Economics Letters 113 (2011) 58–61

Contents lists available at SciVerse ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

The application of nonparametric tests to poverty targeting[☆]

Isha Dewan[a], Rohini Somanathan[b,c,*]

[a] Indian Statistical Institute, New Delhi, India
[b] Delhi School of Economics, 110007 Delhi, India
[c] Institute for Advanced Study, Princeton, United States

ARTICLE INFO

Article history:
Received 9 May 2011
Received in revised form
23 May 2011
Accepted 27 May 2011
Available online 6 June 2011

JEL classification:
C14
I32

Keywords:
Poverty targeting
Distribution crossing
Nonparametric tests
Microcredit

ABSTRACT

Many otherwise successful social programs have limited outreach among the very poor. We show how a recently developed nonparametric test can detect this pattern of program participation. We apply the test to data on participation in a microcredit program in India and do find participation increasing in income among the poorest households in the region.

1. Introduction

Benefits from public programs are often complementary to individual skills and assets. Job training for the unemployed relies on some education, microcredit is most useful to borrowers with self-employment opportunities and even cash transfer schemes are administered effectively to those with a stable residence. Extreme poverty may be accompanied by levels of migration, education and information that are not conducive to program participation. In such cases, the probability of participation may first increase and then decrease with income or, equivalently, the distribution function of income for participants may cross that of non-participants from below over some income range.

One may not detect this pattern of participation through a comparison of mean characteristics of the two groups which may be similar because participants occupy the middle of the distribution with non-participants on each side. This paper asks whether we can usefully expand the set of tools used in targeting studies by incorporating methods that focus on the crossing of distributions. Specifically, we explore the effectiveness of a recently developed test for distribution crossing which uses the null hypothesis of equal distributions and an alternative of a single crossing. Numerical simulations on alternative pairs of distributions suggest that the test has high power even when the distributions being considered are very similar.

The approach used in the test also provides us with an estimate of the crossing point. This helps identify appropriate quantiles at which to compare the two distributions. We illustrate how the crossing point procedure can be combined with quantile tests to characterize the population that is best served by a particular program. Our results lead us to conclude that if there is reason to believe, *a priori*, that participation rates first increase and then decrease with income, a combination of tests for distribution crossing and quantile tests can identify this pattern of program participation.

We conclude with an application of the test to baseline survey data from a microcredit program in India. We find evidence that the distribution of members crosses that of non-members from below in spite of there being no statistically significant difference in the mean characteristics of the two groups. The application illustrates how the crossing point test might be integrated with other approaches in order to understand program participation among the poor.

☆ Thanks to Charlie Brown, Julie Cullen, Esther Duflo, Roger Gordon, Subhash Kochar, B.L.S. Prakasa Rao, Charles Manski and E. Somanathan for comments, and to the Consultative Group to Assist the Poorest at the World Bank for funding the collection of the survey data used in the last section of the paper. Rohini Somanathan would also like to thank the Leon Levy Foundation for support of her membership at the Institute for Advanced Study during 2010–2011 during which period the final version of this paper was written.

* Corresponding author at: Delhi School of Economics, 110007 Delhi, India. Tel.: +91 011 2766533; fax: +91 011 27667159.
E-mail addresses: isha@isid.ac.in (I. Dewan), rohini@econdse.org, rohini@ias.edu (R. Somanathan).

0165-1765/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.econlet.2011.05.042

2. Methods

We denote the population distributions of program participants and non-participants by $F(x)$ and $G(x)$ respectively and are interested in testing the null hypothesis of equal distributions against the alternative that the distribution of participants crosses that of non-participants from below. This would be equivalent to the probability of participation first increasing and then decreasing with income.

The corresponding empirical distribution functions are $F_n(x)$ and $G_m(x)$ where n and m are the respective sample sizes. There are therefore a total of $N = m + n$ order statistics in the combined sample. The Kolmogorov–Smirnov test that is commonly used to test for the equality of two distributions has two well known disadvantages in this setting. First, the low power of this test makes it unlikely that we reject the null when the population distributions of the two groups are similar. Second, the rejection of the null does not provide much information about the alternative. We focus on a test for single crossing of the two distributions proposed by Chen et al. (2002).[1]

We start by summarizing the approach used in deriving the test statistic. If x is income, our hypotheses are as follows:

$H_0 : F(x) = G(x)$

$H_A : F(x) < G(x)$ when $x < x^*$ and

$G(x) < F(x)$ when $x > x^*$.

Notice that under the null hypothesis, the difference between G and F is zero everywhere, while under the alternative, it is positive for all x to the left of the crossing point and negative at all values to the right. At the crossing point the two distribution functions are equal. Therefore the function

$$\lambda(x) = \sup_{t \le x}[G(t) - F(t)] + \sup_{x \le t}[F(t) - G(t)] - |F(x) - G(x)|$$

is maximized at the crossing point, x^*, under H_A.

The sample counterpart of this function is

$$\lambda_N(x) = \sup_{t \le x}(G_m(t) - F_n(t)) + \sup_{x \le t}(F_n(t) - G_m(t)) - |F_n(x) - G_m(x)|.$$

Let $Z_{(1)} < Z_{(2)} < \cdots, Z_{(N)}$ denote the order statistics of the combined sample of X's and Y's. If the largest value of $\lambda_N(x)$ is attained at the jth order statistic, then $Z_{(j)}$ is the estimated crossing point. The proposed test statistic is

$$J_N = \sqrt{\frac{mn}{N}} \max_{0 \le j \le N} \lambda_N(Z_{(j)}). \quad (1)$$

The asymptotic distribution and selected critical values are presented in Chen et al. (2002). Notice that large values of J_N are unlikely under the null of equal distributions.

In some cases, it may be useful to supplement the test with tests for equality of population quantiles to the left of the estimated crossing point. These would provide us with the range of income levels over which program participation is relatively low. Close to the crossing point, the quantiles for the two distributions would be similar, the difference becoming larger for lower levels of income. We illustrate this procedure in the next section with results from simulations on selected distributions.

[1] Other similar tests have been proposed by Deshpande and Shanubhogue (1989) and Hawkins and Kochar (1991), but appear to have lower power in Monte Carlo exercises and also do not provide an estimate of the crossing point, which constitutes an upper bound on the incomes of those neglected by the program.

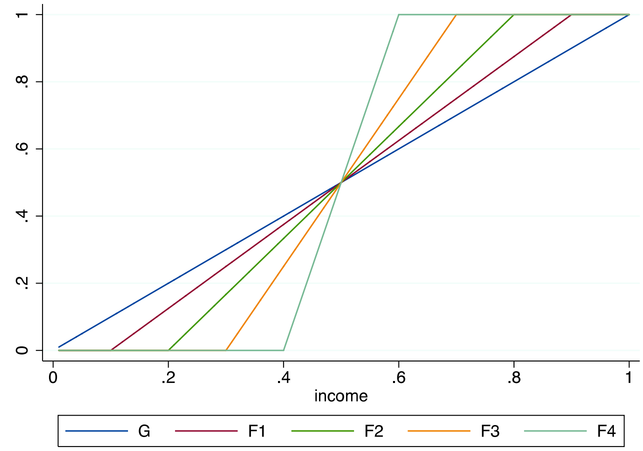

Fig. 1. $G(x) \sim U(0, 1)$, $F(x) \sim U(a, b)$.

3. Simulation results

We use two sets of parametric distributions to illustrate and evaluate the above approach. In each case, we start with a base distribution of non-participants, $G(x)$, and alternative distributions of participants $F(x)$, all of which cross $G(x)$ at the median income level. We apply the crossing point test to samples from the base distribution and each variant of the participant distribution. In each case, we also use sign tests to test hypotheses about differences in quantiles of the two distributions to the left of the crossing point.[2]

In our first case, illustrated in Fig. 1, $G(x)$ is uniformly distributed on (0, 1) and $F(x)$ has uniform distribution on (a, b). The distributions intersect at $\frac{a}{1+a-b}$. We choose alternative parameters a and b for which $a + b = 1$. This ensures that the distributions cross at median income.

Our second case is from the family of two-parameter exponential distributions.

$G(x) = 1 - e^{-\lambda_1(x-\theta_1)}, \quad x \ge \theta_2, \lambda_2 > 0$

$F(x) = 1 - e^{-\lambda_2(x-\theta_2)}, \quad x \ge \theta_1, \lambda_1 > 0.$

The two distributions cross at $x^* = \frac{\lambda_1\theta_1 - \lambda_2\theta_2}{\lambda_1 - \lambda_2}$ which is the median income if

$$\frac{\lambda_1\lambda_2(\theta_1 - \theta_2)}{\lambda_1 - \lambda_2} = \ln 2. \quad (2)$$

For $G(x)$, we set the scale parameter $\lambda_1 = 1$ and the location parameter $\theta_1 = 0$. For $F(x)$ we consider various departures in location and scale parameters subject to (2). The resulting distributions are shown in Fig. 2.

We are especially interested in the performance of the test when the null and alternative hypotheses are very similar to see whether it can still detect the crossing with reasonably high probability.

For each of the two cases described above, we use equally sized samples of two sizes: (i) $n = m = 50$ and (ii) $n = m = 100$, and perform 5000 iterations. For the crossing test we use the simulated 5% critical point of 1.529 given in Chen et al. (2002). We perform sign tests for quantiles of order $p = .1, .2, .3$ and.4 and use 5% asymptotic critical values based on the standard normal distribution. Power comparisons of these tests based on these simulations are shown in Tables 1 and 2.

Not surprisingly, the power of all tests increases as we move away from the base distribution, and the difference between

[2] Sign tests are probably the most commonly used nonparametric procedures for testing for population quantiles. See, for example, Gibbons and Chakraborti (1992).