Multidimensional Poverty with Missing Attributes*

Avidit Acharya and John E. Roemer†

May 2010

Abstract

We show how to minimize the probability of misclassifying individuals as being poor or not poor when data on some of their relevant attributes are missing, but an estimate of the population distribution of attributes is available.

JEL Classification Codes: D63, I32

Key words: Multidimensional poverty, classification error

1 Introduction

There is now almost no objection to the view that poverty is a multidimensional concept.[1] The *Human Poverty Index* created by the United Nations Development Programme (UNDP), for example, takes into account attributes such as adult illiteracy, life expectancy, undernourishment, and access to safe drinking water, in addition to income. The literature on multidimensional poverty, however, has focused its attention on aggregate measures, largely ignoring the practical problems that arise in policy planning and implementation.[2] One such problem is that when citizens have

*We would like to thank Juan Ortner, Juan Moreno-Ternero, and Yves Sprumont for valuable discussions and comments.

†Acharya (corresponding author): Assistant Professor, Department of Political Science, Stanford University, Stanford CA 94305-6044 (Tel: 650-721-1492, email: avidit@stanford.edu). Roemer: Professor, Departments of Political Science and Economics, Yale University, PO Box 208301, New Haven, CT 0+630-8301 (Tel: 203-432-5249, email: john.roemer@yale.edu).

[1]See Atkinson (2003), Bourguignon and Chakravarty (2009), and works cited therein.

[2]See Bourguignon and Chakravarty (2003), Tsui (2002), and Duclos, Sahn and Younger (2006) for discussions of poverty measurement.

to be classified—based on a number of different attributes of well-being—as living in or above poverty, it may be that for any given individual, not all of that individual's attributes are observable.

More concretely, consider the problem of a government official who has the responsibility of doling out certain public benefits only to those individuals who qualify as living in poverty. The bureaucrat is told that in addition to income, an individual's life expectancy (health status, more generally) is an important attribute of his well-being. A poor individual is one whose attributes fall below a specified poverty frontier, as proposed by Duclos, Sahn and Younger (2006). The bureaucrat has at her disposal only an estimated distribution of income and life expectancy in the population. She may have constructed the distribution by taking a random sample of the population and measuring their health status to predict their life expectancy, or by estimating the joint distribution of age-adjusted income and longevity among the recently deceased. Importantly, it may be impossible to measure, or too costly to estimate, the longevity of *every* living individual who may qualify for the benefit.[3] On what basis should the bureaucrat decide whether or not a particular individual qualifies for the benefit?

Suppose the official ignores life expectancy altogether and classifies individuals based solely on an income poverty line. Given a downward sloping poverty frontier, she will be making two kinds of error. The first is a type I error: some individuals will be classified as poor because their incomes are low, but they are in fact above the poverty frontier because their life expectancies are high. The second is a type II error: some individuals who would qualify for the benefits would not receive them because although their incomes are high, they fall below the poverty frontier due to their low life expectancies. These two kinds of error are depicted in the two shaded regions of Figure 1.

If the official must choose a poverty line in income, then she may want to choose one that minimizes the sum of the type I and type II error. That is, she may want to classify individuals according to their observable attributes in such a way as to minimize the expected number of misclassified individuals. Since the official has an estimate of the joint distribution of attributes in the population, it is easy to calculate such a poverty line. The main result of this paper characterizes the solution to this minimization problem.

[3] There are at least two possible causes for missing attributes. Either it is too costly to gather the necessary information on every relevant attribute for the entirety of citizens; or, some attributes, by nature, are unobservable. Our analysis is relevant to both of these situations.

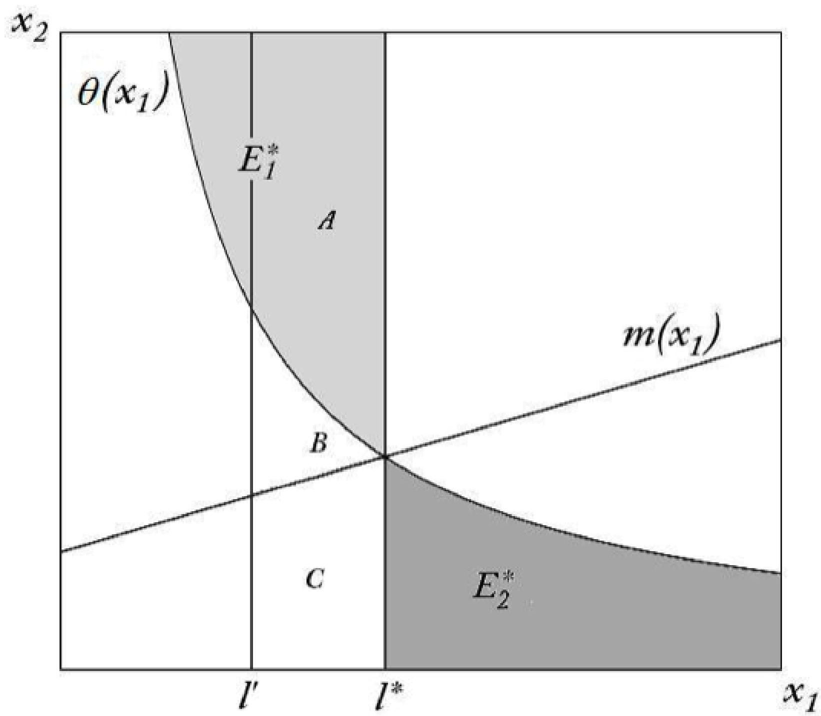

Figure 1. Type I and II errors in shaded regions.

The problem of classifying individuals as being poor or not poor is particularly important for policy. In India, for example, the distribution of many government benefits relies on the identification of those who live below the poverty line. In order to facilitate policy implementation, the government distributes "below the poverty line" (BPL) and "above the poverty line" (APL) cards to citizens based on their income. Citizens must present their cards to government officials distributing benefits such as subsidies on basic commodities. The question of who should receive a BPL card is thus important for policy-makers. But in a multidimensional setting it is complicated by the fact that not all of the attributes of well-being can be ascertained for the entirety of citizens.

2 Minimizing the Misclassified

Each individual has $n \geq 2$ attributes that are relevant for classifying the individual as poor or not poor. These attributes are given by the vector $\mathbf{x} = (x_1, \ldots, x_n)$. Each attribute takes a positive value and the distribution of types in society is summarized by the probability measure $\mathbb{F}$ with density f with full support on $\mathbb{R}^n_{++}$. Assume that for each individual, the first $k < n$ attributes $\mathbf{x}^o = (x_1, \ldots, x_k)$ are observable, while attributes $\mathbf{x}^u = (x_{k+1}, \ldots, x_n)$ are unobservable. We use both $(\mathbf{x}^o, \mathbf{x}^u)$ and (x_i, x_{-i}) to denote $\mathbf{x}$. In the latter case, $x_i > 0$ is a component of $\mathbf{x}$ while x_{-i} is the usual notation for the remaining components. We also use the notation $x^o_{-1} = (x_2, \ldots, x_k)$ so that $\mathbf{x}^o = (x_1, x^o_{-1})$. For any two vectors, $\mathbf{x}$ and $\tilde{\mathbf{x}}$, we write $\mathbf{x} > \tilde{\mathbf{x}}$ to mean $x_i \geq \tilde{x}_i$ for all i, with strict inequality for some i. For any set $X \subset R^{n-k}_{++}$, define the distribution

$$\mathbb{F}\left(X|\mathbf{x}^o\right) = \frac{\int_X f\left(\mathbf{x}^o, \tilde{\mathbf{x}}^u\right) \mathrm{d}\tilde{\mathbf{x}}^u}{\int_{\tilde{\mathbf{x}}^u \geq 0} f\left(\mathbf{x}^o, \tilde{\mathbf{x}}^u\right) \mathrm{d}\tilde{\mathbf{x}}^u}, \tag{1}$$

which is the cumulative distribution of unobservable attributes in X for the cohort $\mathbf{x}^o$. We make the following assumption about the distribution $\mathbb{F}$.

Assumption 1. $\mathbf{x}^o > \tilde{\mathbf{x}}^o$ *implies* $\mathbb{F}\left(X|\mathbf{x}^o\right) < \mathbb{F}\left(X|\tilde{\mathbf{x}}^o\right)$ *for all sets* $X \subset \mathbb{R}^{n-k}_{++}$ *s.t.*

$$\mathbf{x}^u \in X,\ \tilde{\mathbf{x}}^u \in \mathbb{R}^{n-k}_{++} \text{ and } \tilde{\mathbf{x}}^u < \mathbf{x}^u \text{ together imply } \tilde{\mathbf{x}}^u \in X$$

In the social choice literature, a set $X \subset \mathbb{R}^{n-k}_{++}$ that satisfies the centered property above is said to be *comprehensive*. Assumption 1, therefore, is a generalized first order stochastic dominance assumption, which states that we only consider distributions where higher cohorts of observable attributes have unambiguously better distributions of the unobservable attributes.

We next assume that if all attributes were observable, then an individual with attributes $\mathbf{x}$ would be classified as poor if and only if $\lambda(\mathbf{x}) < 0$ for some real valued function λ. We think of $\lambda(\mathbf{x}) = 0$ as representing a "poverty frontier," as described by Duclos, Sahn and Younger (2006). These authors introduced the notion of a poverty frontier as a compromise between the *union* and *intersection* approaches described in Bourguignon and Chakravarty (2009, 2003). We make the following assumption about the shape of the frontier.

Assumption 2. *The following hold:*

(i) $\lambda(\mathbf{x}) = \lambda\left(\tilde{\mathbf{x}}\right) = 0$ *implies* $\lambda\left(\alpha\mathbf{x} + (1-\alpha)\tilde{\mathbf{x}}\right) > 0$ *for all* $\alpha \in (0, 1)$,

(ii) $\mathbf{x} > \tilde{\mathbf{x}}$ *implies* $\lambda(\mathbf{x}) > \lambda\left(\tilde{\mathbf{x}}\right)$, *and*

(iii) for all $x_i > 0$, *there is* $x_{-i} \in \mathbb{R}^{n-1}_{++}$ *s.t.* $\lambda\left(x_i, x_{-i}\right) = 0$.

Part (i) of the assumption is a strict convexity assumption, and part (ii) is a strict monotonicity assumption. These assumptions guarantee that $\lambda(\mathbf{x}) = 0$ is downward sloping. Part (iii) guarantees that there is no level of any attribute below which an individual always qualifies as being poor, independent of his other attributes. This part may seem objectionable but is not necessary for our analysis, and plays no qualitative role. It enables us to simplify the statements of our results.

Since only $\mathbf{x}^o$ is observable, we cannot rely on λ to characterize individuals as poor or not poor. Instead, our goal is to find a real valued function μ such that an individual with attributes x is classified as poor if and only if $\mu(\mathbf{x}^o) < 0$. If we use such a function μ that ignores the unobservable attributes $\mathbf{x}^u$, we will be making two kinds of error. Some individuals with high values in $\mathbf{x}^u$ will be classified as poor when in fact they are not poor. Other individuals with low values in $\mathbf{x}^u$ will be classified as not poor when in fact they are poor. The error regions are respectively

$$E_1(\mu|\lambda) = \left\{\mathbf{x} \in \mathbb{R}^n_{++} \mid \mu(\mathbf{x}^o) < 0 \text{ and } \lambda(\mathbf{x}) \geq 0\right\}$$
$$E_2(\mu|\lambda) = \left\{\mathbf{x} \in \mathbb{R}^n_{++} \mid \mu(\mathbf{x}^o) \geq 0 \text{ and } \lambda(\mathbf{x}) < 0\right\}$$

The total expected fraction of misclassified individuals is

$$\varepsilon(\mu|\lambda) \equiv \mathbb{F}(E_1(\mu|\lambda) \cup E_2(\mu|\lambda)) = \int_{E_1(\mu|\lambda)\cup E_2(\mu|\lambda)} f(\mathbf{x})\mathrm{d}\mathbf{x}. \tag{2}$$

Our problem is to find a function μ that minimizes this classification error for given λ. The only restriction we impose is that μ be strictly increasing. Thus we are minimizing (2) over the set of all strictly increasing functions.

3 An Example with Two Attributes

Suppose that there are only two attributes. One attribute is observable and the other is not. Because μ must be strictly increasing, characterizing an individual as poor if and only if $\mu(x_1) < 0$ is equivalent to finding $\ell \geq 0$ such that an individual is poor if and only if $x_1 < \ell$. Further, since λ is strictly decreasing throughout the x_1-axis, we can invert $\lambda(x_1, x_2) = 0$ to give us the function $x_2 = \theta(x_1)$ such that $\lambda(x_1, \theta(x_1)) = 0$ for all $x_1 > 0$. In Figure 1 we depict θ, and the error regions E_1^* and E_2^* at the solution to the minimization problem. Our problem is to find $\ell \geq 0$ to minimize

$$\int_0^\ell \int_{\theta(x_1)}^\infty f(x_1, x_2)\mathrm{d}x_2\mathrm{d}x_1 + \int_\ell^\infty \int_0^{\theta(x_1)} f(x_1, x_2)\mathrm{d}x_2\mathrm{d}x_1 \tag{3}$$

If we take the derivative of (3) with respect to ℓ and set it equal to zero, we have

$$\int_{\theta(\ell)}^\infty f(\ell, x_2)\mathrm{d}x_2 - \int_0^{\theta(\ell)} f(\ell, x_2)\mathrm{d}x_2 = 0.$$

Add $2\int_0^{\theta(\ell)} f(\ell, x_2)\mathrm{d}x_2$ to both sides, rearrange and apply equation (1) to get

$$\mathbb{F}(\theta(\ell)|\ell) = \frac{1}{2}, \tag{4}$$

where we slightly abuse notation and use $\mathbb{F}(x_2|x_1)$ to mean $\mathbb{F}\left((0, x_2]|x_1\right)$. This equation characterizes the value of ℓ that minimizes (3).

The solution to (4) is unique. Let ℓ^* be the solution. For any $\ell' < \ell^*$, we have

$$\mathbb{F}\left(\theta(\ell')|\ell'\right) > \mathbb{F}(\theta(\ell^*)|\ell') > \mathbb{F}(\theta(\ell^*)|\ell^*) = \frac{1}{2},$$

where the first inequality holds because θ is a decreasing function and the second is due to Assumption 1. Therefore, it cannot be that any $\ell' < \ell^*$ satisfies (4). In like manner, we can show that for $\ell' > \ell^*$ we have $\mathbb{F}(\theta(\ell')|\ell^*) < \frac{1}{2}$ and therefore no $\ell' > \ell^*$ can be a solution to the minimization problem.

Finally, we show that a solution to (4) exists. To that end, define $m(x_1)$ to be the median of the distribution $\mathbb{F}(x_2|x_1)$. Observe that m is a strictly increasing continuous function. By Assumption 2, θ is continuous, strictly decreasing and onto. Therefore, m intersects θ and that, by definition, is the point at which $\mathbb{F}(\theta(\ell)|\ell) = \frac{1}{2}$.

Remark 1. We mentioned in the previous section that Assumption 2(iii) plays no significant role in our analysis. One may object that we used it in proving the existence of a solution to (4). In particular, the function $\tilde{\lambda}(x_1, x_2) = \gamma \log\left(x_1 - \underline{x}_1\right) + \log\left(x_2 - \underline{x}_2\right) - p$ for fixed $\underline{x}_1$, $\underline{x}_2$, p, $\gamma > 0$ is a reasonable choice for a poverty frontier, but does not satisfy Assumption 2(iii) because for $0 < x_i \leq \underline{x}_i$ there is no value of x_{-i} that makes $\tilde{\lambda}(x_1, x_2) = 0$.[4]

However, even for choices of poverty frontier such as $\tilde{\lambda}$ we are still able to find a solution to the minimization of (3). If a solution to (4) exists, then that is the solution to the minimization problem. If a solution does not exist, then there are only two possibilities: either $\mathbb{F}(\theta(\ell)|\ell) > \frac{1}{2}$ for all $\ell > 0$, or $\mathbb{F}(\theta(\ell)|\ell) < \frac{1}{2}$ for all $\ell > 0$. In the first case, the solution is $\ell^* = \infty$ (and $\mu(x_1)$ can be asymptotically increasing to 0) while in the second case it is $\ell^* = 0$ (and $\lim_{x_1 \to 0^+} \mu(x_1) > 0$ for any strictly increasing μ). That is, in the first case we classify every member of society as being poor, and in the second we classify no individual as being poor.

[4]The function $\tilde{\lambda}$ is the Stone-Geary function that hypothesizes subsistence levels of $\underline{x}_1$, $\underline{x}_2$ for each attribute. These subsistence levels can be interpreted to be the absolute poverty lines chosen in the union approach to multidimensional poverty measurement. In other words, an individual will always be characterized as poor if any x_i does not exceed $\underline{x}_i$, since for that individual $\lambda(x_1, x_2) = -\infty$ independent of x_{-i}.

Remark 2. An implication of equation (4) is that unequivocally better off societies have lower poverty thresholds. Suppose that two societies summarized by their distributions $\mathbb{F}$ and $\tilde{\mathbb{F}}$ use the same poverty frontier λ. Society $\mathbb{F}$ is better off than society $\tilde{\mathbb{F}}$ in the sense that $\mathbb{F}(\cdot|x_1) < \tilde{\mathbb{F}}(\cdot|x_1)$ for all $x_1 > 0$. If $\tilde{\ell}^*$ is the poverty threshold in society $\tilde{\mathbb{F}}$ then $\frac{1}{2} = \tilde{\mathbb{F}}(\theta(\tilde{\ell}^*)|\tilde{\ell}^*) > \mathbb{F}(\theta(\tilde{\ell}^*)|\tilde{\ell}^*)$ by Assumption 1.[5]

The function m defined above is also depicted in Figure 1, as is a hypothetical $\ell' < \ell^*$. The classification error under ℓ' differs from the classification error under the optimal ℓ^* by the regions A, B and C. More precisely, $E_1' \cup A = E_1^*$ and $E_2' = E_2^* \cup B \cup C$. By definition of m we have $\mathbb{F}(C) = \mathbb{F}(A \cup B) = \mathbb{F}(A) + \mathbb{F}(B)$ and the full support hypothesis ensures $\mathbb{F}(B) > 0$ so that $\mathbb{F}(A) < \mathbb{F}(B) + \mathbb{F}(C) = \mathbb{F}(B \cup C)$. This establishes that $\ell' < \ell^*$ cannot minimize the classification error. A similar argument can be used to show the $\ell' > \ell^*$ cannot do so either.

4 The Main Characterization Result

Now we return to the general case of n attributes only $k < n$ of which are observable. Fix observable attributes $x_{-1}^o = (x_2, \dots, x_k)$ and define the set

$$\Theta(x_1|x_{-1}^o) = \left\{ \mathbf{x}^u \in \mathbb{R}_{++}^{n-k} \mid \lambda(x_1, x_{-1}^o, \mathbf{x}^u) < 0 \right\}.$$

If there is only one observable attribute, then no others can be fixed and this set can be denoted $\Theta(x_1)$. The following is our main characterization result.

Theorem 1. *The value of x_1 solving $\mathbb{F}\left(\Theta(x_1|x_{-1}^o)|(x_1, x_{-1}^o)\right) = \frac{1}{2}$ is unique (for each $x_{-1}^o \in \mathbb{R}_{++}^{k-1}$ when there is more than one observable attribute). Denote it $x_1^*(x_{-1}^o)$ as a function of the observable attributes held fixed. Then any increasing function μ that characterizes an individual with observable attributes (x_1, x_{-1}^o) as poor if and only if $x_1 < x_1^*(x_{-1}^o)$, is a solution to the minimization of $\varepsilon(\mu|\lambda)$ over increasing functions.*

Proof. Fix $x_{-1}^o \in \mathbb{R}_{++}^{k-1}$ if there is more than one observable attribute. We first minimize the fraction of misclassified individuals in the set of those individuals having attributes x_{-1}^o. For any $\ell \geq 0$, the fraction of misclassified individuals in this set is

$$\int_{\ell}^{\infty} \int_{\Theta\left(x_1|x_{-1}^o\right)} f(x_1, x_{-1}^o, \mathbf{x}^u) \mathrm{d}\mathbf{x}^u \mathrm{d}x_1 + \int_0^{\ell} \int_{\mathbb{R}_{++}^{n-k} \setminus \Theta(x_1|x_{-1}^o)} f(x_1, x_{-1}^o, \mathbf{x}^u) \mathrm{d}\mathbf{x}^u \mathrm{d}x_1.$$

[5] Using subscripts to denote first derivatives, note that $\frac{\mathrm{d}}{\mathrm{d}\ell}\mathbb{F}(\theta(\ell)|\ell) = \mathbb{F}_\theta \frac{\mathrm{d}}{\mathrm{d}\ell}\theta + \mathbb{F}_\ell < 0$ because $\mathbb{F}_\theta > 0$, $\frac{\mathrm{d}}{\mathrm{d}\ell}\theta < 0$, and (by Assumption 1) $\mathbb{F}_\ell < 0$.

Taking derivatives with respect to ℓ and setting it equal to 0,

$$-\int_{\Theta\left(\ell|x^o_{-1}\right)} f\left(\ell, x^o_{-1}, \mathbf{x}^u\right) \mathrm{d}\mathbf{x}^u + \int_{\mathbb{R}^{n-k}_{++}\setminus\Theta(x_1|x^o_{-1})} f\left(\ell, x^o_{-1}, \mathbf{x}^u\right) \mathrm{d}\mathbf{x}^u = 0.$$

Now add $2\int_{\Theta(\ell|x^o_{-1})} f(\ell, x^o_{-1}, \mathbf{x}^u)\mathrm{d}\mathbf{x}^u$ to both sides and rearrange to get

$$\mathbb{F}(\Theta(\ell|x^o_{-1})|(\ell, x^o_{-1})) = \frac{1}{2}. \tag{5}$$

We show that for each x^o_{-1} the value of ℓ solving (5) is unique. Let ℓ^* be a solution to (5). For $\ell' < \ell^*$,

$$\mathbb{F}(\Theta(\ell'|x^o_{-1})|(\ell', x^o_{-1})) > \mathbb{F}(\Theta(\ell^*|x^o_{-1})|(\ell', x^o_{-1})) > \mathbb{F}(\Theta(\ell^*|x^o_{-1})|(\ell^*, x^o_{-1})) = \frac{1}{2},$$

where the first inequality follows from $\Theta(\ell^*|x^o_{-1}) \subsetneq \Theta(\ell'|x^o_{-1})$ by Assumption 2, and the second inequality is due to Assumption 1 since $\Theta(\ell^*|x^o_{-1})$ is clearly comprehensive. A similar argument in the case of $\ell' > \ell^*$ establishes that $\mathbb{F}(\theta(\ell'|x^o_{-1})|(\ell', x^o_{-1})) < \frac{1}{2}$ and hence such an ℓ' cannot be a solution either.

We now argue that a solution to (5) exists. By Assumption 2, we can invert $\lambda(\mathbf{x}) = 0$ to give us a positive valued function $\theta(x_{k+1}, \ldots, x_{n-1}|x_1, x^o_{-1})$ such that

$$\lambda(x_1, \ldots, x_{n-1}, \theta(x_{k+1}, \ldots, x_{n-1}|x_1, x^o_{-1})) = 0.$$

Now define the set

$$\Theta_\kappa(x_1|x^o_{-1}) = \left\{(x_{k+1}, \ldots, x_n) \in \mathbb{R}^{n-k}_{++}|\theta(x_{k+1}, \ldots, x_{n-1}|x_1, x^o_{-1}) + \kappa > x_n\right\},$$

and note that for κ sufficiently negative $\mathbb{F}(\Theta_\kappa(x_1|x^o_{-1}|x_1, x^o_{-1}) < \frac{1}{2}$ while for κ sufficiently high $\mathbb{F}(\Theta_\kappa(x_1|x^o_{-1})|x_1, x^o_{-1}) > \frac{1}{2}$. Clearly $\mathbb{F}(\Theta_\kappa(x_1|x^o_{-1})|x_1, x^o_{-1})$ is continuous and strictly increasing in κ, so that there is a unique $\kappa^*(x_1, x^o_{-1})$ such that

$$\mathbb{F}(\Theta_{\kappa^*(x_1,x^o_{-1})}(x_1|x^o_{-1})|x_1, x^o_{-1}) = \frac{1}{2}.$$

Define the function

$$m(x_{k+1}, \ldots, x_{n-1}|x_1, x^o_{-1}) = \theta(x_{k+1}, \ldots, x_{n-1}|x_1, x^o_{-1}) + \kappa^*(x_1, x^o_{-1}),$$

and note that $m(x_{k+1}, \ldots, x_{n-1}|x_1, x^o_{-1})$ is increasing in x_1. $\theta(x_{k+1}, \ldots, x_{n-1}|\cdot, x^o_{-1})$ is strictly decreasing in x_1 and onto (by Assumption 2), so m intersects θ at the same value of x_1 for every value of $(x_{k+1}, \ldots, x_{n-1})$ still holding x^o_{-1} fixed. That is precisely the point at which $\kappa^*(x_1, x^o_{-1}) = 0$, and thus a solution to (5) exists. □

Denote by $\ell^*(x^o_{-1})$ the solution to (5), which makes explicit the dependence of ℓ^* on the fixed observable attributes. Take any strictly increasing $\mu(\mathbf{x}^o)$ such that $\mu(\ell^*(x^o_{-1}), x^o_{-1}) = 0$. Obviously, such a μ minimizes the total classification error $\varepsilon(\mu|\lambda)$ since it minimizes the classification error for every x_1-cohort.

5 Empirical Illustration

We now illustrate our approach using survey data that Acharya, Roemer and Somanathan (2015) collected in Uttar Pradesh, India, in 2008-09. They asked each of 4,675 respondents a number of questions related to income and well-being. For example, they asked individuals whether they were homeowners; what their houses were made of, and how many rooms they had; the amount of land they owned (and cultivated); and the quantity of certain items that they owned, such as televisions, phones, cameras, fans, chairs, charpoys, etc. We did a factor analysis of the answers to these questions, and found that most information loaded onto a single factor. We divided the results for this factor by the square root of household size, and took the log of the outcome. We call this quantity x_1. This is our income measure. We let x_2 denote the number of years of schooling of the respondent. We will assume, just for the sake of illustration, that we can observe x_1 but not x_2.[6]

We ordered the respondents by their values of x_1, dividing them into 47 groups, each consisting of 100 individuals (except the final, wealthiest, group which consisted of 75 individuals). For each group, we calculated the mean value of x_1 and the median value of x_2. These numbers were used to approximate the function $m(x_1)$ depicted in Figure 2. The figure shows that the distribution of (x_1, x_2)-types violates Assumption 1, since m is not a strictly increasing function. However, this will not affect our analysis (since m will intersect our poverty frontier at a unique point anyway).

We use the function $\tilde{\lambda}(x_1, x_2)$ defined in Remark 1 to illustrate a poverty frontier. Invert this function to get a function $\tilde{\theta}(x_1)$ such that $\tilde{\lambda}(x_1, \tilde{\theta}(x_1)) = 0$ for all $x_1 > \underline{x}_1$. $\tilde{\theta}(x_1)$ is depicted in Figure 2 for $p = \log 3$, $\gamma = 2$, $\underline{x}_1 = 0$, and $\underline{x}_2 = -1$. The reason we take $\underline{x}_2 < 0$ is because we wish everybody above a certain income $\overline{x}_1$ to be classified as not poor, regardless of how many years of school they have had. These parameter choices are only illustrative.

Despite the violations of Assumptions 1 and 2, the poverty line in income that minimizes the total number of misclassified individuals is still the point at which $\tilde{\theta}(x_1)$ intersects $m(x_1)$, which we have labeled $\ell^* = \overline{x}_1$ $(= \sqrt{3})$. The total number of misclassified individuals is 401. All of these individuals are classified as living in

[6] One might wonder why we think that x_2, education, is observable but x_1, our income proxy, is not, especially given the fact that education was observable while income was not in the Acharya-Roemer-Somanathan survey. (This is why we had to construct an income proxy using answers to a number of different questions about property ownership.) The reason that we have chosen to assume that income is observable but education is not is because we feel this is the more natural assumption, since policy-makers currently use income measures to decide who lives below or above poverty.

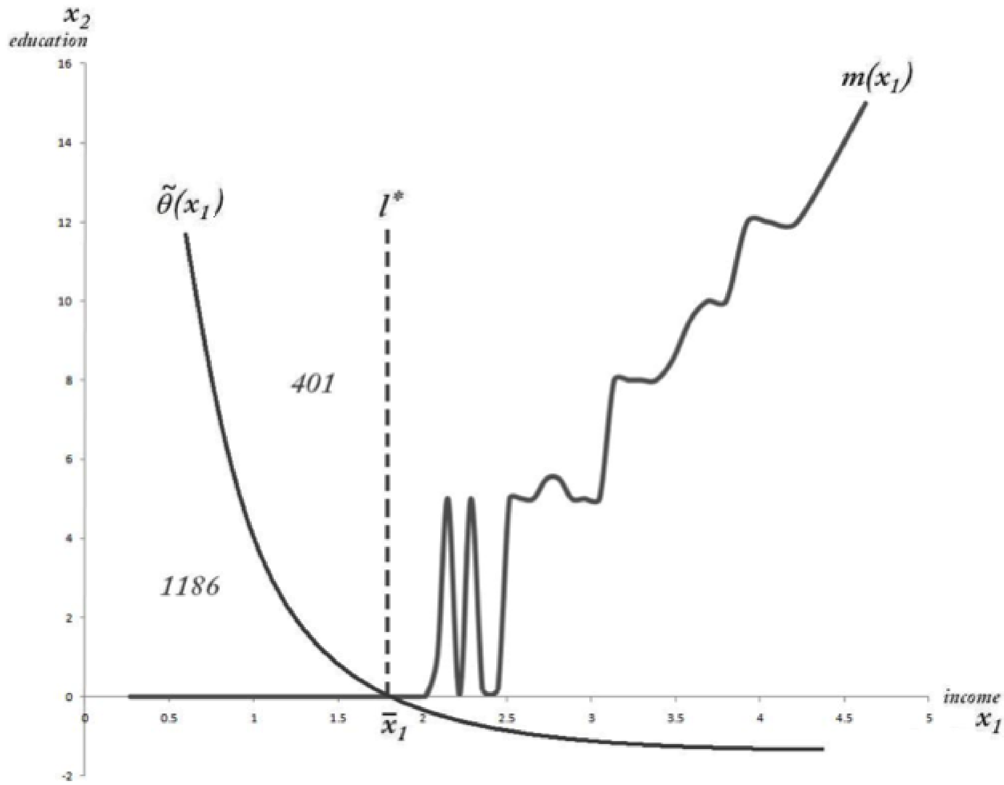

Figure 2. Minimizing the misclassified in Uttar Pradesh.

poverty, even though according to our choice of $\tilde{\lambda}$ they are not. The total number of individuals living in poverty according to $\tilde{\lambda}$ is 1186. There is no type II error.

Therefore, suppose that all 401 + 1,186 = 1,587 individuals that we classify as poor were to receive either a "below the poverty line" (BPL) card or an *Antyodaya anna yojana* (AAY) card.[7] How does this this compare to the number of respondents in the Acharya-Roemer-Somanathan survey who said that they possessed one of these cards? A total of 1,205 individuals in the survey said that they were either BPL or AAY cardholders.[8] We find that 922 individuals with values of x_1 below ℓ^* are not

[7]A BPL card entitles the cardholder to food grains (rice and wheat) and fuel (kerosene) at subsidized prices. To qualify, an individual must have an income below a certain threshold, typically Indian Rupees 10,000 per annum. An AAY card has similar benefits and is given to "temporarily ill persons, disabled persons, or persons aged 60 years or more with no assured means of subsistence or societal support."

[8]That is, 74% of respondents in the sample said they held an "above the poverty line" (APL) card or no card. This is not too different from the 83.5% of individuals reported by *The Hindu* as having an APL card or no card (Madhura Swaminathan. "Public Distribution System and Social

BPL or AAY cardholders. Similarly, 540 individuals, who do not have one of these cards, do not have values of x_1 below ℓ^*.[9]

6 Discussion

There are reasons to expect policy-makers to want to minimize the expected fraction of misclassified individuals—as a normative goal, or because budgets are limited, or for political reasons. We have characterized the solution to this minimization problem. Our approach is easily generalizable to alternative normative considerations, and alternative measures of poverty. For example, one might want to classify an individual with observable attributes $\tilde{\mathbf{x}}^o$ as poor if and only if $\lambda(\tilde{\mathbf{x}}^o, \mathbb{E}(\mathbf{x}^u|\tilde{\mathbf{x}}^o)) < 0$, where $\mathbb{E}(\mathbf{x}^u|\tilde{\mathbf{x}}^o)$ denotes the conditional expectation of $\mathbf{x}^u$ given $\tilde{\mathbf{x}}^o$. (Such a policy would not generally minimize the total *expected* number of misclassified individuals.) Alternatively, suppose that type II errors (error regions E_2) are more socially costly than type I errors (regions E_1) and thus should receive more weight. That is, misclassifying a poor individual as being not poor is more socially costly than misclassifying an individual that is not poor as being poor. In this case, one would want to find μ to minimize $\int_{E_1(\mu|\lambda)} f(\mathbf{x})\mathrm{d}\mathbf{x} + \alpha \int_{E_2(\mu|\lambda)} f(\mathbf{x})\mathrm{d}\mathbf{x}$, where $\alpha > 1$, instead of minimizing $\varepsilon(\mu|\lambda)$ in (2). In general, for a given any poverty measure, our approach can be generalized to minimizing the (weighted) average number of misclassified individuals, according to that measure.

Exclusion" *The Hindu*, May 7, 2008.)

[9] There is apparently so much corruption in deciding who gets BPL or AAY cards that the government of the state of Madhya Pradesh proposed writing "I am poor" on the houses of individuals who have these cards to create a disincentive for the affluent from paying bribes to receive them. See "'I am poor' tag to keep affluent from BPL cards," http://www.khabarexpress.com/.

References

Acharya, Avidit, John Roemer and Rohini Somanathan. 2015. "Caste, Corruption and Distribution in India." forthcoming in *Research in Economics*.

Atkinson, A.B. 2003. "Multidimensional Deprivation: Contrasting Social Welfare and Counting Approaches." *Journal of Economic Inequality* 1(1):51–65.

Bourguignon, François and Satya R. Chakravarty. 2003. "The Measurement of Multidimensional Poverty." *Journal of Economic Inequality* 1(1):25–49.

Bourguignon, François and Satya R. Chakravarty. 2009. "Multi-dimensional Poverty Orderings: Theory and Applications." in *Arguments for a Better World: Essays in Honor of Amartya Sen: Volume I: Ethics, Welfare and Measurement*, ed. by K. Basu and R. Kanbur.

Duclos, Jean-Yves., David E. Sahn and Stephen D. Younger. 2006. "Robust Multidimensional Poverty Comparisons." *Economic Journal* 116(514):943–968.

Tsui, Kai-yuen. 2002. "Multi-dimensional Poverty Indices." *Social Choice and Welfare* 19:69–93.