Available online at www.sciencedirect.com
ScienceDirect
Economics Letters 100 (2008) 9–12
economics letters
www.elsevier.com/locate/econbase

Miss-targeted or miss-measured?

Martin Ravallion *,[1]

Development Research Group, The World Bank, 1818 H St., NW Washington, DC 20433, United States

Received 27 July 2006; received in revised form 3 October 2007; accepted 24 October 2007
Available online 1 November 2007

Abstract

A method is proposed for testing the robustness of the assessed targeting performance of an anti-poverty program to the fact that program administrators have a broader concept of "poverty" than the economist/evaluator. An application is given to China's main urban anti-poverty program.

Keywords: Cash transfers; Targeting; Measurement errors; Multidimensional welfare; China

JEL classification: I38; O15

1. Introduction

In assessing the targeting performance of anti-poverty programs, common practice is to include a question on program participation in a survey that also asks about incomes. Armed with such data, one then measures the proportion of participants who are poor and the program's coverage of the poor to quantify the errors of "exclusion" and "inclusion".[2] These calculations appear to have influenced numerous program assessments in practice.

This paper addresses a seemingly common, but routinely ignored, problem in this practice. The problem is that the concept of "poverty" underlying a program's objectives often appears to be broader than the way "income" is normally defined and measured from surveys, i.e., there are other welfare-relevant variables in deciding eligibility besides current income. However, while the program's administrators can typically list this broader set of variables, they are often rather vague about the precise weights attached to these other variables. The problem for the evaluator is that the program's apparent "miss-targeting" could simply reflect the fact that the survey-based measure of "income" is not a sufficient statistic for deciding who is "poor".

The idea of the method of addressing this problem proposed here is to calibrate a broader welfare metric to the observed program assignment and the qualitatively known program objectives, under the counterfactual of (stochastically) perfect targeting. Targeting performance is then measured based on this counterfactual and compared to conventional measures using the narrower income concept. An application is provided to China's main cash transfer program for urban poverty reduction.

2. A counterfactual model of targeting

There is more than one conceptually defensible way to measure "income." Program administrators may have good reason for putting higher weight on certain observables than is implicit in current incomes. A potentially important source of differences between survey-based incomes and the measures used to target a program is the *time period* over which income is measured. Current income can differ from long-term income. For example, a young well-educated family may have low current income but be on a rising trajectory with good future prospects and (hence) not deemed eligible for the program. Or a family may have a temporarily low income (due say to a spell of unemployment) but still not be deemed sufficiently poor in terms of their standard of living, as indicated by their consumer durables and housing, to warrant public action. Another source of error is in the weighting of household size and demographic composition in forming the metric

* Tel.: +1 202 473 6859; fax: +1 202 522 1153.
E-mail address: mravallion@worldbank.org.

[1] For helpful discussions and other forms of assistance, the author is grateful to Shaohua Chen, Emanuela Galasso, Michael Lokshin, Philip O'Keefe and Youjuan Wang. The findings, interpretations and conclusions of this paper are those of the authors, and should not be attributed to their employers, including the World Bank.

[2] Numerous examples can be found in the literature. In a meta-study, Coady et al. (2004) give such estimates of targeting performance for over 100 programs in developing countries.

0165-1765/$ - see front matter
doi:10.1016/j.econlet.2007.10.020

of economic welfare. Standard methods of setting equivalence scales are known to face serious identification problems (see, for example, Browning, 1992). For such reasons, we do not know in practice whether signs of miss-targeting in a program reflect the reality or stem from a flawed welfare metric used by the evaluator.

It can be assumed that there is an agreed set of variables that are deemed relevant to assessing "welfare" (or "need") for the specific program. Let these variables be (Y, X), where Y is observed current household income and X is a vector of other relevant variables, including household size and demographic composition, and variables relevant to past or future incomes (such as education or financial wealth). While there is agreement on the variables in (Y, X), there is considerable scope for disagreement on the weights to be used on those variables in defining a composite welfare indicator. Statements of the program's objectives are assumed to identify a list if attributes in qualitative terms, but to not be precise about the weights.

In assessing program performance in reaching the poor, a natural counterfactual is that based on the set of weights that best predict the observed program assignment. To see how this can be implemented empirically, consider a means-tested program that aims (more or less explicitly) to fill the poverty gaps, so as to bring everyone up to some minimum level of living. The counterfactual postulates that the program is perfectly targeted according to a latent money metric of welfare (Y^*) that depends on (Y, X) but there are still random latent factors relevant to Y^*. A reasonable functional form for Y^* might be:

$$\ln Y_i^* = \alpha \ln Y_i + \pi X_i + \varepsilon_i \tag{1}$$

where ε_i is an error term with zero mean and variance σ_ε^2. (Note also that X may reflect measurement error in Y.) A household is eligible for the program if (and only if) $Y_i^* < Z_i$, which is the local poverty line (depending on where household i lives). I assume that any household who is deemed eligible accepts the transfer.

This gives us an estimable model of eligibility under the counterfactual of perfect targeting based on Y^*. Assuming that ε_i is normally distributed, the probability of participating in the program is given by:

$$\Pr\left(Y_i^* < Z_i\right) = F\left[(\ln Z_i - \alpha \ln Y_i - \pi X_i)/\sigma_\varepsilon\right] \tag{2}$$

where F is the standard normal distribution. The parameters in Eq. (2) can be estimated as a probit regression. The predicted value (propensity score) is a monotonic increasing function of the expected value of the (proportionate) poverty gap, $E(\ln Z_i/Y_i^*)$. The predicted score can be interpreted as the welfare measure corresponding to the weights actually used by the program's administrators to assign eligibility. Note that the share of the population that is eligible (all i for which $Y_i^* < Z_i$) is given by the actual program participation rate, under the counterfactual.

If one finds little or no miss-targeting based on the propensity scores then one can question the robustness of any evidence of miss-targeting found by standard methods using survey-based incomes. On the other hand, if there is only a modest improvement in measured performance then the assessment is revealed to be robust.

Two remarks are in order about this approach. Firstly, with a sufficiently large set of X's, targeting performance under the

Table 1
Probit for program participation

	Coefficient	*t*-ratio
Log income per capita	0.2725	1.09
Squared log income per capita	−0.0668	−4.08
Log household size	0.2326	5.63
Log area of dwelling	−0.1813	−5.24
Year of dwelling construction	0.0009	1.10
Type of house. (Default: one story house)		
Cheap dwelling	0.0064	0.11
Apartment	−0.0474	−0.68
Single house	−0.2276	−1.54
Ownership of house. (Default: rented public dwelling)		
Rented private house	−0.1340	−1.92
Self designed and owned	−0.1465	−2.00
Inherited or bought a long time ago	−0.0612	−1.08
Owned apartment/house	−0.2832	−5.23
Owned designated low income dwelling	−0.0469	−0.71
Owned through employer subsidy	−0.1495	−4.54
Other	−0.0826	−0.94
Type of toilet. (Default: old style)		
Public	−0.1127	−1.83
Regular toilet	−0.0205	−0.60
Other	−0.2145	−2.61
Type of fuel. (Default: electricity)		
Piped gas	0.1061	1.2
Liquefied petroleum gas	0.0651	0.76
Coal	0.3273	3.47
Other	−0.3639	−1.86
Type of heating. (Default: no heating)		
Heating equipment	−0.0203	−0.41
Air conditioning.	−0.3379	−6.11
Electric heading	−0.1800	−2.49
Type of bath. (Default: Integrated bathroom)		
Shower or bathtub	−0.1574	−1.18
Other	−0.0273	−0.2
Additional house owned. (Default: none)		
One	−0.0444	−0.75
Two or more	0.2005	1.43
Sharing dwelling with other family	0.1438	2.02
Sharing house with another family	−0.0048	−0.05
No computer	0.3191	7.57
Male household head	−0.1039	−3.76
Age of head (Default: head age > 60)		
Age of head < 20	0.5902	1.56
20–30	−0.2952	−2.29
30–40	−0.0936	−1.53
40–50	0.0261	0.48
50–60	0.0499	1.04
Health (Default: head is healthy)		
Disabled	0.8504	16.41
Sick	0.3232	8.88
Years of schooling of head	−0.0149	−3.37
Assets (Default: Financial assets < 10,000 Yuan)		
Financial assets 10,000–30,000	−0.4156	−8.17
Financial assets 30,000–50,000	−0.4542	−4.46
Financial assets 50,000–100,000	−0.2933	−2.3
Income adequacy (Default income less than needed)		
Just right	−0.2911	−9.56
Surplus	−0.4041	−5.44
Income has improved		
No change	−0.2563	−6.82
Worse	−0.4071	−9.89
Wage ratio (income share from wages)	−0.5899	−16.12
Share of retired people in household	−1.2260	−11.07
Share of home workers	−0.2957	−2.56
Share of unemployed	0.2130	2.79

Table 1 (*continued*)

	Coefficient	*t*-ratio
Share of students	0.5245	6.14
Share of children	−0.1558	−1.71
Head's employer (dummy variables)	Yes	
Head's occupation (dummy variables)	Yes	
Sector of employment (dummy variables)	Yes	
City dummy variables	Yes	
Constant	0.3995	0.22
# of obs.	76443	
Pseudo R^2	0.4718	

Source: Chen et al. (2006), which includes complete results.

counterfactual will be excellent. In the limit, participation will be predicted perfectly. However, it appears that the set of X's that are deemed relevant to any program in practice is typically rather small. The point of this exercise is then to assess performance against that specific set of X's, as identified by program administrators. If targeting performance is found to be poor based on (Y, X) then one might explore the issue further by looking for any observable indicators of the miss-targeting and augmenting the probit model accordingly.

Secondly, this approach still leaves scope for questioning whether the weights used by program administrators are appropriate. All that is proposed here is to provide a counterfactual analysis of targeting in which the set of weights on the agreed welfare indicators is at least consistent with the program's actual assignment. Making those weights more explicit will also facilitate their assessment in practice.

3. Application to China's *Dibao* program

The "Minimum Livelihood Guarantee Scheme" — popularly known as *Dibao* (DB) — has been the Government of China's main response to the challenges of social protection in its rapidly changing economy. *Dibao* aims to provide a transfer to all registered urban households with "incomes" below a poverty line set at the municipal level, sufficient to bring them all up to that line.[3] In 2003 participation had reached 22 million people, representing 6% of urban residents (World Bank, 2007).

The program does not rely solely on self-reported incomes. Local authorities and neighborhood committees try to assure that recipients are "genuinely eligible," taking account of other factors including financial assets, consumer durables and housing conditions. There is also a community-appeals process, which includes the posting of applicants' names in a public place for two weeks. It appears to be quite unlikely that local authorities and community groups involved in assigning DB measure current income the same way households would probably report their income in a survey.

The following analysis uses the Urban Household Short Survey (UHSS) done by China's National Bureau of Statistics in 2003/04 for the 35 largest cities, giving a total sample of 76,000, varying from 450 (in Shenzhen) to 12,000 (in Beijing).[4] For these cities, the definitions of geographic areas in the UHSS coincide with those for the DB lines. The UHSS includes *Dibao* participation and income, although one must expect measurement errors in the income numbers.

From the survey one can also assemble a more or less complete list of the household characteristics (in addition to current income) that are identified by the program's administrators as being relevant to the program's welfare objectives, including financial wealth, education attainments, consumer durables and housing conditions. The X's also include $m-1$ dummy variables for the m municipalities, each of which can have its own poverty line.

The probit for program participation is in Table 1.Reported income is found to be a (highly) significant predictor of DB participation. However, one also finds a number of significant X's, consistent with the variables identified in qualitative terms by program administrators. Controlling for income per capita, DB participation is more likely for larger households, living in smaller dwellings, who do not own their dwelling, have an "old style" toilet, are still using coal for cooking, have no heating, no computer, have a female head of household, have a disabled or sick head of household, or a head with little schooling or who works in services or social security/welfare, or a head who is retired, works at home, has been laid off or is unemployed. At given reported income (and other X's), DB households have lower financial wealth, are more likely to feel that their income is "less than they need to make ends meet," are more likely to think that their income has improved, have a lower share of wages in income, have more unemployed or students in the household but fewer retired people. It is clear that the program is putting heavier weight on certain characteristics than is implicit in household income per person as reported in the UHSS.

Table 2 gives the "classic" two-by-two breakdown of targeting performance based on surveyed incomes. One finds that

Table 2
Leakage and coverage of the *Dibao* program based on observed incomes

% of population	Net income below DB line		Total
	Yes	No	
Receiving DB	2.22	1.69	3.91
Not receiving DB	5.48	90.60	96.09
Total	7.71	92.29	100.00

Note: n=76,443.

Table 3
Counterfactual leakage and coverage under perfect targeting based on the propensity scores

% of population	Eligible based on propensity score		Propensity score missing	Total
	Yes	No		
Receiving DB	1.97	1.92	0.01	3.91
Not receiving DB	1.93	93.79	0.37	96.09
Total	3.91	95.71	0.38	100.00

Notes: n=76,443. Column total include cases in which the data for estimating probability of participation are missing, as indicated.

[3] "Registered" urban residents in China are those with an official registration for urban residence. There are also non-registered urban residents, who are often recent migrants from rural areas.

[4] The UHSS is described more fully in Chen et al. (2006).

slightly less than 8% of the population had a net income (observed income *minus* DB receipts) below the relevant DB line. However, sizeable leakage to ineligible households is indicated by Table 2. About 40% of DB recipients are ineligible according to these data (0.43 = 1.69/3.91). Almost three-quarters of those who are eligible are not being covered by the program (0.71 = 5.48/7.71).

To the extent that these calculations reflect measurement errors in incomes *or* a broader concept of "income" that is motivating the program's targeting, it can be argued that the program is doing a better job of reaching the poor than Table 2 suggests. Table 3 gives the results analogous to Table 2 for the counterfactual model of eligibility, using the predicted propensity scores from Table 1 to rank households, instead of using the survey-based incomes alone. The overall program participation rate is (of course) taken to be the same under the counterfactual. So the target group is identified by the 3.91% highest propensity scores.

One finds that the coverage rate is higher than that based on the survey incomes, with 50% of eligible households receiving DB (as compared to 28% based on Table 2). The extent of leakage to the non-poor is slightly higher, however, with half of those receiving DB being ineligible based on the predicted probabilities of participation (as compared to 43% based on survey incomes, as in Table 2). The eligible population (the highest 4% of predicted values) receives over 60% of DB payments.

On the basis of these results, it cannot reasonably be argued that the extent of miss-targeting in Table 2 is mainly due to discrepancies between survey incomes and the latent welfare metric used by the program in practice. One finds substantial leakage to those who should not be eligible, and incomplete coverage of those who should be, even when income and other relevant household characteristics are weighted optimally from the point of view of predicting program participation.

References

Browning, Martin, 1992. Children and household economic behavior. Journal of Economic Literature 30, 1434–1475.

Chen, Shaohua, Ravallion, Martin, Wang, Youjuan, 2006. Dibao: A Guaranteed Minimum Income in China's Cities? Policy Research Working Paper, WPS 3805, Washington DC, World Bank. (Available at: http://econ.worldbank.org/docsearch.)

Coady, David, Grosh, Margaret, Hoddinott, John, 2004. Targeting Transfers in Developing Countries: Review of Lessons and Experience. Washington DC, World Bank.

World Bank, 2007. Urban Di Bao in China: Building on Success. World Bank, Human Development Unity, South Asia, World Bank.