1] = E[M(Y) \mid Y > Q_1(1 - d_1), V = 1]$. The upper bound in (9) is 
attained for this feasible distribution, proving that it is sharp under assumptions A1, A4, A5, and

34
A9. As before, add assumptions A2 and A3 and argue as in the proof of Theorem 1 to show that $u_3$ is sharp under assumptions A1 through A5 and A9.

Next, apply the law of total probability, A8, A5, and $E[M(Y_0) \mid Z = 0] = E[M(Y) \mid Z = 0]$ to get

$$
E[M(Y_1)] = E[M(Y_1) \mid Z = 1]P(Z = 1) + E[M(Y_1) \mid Z = 0]P(Z = 0)
\geq E[M(Y) \mid Z = 1]P(Z = 1) + E[M(Y_0) \mid Z = 0]P(Z = 0)
= E[M(Y) \mid Z = 1]P(Z = 1) + E[M(Y) \mid Z = 0]P(Z = 0) = E[M(Y)].
$$

Consider a distribution for the data for which $E[M(Y_1) \mid Z = 0] = E[M(Y_0) \mid Z = 0]$. The lower bound $\lambda_3$ is attained for this feasible distribution, proving that it is sharp under assumptions A1 through A5 and A8. This completes the proof.  

\[ \square \]

REFERENCES


