



DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

OPTIMAL INCOME SUPPORT TARGETING

Stefan De Wachter and Sebastian Galiani

Number 41

October 2000

Manor Road Building, Oxford OX1 3UQ

Optimal Income Support Targeting*

Stefan De Wachter[†] Sebastian Galiani[‡]

September 8, 2000

Abstract

This paper considers the practical problem of distributing a fixed budget for poverty alleviation to a population whose poverty status is not directly observable. Some information on the relationship between poverty status and a number of observable and verifiable characteristics is assumed to be available in the form of a household survey. The solution we propose differs from other academic work in that it explicitly accounts for administrative constraints on the shape of the transfer function and is computationally more straightforward. It improves on the techniques that are commonly used in practice by taking both the concavity of the social welfare function and the entire conditional distribution of poverty status into account, and by endogenously determining the optimal transfer levels. Although the superiority of our allocation rule over other techniques is tautological, we explore the magnitude of the improvement in an artificial dataset. Finally, we provide an intuitive discussion of the defects of currently operational methods.

JEL-classification: H2, C4

Keywords: targeting, discriminant analysis, poverty

1 Introduction

In several developing countries, social policy directed towards poverty relief consists (partly) of transfer programmes, in which poor households receive direct monetary transfers. The main difficulty faced by governments implementing such income support schemes (and in

*This version has benefited from comments by F. Alvarez, F. Bourguignon, E. Dal Bo, P. Gertler, D. Heymann, N. Lustig, R. Spady and seminar participants at Di Tella University. The first author thanks the European Commission for financial support through a Marie Curie fellowship.

[†]Department of Economics, University of Oxford, Oxford OX1 3UL, UK. stefan.dewachter@wolfson.ox.ac.uk

[‡]Universidad Torcuato Di Tella, Miñones 2159, Capital Federal, Argentina (1428). sgaliani@utdt.edu

general any policy specifically directed towards the poor), is the unobservability of income¹, which is supposed to be the basis for eligibility for support. Often however, some sample information (e.g. a household survey) is available about the relationship between a number of easily observable and verifiable personal characteristics on the one hand and income or consumption on the other. This information can then be used to design a mechanism for deciding which transfer levels should apply and which households will be eligible².

How exactly such information should be translated into an assignment rule, however, is not clear. In practice a range of statistical classification techniques are employed to associate prespecified transfer levels with particular target groups, although the appropriateness of these techniques in the context of income support targeting has not been established. Examining this statement more closely, this paper is based on the observation that several features of the targeting problem are not present in “standard” classification problems and that therefore methods designed to deal with the latter are not necessarily appropriate solutions to the former. We make these ideas more precise by formulating the targeting problem as a social welfare maximization problem subject to two constraints. The first of these is the commonly encountered constraint of a policymaker who treats the transfer problem separately from all other social policy issues. Hence, the total budget available for direct income support is taken as given. Note that this is not a limitation of our method specifically³. Secondly, in order to achieve some degree of administrative simplicity⁴, we constrain the transfer function to be discrete-valued⁵, i.e. the population is divided in several segments within which each household receives the same amount⁶. As the final element of the problem statement, the uncertainty regarding household income is captured by the assumption that the social planner knows the distribution of income conditional on the observed characteristics. In most practical settings, this distribution is not known, but it can be estimated from a survey.

¹Ideally one should have some indicator of “material wealth” (lifetime income) that reflects a household’s true access to resources. We do not address this problem in particular and focus on targeting constraints in program design that are present irrespective of whether an ideal indicator is available. Therefore “income” or “consumption” will be used throughout to mean an indicator of material wealth.

²For example, in Colombia, areas of poverty were identified as part of the national development plan. Targets of food subsidies were then narrowed down to households with children under five years old or pregnant or lactating women. This reduced the number of possible beneficiaries and thus lowered administrative and fiscal costs. Little leakage or fraudulent coupon use was apparent (World Bank (1986) [14]).

³The alternative would be to imbed the income support issues in a general equilibrium policy model. Note that in our partial equilibrium approach comparison with other policies is possible since the shadowcost of the budget is easily calculable.

⁴By this we refer to “transparency” of the support program in general, and in particular to the complexity of (or amount of resources needed for) physically organizing the transfers, checking for fraud, counteracting corruption by local administrators etc. Importantly, in current practice transfers are also discrete.

⁵Even if administrative simplicity is not an issue, this is in principle not a restrictive assumption since the number of values the transfer function can be allowed to take is arbitrary.

⁶Otherwise stated, the problem is reduced to determining which transfer levels should apply and which households are eligible for each level.

The solution to our problem is an operational targeting rule which is easiest to interpret in the case where only one transfer level is allowed. For a given transfer level, the social planner will distribute money to those groups⁷ that are, conditional on their observable characteristics, *on average* poorest in *social welfare* terms. He will continue doing so until the budget is depleted. It turns out that the optimal target group is a simple function of the transfer level and hence this level can be chosen so as to maximize social welfare (in practice by a simple numerical maximization procedure).

Two features of this solution deserve special attention. Firstly, since the increase in social welfare for all households with a particular configuration of personal characteristics is what determines the recipient group of a particular (in this example the only) transfer level, the targeting rule takes both the shape of the social welfare function (and in particular the possibly high concern for the very poor) and the entire conditional distribution of income into account. This allows the social planner to fine-tune targeting to his sensitivity to "missing the very poor" and "wasting money on the rich" and leads us to dub our method "weighted classification (WT)". Secondly, the fact that transfer levels and the corresponding target groups are determined *simultaneously* stresses the importance of the interrelationship between ability of precise targeting on the one hand and choice of transfer levels on the other. The incorporation of both these features in an easily implementable transfer rule is the main contribution of this paper. Additionally, we indicate how one can incorporate "indirect" poverty alleviation measures⁸ and explain how the transfer functions employed by most currently operational income support programs can be seen to be a special case of the rule we propose in this paper.

Several income support schemes are and have been running over the past decade in a large number of Latin-American countries. The type of targeting rules used in programs such as Subsidio Unico Familiar (Chile) or Progresa (Mexico) and some others are called Proxy Means Tests (PMT) in the economics literature - see amongst others Grosh (1994) [9]⁹. Proxy Means Tests proceed by training some classifier on the available sample (or using other information) and then using it to allocate prespecified amounts of money to households in the population¹⁰. In Section 2 (Remark 1 and Example 1), we explain in more detail in which sense PMT based programs regard income support targeting as a classification problem. Since our framework encompasses classification techniques as special cases, it offers

⁷All households in a *group* have the same observable characteristics.

⁸Such as special concern for families with children at primary school age.

⁹Besley and Kanbur (1993) [3] provide an interesting brief discussion of this method and call it "targeting using indicators".

¹⁰A commonly used procedure is as follows. First, the population is divided in a prespecified number of groups according to income (where each group is assigned a transfer level (possibly zero)). Then the conditional mean of income given the personal characteristics is estimated from a sample using regression techniques. Finally the estimated regression function is used to calculate a prediction ("proxy") of the income of each candidate recipient, and this prediction is used to decide which transfer level applies for each applicant. (Before the actual transfer, the administrative body may verify that the covariates were truthfully reported). See Remark 1 below for a more formal explanation.

an intuitive way of discussing the shortcomings of PMTs (see Example 2 and the discussion thereafter). In particular, it will be shown that the two features of our solution mentioned above are not accounted for in PMTs.

The academic literature has given a fair amount of attention to the targeting problem. A large number of studies address various aspects of targeting in a general sense. One of the earlier occurrences of the idea is in Akerlof (1978) [1], who discusses the option of using household characteristics to identify the poor in the context of optimal income taxation. Besley and Kanbur (1988) [4] and Kanbur (1987) [11] are two examples of more empirically oriented studies¹¹.

Most related to our work is a paper by Glewwe (1992) [8], who uses a formulation of the targeting problem very similar to ours. The optimal transfer function (which is restricted to be a polynomial in the household characteristics, truncated at zero (as opposed to a piecewise constant function in our case)) is obtained by minimizing a poverty index in the sample (where poverty status is observable) subject to a budgetconstraint. The minimization problem is solved numerically; hence estimation of the conditional density of poverty status and determination of the transfer function are not treated separately. Such a procedure is somewhat less transparent and considerably more computationally demanding than ours. Additionally, the resulting transfer function will only (approximately) satisfy the budgetconstraint¹² in the population if the sample dataset is representative¹³. In spite of these shortcomings, Glewwe (1992) [8] convincingly demonstrates the power of efficient targeting (whereby transfers and corresponding beneficiaries are jointly determined) in an application to a dataset from Côte d’Ivoire.

Atkinson (1995) [2] highlights several theoretical arguments that are implicit in the use of targeting programs and shows that the assessment of the relative efficiency of (target versus non-target) programs depends on the formulation of social objectives and the constraints¹⁴ under which different types of programs can operate. Here, we do not attempt to discuss these issues, but instead focus on how to optimally design a targeting program (without suggesting that targeting programs are the best choice in all circumstances).

This paper is organized as follows. First we state and solve a simple version of the targeting problem and explain precisely why PMT based programs can be improved upon. Then, in Section 3, we compare (a limited version of) our method with an example of a currently used technique in an artificial dataset and provide an intuitive discussion of the particular features of the results. Section 4 then generalizes and proves the result from Section 2. An illustration on another artificial dataset is presented in Section 5 and Section

¹¹This literature review does not intend to be exhaustive.

¹²If the sample size is N , the population $\kappa \cdot N$ and the total budget for poverty alleviation B , then the budget constraint for the minimization problem in Glewwe’s setup would be B/κ .

¹³In contrast, our method is not affected by nonrandom sampling in terms of household characteristics as long as the sampling of income *conditional* on these covariates is random. In other words, for our method only an estimate of the density of income conditional on household characteristics is required.

¹⁴E.g. stigma, incentive effects, administrative infrastructure to collect taxes, etc.

6 concludes. The appendix contains a description of the algorithm implementing our method and details of the (parametric) conditional density estimation method.

2 The optimal transfer function: a simple case

As a formalization of the discussion in the introduction, we now present a mathematical description of the problem faced by a social planner having to distribute a fixed amount of money to households with unobservable poverty status. For expositional purposes, this section will restrict itself to the simplest case, where the transfer can only take two values (zero and a value to be chosen optimally); this will also make the comparison with “standard” classification procedures (as used in programs based on Proxy Means Tests) more straightforward. The generalization to a transfer function that can take any (fixed) number of values can be found in Section 4. The solution we obtain is an easily interpretable, fully operational and easy-to-implement targeting rule.

2.1 Setup and notation

We assume that there is an index of “richness” R (cfr. permanent consumption) that measures a household’s material well-being. It is not obvious how to actually construct such a measure, but in most applications the standard practice seems to be to use current consumption (which is of course only an approximation of true material well-being). We use the **convention** of referring to R as “income”. R is what generates utility and is (for the time being) also the only relevant factor in terms of social welfare. A household survey is available which contains R together with a number of personal characteristics X (which we also call “covariates”), which are easily observable and verifiable in the rest of the population (which will apply for income support)¹⁵.

It is useful to remark here that these covariates should preferably be difficult to modify by the household concerned. Some variables may be good predictors of poverty but if they are easily modifiable the target mechanism may induce undesirable behavior on the population. Hence, there is a trade-off between classification accuracy and incentives (one of the few studies addressing this issue is Kanbur et al. (1995) [12]). This tradeoff can in principle be embedded in our setup and hence evaluated¹⁶.

The problem is to divide a budget B over a population of size “measure 1”, distributed as $F(R, X)$. We assume a utilitarian social welfare function where the social welfare contribution of a household with “richness” R is $v(R)$ with $v(\cdot)$ a continuous weakly concave function.

¹⁵Note that the household survey is anonymous and, however large it is, cannot be used for redistribution purposes.

¹⁶For example, a program might use the proportion of employed persons per household as a covariate. This may trigger a decrease in household labor supply and hence social welfare. If an estimate of this effect can be obtained, one can compare the induced welfare reduction to the gain from using the variable as a covariate in the targeting program.

In Section 4, we will consider the more general case where $v = v(R, X)$. As mentioned in the introduction, we restrict the transfer function to take only a small number of values. In this section, we specialize this restriction further by assuming that the transfer will be either a single fixed amount of money or nothing. In other words, the transfer function $t(X)$ will be of the form

$$t(X) = t.\Delta(X)$$

where t is a (as yet unknown) number and $\Delta(X)$ is a (as yet unknown) decision function that specifies whether a person with vector of characteristics X will receive the benefit or not. This additional assumption is purely for expositional purposes and will be relaxed in Section 4 to transfer functions that can take more than two values.

All density functions are denoted by $f(\cdot)$ and distribution functions by $F(\cdot)$ where the arguments indicate which distribution is meant. $1\{\cdot\}$ is the indicator function, defined by $1\{\cdot\} = 1$ if the statement between curly brackets is true and $1\{\cdot\} = 0$ otherwise.

Remark 1 *In this simple setting, a Proxy Means Test based program would additionally specify a poverty line R^{pov} and declare everybody with income below R^{pov} eligible for a **pre-specified** transfer \bar{t} . Often the available budget will satisfy $B = \bar{t}.F(R^{pov})$ so that in principle everybody below the poverty line could receive a transfer. Since true income is not observable, the transfer is allocated to an individual with characteristics X if the income proxy $\hat{R}(X)$ (e.g. an estimated regression function) falls below the poverty line¹⁷. This is equivalent to classifying the population into poor ($R < R^{pov}$) and rich ($R > R^{pov}$) households: the income proxy $\hat{R}(X)$ and R^{pov} form the classification device.*

2.2 Problem statement and interpretation

The social planner is faced with the problem of choosing the binary transfer function $t(X)$ (which can take values 0 and t) that maximizes social welfare (the sum of all individual social welfare contributions $v(R)$) subject to the constraint that the total amount transferred should not exceed the budget B . Formally, this is written as follows

$$\begin{aligned} \max_{t(X)} \int_X \int_R [v(R + t(X)) - v(R)] dF(R, X) \\ s.t. \begin{cases} t(X) = t.\Delta(X) \\ \int_x t.\Delta(x) dF(x) = B \end{cases} \end{aligned} \tag{1}$$

¹⁷A practical example along these lines is the index of quality of life based on linear regression used for targeting purposes by the Argentinian government. The classification method described here is just an illustrative example: other well-known statistical classification techniques are also commonly used: Progres, for instance, uses discriminant analysis.

where $F(R, X)$ is given

where $\Delta(x)$ is the binary function defined above. Determining the optimal $t(X)$ is reduced to determining the optimal transfer level t and the target group¹⁸. The relationship with PMT based programs is further discussed in the following

Example 1 (“Standard” binary classification problem) For **given** t , problem (1) is a generalization of a definition of optimality for a “standard” binary classification setup. To make this statement more precise, consider the groups “rich” and “poor”, defined by whether income falls above resp. below the poverty line R^{pov} . The transfer level t is set to a value \bar{t} that satisfies $R^{pov} = F_R^{-1}(B/\bar{t})$ (hence R^{pov} is the income level for which there is enough money available to give everybody below it a transfer \bar{t}). To make the analogy exact, assume that $f(R) = 0$ for $R \in [R^{pov} - \bar{t}, R^{pov}]$ ¹⁹. Now specialize the SWF to be $v(R) = R \cdot 1\{R < R^{pov}\} + R^{pov} \cdot 1\{R \geq R^{pov}\}$ ²⁰. In this setup, the change in social welfare will be equal to $(\bar{t}$ times) the number of correctly classified poor people. Further, choosing a classifier that maximizes social welfare (as specified here) given the budget constraint can be shown²¹ to be equivalent to choosing a classifier that minimizes the number of misclassified rich people given the number of misclassified poor people. The latter is exactly the criterion used by Welsh (1939) [13] to derive the well-known Bayes classifier (see also Fix and Hodges (1958) [6] for an excellent discussion of issues concerning implementation).

2.3 Solution to problem (1)

Since the social planner can only use X as the basis for allocating the transfer, he will, for any transfer level t , first select as recipients the group of households that (according to $F(R|X)$) induces the highest total social welfare contribution per dollar distributed. The next group of recipients will be households that have characteristics vector X with the second highest impact, and so on until the budget is depleted. This is formalized in our main result:

Theorem 1 *The decision function $\Delta(x)$ that is the solution to problem (1) is of the form*

$$\Delta(x) = 1 \left\{ \int_0^\infty [v(R+t) - v(R)] dF(R|X=x) \geq \lambda \cdot t \right\} \quad (2)$$

¹⁸As mentioned before, programs that use PMT set t exogenously. In order to compare our method with such procedures, we explain below that the solution to the ‘entire’ problem 1 is ‘recursive’ in the sense that the optimal $\Delta(x)$ for given t depends on t in a simple way so that one can first solve the problem for fixed t and then iterate, if necessary, to determine the optimal t .

¹⁹That is, there is nobody with income in the interval $[R^{pov} - \bar{t}, R^{pov}]$; this artificial assumption is only made to exactly (instead of approximately) construct the setting of the standard classification problem.

²⁰That is, $v(\cdot)$ is increasing linearly (45° line) up to a level R^{pov} and stays constant at the value R^{pov} for $R > R^{pov}$. This is (minus) the poverty index FGT1 from Foster, Greer and Thorbecke (1984) [7].

²¹Example 2 below formally proves this statement in an indirect way. A direct proof (simply showing the equivalence of the objectives) is of course also possible.

for any transfer level t . λ is chosen to satisfy the budget constraint. The optimal transfer level can be determined by numerical optimization (if not exogenously fixed).

Note that it is not necessarily the *number* of successfully targeted “poor” (as defined by some poverty line) households that is being maximized: the impact (as measured by $v(\cdot)$) of the *degree* of poverty is what matters. If, for instance, the social welfare function is very concave (and hence the concern for the extremely poor high) it may be that a considerable amount of money is deliberately wasted on relatively well-off individuals because they happen to have the same characteristics as the very poor.

Example 2 (*The standard classification problem continued*) Refer back to Example 1 for the setup. Applying theorem 1 to this case, the decision function is

$$1\left\{\int_0^{R^{pov}} [R + t - R] dF(R|X) \geq \lambda.t\right\}$$

or

$$1\left\{\int_0^{R^{pov}} f(R|X) dR \geq \lambda\right\}$$

Applying Bayes’ rule one obtains

$$1\left\{\frac{1}{f(X)} \int_0^{R^{pov}} f(X|R).f(R) dR \geq \lambda\right\}$$

or

$$1\left\{\int_0^{R^{pov}} f(X|R).f(R) dR \geq \lambda \left(\int_0^{R^{pov}} f(X|R).f(R) dR + \int_{R^{pov}}^{\infty} f(X|R).f(R) dR \right)\right\}$$

Using $f(x|R < R^{pov}) = \frac{1}{F(R^{pov})} \int_0^{R^{pov}} f(X|R).f(R) dR$ (analogously for the other term) and rearranging results in

$$1\left\{\frac{f(x|R < R^{pov})}{f(x|R > R^{pov})} \geq c\right\}$$

where $c = \frac{\lambda}{1-\lambda} \frac{1-F(R^{pov})}{F(R^{pov})}$ is a constant. This is the ‘Bayes classifier’ (see Welch (1939) or e.g. Hand (1981)). $f(x|R < R^{pov})$ is the density of the characteristics of the ‘group’ of poor people, and $f(x|R > R^{pov})$ that of the rich. This is hardly surprising, since the Bayes classifier satisfies a definition of optimality which (as stated in Example 1) is in this setup equivalent to that in Problem 1.

Examples 1 and 2 clearly illustrate in which sense PMT based programs may be suboptimal and how our method constitutes a direct improvement. Firstly, PMT based programs do not always consider the entire distribution of income and the covariates (which is needed to construct the optimal classifier). Secondly, they do not determine the choice of transfer level simultaneously with the target group. Additionally, in cases where the preferences of the social planner are not as those in Example 1, they fail to take the concavity of the social welfare function into account.

Theorem 1 does not specify the optimal transfer level. For a given population (or group of applicants) it can be found by numerically maximizing social welfare function (1) (replacing the integral over X by summing over the population) w.r.t. t using targeting rule (2) to determine the recipients. This requires knowledge of the conditional distribution $F(R|X)$ which can be estimated from a household survey (Appendix 2, Section 7.2 describes the method used in this paper, but this choice is by no means compelling). Of course, if this estimate is very bad, the resulting transfer rule may well be far from optimal. This may for certain datasets be a difficult modelling issue which is best treated separately. It should be noted that the marginal distribution of household characteristics is of no importance. This implies that sampling anomalies (like clustering, stratification, selection bias) in the household survey do not pose a problem as long as there is no bias in the sampling of income conditional on the characteristics.

3 Comparison with Proxy Means Tests: an artificial dataset

In some sense, experimentally demonstrating the superiority of our targeting rule over other allocation rules (that satisfy the same constraints) is somewhat unnecessary, since it is the optimal one *by definition* (see expression 1) if the distribution $F(R|X)$ is known. Nevertheless, some examples of its performance may give an indication of *the extent* to which it may outperform other targeting rules and allows to illustrate the features of the solution in an intuitive way.

The artificial datasets we constructed contain 3 observables, dubbed “region”, “number of children”, and “education”. Region is Bernoulli distributed (2 regions: rural and city). Education is modelled as a conditionally (on region) lognormal (continuous) variable and number of children is conditionally (on region and education) binomial. Income is specified to be conditionally lognormal where each covariate is allowed to affect both mean and dispersion of the income distribution. The parameters are then calibrated to generate some ‘reasonable looking’ data. Finally, the continuous education variable is coded into 5 categories, resulting in datasets with only discrete covariates. It turns out that discrimination is quite difficult in this artificial environment.

The experiments we report here are only a selection out of a series (with various combi-

nations of B , t , and social welfare function; t is always considered to be exogenously specified to allow a comparison with the Proxy Means Test) and were performed as follows. First, a "household survey" of size 3000 is generated (containing income). This sample is then used to obtain conditional density estimates $\hat{f}(R|X)$. Finally, the resulting estimates are plugged into the allocation rule from Theorem 1 to perform the targeting exercise in a "population" of 30000. Note that in the estimation phase we do not use any "insider information" about the data generating process. Instead, we perform a model fitting exercise in the usual way. All calculations were done using the *Ox* programming language (v2.20) of Doornik (1999) [5].

The parameters for the first set of experiments were chosen as follows: first, we fix the (single) transfer level and set the poverty line at the 10th (\$529), 15th (\$626) and 25th (\$680) percentile of the income distribution, respectively. The total budget is in each case chosen so that if income were observable, everybody below the poverty line would receive a transfer²². For each case we consider 3 different 'social welfare functions': (the negative of) the poverty indices FGT1, FGT2 and FGT4 from Foster, Greer and Thorbecke (1984) [7]; these are defined as

$$FGT(n) = \int_0^{R^{pov}} \left| \frac{R - R^{pov}}{R^{pov}} \right|^n dF(R)$$

so in terms of our notation, $v_{FGT(n)}(R) = -1\{R < R^{pov}\} \left| \frac{R - R^{pov}}{R^{pov}} \right|^n$; as n increases, the SWF becomes more concave: concern for the very poor becomes stronger. Table 1 reports the realized/expected²³ increase in social welfare (as a percentage of total potential increase in social welfare (i.e. if income were observable)) achieved by our weighted targeting rule ("WT") on the one hand and a targeting rule based on regression (called "OLS" here) on the other²⁴. We specified the conditional mean in both cases in the same way.

| poverty line | <i>10th perc</i> | | <i>15th perc</i> | | <i>25th perc</i> | |
|---------------------|------------------|------------|------------------|------------|------------------|------------|
| SWF | WT | OLS | WT | OLS | WT | OLS |
| FGT(1) | 33 / 30 | 24 | 46 / 42 | 39 | 60 / 60 | 57 |
| FGT(2) | 45 / 41 | 17 | 52 / 47 | 32 | 66 / 64 | 55 |
| FGT(4) | 75 / 75 | 6 | 76 / 72 | 15 | 82 / 77 | 38 |

Table 1: Increase in SW (percentage of total possible increase)

²²Note again that this is done to allow comparison with a PMT and implies that our choice of the transfer level may not be the optimal one.

²³As calculated with the estimated density. The total realized and expected increase are divided by the total possible increase (this can be done because we also generated income data for the population using our DGP; of course these were not used for targeting).

²⁴The latter works as follows: use the sample to estimate a regression function $\hat{E}(R|X)$; then calculate the conditional average income in the population. Assign the transfer to the $\frac{B}{t}$ households that have the lowest estimated average income.

First analyzing the experiments with the poverty line at the 10th percentile of the income distribution, the first column of Table 1 shows that OLS only achieves 24% of the total potential increase in social welfare when measured by the linear SWF FGT1, decreasing to 6% for FGT4. WT, on the other hand, performs considerably better in all three cases, and is more than 10 times as efficient when the concern for the very poor is highest.

Another way of looking at these results would be to compare the estimated income distributions of benefit recipients for different setups. However, since this is not possible for OLS, Figure 1 instead graphs the realized distributions²⁵. Note that OLS does not involve any SWF, and therefore the distribution is the same for each SWF. Of course the realized distributions are in the case of WT only an approximation of the estimated ones, which determine the allocation rule.

The distribution for WT with FGT1 in Panel B is somewhat similar to that obtained through OLS (Panel A). This is the case because under FGT1 the problem solved by WT is a “standard” binary classification problem and, as shown in Section 2, WT then produces (approximately) the same solution as the Bayes classifier. OLS is also a binary classification method and the only difference with WT is that the latter is the optimal classifier. The extent to which OLS is a “non-optimal” classification method is in other words the only factor to cause a discrepancy in the target populations.

Introducing a second factor, extra concern for the very poor, alters the outcome dramatically. For FGT2, WT generates an income distribution for benefit receivers that is far more disperse than that implied by OLS. Since under this SWF the concern for the very poor is much higher, WT chooses to target groups of households that have (conditional on their observable characteristics) a thicker left tail in their income distribution. In our artificial population, it appears to be the case that several extremely poor households have the same observable features than some relatively well-off ones. Even though (some of) these configurations of characteristics (i.e. values of the vector X) may well be associated with a higher conditional mean income (and hence will not be targeted by OLS), they also have a higher dispersion, and hence fatter tails. Since the left tail receives a much higher weight (and the right tail is immaterial) under FGT2 (and FGT4), these households are preferred by WT.

Increasing the poverty line to the 15th (\$626) and 25th (\$680) income percentile (and the budget accordingly) leads to similar conclusions, although to a somewhat lesser extent. Especially for FGT1, under which targeting any person below the poverty line is weighted equally, the inferiority of OLS is less striking: classification becomes easier for higher poverty lines and even a suboptimal classification method appears to perform relatively well.

A final thing to note in Table 1 is the consistent similarity of the realized and expected performance of WT, indicating a good quality of the conditional density estimates²⁶.

²⁵The distributions for FGT2 and FGT4 under WT are identical in this case and hence only one is shown. This happens because the covariates only take discrete values, introducing some “discontinuity”. This is certainly not the case in general. In order to plot identical axes, all observations in the far right tail (above \$1250) in Panel B and C were removed.

²⁶This is just indicative; note that our “population” is not a “population” in the statistical sense; the

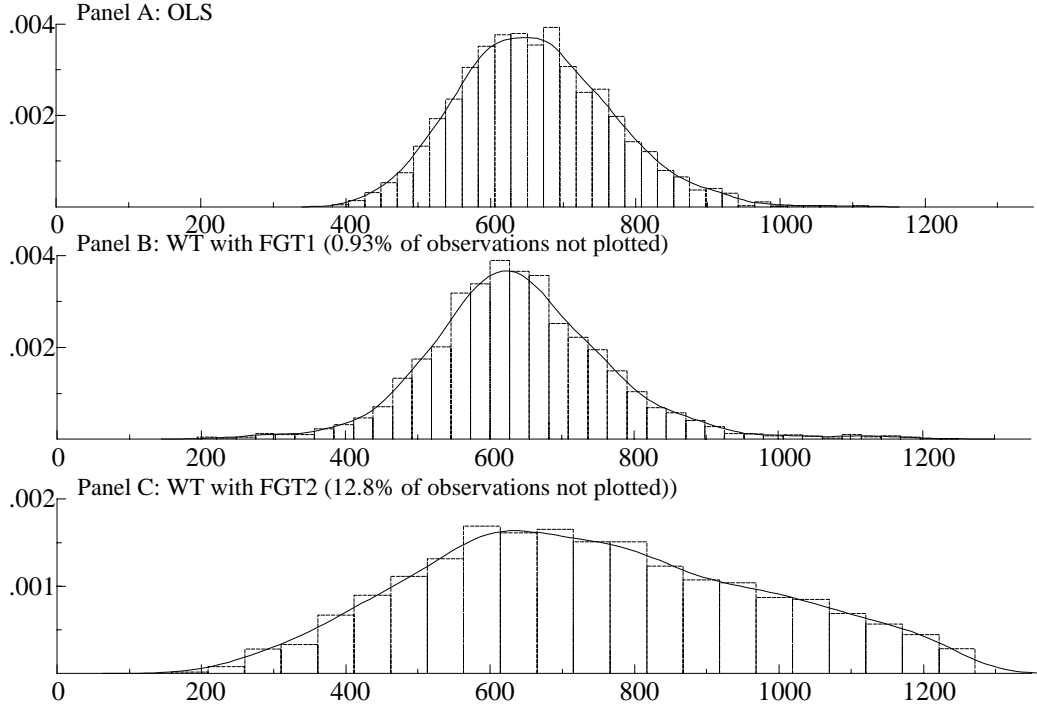


Figure 1: *Histogram and kernel density estimates of the income distribution of benefit recipients ($R^{pov} = 592$; $t = 50$)*

Keeping all other parameters (including the budgets) the same as in the first experiment, we now increase the transfer level by 80%²⁷ (to \$90). The corresponding results are reported in Table 2.

| poverty line | 10th perc | | 15th perc | | 25th perc | |
|---------------|-----------|-----|-----------|-----|-----------|-----|
| | WT | OLS | WT | OLS | WT | OLS |
| SWF | | | | | | |
| FGT(1) | 23 / 23 | 14 | 35 / 32 | 25 | 50 / 48 | 45 |
| FGT(2) | 27 / 31 | 8 | 33 / 34 | 16 | 42 / 42 | 34 |
| FGT(4) | 42 / 59 | 2 | 61 / 65 | 6 | 67 / 63 | 18 |

Table 2: *Increase in SW (percentage of total possible increase)*

The overall performance of both methods seems to have declined. This is normal because the total potential gain has sharply increased: more money is given to fewer (and therefore realized performance is therefore not, strictly speaking, a "fixed" (in the classical inference sense) quantity.

²⁷This means that there is not enough money available to distribute to everybody below the poverty line.

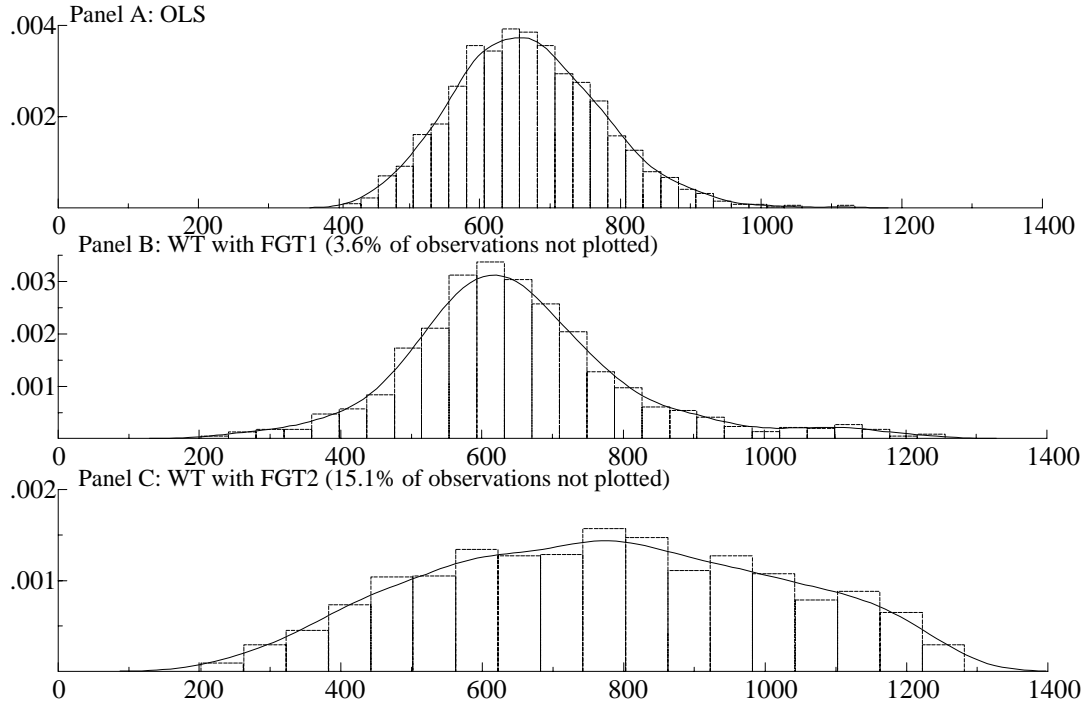


Figure 2: *Histogram and kernel density estimates of the income distribution of benefit recipients ($R^{pov} = 592$; $t = 90$)*

ideally poorer) people. The same comments as for the previous experiment are in order. Especially notice the extremely poor performance of OLS in situations with low poverty lines and a high concern for the poor: WT performs up to 20 times better! The corresponding income distributions for receivers of the benefit (case of poverty line at 10th income percentile; FGT2 and FGT4 are again identical in this dataset) are presented in Figure 2. These explain the very poor performance of OLS for FGT4: since the benefit level has increased, reaching just a few of the very poorest households has a large impact on social welfare. With OLS almost nobody below $592 - 90$ receives a transfer²⁸.

One may argue that the feature of the DGP that is driving our extreme results, conditional heteroskedasticity, is not very likely to be present in most actual situations, at least not as strongly as in this artificial environment. Nevertheless, WT is in such cases still a valuable tool in that it provides an upper bound on what can be achieved through direct transfers.

²⁸It is especially for individuals below 500 (and the lower the better of course) that the increase in SWF is greatest: if income is say 550, only half of the transfer generates an increase in social welfare (see the shape of $v(\cdot)$).

Moreover, the present experiments have assumed the transfer level to be exogenously fixed to enable comparison with Proxy Means Test-type of methods. The reader will recall that the most important virtue of our method is that the transfer level(s) and corresponding target groups can be determined simultaneously. This allows, for instance, transfer levels to be “automatically” increased as accurately targeting the poorest appears feasible; again the extent of this tendency will depend on how high the concern for the very poor is set to be.

After deriving the optimal transfer rule in the general setting in Section 4, we present an application featuring endogenous determination of transfer levels in Section 5.

4 Generalizations

A number of features commonly encountered in practice have not yet been incorporated in our setup. This section will therefore generalize our main result to account for this. Firstly, it may in most applications be too restrictive to impose only one transfer level to be used. Therefore our generalization will allow for any finite number K of transfer levels.

Secondly, in recent programmes such as Progresa (Mexico) special attention has been given to families with special needs. A typical example is children at primary school age. In a sense, this concern can be seen as a short-cut to a dynamic poverty reduction strategy, because education is typically a strong determinant of future income. We can easily modify our allocation rule to take this intertemporal feature into account. Indeed, the example cited above can be rephrased by considering the social welfare contribution of households with equal income to be lower the more children at primary school age there are in the household. More generally, the social welfare contribution $v(R)$ can be generalized into $v(R, X)$ where the dependency can be on just a few characteristics (children,...). The exact functional specification is of course ad-hoc, as is the current practice of implementing these features.

These two aspects lead to the following generalization of Theorem 1:

Theorem 2 (*Optimal allocation rule in the general case*) *Let $T = \{0, t_1, t_2, \dots, t_K\}$ be the set of possible transfer levels. Define the ‘eligibility index’ $I(t_k, x)$ as*

$$I(t_k, x) = \int_0^\infty [v(R + t_k, x) - v(R, x)] dF(R|X = x) - \lambda \cdot t_k$$

Then the optimal allocation rule is of the form

$$t(x) = \arg \max_{t_k \in T} I(t_k, x)$$

The scalar λ is chosen so as to satisfy the budget constraint.

Proof. We assume that all variables are continuously distributed. Denoting by Ω the space of personal characteristics x , we look for subspaces $\Omega_k \subset \Omega$ corresponding to transfer

levels t_k , $k = 1, \dots, K$ that define who will receive which benefit. Otherwise stated, a household with characteristics x receives transfer t_k if $x \in \Omega_k$. We now fix $t_k \forall k$ and look for an expression defining Ω_k .

Using this notation, the problem is reformulated as

$$\begin{aligned} \max_{t_k, \Omega_k} \quad & \sum_{k=1}^K \int_{\Omega_k} \int_R v(R + t_k, X) - v(R, X) dF(R, X) \\ \text{s.t.} \quad & \sum_{k=1}^K \int_{\Omega_k} t_k \cdot dF(X) \leq B \end{aligned}$$

This is equivalent to maximizing

$$\sum_{k=1}^K \int_{\Omega_k} \int_R v(R + t_k, X) - v(R, X) dF(R, X) + \lambda \cdot (B - \sum_{k=1}^K \int_{\Omega_k} t_k \cdot dF(X))$$

Rewriting:

$$\sum_{k=1}^K \int_{\Omega_k} \left\{ \int_R [v(R + t_k, X) - v(R, X) - \lambda t_k] dF(R|X) \right\} dF(X) + \lambda \cdot B$$

This will be maximized for given λ and t_k , $k = 1, \dots, K$ by choosing to assign x to Ω_k if, at this x , the expression between $\{\}$ is maximal (over all k). In other words, if the 'expected' increase in social welfare obtained by giving the transfer t_k to a person with characteristics x exceeds some threshold level $\lambda \cdot t_k$ by more than it exceeds any other threshold level $\lambda \cdot t_j$:

$$\Omega_k = \left\{ x \in \Omega : \int_R [v(R + t_k, X) - v(R, X) - \lambda t_k] dF(R|X) \geq \int_R [v(R + t_j, X) - v(R, X) - \lambda t_j] dF(R|X) \forall j \neq k \right\}$$

This defines the form of the decision function. For given transfer levels t_k , one can calculate λ (which is the only thing left to specify the decision function completely) iteratively to satisfy the budgetconstraint. ■

The allocation rule says that the transfer given to a household with characteristics X should be t_k if the “expected **excess** increase in social welfare” is highest when this transfer level is used for the household concerned. This is a natural extension of Theorem 1.

Figure 3 illustrates the logic behind Theorem 2 for a situation where there are only three (equally large) groups A , B and C , with covariates X_A, X_B and X_C respectively and two transfer levels. Then define the average increase in social welfare by giving a transfer t to group A as $g_A(t) \equiv \int_0^\infty [v(R + t, X_A) - v(R, X_A)] dF(R|X = X_A)$ and similarly for the other groups. The graph now shows the optimal value for λ and the two transfer levels (so it is assumed). Then it is easy to see that group A will not receive any transfer and groups B and C will both receive t_1 . Although budget-feasible, it would clearly not be optimal to not give a transfer to group B and use it instead to finance giving a transfer $t_2 (= 2 \cdot t_1)$ to group C . Further note that because of the concavity of $g(\cdot)$ the targeted groups will be unique for given transfer levels.

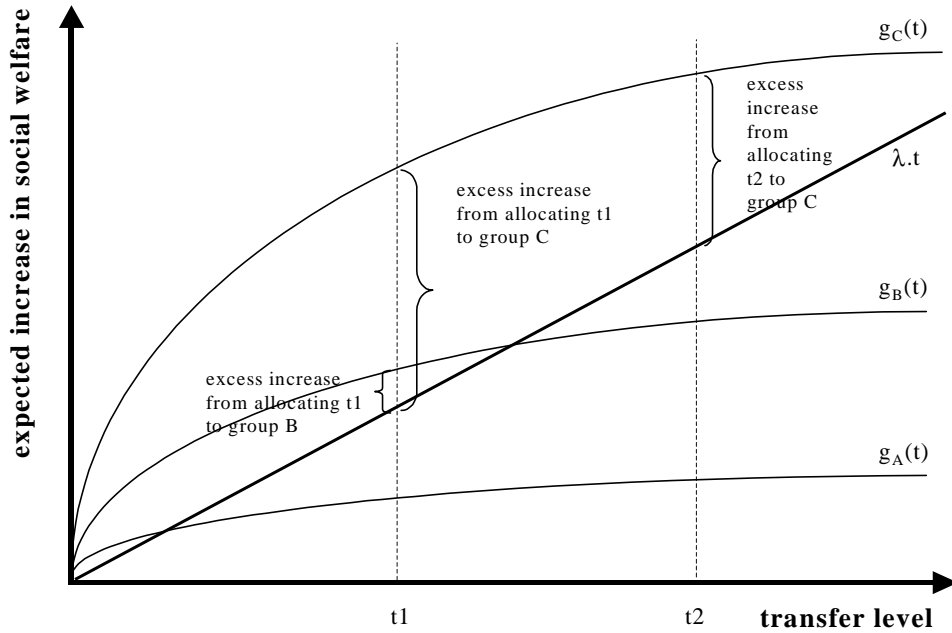


Figure 3: *An illustration of the general transfer rule*

5 An application

This section discusses some results from applying the general transfer rule in another artificial setup. The main aim is to provide an intuitive discussion of the features of the general solution and to examine the restrictiveness of the assumption of a discrete-valued transfer function.

The datasets we use are identical to those in Section 3, but without coding the variable “education” into discrete categories. This has the advantage that the population consists of more different groups of people. The “household survey” has again 3000 observations (used to re-estimate the conditional density) but the “population”-size²⁹ is reduced to 10000 for reasons of computational speed³⁰. The SWF is FGT2, the poverty line is at \$626 (roughly the 15th percentile of the income distribution) and the budget \$75000³¹.

The results of the experiments, in which we recalculate the optimal transfer levels (see Appendix 1 for technical details) increasing the number of levels from 1 to 5, are summarized in Table 3. Each row represents a different number of transfer levels (i.e. a different experiment). The first column shows the number of households that do not receive any

²⁹ As before, the population consists of drawings out of the same DGP as the sample.

³⁰ All computations reported in this section took less than one day on a Celeron 300Mhz PC.

³¹ The actual numerical values have of course no meaning in themselves, and are only presented to increase readability. They are comparable to those used for the experiments in Section 3.

transfer. In columns 2 to 6, each cell contains the optimal transfer levels and the size of the corresponding recipient group (between brackets). The 7th column presents the increase in social welfare (as calculated using the estimated densities, not the “realized” SW^{32}).

| (#nonrec) | t_1 (#rec) | t_2 (#rec) | t_3 (#rec) | t_4 (#rec) | t_5 (#rec) | ΔSW |
|-----------|--------------------|-------------------|-------------------|------------------|-----------------|-------------|
| (8889) | 67.5 (1111) | N/A | N/A | N/A | N/A | 11.716 |
| (8685) | 42.5 (1036) | 111 (279) | N/A | N/A | N/A | 12.326 |
| (8562) | 30 (832) | 67 (463) | 133 (143) | N/A | N/A | 12.528 |
| (8483) | 23 (659) | 48.5 (505) | 81.5 (246) | 143 (107) | N/A | 12.600 |
| (8430) | 18.5 (560) | 39.5 (466) | 63.5 (323) | 102 (166) | 159 (55) | 12.642 |

Table 3: transfer levels, number of recipients and increase in social welfare

Interestingly, these results suggest that there are “diminishing returns” in adding flexibility to the transfer function. The increase in payoff (in SW terms) from allowing an extra level clearly decreases as the number of levels increases. This finding is visualized in Figure 4, which also includes an estimate of the “asymptotic increase”, obtained by calculating the increase in SW for a “nearly continuous transfer function” (398 transfer levels, ranging from 1 to 200 in steps of $1/2^{33}$).

In order to interpret the magnitude of the increase in social welfare, it is useful to know that the (realized) change in SW achieved by using the “OLS” targeting method from Section 3 in this experimental setup (with 67.5 as the only transfer level) is 8.7.

A comparison of the realized income distributions of the different recipient groups can illustrate the mechanics behind the solution³⁴. Figure 5 presents QQ-plots (up to the 75th percentile) of the income distributions of different recipient groups for the cases with 3, 4 and 5 transfer levels³⁵. Panel A shows the quantiles of the income distribution group receiving the lowest level against those of the group receiving nothing, for the case of 3 transfer levels. Apart from the very lowest quantiles, the income distribution of the group receiving \$30 is far more “shifted to the left” than that of non-receivers. At first sight, it may seem strange that relative to the total groupsize (and hence also in absolute terms), there are more non-receivers with an income below \$400 than households receiving the lowest transfer level. Two

³²The experiments in Section 3 showed that these were very similar. In a real application, the realized increase is of course not observable.

³³As opposed to the results in Table 3, these levels are just prespecified (as in Section) and not endogenously determined. One could think of this limit as the “most continuous transfer function” if the smallest currency unit in our artificial society is $1/2$. In that case, the levels are optimally chosen.

³⁴Note that these distributions would not be observable in a real application. We follow this procedure to keep the analogy with a real application, but adding a “look behind the scenes”. The same remark as in Section 3 applies concerning the role of the realized income distributions in the interpretation of the optimal targeting rule.

³⁵Note that we compare (pre-transfer) income *distributions* conditional upon receiving a particular transfer. Even if one distribution stochastically dominates another, this does not say anything about the *absolute* difference in number of households below a particular income level. The reader can refer to the groupsizes reported in Table 3 to get an idea of absolute quantities.

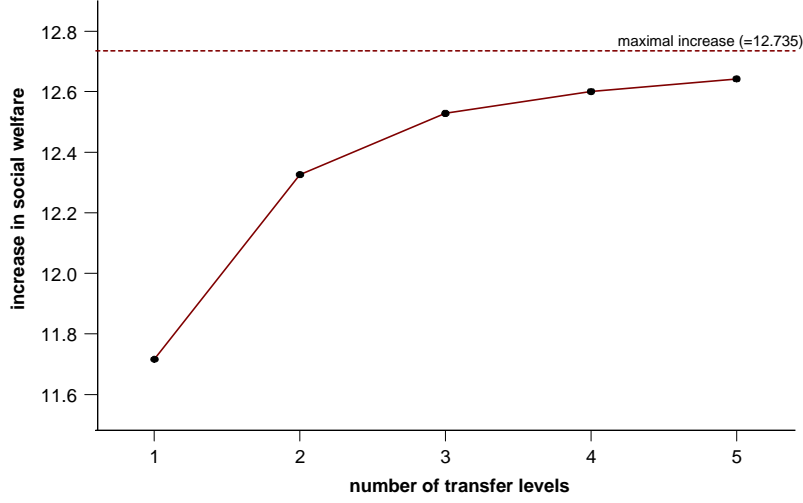


Figure 4: *The effect of increasing the number of transfer levels*

explanations can account for this feature: first of all, it may be that these households are difficult to target (cfr. the discussion in Section 3); secondly, the very poorest households that can be targeted relatively well receive a higher transfer. This is apparent from Panel B³⁶, which plots the group receiving \$30 against recipients of higher transfer levels. Especially the group receiving \$133 has a very thick left tail compared to the \$30 group.

Panels C and D are the counterparts of Panel B for the cases with 4 and 5 transfer levels, respectively³⁷. In Panel C, higher transfer levels roughly imply fatter left tails of the income distribution.

This is also true for Panel D, except for the two highest groups. The fact that the income distributions of these groups do not “behave correctly” can be attributed to two factors. The first is the sampling error due to the fact that our “population” (size 10000) is not infinite³⁸ and hence distributions calculated on the basis of it are not exact (assuming that the initial problem was to find the optimal transfers for the population (in the statistical sense) described by the “true” $F(R, X)$). The second is the error introduced by estimating $F(R|X)$ from the household survey.

In a practical application, it is of course not meaningful to think of the population as a “sample”. The impact of the possible error in estimating the population (in the “physical” sense) distribution can be judged by constructing confidence intervals on the increase in

³⁶The division into two panels is done to enhance visual clarity.

³⁷We do not display the QQ-plot of nonreceivers versus the lowest level group for these cases. It is qualitatively similar to that of the case with 3 levels.

³⁸Note that the group receiving the highest transfer in Panel D consists of only 55 observations.

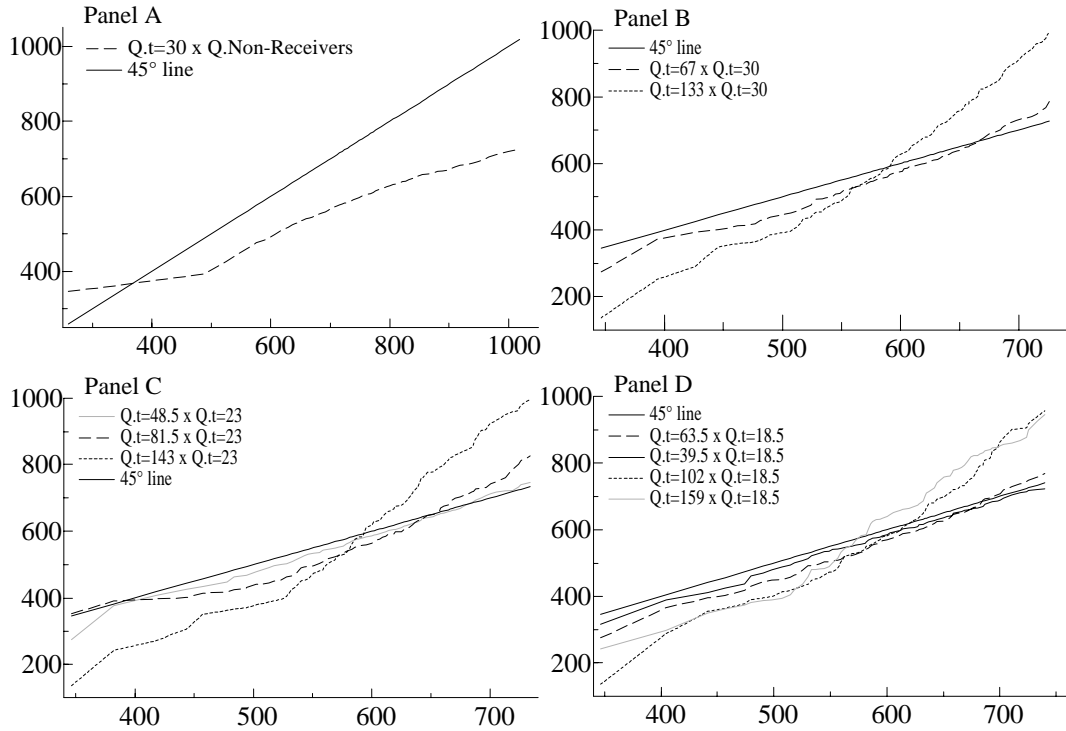


Figure 5: Comparing the realized income distributions of the recipient groups for the case with 3 (Panels A and B), 4 (Panel C) and 5 (Panel D) levels for the case with 3 (Panels A and B), 4 (Panel C) and 5 (Panel D) levels

social welfare for various levels. This could serve as an additional tool to determine the optimal number of levels³⁹. Unfortunately, calculating such confidence intervals is not a trivial issue and is left as a topic for further research⁴⁰.

Now reconsidering Figure 4, one could ask the question what the optimal number of transfer levels should have been in our example. If increasing flexibility is costly, it might have been reasonable to argue for the use of 3 levels on the basis of this graph alone. This suggests that a possible procedure could be to base the choice of the number of transfer levels on graphs like Figure 4, recalculated for various plausible estimates of $F(R|X)$.

³⁹Note that this uncertainty is not part of our initial problem setup and can therefore not be consistently incorporated in the determination of the optimal levels. Since the estimation uncertainty is not known with certainty either, there seems at first sight to be little advantage in incorporating it in the problem setup.

In practice, one could compare transfer functions obtained with various reasonable estimates of $F(R|X)$.

⁴⁰If a parametric conditional density estimation method is used, one could in principle calculate asymptotic standard errors using the delta method. Serious practical problems may occur due to the possible nondifferentiability of the social welfare function and time-consuming computation of the derivatives (if they exist) for cases with many transfer levels.

6 Conclusion

In this paper, we have outlined an optimal operational method for distributing a fixed budget among a population whose poverty status is not directly observable. Our main contribution is the development of a technique that allows joint determination of transfer levels and their corresponding target groups. It is obtained as the solution to a social welfare maximization (or poverty index minimization) problem under the constraint that the transfer function should be piecewise constant. This ensures that implementation is likely to be feasible and transparent even when administrative resources are limited. Importantly, the assumption of discrete-valued transfer functions turned out to be not at all restrictive in an artificial dataset.

Since our method was shown to include currently used techniques based on Proxy Means Tests as a special case, a comparison with these techniques was straightforward and stressed their three defects. Firstly, Proxy Means Tests consider the income support targeting problem as a “standard” statistical classification problem without (in most cases) using the optimal solution to that problem (the Bayes classifier). Secondly, since most governments measure poverty by a concave poverty index, the relevant problem to be solved is a “weighted” classification problem instead of the “standard” one⁴¹. Finally, PMT based support programs determine transfer levels exogenously whereas they should depend on the accuracy with which the target groups can be reached.

Although our method constitutes an improvement in each of these three aspects, its main drawback is the somewhat greater computational effort needed for its implementation. We have illustrated, however, that our allocation rule can also be used with prespecified transfer levels. Even with endogenously determined levels, all calculations, although somewhat more time-consuming, are straightforward and transparent. They never become computationally prohibitive.

Several interesting questions remain unanswered, especially those of a more practical nature. The selection of covariates for targeting and its relationship with incentives and possible feedback effects on social welfare is the most important issue. It is possible to calculate within our framework the cost of excluding a variable for targeting. Case-specific studies are needed to evaluate whether this cost exceeds the social cost due to incentive effects when including the variable. Another important issue pertains the choice of the social welfare function. One option could be to elicit a policy maker’s preferences by computing transfer functions for various SWFs. Thirdly, the choice of the “correct” number of transfer levels would be made easier with a formula for evaluating the impact of estimation uncertainty.

Finally, this paper has only presented a theoretical framework. The important next step will hopefully be the evaluation of the performance of our method in an actual application.

⁴¹This is by far the least substantial of our contributions, because one could easily perform a PMT replacing R by $v(R)$ in the household survey (and changing the “poverty line” accordingly). It is important, however, to be aware of the relationship between targeting rule and choice of SWF or poverty index.

7 Appendices

7.1 Appendix 1: Computational details

Without going into all the details of our implementation⁴², this Appendix outlines the structure of the algorithm with which our experiments were performed. The question is to implement our transfer rule for a particular database of households, taking as given the conditional distribution of income. From Theorem 2, it is clear that determination of the recipient groups of a vector of transfer levels requires calculation of the parameter λ . First we explain how our algorithm computes this parameter for a given choice of transfer levels. Then we discuss how to search for the optimal transfer levels. Throughout, the main themes are the possible presence of discontinuities in the objective functions and the concern for computational speed.

7.1.1 Calculating λ for given transfer levels

For given transfer levels, changing the value of λ will alter the recipient groups of the respective transfer levels and hence the total amount spent. Since this amount has to equal the budget, it is easy to define a function $H(\lambda, B)$ which has a maximum at the “correct” value of λ . This function can then be maximized numerically; in practice we use a gridsearch algorithm to avoid local maxima problems and problems due to discontinuities. Note that for a population where no covariate is effectively continuously distributed (as in any practical situation), $H(., .)$ is not likely to be continuous at all λ ; it may happen that no value of λ exists for which the entire budget is depleted exactly⁴³⁴⁴.

7.1.2 Calculating the transfer levels

Since we are now able to calculate recipient groups for any configuration of transfer levels and hence the corresponding increase in social welfare, finding the optimal levels becomes an unconstrained maximization problem. However, since for each K -vector of transfer levels, determining the recipient groups involves calculating K integrals for each household in the database, the computational burden could easily become prohibitive for a desktop PC. Secondly, the objective function may contain discontinuities, rendering standard gradient algorithms less reliable.

⁴²All computercode is available upon request from SDW.

⁴³In that case it is important to choose a value of λ at “the right side” of the budget constraint without specifying H in a way that a bad starting value would jeopardize convergence.

⁴⁴Since calculation of λ is needed for each new set of values for the transfer levels (which in turn will be chosen iteratively; calculation of λ is therefore a maximization problem “within” a maximization problem), it is important to ensure that the routine is both fast and reliable. In our implementation, this is achieved by dynamically updating the initial grid.

Our solution is to construct an algorithm that performs a (multidimensional) gridsearch on a lattice. That is, the transfer levels are restricted to take values on a lattice (e.g. the integers). For each household i in the population, the quantity $\int_0^\infty [v(R+t, x_i) - v(R, x_i)] dF(R|X = x_i)$ can then be calculated for each t on the lattice. The resulting matrix is then used as an input to the gridsearch algorithm. The search ends when the grid is as fine as the lattice. For smooth SWFs and smooth conditional densities of income, the objective function can be expected to be relatively smooth (in spite of small discontinuities). Of course, for very badly behaved cases, the lattice may be made finer.

Starting values are relatively easy to obtain. First, the exercise is done for one transfer level. This takes only a few seconds and hence a very wide starting grid can be chosen. This solution can then be used to construct the starting grid for the case with two levels, and so on. Note that the lattice does not have to be recalculated when the number of transfer levels is changed.

7.2 Appendix 2: A flexible parametric method for estimating conditional densities

This appendix describes the method for constructing the conditional density estimates used in the paper. Note that this is just an example of how this can be done in practice; the allocation rule we propose is the optimal one taking the (true) conditional density as given and any estimation method / parametrization may be used. Of course, if the estimates are very 'bad', it is not clear what the properties of the resulting allocation rule will be.

We parametrize the conditional density $f(R|X)$ as a (uniform) mixture of $J+1$ normal densities:

$$f(R|X) = \frac{1}{J+1} \sum_{j=0}^J \frac{1}{\sqrt{2\pi s(j, X)}} \exp\left(-\frac{(R - m(j, X))^2}{2s(j, X)}\right)$$

where the *mean* $m(j, X)$ of *subdensity* j is parametrized as

$$m(j, X) = b_0(X) + b_1(X) \cdot \frac{j}{J+1}$$

and the *variance* $s(j, X)$ of *subdensity* j as

$$s(j, X) = \exp \left[\theta_0(X) + \theta_1(X) \frac{j}{J+1} + \theta_2(X) \left(\frac{j}{J+1} \right)^2 + \theta_3(X) \left(\frac{j}{J+1} \right)^3 \right]$$

This choice of $m(j, X)$ and $s(j, X)$ is of course not the only possible one: anything flexible enough suffices. The parametrization of the functions $b_0(X)$, $b_1(X)$, $\theta_0(X)$, $\theta_1(X)$, $\theta_2(X)$ and $\theta_3(X)$ is left as a 'modelling'-exercise. Not all functions need to be nonzero nor should each function contain the same variables.

For example, it is easy to see that if the data are jointly normally distributed, then only $b_0(X) = X'\beta$ and $\theta_0(X) = \sigma^2$ (where β and σ^2 are constants to be estimated) is the appropriate parametrization: f will then be the normal density. If the only deviation from this setup is conditional heteroskedasticity, then an appropriate model choice is to also include $\theta_i(X)$ $i = 1, 2, 3$ as a linear function of X with some unknown parameters. Typically, income distributions will be very skewed, in which case also $b_1(X)$ should be included, possibly with dependence on some or all covariates (in case of conditional skewness).

Although this parametrization may be judged to be ad-hoc, it performs well in the examples we tried. Another advantage is that estimation of the parameters can be done by simply maximizing the likelihood. The only aim of this exercise is smoothing and “filling in the gaps left by the dataset” and not “estimating fundamental parameters” that need to be used for further research; therefore we judge this method to be sufficiently sound for the given purposes.

As a final remark, note that if the regression function $E(R|X)$ is (thought to be) linear in X , then $b_0(X)$ and $b_1(X)$ should be chosen linear in X .

References

- [1] Akerlof, G. (1978): "The Economics of Targeting as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning", *American Economic Review*, vol. 68, pp. 8-19.
- [2] Atkinson, T. (1995): "On Targeting Social Security: Theory and Western Experience with Families" in *Public Spending and the Poor*, van de Walle, D. and Nead, K. (eds.), Johns Hopkins.
- [3] Besley, T. and Kanbur, R. (1993): "The Principles of Targeting", in *Including the Poor*, Lipton, M. and Van Der Gaag, J. (eds.), The World Bank.
- [4] Besley, T. and Kanbur, R. (1988): "Food Subsidies and Poverty Alleviation", *Economic Journal*, vol. 92, pp. 701-19.
- [5] Doornik, J.A. (1999), *Ox: An Object-Oriented Matrix Language* (3rd edition), London: Timberlake Consultants Press.
- [6] Fix, E. and J.L. Hodges (1951), *Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties*, reprinted in Silverman, B.W. and M.C. Jones (1989), E. Fix and J.L. Hodges (1951): An important contribution to nonparametric discriminant analysis and density estimation, *International Statistical Review*, 57, 3, pp. 233-247.
- [7] Foster, J., J. Greer and E. Thorbecke (1984), A class of decomposable poverty measures, *Econometrica*, 52, pp. 761-765.

- [8] Glewwe, P. (1992), Targeting Assistance to the Poor: Efficient Allocation of Transfers when Household Income Is Not Observed, *Journal-of-Development-Economics*, 38(2), pp 297-321.
- [9] Grosh, M (1994), Administering Targeted Social Programs in Latin America: From Platitudes to Practice, World Bank.
- [10] Hand, D.J. (1989), *Discrimination and classification*, Wiley (Chichester).
- [11] Kanbur, R. (1987): "Transfer, Targeting, and Poverty", *Economic Policy*, vol. 4, pp. 112-36.
- [12] Kanbur, R, M. Keen and M. Tuomala (1995), Labor Supply and Targeting in Poverty-Alleviation Programs in van de Walle, D. and Nead, K. (eds.): *Public Spending and the Poor: Theory and Evidence*, John Hopkins.
- [13] Welch, B.L. (1939), Note on discriminant functions, *Biometrika* 31, 218-220.
- [14] World Bank (1986), *Poverty and Hunger: Issues and options for food security in developing countries*, Washington D.C.