Uncovering Household Self-Targeting with Machine Learning

Moya Chin*

EXTENDED ABSTRACT

Please do not cite or circulate without permission.

December 16, 2019

Abstract

I investigate the extent to which households in Colombia manipulate their eligibility for a social program. Eligibility is determined by a poverty score that is calculated from answers to a household survey, the formula for which was released four years after the start of the program. Because proxy-means testing systems can potentially predict household poverty poorly, households have incentives to manipulate their eligibility. I find that, as in Camacho and Conover (2011), there is a significant discontinuity at the eligibility cutoff and that one method households use to manipulate their eligibility is by having their poverty score overwritten. I then use machine-learning techniques to predict households' actual poverty level. I find that households who manipulate their score are more likely to be poorer than households with the same score and are more likely to be poorly predicted by the government's poverty score. These findings suggest that not all proxy-means testing systems predict household poverty well and when they do not, households self-target by manipulating their eligibility.

*Harvard University, Department of Economics, mchin@g.harvard.edu. I am grateful to Melissa Dell, Juan Sebastián Galán, Gautam Rao, and Jonathan Roth. This project received numerous helpful comments from participants at the Harvard Development Tea and the Applied Microeconomics seminar.

1. Introduction

Targeting households for social welfare programs has remained a constant issue for governments. In developing countries, where the poor are often concentrated in the informal sector, governments must deal with costly income verification and agency problems that may arise among those who determine eligibility (Niehaus and Sukhtankar, 2013; Olken and Pande, 2012; Reinikka and Svensson, 2004).

To minimize these issues, many governments use proxy-means testing systems (PMTs). In a PMT, household information, such as asset ownership and demographics, is collected in order to predict a household's poverty status and determine eligibility. However, these systems may be subject to manipulation, as households may engage in lying or fraud to become eligible. This can increase or decrease targeting efficiency, depending on whether households who qualify through manipulation are more or less likely to be eligible. This project studies the implementation of a PMT for a large social program in Colombia before and after the targeting rule was made public. The resulting manipulation by households to become eligible allows me to study which households self-target and when they do so.

While many PMTs use sophisticated econometric techniques, there is far from consensus on their prediction quality (Coady et al., 2004; Kidd and Wylde, 2011). PMTs often rely on techniques that (1) are un-regularized, which contributes to over-fitting of the data, and (2) do not identify interactions in the data, which leads to model mis-specification. These issues can result in poor out-of-sample predictions and have motivated growing interest in machine learning methods (Blumenstock, 2016; Mullainathan and Spiess, 2017). This project applies machine learning methods to evaluate the prediction quality of a PMT used in Colombia.

Because PMTs can potentially predict household poverty poorly and the benefits of program eligibility are high, households face large incentives to manipulate their way into eligibility. Following the release of the formula for the eligibility rule, there were increased opportunities for households to engage in manipulation. The second goal of this project is to document whether the prediction quality of the government's PMT influences whether households engage in manipulation.

2. Empirical strategy

I revisit Camacho and Conover (2011) and study Sisbén I, a national social program implemented between 1994 and 2003 in Colombia. A household's poverty status is summarized by a poverty score, an index that ranges from 0 to 100, representing the probability that the household's consumption level is above the poverty line (lower scores indicate higher poverty). Various government welfare programs then use the poverty score to determine eligibility. The formula for calculating the poverty score was released to local officials in mid-1997.

I use the administrative Sisbén survey, which contains the answers from the household survey and each household's poverty score. Because Sisbén does not contain consumption measures, I use the 1993, 1997, and 2003 Quality of Life surveys to train machine learning algorithms to create a prediction algorithm that can be applied to the Sisbén survey. I estimate an ensemble of a random forest, boosted tree, and lasso.

3. Results

First, as in Camacho and Conover (2011), I find a significant discontinuity in the density of households at the eligibility cutoff after 1998, when the poverty score formula was released, suggesting that poverty scores are being manipulated (Figure 1). I document that one method households use to manipulate their eligibility is by having the poverty score altered. I do not find evidence that these irregularities are due to mis-calculation or other errors.

Second, households who had their poverty scores altered, compared to households with the same original score but who did *not* have their scores altered, are more likely to be poor as predicted by a machine-learning algorithm (Figure 2). I find that the government's poverty score predicts poverty poorly, particularly for the poorest households (Figure 3).

Third, households for whom the government's poverty score under-predicts their poverty are more likely to have their score altered (Figure 4). This effect is seen both before and after the score was released, suggesting that households are able to predict when their poverty will be predicted poorly by the government.

References

Blumenstock, Joshua (2016). "Fighting poverty with data." *Science* 353.6301, pp. 753–754.

Camacho, Adriana and Emily Conover (2011). "Manipulation of Social Program Eligibility." *American Economic Journal: Economic Policy* 3.2, pp. 41–65.

Coady, David, Margaret Grosh, and John Hoddinott (2004). "Targeting Outcomes Redux." *World Bank Research Observer* 19.1, pp. 61–85.

Kidd, Stephen and Emily Wylde (2011). "Targeting the Poorest: An assessment of the proxy means test methodology." *AusAID Working Paper*.

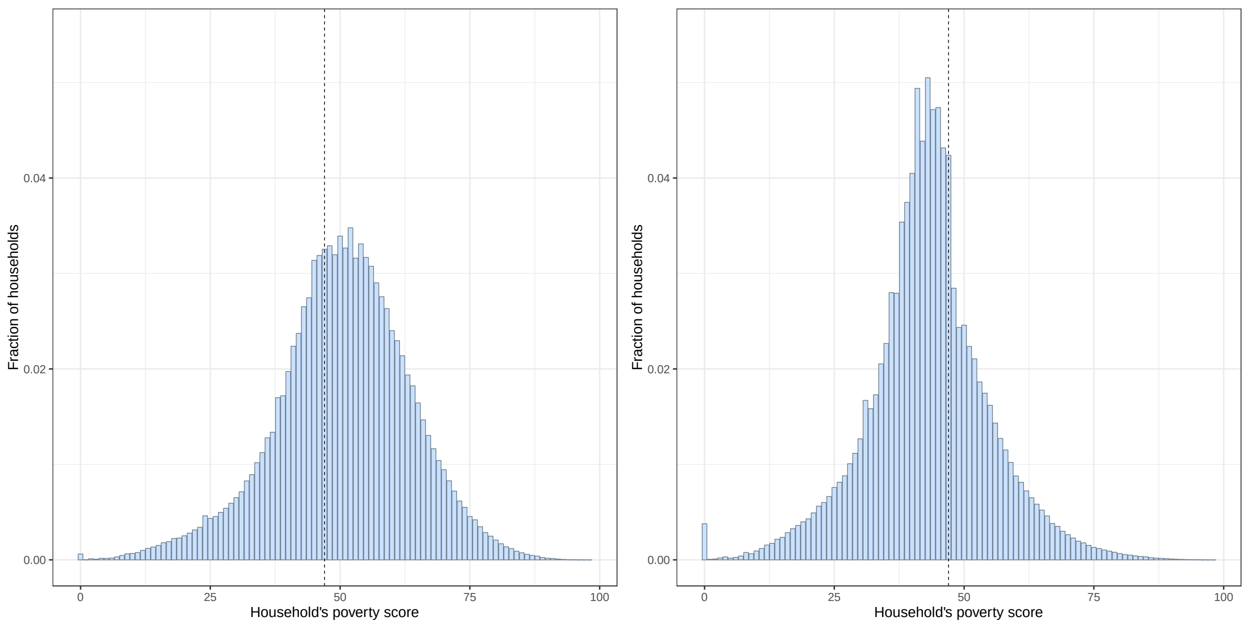
Mullainathan, Sendhil and Jann Spiess (2017). "Machine Learning: An Applied Econometric Approach." *Journal of Economic Perspectives* 31.2, pp. 87–106.

Niehaus, Paul and Sandip Sukhtankar (2013). "Corruption Dynamics: The Golden Goose Effect." *American Economic Journal: Economic Policy* 5.4, pp. 230–269.

Olken, Benjamin A. and Rohini Pande (2012). "Corruption in Developing Countries." *Annual Review of Economics* 4, pp. 479–509.

Reinikka, Ritva and Jakob Svensson (2004). "Local Capture: Evidence from a Central Government Transfer Program in Uganda." *Quarterly Journal of Economics* 119.2, pp. 679–705.

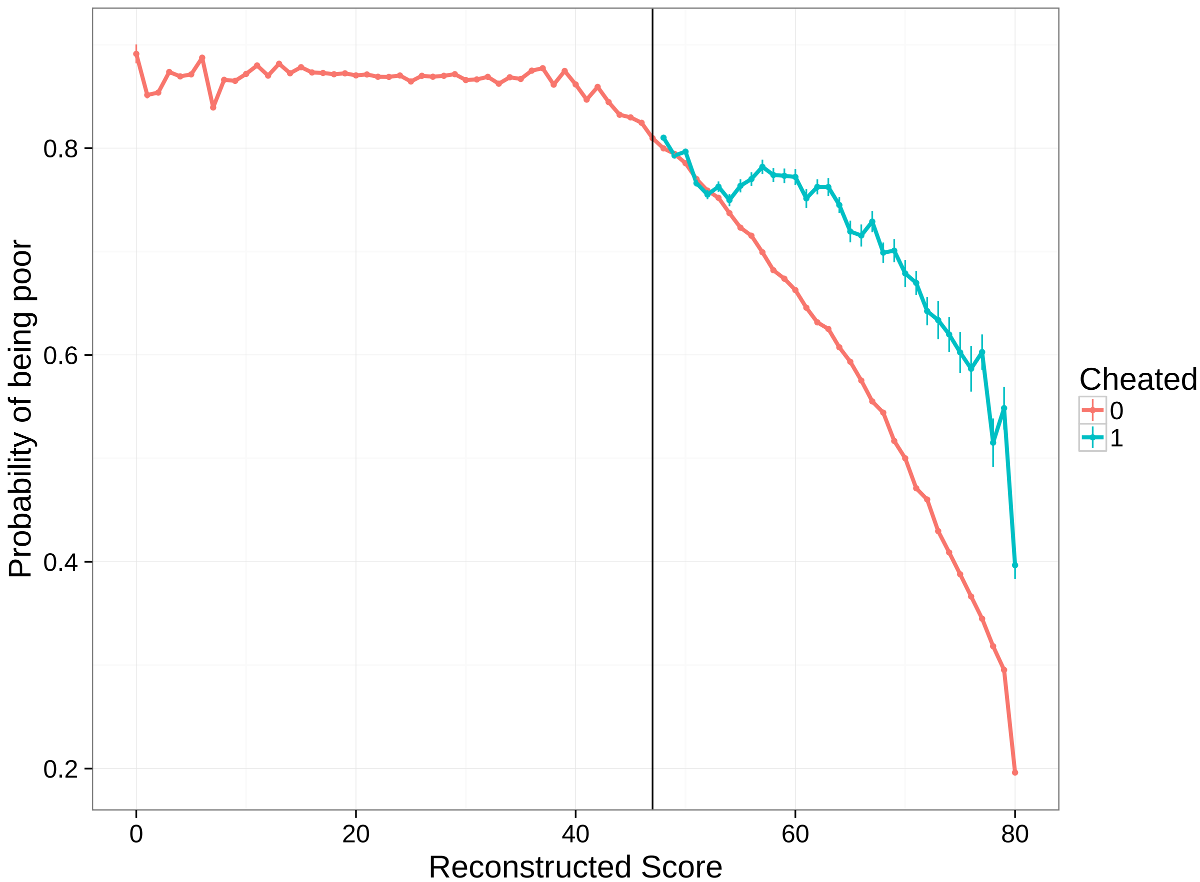
Figure 1 Density of poverty scores, before and after 1998



(a) 1994-1997 **(b)** 1998-2002

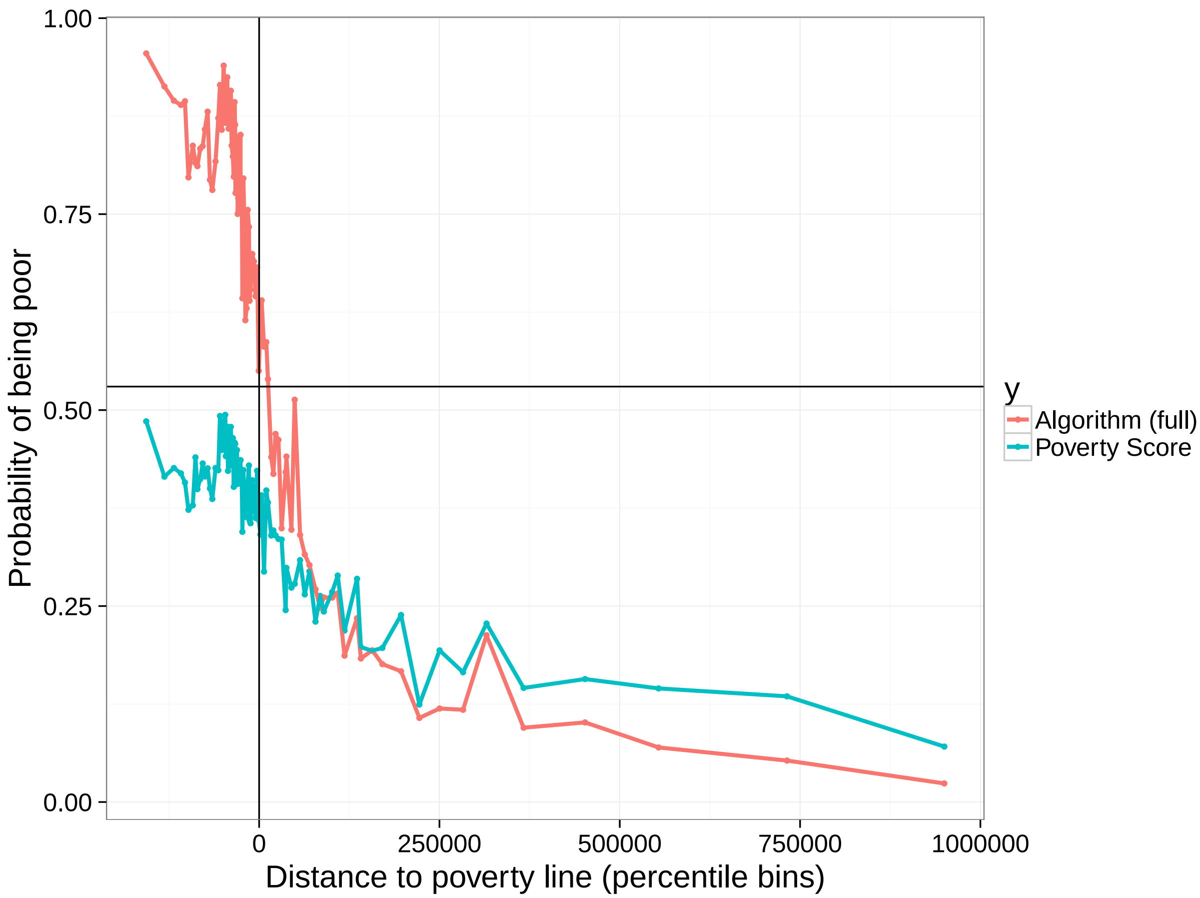
Data from Sisbén survey. Density of poverty scores assigned to households in Sisbén. The poverty score algorithm was released to local officials in late 1997. Households with a score below 47 (the dotted vertical line) are eligible for a variety of social programs.

Figure 2 Households whose scores are manipulated are more likely to be poor than households whose scores are not manipulated



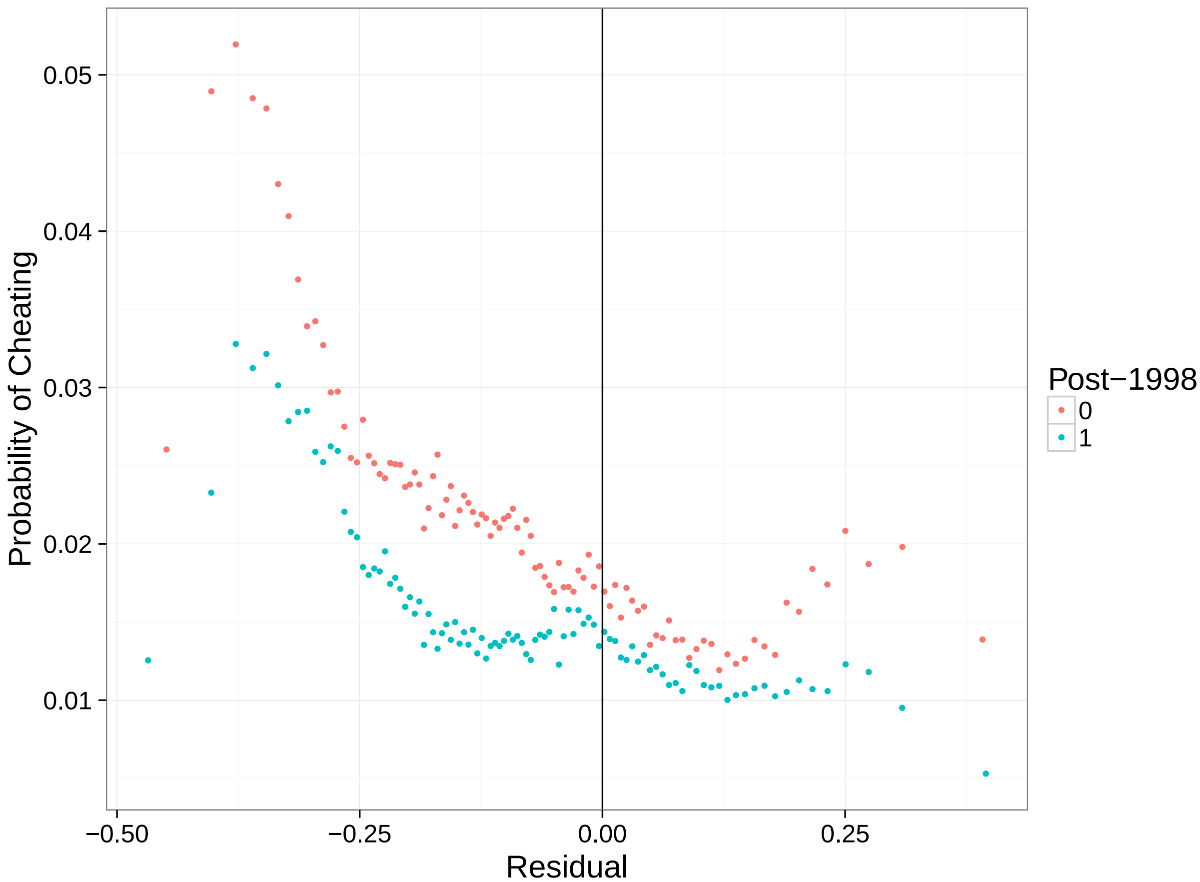
Data from Sisbén survey, for surveys conducted after 1998. Horizontal axis is the household's reconstructed poverty score, prior to any manipulation. Vertical axis is the household's probability of being poor, as predicted by the machine learning algorithm. Outcomes plotted separately for households whose poverty score was manipulated (blue line) and households whose poverty score was not manipulated (red line). Households to the left (right) of the vertical line are (not) eligible for social programs.

Figure 3 Prediction quality by household consumption



Data from Quality of Life surveys. Horizontal axis is the household's actual distance from the poverty line. Vertical axis is the household's predicted probability of being poor (either by the machine learning algorithm, in red, or by the government's poverty score, in blue). Binned by percentile, so that each point in the graph represents 1% of households. Households to the right (left) of the vertical line are above (below) the poverty line. Households above (below) the horizontal line are (not) eligible for social programs.

Figure 4 Households who manipulate their score are more likely to be considered richer by the government's poverty score



Data from Sisbén survey, for surveys conducted before 1998 (in red) and after 1998 (in blue). Horizontal axis is the household's residual, which is defined as the government's poverty score minus poverty as predicted by the machine learning algorithm. Households with a negative (positive) residual, ie. to the left (right) of the vertical line, are *poorer* (richer) than the government's poverty score would predict. Binned by percentile, so that each point in the graph represents 1% of households.