DEPARTAMENTO DE ECONOMÍA
DOCUMENTO DE TRABAJO

"Optimal Income Support Targeting"

Stefan De Wachter y Sebastián Galiani

D.T.: N° 39 2001

Vito Dumas 284, (B1644BID) Victoria, Pcia. de Buenos Aires, Argentina
Tel: 4725-7053 Fax: 4725-7010
Email: economia@udesa.edu.ar

Optimal Income Support Targeting[*]

Stefan De Wachter[†]
University of Oxford

Sebastian Galiani
Universidad de San Andres

April 10, 2005

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. The solution we propose 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. We provide an algorithm to calculate the optimal transfers for any population of benefit applicants. Finally, we explain how our method is a generalization of statistical classification techniques and thus provide an intuitive discussion of the defects of currently operational methods.

JEL-classification: H2, C4
Keywords: classification, poverty, program design, targeting

[*]This version has benefited significantly from detailed feedback by John Wilson (the editor) and two anonymous referees. We also acknowledge comments by F. Alvarez, M. Berliant, F. Bourguignon, G. Cruces, E. Dal Bo, P. Gertler, E. Schargrodsky, R. Spady and seminar participants at Di Tella University, UC Berkeley, the World Bank and the University of Sheffield. The first author thanks the European Commission for financial support through a Marie Curie fellowship.

[†]corresponding author: Stefan De Wachter, Department of Economics, Manor Road, Oxford OX1 3UQ, UK; stefan.dewachter@economics.ox.ac.uk

1 Introduction

The main determinants of the impact of antipoverty schemes are the magnitude of the aid flow and the accuracy with which it reaches those with the greatest need. Although the former dominates news headlines and popular debate, the call for an efficient and systematic approach to tackling the latter has become more pressing over time (see among others World Bank (1990) and World Bank (2001)). The first best solution to a redistribution problem requires the knowledge of households' material wealth[1]. However, the latter is not observable in most practical situations faced by governments in developing countries. Hence, the implementation of a "second best" solution is the key issue in the design of antipoverty programs in general, and of direct income support schemes, in which poor households receive direct monetary transfers, in particular.

In this paper we study the problem of how to optimally distribute a fixed budget for poverty alleviation to a population whose poverty status is not directly observable. Indirect information, e.g. in the form of 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. How exactly such information should be translated into an assignment rule, however, is not obvious. 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.

Several income support schemes have been running over the past decades in a large number of developing countries. We will refer to the type of targeting rules used in programs such as Subsidio Unico Familiar (Chile) or Progresa (Mexico) and some others by "Proxy Means Tests" (PMT) (as in some of the literature - see e.g. Grosh (1994)[2]). Proxy Means Tests typically consist of training a classifier on the available sample (or using other information) and then using it to allocate *prespecified* amounts of money to households in the population. A commonly used procedure is as follows. First, the administrative body sets a poverty treshold - a household will be regarded as poor if its income is below the treshold - and a transfer level. Second, the conditional mean of household income given a set of observable characteristics is estimated using a houshold survey in which income data are available. Finally, to decide which households are poor, the estimated regression function is used to calculate a prediction ("proxy") of the income ("means") of each candidate recipient, and this prediction decides whether the applicant is eligible for a transfer. Before the actual transfer, the administrative body may verify that the covariates were truthfully reported.

[1] In practice, the best indicators of "material wealth" (lifetime income) are income or consumption. Although there is a debate about the relative merits of various indicators, we do not address this problem in particular. Instead we 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 (see Chaudhuri and Ravallion (1994) for related comments).

[2] Besley and Kanbur (1993) provide a discussion of this method and call it "targeting using indicators".

Our work is motivated by 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 a budget constraint and a shape constraint on the transfer function. The latter is specified to be piecewise constant with a prespecified number of levels. The solution to this problem is the optimal transfer function.

The main contributions of this work can be summarized as follows. Firstly, our solution embeds currently operational targeting procedures (i.e. PMTs) as a special case. This allows us to describe precisely which unnecessary assumptions the latter implicitly impose on the problem and to explain how they can be improved upon.

Secondly, our targeting rule can be designed to exactly fit the administrative format of PMTs and is therefore directly applicable in practice. Moreover, the problem formulation is flexible enough to incorporate the usual program design features (equivalence scales, extra credits for special needs such as young children). Since practical implementation of our method entails no additional difficulties over those currently encountered, and since it always performs at least as well as PMTs in the metric defined by the social planner's objective, it has the potential to become the preferred tool in practical work.

Finally, and in our view most importantly, our solution provides the option of optimal endogenous determination of transfer levels. This feature is not present in current targeting practice. Nevertheless, it is intuitively clear that the "best" transfer levels depend on the accuracy with which target groups can be reached. It is therfore natural to choose levels and target groups simultaneously. The construction of a computer algorithm to this end is one of the central practical contributions of our work.

A large number of studies address various aspects of indicator targeting. One of the earlier occurrences of the idea is in Akerlof (1978), who discusses the option of using household characteristics to identify the poor in the context of optimal income taxation. Further developments can be found in Besley and Kanbur (1988), Besley and Kanbur (1993), Kanbur (1987b), Kanbur and Keen (1989), Kanbur, Keen, and Tuomala (1995) and Nichols and Zeckhauser (1982). Only a surprisingly small number of papers have considered practical implementation of the targeting problem, two notable exceptions being Ravallion and Chao (1989) and Glewwe (1992).

The rest of the paper is organized as follows. First we state and solve the targeting problem mathematically. Then, in Section 3, we explain precisely why PMT based programs can be improved upon and provide an intuitive discussion of these insights using an example with an artificial data. Section 4 presents two small applications using Peruvian and Argentinian household surveys and Section 5 concludes. The Appendix contains a description of the algorithm implementing our method, details of the conditional density estimation method and the datasets used.

2 The optimal transfer function

We now present a mathematical description of the problem faced by a social planner having to distribute a given budget to households with unobservable poverty status.

2.1 Set-up and problem statement

Assume that the index of material wealth W measures household 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. We use the convention of referring to W as "income". Each household is characterized by its income W and by a number of easily observable and verifiable characteristics ("covariates") X. Because these characteristics will form the basis for the transfer mechanism, they should be difficult to modify by households. 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 targeting accuracy and incentives[3].

The social planner's problem is to divide the budget B over a population with characteristics distributed according to $F(X)$. Instead of observing W directly, the social planner only knows the conditional distribution $F(W|X)$ of household wealth given the observables[4]. In practice, lack of knowledge of $F(W|X)$ is remedied by the availability of a household survey containing both W and X. This household survey can only be exploited to obtain an estimate of $F(W|X)$; it is anonymous and hence useless for direct redistribution purposes.

The social planner's preferences are represented by a utilitarian social welfare function (SWF) where the social welfare contribution of a household is denoted by $v(W, X)$ with $v(.,.)$ continuous and weakly concave in W. The presence of the covariates X in the social welfare contribution function may seem odd. However, 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. The idea behind including X in v is to take this intertemporal feature into account. Indeed, the example cited above can be rephrased by regarding the social welfare contribution of households with equal income to be lower the more children at primary school age there are in the household. The exact functional specification is of course ad-hoc, as is the current practice of implementing these features.

We restrict the transfer function to take a fixed number of values $t_1, ..., t_K$ for reasons of administrative feasibility and efficiency. This is reminiscent of an argument commonly used in the taxation literature: Hettich and Winer (1988, p. 706), for example, rationalize the typically small number of tax bands observed in practice by arguing that administration

[3] One of the few studies addressing this issue is Kanbur, Keen, and Tuomala (1995).

[4] In what follows, density functions are denoted by $f(.)$ and distribution functions by $F(.)$ where the arguments indicate which distribution is meant.

of personalized tax rates would be too costly. With this assumption, the social planner's problem of choosing the optimal transfer function $t(x)$ amounts to selecting K transfer levels and their respective non-overlapping recipient groups.

Problem 1 *Let Ω_k denote the recipient group of transfer level t_k; $k = 1, ..., K$. The social planner chooses the transfer rule $t(x)$ by solving*

$$\max_{(t_k,\Omega_k)\ k=1,...,K} \sum_{k=1}^{K} \int_{\Omega_k} \left\{ \int_W v(W + t_k, X) - v(W, X)\ dF(W|X) \right\}\ dF(X) \tag{1}$$

$$s.t.\ \sum_{k=1}^{K} \int_{\Omega_k} t_k\ dF(X) \leq B \tag{2}$$

2.2 Optimal transfer function

Although no closed-form solution to Problem 1 exists, the following result provides the motivation for a simple numerical algorithm to determine recipient groups associated with given transfer levels. Once this is possible, the choice of optimal transfer levels becomes an unconstrained numerical maximization problem that is relatively easy to implement.

Theorem 1 *(Optimal allocation rule for fixed transfer levels) Let $T = \{0, t_1, t_2, ..., t_K\}$ be the set of fixed possible transfer levels. Define the eligibility index $I(t_k, x)$ as*

$$I(t_k, x) = \int_0^\infty [v(W + t_k, x) - v(W, x)]\ dF(W|X = x) - \lambda.t_k \tag{3}$$

for some appropriately chosen scalar λ. Then the optimal allocation rule is given by

$$t(x) = \arg\max_{t_k \in T} I(t_k, x) \tag{4}$$

Proof. Assume that all variables are continuously distributed. Denoting the population by Ω, we look for an expression defining subspaces $\Omega_k \subset \Omega$ corresponding to fixed transfer levels t_k, $k = 1, ..., K$. These subspaces define who will receive which benefit: a household with characteristics x receives transfer t_k if $x \in \Omega_k$. The problem is

$$\begin{array}{c} \max_{\Omega_k} \sum_{k=1}^{K} \int_{\Omega_k} \int_W v(W + t_k, X) - v(W, X)\ dF(W, X) \\ s.t.\ \sum_{k=1}^{K} \int_{\Omega_k} t_k.dF(X) \leq B \end{array} \tag{5}$$

By Lemma 1 in the Appendix, there exists a number λ such that the above problem is equivalent to maximizing

$$\sum_{k=1}^{K} \int_{\Omega_k} \int_W v(W + t_k, X) - v(W, X)\ dF(W, X) + \lambda.(B - \sum_{k=1}^{K} \int_{\Omega_k} t_k.dF(X)) \tag{6}$$

Rewriting:

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

This will be maximized for given λ and t_k, $k = 1, ..., K$ 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 treshold level $\lambda.t_k$ by more than it exceeds any other treshold level $\lambda.t_j$:

$$\Omega_k = \left\{ \begin{array}{l} x \in \Omega : \int_W \left[v(W+t_k, X) - v(W,X) \ - \lambda t_k \right] \ dF(W|X) \geq \\ \int_W \left[v(W+t_j, X) - v(W,X) \ - \lambda t_j \right] \ dF(W|X) \ \forall j \neq k \end{array} \right\}$$

This defines the form of the decision function. ■

Theorem 1 does not specify the value of λ in closed form. However, since targeting rule (4) will produce different recipient groups and hence different total expenditure for each value of λ, one can choose λ so as to equate expenditure to the budget. Because λ is a scalar, this is a one-dimensional, and therefore computationally simple problem.

We now add a few ingredients that were omitted from Problem (1) for expositional clarity. In practice one usually takes the household rather than the individual as the unit to be targeted. Household utility will be measured based on per capita equivalent income or expenditure. This requires the following modifications to the above set-up. W denotes equivalent per capita household income or expenditure (total income W^{tot} adjusted by some fixed equivalence scale $\rho(X)$, i.e. $W = W^{tot}/\rho(X)$). $v(W, X)$ denotes total social welfare generated by a household with per capita income W and can (for instance) be constructed as its members' individual utility (evaluated at per capita income) times the number of members.

When the program chooses to administer only a single transfer level, this transfer is typically expressed in per capita terms. Often some adaptation is made to account for varying family sizes, for instance using some fixed multiplicative scale $\phi(X)$ depending on the number of members of the household or its composition[5]. The eligibility index in (3) now becomes

$$I(t_k, x) = \int_0^{\infty} \left[v(\frac{W.\rho(x) + t_k.\phi(x)}{\rho(x)}, x) - v(W, x) \right] \ dF(W|X = x) - \lambda.t_k.\phi(x) \qquad (7)$$

With this modification and a corresponding change in the budget constraint all results go through as before. Using this notation, t_k remains the "base" transfer level for group k.

[5] When using endogenously determined transfer levels and a sufficiently large number of levels, this modification is probably unnecessary as the method will automatically choose the optimal levels. When only one or two levels are used, the transfer scaler $\phi(.)$ can be useful.

2.3 Interpretation

For the simplest case of a single transfer level ($K = 1$), the optimal allocation rule (4) reduces to

$$t(x) = t\ 1\left\{\int_0^\infty \left[v(W+t,x) - v(W,x)\right]\ dF(W|X=x) \geq \lambda.t\right\} \tag{8}$$

where the indicator function $1\{.\}$ equals 1 if the statement between $\{\}$ is true and 0 otherwise. This says that the single transfer t should be allocated first to households with characteristics X which, according to $F(W|X)$, will benefit most, then to households with the second highest impact on social welfare, and so on until one reaches households for which the expected increase in social welfare does not exceed λt anymore.

For the general case where $K > 1$, expression (4) states 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. Figure 1 illustrates the idea for a situation where there are three equally large groups A, B and C, with covariates X_A, X_B and X_C respectively and two transfer levels. Define the average increase in social welfare by giving a transfer t to group A as $g_A(t) \equiv \int_0^\infty \left[v(W+t,X_A) - v(W,X_A)\right]\ dF(W|X=X_A)$ and similarly for the other groups. Suppose that the graph shows the optimal value for λ and the two transfer levels. Then one sees that group A will not receive any transfer and groups B and C will both receive t_1. Although budget-feasible, it would not be optimal to take away the transfer from group B and use it to finance a transfer t_2 $(= 2.t_1)$ to group C. Further note that because of the concavity of $g(.)$ the targeted groups will be unique for given transfer levels.

Two features of the solution deserve special attention. Firstly, since the expected increase in social welfare for all households with a particular configuration of personal characteristics is what determines the recipient group of a particular 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[6]) 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 in Problem 1 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.

Theorem 1 does not specify the optimal transfer levels: it only defines the optimal recipient groups for a given configuration of transfer levels. Hence, it applies in a program in which transfer levels are exogenously specified as well. In order to calculate optimal transfer levels for a given population of applicants, one can numerically maximize social

[6] Indeed, it is not necessarily the *number* of succesfully targeted poor (as defined by some poverty line) households that is being maximized: the impact (as measured by $v(.)$) of the *degree* of poverty is what matters.

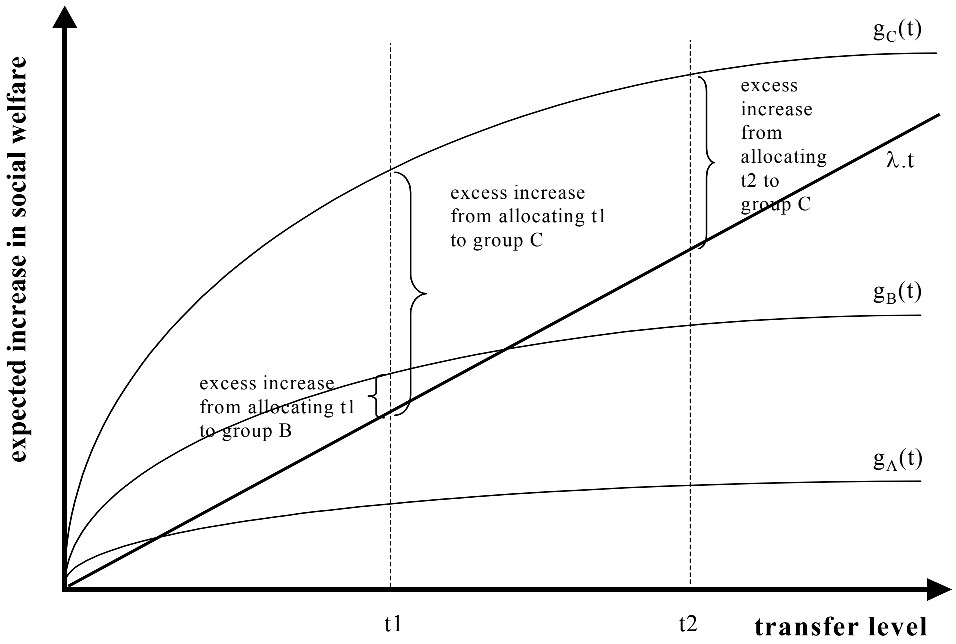

Figure 1: An illustration of the optimal transfer rule

welfare function (1) - replacing the integral over X by summing over the population - with respect to t_k, using targeting rule (4) to determine the recipients. Details can be found in the Appendix.

A required input in Theorem 1 is the conditional distribution $F(W|X)$. It can be estimated from a household survey (Section 6.4 in the Appendix describes the method used in this paper, but this choice is by no means compelling). If this estimate is inaccurate, the resulting transfer rule may well be far from optimal relative to a situation where the true distribution is known. Determination of the best method for estimation of $F(W|X)$ for a given dataset is a difficult problem which we do not attempt to solve in this paper[7].

2.4 Related literature

In this section, we discuss the most important studies that aim to provide an implementable method for targeting. Although these papers are notable contributions in their own respect, here we only focus on the differences between the solution they propose to income support targeting and ours.

Ravallion and Chao (1989) (see also Ravallion (1989), Datt and Ravallion (1993), Ravallion (1993) and Ravallion and Sen (1994) for applications and extensions) present an algorithm to implement the targeting problem formulated by Kanbur (1987a). Their set-up is similar in spirit to that in Problem 1, with the following important differences.

[7]Given the typical sample size of household surveys available in developing countries, fully nonparametric conditional density estimation methods are unlikely to be a viable possibility. Seminonparametric approaches with a flexible parametric specification of the regression for the mean, variance and skewness may be more promising.

Firstly, the social welfare function is fixed to be the Foster, Greer, and Thorbecke (1984) poverty index FGT2. This assumption is necessary because the algorithm is based on the first-order conditions of the maximization problem and the linearity of the derivative of FGT2 turns out to be an essential feature. The drawback is that the type of flexibility in the objective function that our method can incorporate cannot be accommodated. Secondly, the optimal transfer rule is not constrained to any class of functions at all: the algorithm calculates a separate transfer level for each value of the covariates X. This fully non-parametric approach allows for added flexibility in principle, but puts practical restrictions on the number and type (continuous / categorical) of indicators that can be used. If for instance three indicators are used that can take 10 different values each, one needs to calculate 1000 transfer levels. Apart from the computational problems this entails, it is not clear that such flexibility is an advantage from an administrative point of view. As Ravallion and Chao's (1989) choice of application suggests, their method is primarily designed for situations where the population consists of a small number of groups. As a final point to note, estimation of the conditional income distributions is treated separately as in our work[8].

Glewwe's (1992) approach, as opposed to Ravallion and Chao's work, does allow for flexibility in the objective function. The starting point of his paper is similar to our Problem 1 without the piecewise constancy assumption on $t(.)$. Details of implementation, however, differ substantially. Firstly, where our Theorem 1 holds the key to computational feasibility, Glewwe immediately proposes a "brute force" numerical approach to the original constrained maximization problem. It is far from clear whether this is a workable approach in general. In an application, Glewwe (1992) focusses on a specific choice of functional form, in which the optimal transfer rule is formulated to be a polynomial in the covariates, truncated at zero. It is questionable whether a continuous transfer function is a desirable feature from the point of view of transfer administration. In addition to the practical and logistic problems this entails, one might worry about the "political" difficulty of distributing different amounts to people that are quite similar (e.g. living in the same community) but not *exactly* identical. Secondly, instead of using the conditional distribution of income given the covariates as an input, Glewwe (1992) solves for the optimal transfer function by numerically maximizing the social planner's objective function over the sample. *De facto*, when casting the problem in terms of our Problem 1, such a procedure uses the empirical conditional distribution and skips the modelling stage. This confounds the transfer choice problem with the statistical task of smoothing - see Section 4 for related remarks. Thus, only in an ideal world where administration costs are irrelevant, computing power unbounded, and household surveys very large, the solution to (a slightly generalized version of) Glewwe's problem weakly dominates our proposal.

[8]The method used is a type of histogram estimator (linear interpolation of the empirical conditional distribution functions). This fully nonparametric procedure performs only a minimal amount of smoothing and is suitable for large datasets and a small number of groups. As with our targeting rule, other estimators for conditional densities can in principle be used. See Section 4 for related comments.

3 Comparison with Proxy Means Tests

The exposition in Section 2 has shown that (4) is the best possible piecewise constant targeting rule. In current income support targeting practice, a group of methods referred to as Proxy Means Tests (PMTs) are the "industry standard". Any proposal for a new methodology should therefore provide evidence of its comparative virtues over PMTs.

Subsection 3.1 develops a formal argument to this end. We first show how PMTs fit in our framework and hence how they can be interpreted as a special case of the optimal rule. In particular, we explain how PMTs mistakenly regard targeting implicitly as a statistical classification problem. By nesting PMTs within the class of optimal transfer rules, we can analyze in detail which implicit restrictions PMTs impose.

Subsection 3.2 then presents a numerical example. The aim is to provide an intuitive discussion of the features of the optimal solution compared to those of PMTs. We use artificial data in order to focus on the issues of interest in a controlled set-up.

3.1 PMTs as a special case of optimal targeting

We begin by describing the mechanics of a PMT based program in detail. In addition to the budget B such a program would specify a poverty line W^{pov} and declare everybody with income below W^{pov} eligible for the prespecified transfer $\bar{t}$. Since income is not observable, the transfer is allocated to an individual with characteristics X if the income proxy $\widehat{W}(X)$ (e.g. a regression function estimated from household survey data in which W is observable) falls below the poverty line. This classifies the population into poor ($W < W^{pov}$) and rich ($W > W^{pov}$) households: the income proxy $\widehat{W}(X)$ and W^{pov} form the classification device. An important difference between the general targeting problem formulated in Problem 1 and PMTs is that in the latter the single transfer level t is fixed by a government agency to some amount $\bar{t}$.

The classification method described above is only one example: other well-known statistical classification techniques are also commonly used in PMTs: Progresa, for instance, uses discriminant analysis. In that case the household survey (with income coded into a binary variable - poor and rich - according to W^{pov}) is used to estimate a discriminant function which is then used as the classifier in the population of benefit applicants.

The reader will have noticed from the above exposition that classification methods, and hence PMTs, only look at whether the individual is poor or rich and not at the *extent* of poverty. This suggests that the social welfare function "implied" in a PMT equally weighs the increase in welfare achieved by transferring one dollar to any poor individual (anyone below W^{pov}) and considers money given to any rich individual as wasted. This can be formalized by (a multiple of) the poverty index FGT1 from Foster, Greer, and Thorbecke (1984). In our notation this becomes

$$v(W) = W.1\{W < W^{pov}\} + W^{pov}.1\{W \geq W^{pov}\} \tag{9}$$

To make the analogy exact, we make the simplifying assumption that $f(W) = 0$ for $W \in [W^{pov} - \bar{t}, W^{pov}]$ (i.e. that there is nobody with income in the interval $[W^{pov} - \bar{t}, W^{pov}]$).

Applying formula (8) using (9), the optimal decision is to allocate the transfer according to

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

or

$$1\{\int_0^{W^{pov}} f(W|X) \; dW \geq \lambda\}$$

Applying Bayes' rule one obtains

$$1\{\frac{1}{f(X)} \int_0^{W^{pov}} f(X|W).f(W) \; dW \geq \lambda\}$$

or

$$1\{\int_0^{W^{pov}} f(X|W).f(W) \; dW \geq \lambda \begin{pmatrix} \int_0^{W^{pov}} f(X|W).f(W) \; dW \\ + \int_{W^{pov}}^{\infty} f(X|W).f(W) \; dW \end{pmatrix}\}$$

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

$$1\{\frac{f(X|W < W^{pov})}{f(X|W > W^{pov})} \geq c\} \tag{10}$$

where $c = \frac{\lambda}{1-\lambda} \frac{1-F(W^{pov})}{F(W^{pov})}$ is a constant. This is the optimal classifier or Bayes classifier (see the original paper by Welch (1939) or a textbook like Hand (1989)). $f(X|W < W^{pov})$ is the density of the characteristics of the group of poor people, and $f(X|W > W^{pov})$ that of the rich.

This derivation shows that the optimal statistical classification method coincides with targeting rule (8) if the transfer level is exogenously given and $v(.)$ is specified as in (9). PMTs regard targeting as a classification exercise, and in that sense could be close to the optimal procedure in this restrictive set-up. Nevertheless, optimality is only achieved by the Bayes classifier (10) and not *necessarily* by any other classification method.

The restrictions that were imposed on the general formulation in Problem 1 in order to obtain a solution that resembles a PMT based program clearly illustrate how the latter may be suboptimal. Firstly, they do not determine the choice of transfer level simultaneously with the target group. Secondly, PMT based programs do not always use the entire distribution of income given the covariates (which is needed to construct the optimal classifier). Finally, in cases where the preferences of the social planner are not as in expression (9), they fail to take the concavity of the social welfare function into account.

While the latter two problems can in principle be dealt with within the framework of PMTs, the endogenous determination of the transfer level cannot. This is likely to be a significant restriction. The optimal solution we propose does take all these features into account without jeopardizing computational tractability.

3.2 Determinants of the superiority of WT: some Monte Carlo investigations

This section summarizes the conclusions of a number of experiments on the performance of the optimal rule using artificial data. Our aim is to provide an intuitive feel for the reasons behind the superiority of WT over PMTs, and to point out in which type of situations one may expect this superiority to be large.

Our randomly generated datasets contain 3 covariates; the distributions from which they are drawn are calibrated to generate "reasonable looking" data. Experiments are performed as follows. First, a "household survey" of size 3000 is generated, containing income. This sample is then used to obtain conditional density estimates $\widehat{f}(W|X)$. Finally, the resulting estimates are plugged into the allocation rule from Theorem 1 to perform the targeting exercise in a "population" of 30000 drawn from the same DGP. In all direct comparisons with a PMT we keep the single transfer level fixed for WT. The PMT we use is dubbed "OLS" and implemented as follows: first use the sample to estimate a regression function $\widehat{E}(R|X)$; then assign the transfer to the households in the population with the lowest predicted income until the budget is depleted. In the interest of space, we present a summary of the results, organized along 3 themes[9].

Shape of the social welfare function One may expect targeting outcomes to vary with the degree of concern for the very poor. To examine these effects, we conduct targeting exercises using as social welfare functions the negative of the poverty indices FGT1 and FGT2 from Foster, Greer, and Thorbecke (1984). These are defined as

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

so in terms of our notation, $v_{FGT(n)}(W) = -1\{W < W^{pov}\} \left| \frac{W - W^{pov}}{W^{pov}} \right|^n$; as n increases, the SWF becomes more concave and concern for the very poor becomes stronger. The following pattern emerges across experiments. Firstly, for FGT1, WT only marginally outperforms PMT in social welfare terms. However, the additional weight on the very poor assigned by FGT2 dramatically widens the performance gap, leading to WT social welfare increases over twice as large as those obtained by PMT. The latter effect becomes even stronger as the poverty line is lowered and the budget is decreased: targeting then becomes more difficult.

Figure 2 displays the realized income distributions of benefit recipients for OLS and WT based on FGT1 and FGT2, generated by a representative experiment. Note that OLS does not involve any SWF and therefore the distribution is invariant to the choice of SWF. The distribution for WT with FGT1 in Panel B is similar to that obtained through OLS (Panel A). This is the case because under FGT1 the problem solved by WT is a

[9] *Ox*-code to repeat the experiments and variations thereof is downloadable from the authors' website, as is a more detailed discussion of the experiments on which this section is based. The *Ox* compiler, documented in Doornik (1999), is available free of charge.

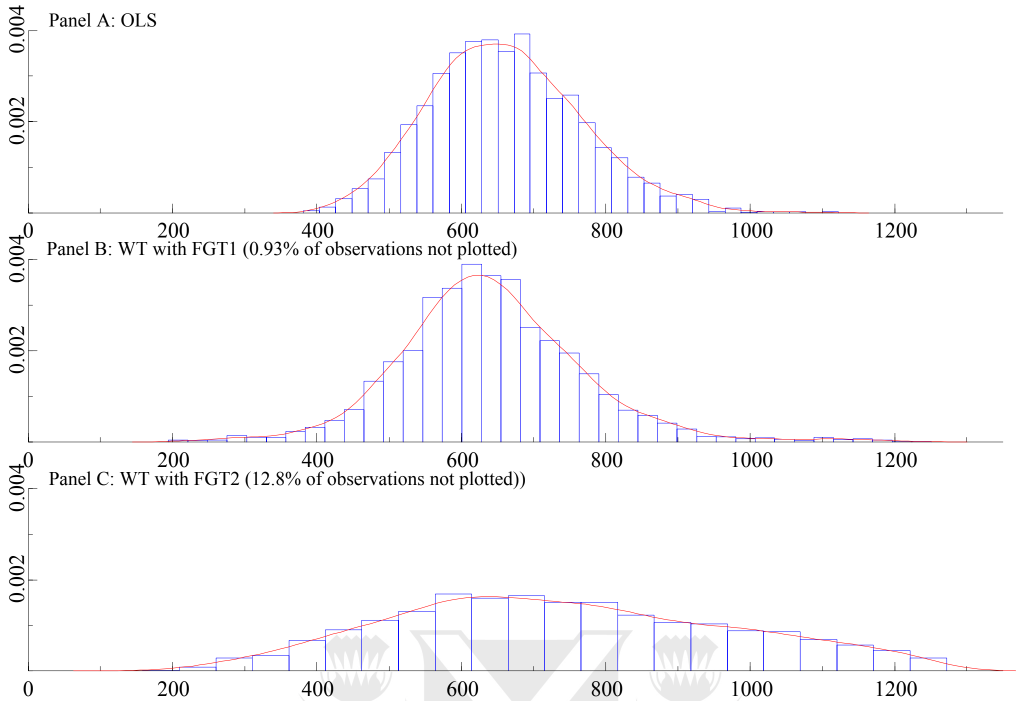

Figure 2: Realized income distributions of benefit recipients in an artificial dataset (poverty line = 592).

"standard" binary classification problem and, as shown in Section 3.1, WT then collapses to the Bayes classifier. OLS is also a binary classification method; the only difference with WT in this case 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 between 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 obtained 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 under FGT2, these households are preferred by WT.

The role of conditional heteroskedasticity One conclusion from the above discussion is that the presence of an inverse relationship between the conditional mean and the thickness of the conditional left tail across groups in the population will increase the superiority of WT. This, however, is a feature of the *data* and not of the targeting method. The artificial data used above were designed to contain some amount of conditional heteroskedasticity. However, one can easily imagine a population where, for instance, income is lognormally distributed conditional on the observables with fixed dispersion and a linear regression for the mean. In such a situation, WT will produce *the same* outcome as OLS. WT cannot improve on OLS in this case because all the relevant information about the population is summarized in the regression function.

The performance differential between standard PMTs and WT will depend on the nature of the population as well as on the budget and the shape of the SWF. There is no reason why this differential should always be of the same magnitude or nature as in this example. However, by construction WT always performs at least as well - up to estimation error or misspecification in the conditional densities - as PMTs in the metric defined by the social planner's objective. Since WT does not require any additional administrative effort and *may* result in considerable gains, we regard this "weak" superiority to be a good reason to prefer WT to PMT in practical applications.

Endogenous multiple transfer levels Compared to the above results based on a single fixed transfer level, the use of endogenously determined multiple levels leads, unsurprisingly, to further increases in the performance of WT. Interestingly however, in the experiments a clear pattern of diminishing returns to the inclusion of transfer levels was detected. Where the use of 2 transfer levels led to a 5% improvement over the reference situation with a single level, use of more than three transfer levels only generated trivial benefits. The limiting case with a nearly continuous transfer function (398 levels) only resulted in a 1.5% increase over the result based on 3 levels. Although this conclusion is certainly dependent on the data and the program parameters, it does suggest that the gains in terms of administrative efficiency and political feasibility resulting from the use of a small number of transfer levels are likely to outweigh the social welfare costs.

4 An application

Until now we have ignored the requirement that the conditional distribution $F(W|X)$ be known. In the experiments presented in the previous section this distribution could be estimated fairly easily and accurately from a small "household survey" so as to approximate the distribution in the larger "population". In practice, the situation is often not as ideal. Typically a large number of indicators are available. The sample then consists of a large number of groups for some of which only very few observations are available. The role of the smoothing / interpolation device used to recover $F(W|X)$ then becomes much more important. In fact, analogously to the standard classification literature[10], each choice of

[10] See e.g. Fix and Hodges (1951) for an old but excellent discussion. The optimal classifier was shown in 1939 to be the Bayes classifier; subsequent research focussed on implementation, i.e. estimation of the distribution of the covariates for the different groups.

estimator for $F(W|X)$ results in a different version of the WT rule.

The central issue in actual applications of WT in situations with small household surveys and many indicators is therefore the construction of a good approximation of the conditional income distribution in the population. A full examination of this problem is outside the scope of this paper, but we now briefly present two applications using actual household surveys. Our aim is to illustrate the potential impact of the choice of estimator on the results as well as to give an example of the simplicity and flexibility of our targeting method.

For simplicity, we perform in-sample targeting throughout. This is not a particularly relevant exercise from the point of view of method design[11], but will suffice as an illustration of our main points. We use two different datasets. The first is the Argentine Permanent Household Survey (2001, 12413 households), the second the Peruvian National Survey of Life Standards (2000, 3949 households). Both are described in the Appendix.

The following conventions are used. Household wealth W is measured by per capita income (Argentina) or expenditure (Peru) and is defined by total income / expenditure divided by the square root of number of members; in the notation of expression (7) this means $\rho(X) = \sqrt{nrMembers}$. Because households vary in family composition, we need to specify how to count the social welfare contribution. We do this by simply multiplying the social welfare contribution evaluated at household per capita income by the number of members. Since income is scaled by $\rho(X)$, this procedure implies that a given dollar amount of transfer is "money better spent" when given to a large houshold than to a small one, ceteris paribus. This is natural as it reflects the economies of scale such as sharing of consumption goods prevailing in larger households.

Because large households would be disadvantaged if only a single level were used, we scale transfers by the same function as used for the equivalence scale, i.e. $\phi(x) = \rho(X) = \sqrt{nrMembers}$. In each case, the base transfers are fixed at roughly 20% of the poverty line and the budgets are set such that roughly 15% of the population will receive a transfer.

For both datasets, we perform the following computations. The benchmark case is targeting by the OLS PMT used before: households are ranked based on predicted log per capita income[12] and a fixed transfer level (scaled for family composition) is distributed to households with the lowest predicted income until the budget is depleted. We compare this to two versions of WT. The first uses the parametric model for the conditional densities outlined in the Appendix (denote this by WT(par)); the second uses the empirical conditional distribution (WT(emp)). If the sample is considered to be the population, the latter case can be seen as a literal application of Theorem 1. On the other hand, it is useless for practical applications, in which the central issue is out-of-sample classification, as it performs no smoothing at all and assigns zero probability to all points not in the

[11] When choosing between alternative estimators of $F(W|X)$ using only a single small household survey, methods commonly used in the classification literature (e.g. cross-validation) are more reliable.

[12] Using a logarithmic transformation for estimation was found to perform considerably better than working on the original scale.

	FGT 2		Log utility	
	fixed	optimal	fixed	optimal
parametric	5%	*14%*	7%	*11%*
empirical	135%	$>$ *200%*	26%	*39%*

Table 1: Equivalent gain over PMT for various versions of WT rules: Peruvian data

sample. The results for WT(emp) are best interpreted as the upper bound of in-sample targeting performance over all possible conditional density estimation techniques.

The social welfare functions are FGT 2 with a poverty line at the 25th percentile of the unconditional per capita household income distribution and a utilitarian social welfare function with logarithmic utility. Transfer levels for WT are chosen in two ways for each case. Either it is done under the same conditions as for the PMT (single fixed level with prespecified scaling for family size) - see column "FIXED" in Tables 1 and 2, or using the additional features of WT (5 endogenous levels *without* scaling) - see column "OPTIMAL". The latter case is perhaps most natural: when the transfer level is made dependent on household size, de facto multiple levels apply anyway and hence it makes sense to let the targeting method tune these levels optimally to the accuracy with which target groups can be reached.

The results are presented in terms of equivalent gain of using WT over PMT, i.e. the percentage budget increase needed by PMT in order to reach the same level of performance as WT with the initial budget. These comparisons are computed by letting PMT always use the fixed base level of transfer with fixed scaling for family size.

Results for the Peruvian dataset are presented in Table 1. Use of WT result in nonnegligible savings, especially when transfer levels are determined endogenously. Two striking features deserve further comment. The first is the relatively small performance increase over PMT when using our parametric smoother to construct $F(W|X)$. For some combinations of program parameters t and B for the "FIXED" case, the superiority of WT disappears (details not reported). The second is the discrepancy between the results obtained by using our smoother and those obtained when using the empirical distributions. Although this discrepancy may be due to the inability of the parametric smoother is to capture variations in conditional higher moments, it is most likely caused by extreme overfitting of WT(emp). Especially for the set-up using FGT(2), only very few observations are available (since only the left tail of the distribution is relevant) for some values of the covariate-vector. This can occasionally lead to very precise location of small groups of very poor households when using the empirical distribution[13]. When, in addition, transfer levels are determined endogenously, WT(emp) will assign very high transfers to these households[14]. A standard PMT is not able to do this.

The results for the Argentinian dataset presented in Table 2 are similar. The only

[13] In fact, this effect disappears when only a few indicators are used: more data are available for each group and the smoothed and empirical versions of WT perform similarly.

[14] This illustrates a potential pitfall when using the empirical distribution from a small dataset to calculate the optimal transfer rule, as for instance done in Glewwe (1992).

	FGT 2		Log utility	
	fixed	optimal	fixed	optimal
parametric	5%	*15%*	7%	*12%*
empirical	62%	*131%*	29%	*44%*

Table 2: Equivalent gain over PMT for various versions of WT rules: Argentinian data

notable difference are the less extreme numbers for WT(emp) with FGT 2. This is not surprising: the dataset is 3 times larger so that the danger of in-sample overfitting is smaller.

WT(par) performs rather modestly when a single scaled fixed transfer level is used. As for the Peruvian data, this means that either the parametric smoother is unable to capture the kind of features in the data that gave rise to the more impressive results in the experiments with artificial data, or these features are not present in the data. Nevertheless, endogenous determination of transfer levels naturally increases the edge of WT over PMT.

In summary, these empirical applications illustrate that (i) WT can incorporate all features of a targeting program (equivalence scales, fixed corrections for family composition, etc) in a single consistent framework, (ii) endogenous determination of transfer levels can increase the performance differential and (iii) WT always performs at least as well as PMT, even though the difference can be small. Perhaps the most important insight gained from this application concerns the construction of an estimate for the conditional densities. Targeting decisions and performance can be sensitive to the choice of smoother (as exemplified by the extreme case WT(emp) discussed above). Moreover, improvements may be available by constructing conditional density estimates using information from various sources (see e.g. the welfare estimator of Elbers, Lanjouw, and Lanjouw (2003)) and making corrections for measurement error (as in Chesher and Schluter (2002)).

5 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 computationally feasible technique that allows joint determination of transfer levels and their corresponding target groups whilst closely fitting the administrative set-up prevailing in currently operational programs. The targeting rule is obtained as the solution to a social welfare maximization (or poverty index minimization) problem under the constraint that the transfer function be piecewise constant. This ensures that implementation is 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.

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 treat 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 one[15]. Finally, PMT based support programs determine transfer levels exogenously whereas they should depend on the accuracy with which the target groups can be reached.

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. Finally, robust out-of-sample targeting requires a sensible estimation method for the conditional densities of income. Extensive experimentation with various methods and testing procedures is required.

6 Appendices

6.1 Proof of the Lagrange multiplier representation

The aim of this Appendix is to show formally that the problem of choosing receiver groups associated with given transfer levels can be written using a "Lagrange multiplier" representation, a fact used in the key step of the proof of Theorem 1:

Lemma 1 *Under assumptions given below, there exists a number λ such that the solution π^* of*

$$\begin{array}{c} \max_{\pi\in\Pi} \sum_{k=1}^{K} \int_{\Omega_k} \int_W v(W+t_k, X) - v(W,X)\ dF(W,X) \\ s.t.\ \sum_{k=1}^{K} \int_{\Omega_k} t_k.dF(X) \leq B \end{array} \tag{11}$$

where Π is the set of partitions $\pi := \{\Omega_0, \Omega_1, ..., \Omega_K\}$ of the population Ω, can be obtained by solving

$$\max_{\pi\in\Pi} \sum_{k=1}^{K} \int_{\Omega_k} \int_W v(W+t_k, X) - v(W,X)\ dF(W,X) + \lambda.(B - \sum_{k=1}^{K} \int_{\Omega_k} t_k.dF(X)) \tag{12}$$

Define the following $2K$ measures (real-valued countably additive set functions) on the measure space (Ω, Σ, μ_F), where Ω represents the population, the sigma-algebra Σ is loosely interpreted as the collection of potential target groups in the population, and

[15] This 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.

μ_F is the dominating measure associated with the distribution function F:

$$\mu_{SW,k}(\Omega_k) := \int_{\Omega_k} \left\{ \int_W v(W + t_k, X) - v(W, X) \; dF(W|X) \right\} dF(X) \quad k = 1, ..., K$$
$$\mu_{BC,k}(\Omega_k) := \int_{\Omega_k} t_k \; dF(X) \qquad k = 1, ..., K$$

where "SW" stands for "social welfare" and "BC" stands for "budget constraint". Using this notation, Lemma 1 states that maximizing

$$\sum_{k=1}^{K} \mu_{SW,k}(\Omega_k) \quad s.t. \; \sum_{k=1}^{K} \mu_{BC,k}(\Omega_k) \leq B$$

is equivalent to maximizing

$$\sum_{k=1}^{K} \mu_{SW,k}(\Omega_k) + \lambda(B - \sum_{k=1}^{K} \mu_{BC,k}(\Omega_k))$$

for some fixed number λ.

Definition 1 *A measure μ is said to be nonatomic on Σ if for every $S \in \Sigma$ and every positive number $b < \mu(S)$ there exists an $S_b \subset S$ for which $\mu(S_b) = b$. A set $E \in \Sigma$ is an atom of μ if for every measurable set $F \subset E$ either $\mu(F) = 0$ or $\mu(F) = \mu(E)$.*

Assumption 1 *$\mu_{SW,k}$ and $\mu_{BC,k}$ are nonatomic, $\forall k$.*

The key to the proof of Lemma 1 is Chernoff's (1951) generalization of Lyapunov's (1940) convexity theorem for the range of a vector measure. This generalization proves convexity for the range of a vector-valued function ψ, which maps a partition of Ω into a $2K$-dimensional vector of positive numbers and is defined as

$$\psi(\pi) := \left[\; \mu_{SW,1}(\Omega_1) \quad \cdots \quad \mu_{SW,K}(\Omega_K) \quad \mu_{BC,1}(\Omega_1) \quad \cdots \quad \mu_{BC,K}(\Omega_K) \; \right]' \tag{13}$$

where $\pi := \{\Omega_0, \Omega_1, ..., \Omega_K\}$ is a partition of Ω. The *range* R of ψ is the subset of $\mathbb{R}^{2K}$ obtained by collecting all values of ψ as its argument ranges across all possible partitions $\pi \in \Pi$ of the population into recipient groups in Σ, i.e. $R := \{\psi(\pi) \in \mathbb{R}^{2K} | \; \pi \in \Pi\}$. Throughout, we use the generic notation of a vector $z \in \mathbb{R}^n$ with components denoted by z_k, that is, $z = (z_1, z_2, ..., z_n)'$. The following Lemma carries the main burden of the proof of Lemma 1 and follows directly from a theorem by Chernoff (1951).

Lemma 2 *The range R of the function ψ is bounded, closed and convex.*

Proof. Define $\psi^+(\pi) := [\mu_{extra}(\Omega_0) \;\; \psi(\pi)']'$ where $0 \leq \mu_{extra}(\Omega_0) \leq A < \infty$ and nonatomic but otherwise arbitrary. By the theorem of Chernoff (1951) (choosing, in his notation, $k = K + 1$, $n_1 = 1$ and $n_i = 2$, $i > 1$), the range R^+ of ψ^+ is bounded, closed and convex (BCC). It is easily shown by contradiction that the projection of R^+ on the space $\{z \in \mathbb{R}^{2K+1} |\ z_1 = 0\}$, given by $\{z \in \mathbb{R}^{2K} |\ [y\ z']' \in R^+\ for\ some\ y\} = R$, is also BCC. ■

The scene is now set for the proof of Lemma 1:

Proof. The proof only considers the case where the budget is insufficient to assign the maximum transfer level to everybody; otherwise the Lemma obviously holds with $\lambda = 0$. The idea of the proof is graphically displayed in Figure 3 for the case of a single transfer level ($K = 1$). The range R of ψ is BCC by Lemma 2. Because all components of ψ are finite measures, $\mu_{SW,k} \leq A$ for some finite A. The set

$$\Upsilon := \{z \in \mathbb{R}^{2K} |\ \sum_{k=K+1}^{2K} z_k \leq B;\ z_k \geq 0;\ z_1, ..., z_K \leq A\}$$

is clearly BCC. The intersection of two BCC sets is BCC, hence the choice set $R \cap \Upsilon$ of the optimization problem in Lemma 1 is BCC. Therefore the objective function $\sum_{k=1}^{K} z_k$ reaches its maximum M on $R \cap \Upsilon$. Call this point z^*. This point corresponds to the solution to problem (11). Because the objective function is increasing as recipient groups are expanded, we may assume that the budget constraint binds at this maximum, i.e. $\sum_{k=K+1}^{2K} z_k^* = B$.

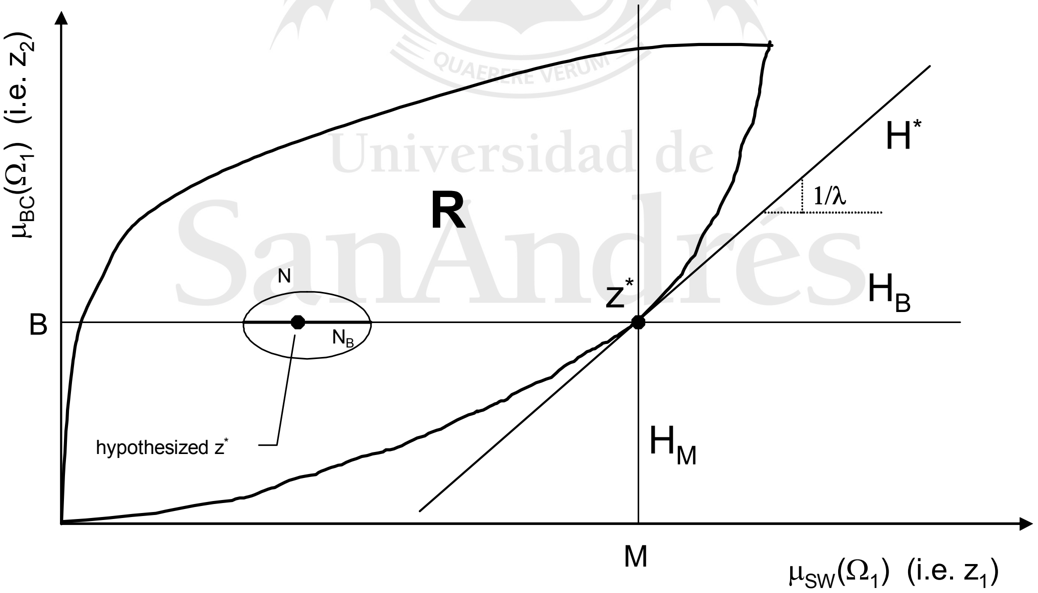

Figure 3: Graphical representation of the key concepts in the proof of Lemma 1 for the case of $K = 1$.

Define the $(2K - 1)$-dimensional "budget constraint" hyperplane H_B

$$H_B := \{z \in \mathbb{R}^{2K} |\ \sum_{k=K+1}^{2K} z_k = B\}$$

and the “iso-social welfare increase” hyperplane H_M

$$H_M := \{z \in \mathbb{R}^{2K} | \sum_{k=1}^{K} z_k = M\}$$

z^* is a boundary point[16] of R. Why? Suppose instead that $z^* \in int(R)$. Then there exists a neighbourhood N of z^* such that $N \subset R$. Since H_B contains z^*, $N_B := N \cap H_B$ is a neighbourhood of z^* relative to H_B. N_B only contains points satisfying the budget constraint. Hence, $\sum_{k=1}^{K} z_k$ cannot be maximized in the interior of N_B. This is a contradiction, and hence z^* must be a boundary point of R.
Consequently, by the Supporting Hyperplane Theorem, there exists a supporting hyperplane H^* to R containing z^*. H^* is defined by constants a_k, b_k and c:

$$H^* := \{z \in \mathbb{R}^{2K} | \sum_{k=1}^{K} a_k z_k + \sum_{k=1}^{K} b_k z_{K+k} = c\}$$

with the property that

$$\forall z \in R : \sum_{k=1}^{K} a_k z_k + \sum_{k=1}^{K} b_k z_{K+k} \leq c. \qquad (*)$$

We now search for restrictions on the coefficients of H^*. By the definition of z^* as the maximizer of $\sum_{k=1}^{K} z_k$ over $R \cap H_B$ with $\sum_{k=1}^{K} z_k^* = M$, $H_M \cap H_B$ is a $(2K-2)$-dimensional supporting hyperplane to $R \cap H_B$ relative to H_B and contains z^*[17]. This is also true for $H^* \cap H_B$. Hence

$$H^* \cap H_B = H_M \cap H_B \qquad (**)$$

(if H^* is not unique, it can be chosen such that (**) holds) From (**) it follows that $H^* \cap H_B \cap H_M = H^* \cap H_B$, which implies two constraints on the coefficients defining H^*. The first is that the matrix G

$$G := \begin{bmatrix} a_1 & \cdots & a_K & b_1 & \cdots & b_K \\ 1 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 1 \end{bmatrix}$$

associated with the system of equations

$$Gz = \begin{bmatrix} c \\ M \\ B \end{bmatrix}$$

[16] Throughout, a “boundary point” of a set A is to be interpreted relative to the lowest dimensional hyperplane containing A.

[17] This follows from the shape of the objective function and is easily proven by contradiction; the statement that z^* is a boundary point of R (and not only of $R \cap H_B$), proven before, is not as direct.

has row rank 2. Hence $a_1 = ... = a_K = a$ and $b_1 = ... = b_K = b$.
The second constraint is $c = aM + bB$.
By (*) we now have that

$$\forall z \in R : a \sum_{k=1}^{K} z_k + b \sum_{k=1}^{K} z_{K+k} \leq aM + bB$$

Dividing through by a and defining $\lambda = -\frac{b}{a}$ gives $\forall z \in R : \sum_{k=1}^{K} z_k + \lambda(B - \sum_{k=1}^{K} z_{K+k}) \leq M$, or otherwise stated that

$$\forall \{\Omega_0, \Omega_1, ..., \Omega_K\} \in \Pi : \sum_{k=1}^{K} \mu_{SW}(\Omega_k) + \lambda(B - \sum_{k=1}^{K} \mu_{BC}(\Omega_k)) \leq M$$

with equality at the solution to (11). ■

6.2 The assumption of nonatomicity

An important assumption behind the convexity result in Chernoff's (1951) theorem is the non-atomicity of the measures $\mu_{SW,k}$ and $\mu_{BC,k}$. In practice, this assumption is always violated, and this section examines how this affects the use of Theorem 1. We begin by examining the most extreme possible example, in which there is a single binary covariate x, i.e. two groups in the population. The small group of households defined by $x = 0$ are rich; allocating the fixed transfer level to each member of this group results in the outcome marked "r" in Figure 4. The large group of households defined by $x = 1$ are poor and are associated with the point marked "p". It turns out that the budget B is too small to make p feasible; therefore the optimal solution is point r. When using Theorem 1 to select recipient groups, it is intuitively clear (see (8)) that the solution will be point 0. Formally, there exists no supporting hyperplane H^* as in Figure 3 that passes through point r.

This extreme example illustrates the type of problems that may arise when applying Theorem 1 in practical calculations. One possible solution to this problem is to convexify the solution set by allowing randomization within the marginal recipient group[18] as is for instance done in the Mexican program Tu Casa. Even without this, in typical situations encountered in practice, problems due to atoms are unlikely to lead to large discrepancies between the computed optimum and the true optimum. Indeed, one usually encounters situations in which a few continuous covariates are available, leading to a population with many thousands of small atoms. The kind of "mistake" one can make in such situations is schematically displayed in Figure 5. When using our algorithm based on Theorem 1, one will select the outcome z^*, even though the exact maximum is reached at z'.

So far we have focussed on the case where transfer levels are fixed. When varying transfer levels, the location of the atoms in the Figures in this section changes, further

[18] In the example, this amounts to allowing some fraction of the group of poor to receive a transfer. Graphically, the dotted line connecting 0 and p then becomes part of the choice set.

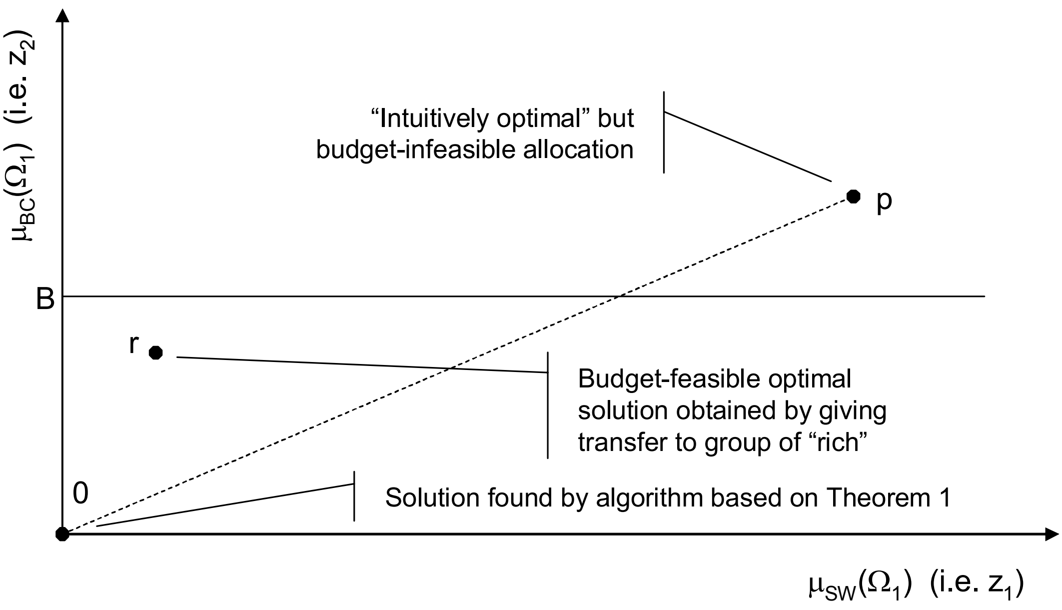

Figure 4: Illustration of the failure of an algorithm base on Theorem 1 to select the correct constrained maximum when the assumption of nonatomicity is not valid.

reducing potential problem due to atoms. For instance, in the example of Figure 4, the ideal outcome can be achieved by lowering the transfer level enough so that allocating transfers to the group of poor becomes budget feasible. It is clear that the value of t at which this becomes possible corresponds to a strong discontinuity in the social planner's objective. Consequently, all numerical searches are performed using custom-designed grid-search algorithms. The user can set the initial range of the grid over which the search is performed as wide as desired, and in situations where one suspects large atoms to be present, one may increase this range accordingly. Details of implementation are discussed in the following section.

6.3 Computational details: brief overview

The main technical contribution of the paper is Theorem 1, which allows us to transform the original constrained maximization problem into two nested unconstrained problems, where the "inner" loop is a one-dimensional search. The objective functions for these searches may be discontinuous, and consequently we have developed tailor-made routines.

6.3.1 Calculating λ for given transfer levels

For given transfer levels, changing the value of lambda 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 lambda. This function is then maximized numerically using a gridsearch algorithm to avoid local maxima problems and problems due to discontinuities. As in the

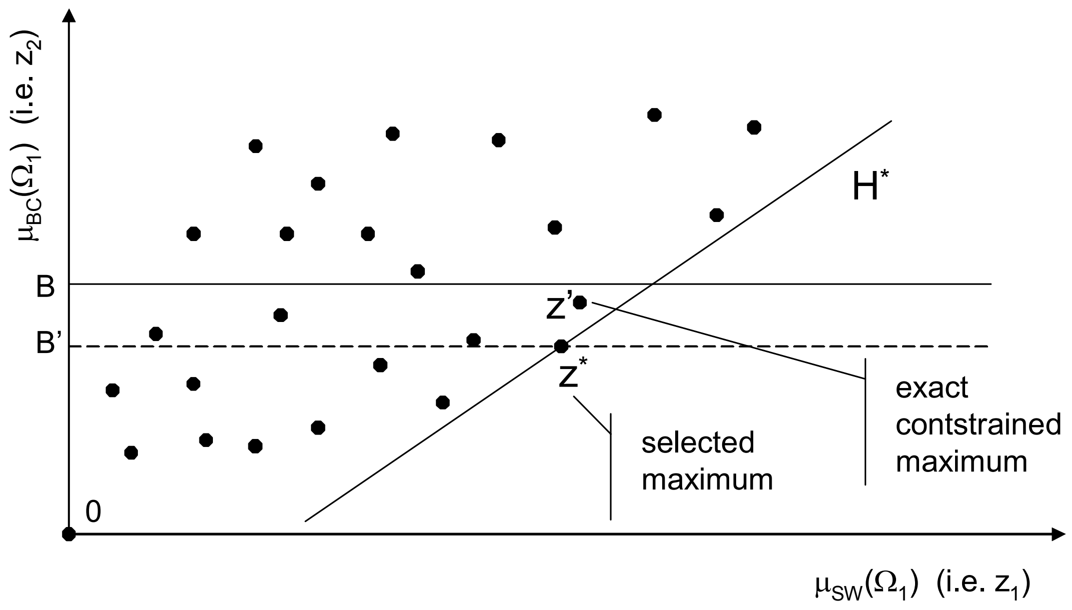

Figure 5: The effect of atoms on algorithm performance in realistic situations.

example depicted in Figure 5, it may happen that no value of λ exists for which the entire budget is depleted exactly.

6.3.2 Calculating the transfer levels

Once we are 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. Additionally, the objective function may contain discontinuities, rendering standard gradient algorithms unreliable.

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. integers 10 to 200). For each household i in the population and each t on the lattice, the quantity $\int_0^\infty [v(W+t,x_i) - v(W,x_i)]\ dF(W|X=x_i)$ is calculated and stored before iteration starts. The resulting matrix is then used as an input to the gridsearch algorithm.

6.4 A flexible parametric method for estimating conditional densities

We now describe the method for constructing the conditional density estimates used in the paper. We note that this is just an example of how this can be done in practice. In principle, any density estimation method can be used as an input to our procedure.

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

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

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

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

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

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

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, 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.

The parameters are estimated by maximum likelihood. The only aim is smoothing and "filling in the gaps left by the dataset" and not "estimating fundamental or causal parameters" that need to be used for further research.

6.5 Description of the household surveys

Peruvian data **Source**: Encuesta Nacional sobre Medición de Niveles de Vida, Mayo 2000, Instituto CUANTO. **Size**: 3949 households. **Variables**: total (yearly) expenditure, age, gender and (binary) skill level of head of the household, number of member and numbers of minors per household, dummies for regions, dummies for water, sewage and electricity access, and quality indicators and size of the house.

Argentinian data **Source**: Encuesta Permanente de Hogares (EPH), Mayo y Octubre 2001, Instituto Nacional de Estadística y Censos. **Size**: 12413 households. **Variables**: average total income in 2001, gender and (binary) skill level of head of the household, number of members and number of minors per household, dummies for regions, dummies for water, sewage and electricity access, and quality indicators and size of the house.

References

Akerlof, G. (1978): “The Economics of Targeting as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning,” *American Economic Review*, 68, 8–19.

Besley, T., and R. Kanbur (1988): “Food Subsidies and Poverty Alleviation,” *Economic Journal*, 98, 701–719.

——— (1993): “The Principles of Targeting,” in *Including the Poor*, ed. by M. Lipton, and J. Van Der Gaag. The World Bank.

Chaudhuri, S., and M. Ravallion (1994): “How Well Do Static Indicators Identify the Chronically Poor?,” *Journal of Public Economics*, 53(3), 367–394.

Chernoff, H. (1951): “An extension of a result of Liapounoff on the range of a vector measure,” *Proceedings of the American Mathematical Society*, 2, 722–726.

Chesher, A., and C. Schluter (2002): “Welfare Measurement and Measurement Error,” *Review of Economic Studies*, 69, 357–378.

Datt, G., and M. Ravallion (1993): “Regional Disparities, Targeting, and Poverty Alleviation in India,” in *Including the Poor*, ed. by M. Lipton, and J. Van de Gaag. World Bank, Washington DC.

Doornik, J. (1999): *Ox: An Object-Oriented Matrix Language (3rd Edition)*. Timberlake Consultants Press.

Elbers, C., J. O. Lanjouw, and P. Lanjouw (2003): “Micro-Level Estimation of Welfare,” *Econometrica*, forthcoming, 57p.

Fix, E., and J. 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), 233–247.

Foster, J., J. Greer, and E. Thorbecke (1984): “A Class of Decomposable Poverty Measures,” *Econometrica*, 52, 761–765.

Glewwe, P. (1992): “Targeting Assistance to the Poor: Efficient Allocation of Transfers When Household Income Is Not Observed,” *Journal of Development Economics*, 38(2), 297–321.

Grosh, M. (1994): “Administering Targeted Social Programs in Latin America: From Plattitudes to Practice,” *mimeo.*

Hand, D. (1989): *Discrimination and Classification*. Wiley.

Hettich, W., and S. L. Winer (1988): "Economic and political foundations of tax structure," *American Economic Review*, 78(4), 701–712.

Kanbur, R. (1987a): "Measurement and Alleviation of Poverty," *IMF Staff Papers*, 34, 60–85.

——— (1987b): "Transfer, Targeting, and Poverty," *Economic Policy*, 4, 112–36.

Kanbur, R., and M. Keen (1989): "Poverty, Incentives, and Linear Income Taxation," in *The Economics of Social Security*, ed. by A. Dilnot, and I. Walker. Clarendon Press, Oxford.

Kanbur, R., M. Keen, and M. Tuomala (1995): "Labor Supply and Targeting in Poverty-Alleviation Programs," in *Public Spending and the Poor: Theory and Evidence*, ed. by D. Van de Walle, and K. Nead. John Hopkins.

Lyapunov, A. A. (1940): "Sur les fonctions-vecteurs complètement additives," *Bulletin of the Academy of Sciences of the USSR, Mathematics Series*, 4, 465–478.

Nichols, A., and R. Zeckhauser (1982): "Targeting Transfers Through Restrictions on Recipients," *American Economic Review*, 72(2), 372–377.

Ravallion, M. (1989): "Land-Contingent Policies for Rural Poverty Alleviation," *World Development*, 17, 1223–1233.

——— (1993): "Poverty Alleviation Through Regional Targeting: A Case Study for Indonesia," in *The Economics of Rural Organization: Theory, Practice and Policy*, ed. by K. Hoff, A. Braverman, and J. E. Stiglitz. Oxford University Press, Oxford.

Ravallion, M., and K. Chao (1989): "Targeted Policies for Poverty Alleviation Under Imperfect Information: Algorithms and Applications," *Journal of Policy Modeling*, 11(2), 213–224.

Ravallion, M., and B. Sen (1994): "Impacts on Rural Poverty of Land-Based Targeting: Further Results for Bangladesh," *World Development*, 22(6), 823–838.

Welch, B. L. (1939): "Note on Discriminant Functions," *Biometrika*, 31, 218–220.

World Bank (1990): *World Development Report 1990* Washington D.C.

——— (2001): *World Development Report 2000/2001* World Bank, Washington D.C.