

Klasen, Stephan; Lange, Simon

Working Paper

How narrowly should anti-poverty programs be targeted? Simulation evidence from Bolivia and Indonesia

Discussion Papers, No. 213

Provided in Cooperation with:

Courant Research Centre 'Poverty, Equity and Growth in Developing and Transition Countries',
University of Göttingen

Suggested Citation: Klasen, Stephan; Lange, Simon (2016) : How narrowly should anti-poverty programs be targeted? Simulation evidence from Bolivia and Indonesia, Discussion Papers, No. 213, Georg-August-Universität Göttingen, Courant Research Centre - Poverty, Equity and Growth (CRC-PEG), Göttingen

This Version is available at:

<https://hdl.handle.net/10419/145296>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

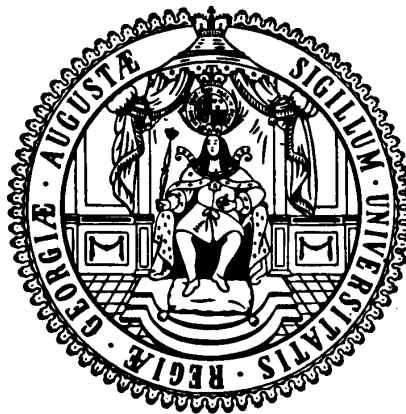
You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Courant Research Centre

‘Poverty, Equity and Growth in Developing and Transition Countries: Statistical Methods and Empirical Analysis’

Georg-August-Universität Göttingen
(founded in 1737)



Discussion Papers

No. 213

**How Narrowly Should Anti-poverty Programs
Be Targeted?
Simulation Evidence from Bolivia and Indonesia**

Stephan Klasen, Simon Lange

August 2016

Platz der Göttinger Sieben 5 · 37073 Goettingen · Germany
Phone: +49-(0)551-3921660 · Fax: +49-(0)551-3914059

Email: crc-peg@uni-goettingen.de Web: <http://www.uni-goettingen.de/crc-peg>

How Narrowly Should Anti-poverty Programs Be Targeted?

Simulation Evidence from Bolivia and Indonesia*

Stephan Klasen[†] Simon Lange[‡]

August 10, 2016

Abstract

A key question in the design of anti-poverty programs is to what extent they should be targeted. Empirical evaluations of targeted transfer schemes and simulation exercises often point to further gains that can be had from targeted transfers *vis-à-vis* universal transfers or from more narrow targeting. Theoretical work, on the other hand, has identified hidden costs associated with targeting—including politico-economic constraints on budgets—but these are frequently ignored in empirical work. In this paper we first argue that common targeting measures can be interpreted as preferences that attach specific weights to true and false positive rates. Based on data from Bolivia and Indonesia, we show that targeting based on an imperfect poverty classifier based on proxy means tests results in very distinct ‘optimal’ beneficiary shares when these measures are used as a decision criterion. Implications from poverty simulations are sensitive to assumptions about the political economy relationship between the beneficiary share and the available budget. In fact, in many situations, optimizing targeting measures will be misleading when the actual goal is to maximize the effect on poverty.

Keywords: welfare and poverty measurement; targeting; transfers; social assistance; proxy means tests; poverty; Bolivia; Indonesia.

JEL Classification Numbers: C52; I38; O21.

*This work was supported by the German Federal Ministry for Economic Cooperation and Development in cooperation with the Kreditanstalt für Wiederaufbau (KfW). We are grateful to John Hoddinott, Julia Johannsen, Fernando Landa, Roland Pardo, Ramona Rischke, Cecilia Vidal, as well as two anonymous referees and Pete Lanjouw, the responsible editor, for helpful comments on earlier versions of this paper.

[†]Economics Department, University of Göttingen.

[‡]Corresponding author. Economics Department, University of Göttingen. Address: Platz der Göttinger Sieben 3, 37073 Göttingen, Germany, telephone: +49-551-39-21679 e-mail: simon.lange@wiwi.uni-goettingen.de.

1 Introduction

Targeting of transfer programs is often seen as ultimately “a matter of cost-effectiveness in securing a particular benefit [...], one of maximizing the poverty-removal benefits accruing from a given burden of cost” (Sen, 1995).¹ Yet *errors of targeting*, undercoverage of the poor and leakage to the non-poor, often feature more prominently in public and academic debates than does poverty removal (e.g. Cornia and Stewart, 1995). As targeted programs have to rely on imperfect strategies to distinguish the poor from the non-poor, there is a trade-off here governed by the beneficiary share, the population proportion included in the program. Broad targeting will assure that a majority of the poor will be included while narrow targeting will limit leakage to the non-poor. But how narrow exactly should programs be targeted?

The likely effects on poverty can be one yardstick: in an early contribution to the literature, Akerlof (1978) shows that targeting specific groups that are needy on average allows for large payments with relatively low marginal rates of taxation. Authors concerned with the design of targeted transfers in developing countries have pointed to limited resources and inequality among the poor as an argument for narrow targeting (Grosh et al., 2008, p. 91), arguing that benefits should be concentrated “as far down the distribution as possible, even when overall poverty is high” since budgets are usually very limited and the poorest of the poor are often many times worse off than the poor near the poverty line.

Proponents of broad targeting, on the other hand, have pointed out that targeting involves extra cost that come in various forms (e.g. Besley and Kanbur, 1993). First, missing the poor may be costly if long-term growth-effects are expected from transfers (Cornia and Stewart, 1995; Jalan and Ravallion, 1998). Second, identifying the deserving poor and verifying their situation may also only be possible through administrative efforts that drive up costs. Third, the administrative infrastructure that has to be in place for effective targeting may at the same time create opportunities for corruption and welfare fraud (Camacho and Conover, 2011; Niehaus et al., 2013). Fourth, narrow targeting may result in more severe incentive-problems, information costs, and welfare stigma (Moffitt, 1983; Sen, 1995).

A final argument against narrow targeting stems from the literature on the political economy of targeted transfer programs. Several theoretical studies find that narrow targeting may undermine political support for a program and may thus result in underfunded programs or even no program at all (De Donder and Hindriks, 1998; Moene and Wallerstein, 2001; Gelbach and Pritchett, 2001, 2002). Real-life examples of this phenomenon are discussed in Besley and Kanbur (1993) and Gelbach and Pritchett (2001, 2002). However, altruism and warm glow considerations are usually ignored in this theoretical work in the interest of clarity, and Atkinson (1995), in particular, argues that all objectives of programs have to be considered in order to shed light on the political economy of targeted transfers.

In this paper we analyze the consequences of drawing on *targeting measures*, single-valued indicators that aggregate over error rates, in order to set the beneficiary share in program design.

¹General reviews are provided by Besley and Kanbur (1993), Sen (1995), and Ravallion (2003).

We ask what implications their use in the design-stage of programs has on poverty effects. We first show that commonly-used targeting measures can be interpreted in a common framework based on *receiver operating characteristic* (ROC)-curves as preferences defined over the extent of coverage and leakage. Using household data from two very different developing countries, Bolivia and Indonesia, we then stipulate two regression-based proxy-means tests (PMTs), a widely-used targeting strategy based on the estimation of a statistical model for welfare that is then applied to the population as a whole in order to obtain predictions (e.g. [Grosh and Baker, 1995](#)).

Our first set of results indicates that policy-makers concerned primarily with coverage of the poor might opt for a measures such as the *targeting differential* (TD), the difference between the shares among the poor that are correctly and incorrectly included ([Galasso and Ravallion, 2005](#)). If less than half the population is poor, maximizing this indicator will result in programs that are broadly targeted, i.e. that deliberately include some of the non-poor in order to ensure high coverage among the poor. Conversely, policy-makers concerned primarily with leakage might opt for an approach that tries to bring down the *total error rate* (TER), the proportion of individuals that are erroneously classified in either direction (e.g. [Cornia and Stewart, 1995](#)). We show that this implies a lesser weight on coverage among the poor and, hence, a larger weight on the proportion of non-poor that are covered. This will result in programs that target a population fraction that is smaller than the fraction in poverty.

While the first part of the paper shows how such measures can be rationalized in a common framework as preferences over errors of targeting, it remains an open question whether the resulting implications for targeting are also well-aligned with poverty effects and in how far this depends on the initial distribution of welfare and the extent of poverty. In the second part, we therefore simulate poverty effects of targeted transfer schemes based on our PMTs under alternative assumptions about the relationship between the extent of targeting and available budgets. While previous simulations have been concerned with the optimal allocation of a fixed budget ([Ravallion and Chao, 1989](#); [Glewwe, 1990](#); [Grosh and Baker, 1995](#)), the literature on the political economy of targeting has demonstrated that the budget is likely an increasing function of the beneficiary share ([Moene and Wallerstein, 2001](#); [Gelbach and Pritchett, 2001, 2002](#)). We therefore also run simulations in which the budget is an increasing function of the beneficiary share. To that end, we present empirical evidence from Chinese data on cash transfer schemes that is consistent with the view that a one-percent increase in the beneficiary share results in an increase by as much as one third of a percent in the budget available to program administrators.

Our results in the second part of the paper suggest that targeting based on optimizing targeting measures can often be misleading if the actual goal is to maximize the effect of a program on poverty. In fact, optimizing the TER or the TD will in most scenarios result in targeting that is too narrow or too broad, respectively, relative to the poverty-minimizing extent of targeting. Hence, a strong focus on either coverage or leakage may come at the cost of foregoing a larger effect on poverty. On the other hand, setting the beneficiary share equal to the initial poverty headcount often achieves an effect on poverty that is close to the maximum effect attainable.

Exceptions to this include situations in which the budget is highly elastic with respect to the targeting rate and the initial poverty gap is narrow (as is the case in Indonesia), in which case maximizing the TD results in poverty reductions that are close to those that are maximally feasible.

Our study contributes to a large literature on targeting strategies and targeting outcomes. However, apart from studies that investigate targeting of single programs in specific contexts, empirical work in this field has often been concerned with comparisons between alternative targeting strategies based on simulations (Grosh and Baker, 1995; Skoufias and Coady, 2007), *ex post* evaluations (e.g. Coady et al., 2004), or, in some cases, randomized control trials (Alatas et al., 2012, 2013). In contrast, we focus on the (largely empirical) question of how narrow programs will be targeted under alternative objective functions. To that extent we employ a commonly-used method, PMTs, that comes in two variants in our study. Our findings indicate, however, that differences between PMTs based only on geographic information and household demographics and sophisticated versions that also incorporate information on asset ownership are negligible while initial poverty and the extent of targeting policy-makers opt for result in large differences in outcomes.

Our study is closely related to work by Ravallion (2009b) who studies the predictive power of targeting measures for poverty effects of China’s urban *Dibao*-program. In fact, we use the same program to estimate the elasticity of the budget with respect to the beneficiary share. With few exceptions—notably, TD —his findings suggest that correlations between targeting measures commonly used in the literature and poverty effects are tenuous at best. Our findings indicate that targeting measures are related to poverty effects only under very specific circumstances and that, therefore, they will often make for poor proxy objectives.

The remainder of the paper is organized as follows: in section 2, we show that ROC-analysis provides a useful framework to rationalize the use of targeting measures that combine information on targeting errors in order to design transfer schemes. Section 3 describes the datasets, the indicator of economic welfare, and the classifier we use in subsequent sections. Section 4 empirically investigates the consequences of making aggregated targeting measures the basis for program design. Section 5 reports results from poverty simulations. Final remarks are offered in section 6.

2 Targeting methods and targeting errors

2.1 Targeting methods

While direct means tests using verifiable data on incomes (e.g. from tax records) are usually feasible in developed countries, identifying the needy is much more challenging in developing countries where a larger share of total income is generated in the urban informal sector and where people in rural areas tend to rely on subsistence agriculture for a living (Deaton, 1997). Verifiable data on incomes or other direct measures of economic welfare are thus often lacking

for a large portion of the population.

In such a setting, one is forced to choose from imperfect methods that do not require complete expenditure or income data. For instance, schemes based on *geographic targeting* channel resources to regions in which the extent of poverty is estimated to be large (Bigman and Fofack, 2000; Schady, 2002). Alternatively, one may try to tap into local knowledge by delegating the selection of beneficiaries to locals, a strategy referred to as *community-based targeting* (Conning and Kevane, 2002; Bardhan and Mookherjee, 2005; Galasso and Ravallion, 2005). Workfare schemes (e.g. Besley and Coate, 1992; Dutta et al., 2012) and the subsidization of goods and services consumed primarily by the poor are *self-targeting mechanisms* (Besley and Kanbur, 1988; Alatas et al., 2013). Participants incur some sort of utility loss that makes participation attractive only for the deserving. In practice we often observe that several of these methods are combined.

A further alternative—and the one that the empirical analysis in this paper is based on—is to consider a number of readily observable and verifiable household characteristics to construct a proxy means test (PMT). PMTs are widely employed—especially in Latin America—with Chile the first country to base targeting of its social pension and disability scheme on a PMT (Lindert et al., 2006). Other examples include Colombia (Canstáneda, 2005), Mexico (Skoufias et al., 2001), Egypt (Ahmed and Bouis, 2002), and Indonesia (Alatas et al., 2012). PMTs employ a limited number of variables in order to predict the welfare-level of households. Grosh and Baker (1995) and Kidd and Wylde (2011) provide systematic assessments of PMTs; Alatas et al. (2012) provide experimental evidence on targeting based on PMTs *vis-à-vis* alternatives that rely on community involvement.

A statistical model, usually a linear model, is first calibrated with data from a representative household expenditure survey. A shortened questionnaire is then administered to a much larger set of households. In conjunction with the statistical model, data from the shortened questionnaire are finally used to derive scores for the identification of beneficiaries. Whether this strategy is suitable for the identification of the deserving poor is largely an empirical question that depends on the underlying model, the quality of the available data, the joint distribution of the target measure and the proxies, and—as will be demonstrated in the present paper—the proportions of the population in poverty and targeted. Coady et al. (2004) provide a meta-study comparing PMTs and alternative targeting strategies in the developing country-context in terms of targeting accuracy. Their findings indicate that while PMTs perform better than alternatives on average, targeting accuracy varies considerably.

2.2 Errors of targeting

All of the above methods result in two *errors of targeting*: non-poor households are erroneously included in the program and poor households are excluded. Two widely used indicators of the accuracy of targeting are the proportion of type I and type II errors. Table 1, a classification matrix, will help us to organize the discussion. n_1 , n_2 , n_3 , and n_4 denote true negatives, false

TABLE 1: Classification matrix: taxonomy of targeting errors

Poverty Status	Beneficiary Status		Total
	Non-beneficiary	Beneficiary	
Non-poor	n_1	n_2	N^{np}
Poor	n_3	n_4	N^p
Total	N^{nb}	N^b	N

positives, false negatives, and true positives, respectively. The *true positive rate* (TPR) and the *false positive rate* (FPR) are defined as the share among the poor that were (correctly) included in the program and the share among the non-poor that were erroneously included:

$$TPR = n_4/N^p \quad \text{and} \quad FPR = n_2/N^{np}.$$

These are a common staple in both *ex ante* and *ex post* evaluations of imperfectly targeted programs and often feature prominently in discussions about the effectiveness of programs (see, for instance, Cornia and Stewart, 1995).

It is often assumed that a good targeting mechanism should result in high coverage among the poor, i.e. a high TPR , and low leakage to the non-poor, i.e. a low FPR . Broadly targeted programs will result in high coverage among the poor yet also considerable leakage to the non-poor while narrowly targeted programs will result in limited leakage yet also low coverage among the poor. Denote the beneficiary share, the proportion of the population that receives the program, by $r = N^b/N$. It is easy to see that both $TPR \in [0, 1]$ and $FPR \in [0, 1]$ and that $TPR, FPR \rightarrow 1$ as $r \rightarrow 1$ and $TPR, FPR \rightarrow 0$ as $r \rightarrow 0$. Hence, the TPR and the FPR can be used to describe the trade-off between coverage and leakage.²

2.3 ROC-curves

A useful tool to illustrate this trade-off further are *Receiver Operating Characteristic* (ROC)-curves, plots of the TPR against the FPR (Thompson and Zucchini, 1989; Wodon, 1997; Johannsen, 2008; Landau et al., 2012). Analytically, we can think of the ROC-curve $g_{\mathcal{X}}$ associated with a particular classifier \mathcal{X} as a function that maps from the space of FPR s to the space of

²Alternative targeting errors are the *undercoverage rate*, the proportion of poor individuals that are not included in the program, and the *leakage rate*, the proportion of beneficiaries that are non-poor. These error rates are also sometimes referred to as exclusion and inclusion errors, respectively, and are given by

$$U = n_3/N^p \quad \text{and} \quad L = n_2/N^b.$$

It is easy to see that $U = 1 - TPR$ while L can be expressed as a function of TPR , FPR , and the headcount ratio H_0 : writing $N^b = n_2 + n_4$ and expanding yields

$$L = FPR(1 - H_0)/(FPR(1 - H_0) + TPR \cdot H_0).$$

Hence, the informational content is the same here as with the TPR and the FPR . However, one difference is that the domain of L depends on H_0 : as $r \rightarrow 1$, $FPR \rightarrow 1 - H_0$. It is thus less clear how to aggregate over both errors in order to arrive at a composite measure such as those discussed in the next section.

TPRs: $g_X : [0, 1] \rightarrow [0, 1]$. This function describes the ‘efficient’ *TPR-FPR*-combinations that the classifier allows for. As explained above, both the *TPR* and *FPR* will increase as more people are admitted into the program. Hence,

$$\frac{dg_X(FPR)}{dFPR} = \frac{dTPR}{dFPR} > 0. \quad (2.1)$$

If the classifier results in an improvement over random targeting, that is, if it conveys some information about who is poor and who is not, the *TPR* will increase at a faster rate initially. This gives rise to a concave ROC-curve. The better the classifier at reproducing the original welfare ranking, the more the curve will bend towards the point $(0, 1)$. With perfect targeting, a classifier that results in a rank correlation between predicted welfare and actual welfare of unity, only the *TPR* would increase initially until all the poor were covered and $TPR = 1$. The *FPR*, on the other hand, would be zero over this range. Only when all poor are beneficiaries will the *FPR* increase until ultimately all households are covered and both *TPR* and *FPR* are unity. A random classifier, on the other hand, would see both *TPR* and *FPR* increase at the same rate and the expected ROC-curve would therefore be the 45 degree line. Hence, the ratio of the two would be unity in expectation. The farther the ROC-curve ‘bends’ towards the top left corner of the plot, the more accurate the targeting.

2.4 Targeting measures

It is fairly common to aggregate both the *TPR* and the *FPR* in order to arrive at an aggregate measure of targeting performance. In fact, given the trade-off described above, it is necessary to do so whenever targeting rates are used in order to decide how narrowly programs are to be targeted. We show here first that two commonly used targeting measures can be interpreted as welfare functions defined over *TPRs* and *FPRs*. Therefore, ROC-analysis provides a useful unifying framework in this respect.

The two measures we consider are the *targeting differential* (*TD*) and the *total error rate* (*TER*). The former is defined as simply the difference between the *TPR* and the *FPR*:

$$TD = TPR - FPR. \quad (2.2)$$

Hence, $TD \in [-1, 1]$ and a greater value would indicate better targeting. This measure has been proposed by Galasso and Ravallion (2005) and has recently been found to be *the only* targeting measure to predict the poverty effect of a cash transfer program in China by (Ravallion, 2009b).

The *TER*, on the other hand, is defined as the number of individuals wrongly classified either way divided by the total population:

$$TER = (n_2 + n_3)/N. \quad (2.3)$$

Note that $TER \in [0, 1]$ and that a lower value suggests better targeting. The *TER* is unity if

every individual is wrongly classified and zero if everyone is correctly identified.

2.5 Targeting measures as objective functions

As in [Wodon \(1997\)](#), assume the society's (or a policy-maker's) objective can be described as maximizing a targeting measure defined over targeting rates subject to the constraint imposed by a particular classifier \mathcal{X} . (We will show presently that both (2.2) and (2.3) can be interpreted in this way.) In general terms, assume that $W = W(TPR, FPR)$ with $\partial W / \partial TPR > 0$ and $\partial W / \partial FPR < 0$. The general problem can then be written as

$$\text{Max}_r \quad W(TPR, FPR) \quad \text{subject to} \quad g_{\mathcal{X}}(FPR) = TPR.$$

Plugging the constraint into the social welfare function for TPR and taking the first derivative yields the first order-condition (FOC) for a maximum as

$$\frac{\partial W}{\partial TPR} \frac{dg_{\mathcal{X}}}{dFPR} \frac{dFPR}{dr} + \frac{\partial W}{\partial FPR} \frac{dFPR}{dr} = 0.$$

Using (2.1), we obtain

$$\frac{dTPR}{dFPR} = -\frac{\partial W / \partial FPR}{\partial W / \partial TPR}.$$

This expression shows that ROC-curves are a useful tool to rationalize the implications of using targeting measures as social welfare functions: the left hand-side of this equation is the slope of the ROC-curve while the right hand-side is the marginal rate of substitution between the FPR and the TPR . Note that $\partial W / \partial FPR < 0$ such that both the right hand-side and the left hand-side are positive. The optimal beneficiary share r^* is found by equalizing the two expressions.

In the case of the targeting differential (2.2), a welfare function can directly be written as $W_{TD}(TPR, FPR) = TPR - FPR$. The problem is then

$$\text{Max}_r \quad TPR - FPR \quad \text{s.t.} \quad g_{\mathcal{X}}(FPR) = TPR \tag{2.4}$$

yielding FOC

$$\frac{dTPR}{dFPR} = 1.$$

If a policy-maker aims to maximize TD , she will settle on the point of the ROC-curve in which the slope is unity.

The TER in (2.3) can equally serve as an objective criterion. Note that the problem of minimizing the TER subject to the constraint imposed by the classifier can be recast as the maximization problem

$$\text{Max}_r \quad H_0 TPR - (1 - H_0) FPR \quad \text{s.t.} \quad g_{\mathcal{X}}(FPR) = TPR. \tag{2.5}$$

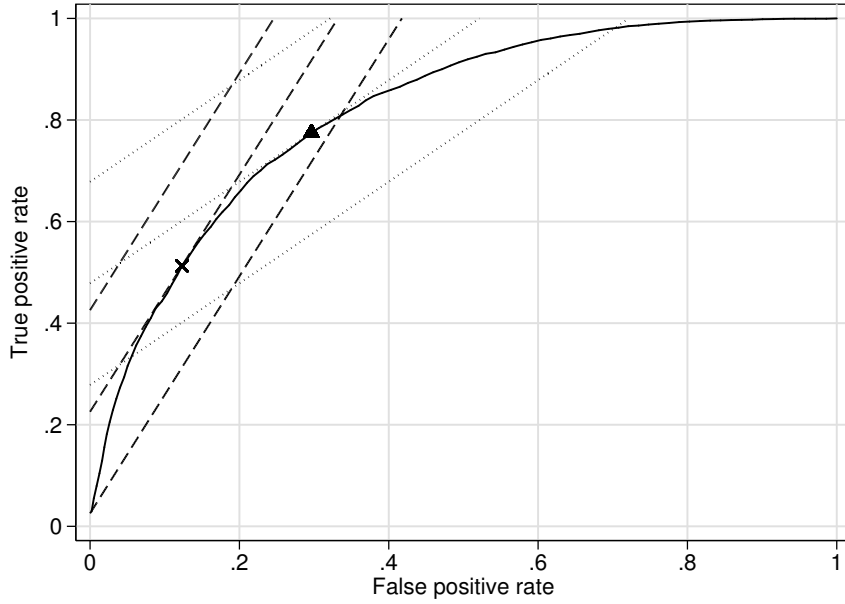


FIGURE 1: ROC-curve and indifference curves based on data for Indonesia and PMT M1 (see next section) and a poverty headcount of 30 percent. The solid line represent the viable TPR - FPR -combinations subject to the classifier used. The dashed (dotted) lines indicate indifference curves pertaining to the use of TER (TD) as a welfare indicator—with social welfare increasing in the TPR and decreasing in the FPR —and the cross (triangle) indicate the optimal solution.

Hence, we can think of policy-makers that are only concerned about the TER as maximizing a social welfare function that is a linear combination of targeting rates—only that this time she attaches weights of H_0 and $-(1 - H_0)$ to the TPR and the FPR , respectively. The respective FOC is

$$\frac{dTPR}{dFPR} = \frac{1 - H_0}{H_0}.$$

If $H_0 < 1 - H_0$, as is often the case in practice, she attaches greater weight to the FPR , that is, leakage receives more attention. She will settle for a point in which the slope of the ROC-curve is greater and the optimal value of r will be smaller; the resulting program would be more narrowly targeted. This is illustrated in figure 1.

Note that other targeting measures frequently used in the literature do not lend themselves to be analyzed in this framework but that this is because *they do not lend themselves to deciding on the optimal beneficiary share*. A large class of targeting measures in this category are based on *concentration curves* that consider the incidence of benefits over the initial distribution of welfare (e.g. Coady et al., 2004; Fujii, 2008; Ravallion, 2009b) such as the share of benefits that goes to the poorest x percent of the population. Various normalizations are discussed by Coady et al. (2004) and Ravallion (2009b). However, while these may be useful in *ex ante*-evaluations

TABLE 2: Results from Google Scholar-search of various terms related to poverty targeting.

Search terms	Time period	Results
targeting AND program AND (transfer OR subsidy OR in-kind) AND ‘coverage’ AND ‘leakage’	1990–2016	16,900
targeting AND program AND (transfer OR subsidy OR in-kind) AND ‘poverty impact’	1990–2016	3,360
targeting AND program AND (transfer OR subsidy OR in-kind) AND ‘poverty effect’	1990–2016	478

The search was conducted using Google Scholar on 7/11/2016.

when it is difficult to judge the poverty effects of a program, they cannot be used to decide how narrowly programs should be targeted as they would always suggest that only the poorest should be targeted irrespective of the value of x . As they are invariant to the size of the transfer, we would end up with a program in which only the poorest household would be targeted and would receive the entire budget as a transfer, a strategy that cannot have much of an effect on poverty and would be impractical in real life.

2.6 On the appropriateness of welfare functions defined over targeting rates

Should we think of policy-makers as maximizing a function that is defined over targeting rates? Or should we assume that the success of a policy is measured by its contribution to poverty reduction? The latter strategy, which we explore more fully in section 5 below, assumes that policy-makers design programs so that they have the maximum effect on poverty subject to a budget constraint *and* the targeting mechanism at hand. While this seems more appropriate at least from a normative perspective, there are two arguments in favor of using the approach based on targeting measures (see [Wodon, 1997](#), for a related discussion).

First, estimating the effect of a targeted program on poverty *ex ante* is bereft with problems. We will come back to this point in more detail in section 5. The main challenges include anticipation of behavioral responses (e.g. [Datt and Ravallion, 1994](#)), ‘hidden costs’ associated with targeting ([van de Walle, 1998](#)), and politico-economic feedbacks ([De Donder and Hindriks, 1998](#); [Gelbach and Pritchett, 2001, 2002](#); [Moene and Wallerstein, 2001](#)).

Second, errors of targeting feature prominently in both academic debates as well as in policy discussions. For instance, an all-text Google Scholar-search for works published since 1990 that contain both the terms ‘coverage’ *and* ‘leakage’ as well as other relevant terms associated with targeting results in about five times more findings than the same search that replaces these terms with ‘poverty impacts’ (see table 2).³ And media outlets frequently run stories about welfare fraud and wastage of public funds devoted to poverty alleviation programs that are met with

³For all three trials we found that the first ten entries were highly relevant to the subject of the present paper.

TABLE 3: Poverty measures of the FGT-class (for poverty lines set at the 10th and 40th percentile of the distribution of real per capita expenditure) and the Gini coefficient.

	$H_0 = 10$		$H_0 = 40$		Gini coefficient
	$FGT(1)$	$FGT(2)$	$FGT(1)$	$FGT(2)$	
Bolivia	0.031	0.014	0.168	0.094	0.550
Indonesia	0.016	0.004	0.105	0.038	0.379

much public attention.

3 Data and regression-based PMTs

In this section we introduce the datasets, explain our choice of welfare indicator and the construction of PMTs, and present regression results to illustrate our approach with data from two developing countries.

3.1 Datasets

The datasets used in this part of the study are the 2011 Bolivian *Encuesta de Hogares* (EH) and the 2012 Indonesian *National Socioeconomic Survey* (Susenas). Both datasets are multi-purpose household surveys that include expenditure modules and information on asset ownership and are representative at the national level. The data are in both cases obtained by two-stage sampling with stratification. See [INE \(2015\)](#) for a report on the 2011 EH and [Surbakti \(1995\)](#) for an overview over the rich history of the Susenas.

3.2 Welfare indicator

In line with much of the literature, we assume that the relevant welfare indicator is real household expenditure per household member. While this is often preferred to readily available alternatives (such as per capita income) in developing country-settings (see [Deaton, 1997](#); [Skoufias and Coady, 2007](#)), it is clearly subject to debate. However, such considerations are beyond the scope of the present paper.⁴

Table 3 reports scale-invariant summary statistics of the distribution of real expenditure per capita in these two countries: the Gini coefficient and various poverty measures of the class of poverty measures proposed by [Foster et al. \(1984\)](#) at generic poverty lines set at the 10th and the 40th percentile of the distribution. Individual-level inflation factors, defined as the inverse of

⁴There are some differences in how the expenditure aggregate is calculated between the Bolivian and the Indonesian data. While the expenditure aggregate that comes from the Indonesian Susenas simply adds total food and non-food expenditure including imputed rents, the Bolivian data also include information about the ownership of durables, estimated values, and the year in which these goods were purchased. This allows us to also estimate service flows from durable goods which are included in the expenditure aggregate.

the sampling probability of each individual, are used throughout. Real expenditure per person is calculated by deflating expenditure by the poverty lines given in the dataset⁵ and the number of household members.⁶ As one would expect, the results suggest that inequality and both the depth of poverty and the extent of inequality among the poor is much more pronounced in Bolivia. The datasets we analyze thus provide us with the opportunity to study the consequences of targeting based on alternative targeting measures in two very distinct settings.

3.3 Proxy-means tests

We investigate the targeting performance of two different imperfect classifiers based on generic, regression-based proxy-means tests (PMTs) for each country. These differ in approach and complexity. Our choice of proxies for the baseline models, which we will refer to as ‘M1’ in what follows, is guided by concerns over verifiability,⁷ incentive-compatibility, and legality. The more ‘sophisticated’ models, which we will refer to as ‘M2’, include variables capturing asset ownership which is probably more difficult to verify and may be less incentive-compatible.

All our models include variables that capture households’ demographic make-up. In the Bolivian PMTs we include total household size as well as the number of individuals in a total of ten gender-age cells. We group together children between zero and five, six and eleven, and twelve and 14 as well as adults between 15 and 59 and above 60. The selection of bins is guided in this case by official school entrance ages—six for primary and twelve for lower secondary—as well as eligibility for Bolivia’s *Renta Dignidad*, a social pension scheme, as we expect these to be correlated with expenditure. In our Indonesian PMTs we opted for more gender-age cells that are motivated by similar considerations. We group together children aged zero to four, five to twelve, 13 to 18 as well as adults aged 19 to 24, 25 to 54, 55 to 64, and 65 and above. We decided to use more fine-grained age bins for Indonesia as the calibration sample is much larger. There is thus less potential for an ‘overfitting’ of the data.

We also always include variables that capture characteristics of the dwelling and service use, including type of floor, wall, and roof material. For our Bolivian models we further include the number of rooms, dummies indicating whether the dwelling’s kitchen is located in a separate room, a separate water connection, and two exclusive dummies indicating in-house or shared toilet. Two further dummies indicate access to waste removal-services and electricity. We believe that these variables are easy to verify during a personal visit to the household. Information on service use may in addition be available from official records. The Indonesian models include information about access to water, lighting, the type of cooking fuel, as well as the availability and type of toilet.

⁵We normalize the poverty lines such that the weighted average across households is unity.

⁶We do not consider the thorny issue of defining equivalence scales that adjust for public goods at the household-level and differences in needs across different demographic groups. While this is potentially an important task in practice (e.g. Drèze and Srinivasan, 1997), there is very little theoretical guidance on these questions (Deaton, 1997, provides a discussion).

⁷Niehaus et al. (2013) study the trade-off between statistical accuracy and enforceability.

While this covers all variables employed in our baseline models, we also employ information on asset ownership in our more ‘sophisticated’ models. These will be employed in a robustness check to our poverty simulations. The asset list for Bolivia includes refrigerators, personal computers, TV sets, microwaves, washing machines, air conditioners, heaters, cars, landline phones, and cell phones which are all coded as binary variables. The asset list for Indonesia is similar: it includes refrigerators, personal computers, TV sets, bicycles, motorcycles, cars, boats, air conditioners, kettles, and phones.

We also include dummies for departments in Bolivia and provinces in Indonesia and interactions of these with a variable indicating rural location in order to capture geographic differences in economic welfare. As the signal of any of the above variables about economic welfare may depend on the locality, we further include interactions of all of the above variables with rural location.

3.4 Sampling and cross validation

We employ Ordinary Least Squares (OLS)-regressions of log real expenditure against proxies in order to evaluate the usefulness of PMTs in the present data. In order to mimic a real-life PMT exercise and to avoid overfitting the data, all investigations are carried out using out-of-sample predictions. In particular, we test the predictive power of models trained using a subsample of the the two datasets as *calibration samples*.⁸ Usual households surveys conducted in Bolivia frequently cover roughly 4,000 households while the 2011 EH covers 8,851 households. We will therefore randomly set aside about 50 percent of the data for calibration. The remaining records are then employed to analyze the performance of PMTs, that is, they serve as the *validation sample*.⁹

The Susenas frequently cover many more households reflecting the fact that Indonesia’s population is much larger and, by almost all standards, more diverse. For instance, the 2012 Susenas provides records for about 286,086 households. We sample a small fraction, ten percent, for validation purposes so that the difference between the calibration sample and a regular Susenas dataset are negligible.

In sampling households for calibration we take into account the original design of the surveys. Households were sampled in a two-stage procedure. In a first step, *primary sampling units* (PSUs) were randomly selected. In a second step, a fixed number of households within PSUs were sampled. We thus also randomly sample PSUs in order to obtain calibration samples. We also ensure that the stratification carries over to the subsamples. In the case of the EH,

⁸In real-life PMT exercises, a household survey is usually use for calibration and a shortened questionnaire is administered to all potential beneficiaries. Hence, the calibration sample is usually somewhat dated compared to the data collected with the shortened instrument and this constitutes an additional source of error. While potentially important, the question of what effects such time lags has on the accuracy of PMTs is beyond the scope of this paper.

⁹While this is not exactly what would be done in practice where households in the calibration sample may also be beneficiaries, we imagine that the number of households in the calibration sample is small in practice compared to the households that potentially qualify.

TABLE 4: Cross-validated regression statistics for log per capita real expenditure.

Model	# of parameters	In-sample R-squared	Out-of-sample R-squared
Bolivia, M1	143.2	0.81 (0.80, 0.81)	0.75 (0.65, 0.78)
Bolivia, M2	162.6	0.85 (0.84, 0.85)	0.81 (0.76, 0.82)
Indonesia, M1	336.4	0.47 (0.47, 0.47)	0.47 (0.45, 0.50)
Indonesia, M2	356.4	0.54 (0.53, 0.54)	0.53 (0.51, 0.56)

Based on weighted OLS regressions using individual inflation factors. 95%-confidence intervals (based on 20 sample splits) in parentheses.

stratification was based on urban-rural location and an indicator of basic needs constructed from the 2001 census, the data from which the sample frame derives (INE, 2015). The Susenas is also stratified based the rural/urban distinction. Hence, we ensure that the proportion of households in rural areas from the sample carries over to the subsamples by sampling individually PSUs within each strata. The Susenas are collected in four waves within one year so we also ensured that the proportion of households interviewed at each of the four points in time over a given year, one-fourth, approximately carries over to the calibration sample. In order to obtain confidence intervals in the ensuing analysis, we rely on 20 random sample splits that adhere to the above conditions.

Finally, an important question when dealing with complex survey data is whether or not weights should be used. Since we are ultimately interested in poverty and targeting rates at the individual-level, we use individual inflation factors in what follows. These are defined as household inflation factors (e.g. Deaton, 1997) multiplied with the number of household members.

3.5 Regression results

Table 4 reports results from the above regression models. We report only indicators of the regression fit, namely, the number of parameters, R-squared-statistics from regressions based on calibration samples, and, for comparison, the squared Pearson correlation coefficient between predicted values (or *PMT scores*, in the terminology of this method) and actual values in the validation sample. Note that the former is actually the squared correlation coefficient between predicted values and actual values of the dependent for the calibration samples. We therefore refer to these quantities as the *in-sample* and *out-of-sample R-squareds*. Again, all quantities are estimated using individual inflation factors as weights and confidence bands on correlation coefficients are derived from repeated cross-validation.

Several observations can be made from table 4: first, we find that regression models result in a much better fit and much higher out-of-sample R-squareds for Bolivia. This could be due to the fact that inequality is higher in Bolivia and that it is therefore easier to tell the poor apart from the rich based on proxies. This explanation would be in line with Coady et al. (2004) who report that, conditional on per capita GDP and accountability, countries with higher inequality

tend to run programs that are better targeted.¹⁰ When we turn to our simulations below, it will be important to keep in mind that similar models result in a much larger fraction of real expenditure explained in Bolivia.

Second, as one would certainly expect, we find that in-sample R-squared-statistics are higher than their out-of-sample counterparts, particularly for Bolivia, and confidence intervals are somewhat wider for the latter. Finally, we find that increasing the complexity of the models results in modest improvements. While this is not the main focus of the present study, it seems that a better out-of-sample correlation could be had with more interactions and, possibly, additional regressors that we did not include so far, and this seems particularly likely for Indonesia.

4 How narrowly should programs be targeted?

In this section and the following we explore the usefulness and implications of using various targeting measures in deciding on how narrow programs should be targeted. In effect, we assume that policy-makers have specific preferences defined over targeting errors as in section 2 rather than over welfare or poverty indicators. In the following section we then explore to what extent simulated effects on poverty coincide with ‘optimal’ beneficiary shares derived from optimizing targeting measures.

We start by investigating the distribution of targeting errors that result from maximizing the above-described targeting measures. To do so, we incrementally change the percentile of the population targeted for each country and validation sample separately and back out error rates resulting from using PMTs to rank households. This results in one *TPR-FPR*-combination for each percentile and this, in turn, can be used to calculate the resulting targeting measures. As explained above, ROC-curves are one way of illustrating the trade-off between coverage among the poor and leakage to the non-poor. Figure 2 plots *TPRs* against *FPRs* that are both averaged by percentile over 20 random sample splits. We draw these curves for situations in which ten, 20, 30, and 40 percent of the population are deemed poor, respectively. We also indicate resulting combinations of *TPRs* and *FPRs* when one of three strategies is used in order to determine the extent to which a hypothetical is targeted: (1) maximizing *TD* as in (2.4), (2) minimizing *TER* (or, equivalently, solving (2.5)), and (3) simply setting the targeting rate equal to the initial poverty headcount. Resulting ‘optimal’ targeting and error rates are also reported in table 5.

Both figure 2 and panel A of table 5 bring out the differences in targeting rates that would be considered ideal based on alternative targeting strategies. Recall that the relative weight on the *TPR* in the objective function in (2.4) and 2.5 are unity and $H_0/(1-H_0)$, respectively. Since the latter is less than unity whenever $H_0 < 0$, targeting should be more broad in order to maximize *TD* and less broad in order to minimize *TER*. This is indeed what we see in figure 2 and in

¹⁰As it stands, it cannot be ruled out that this result is driven by differences in the constructions of the expenditure aggregate which, in the case of Bolivia, includes service flows from durables (rather than spending on durables) and excludes what is likely transitory expenditure (e.g. hospitalization fees). Note that there is no point in trying to harmonize the expenditure aggregates as the underlying surveys are different. For instance, the Susenas collects much more detailed data on food items consumed.

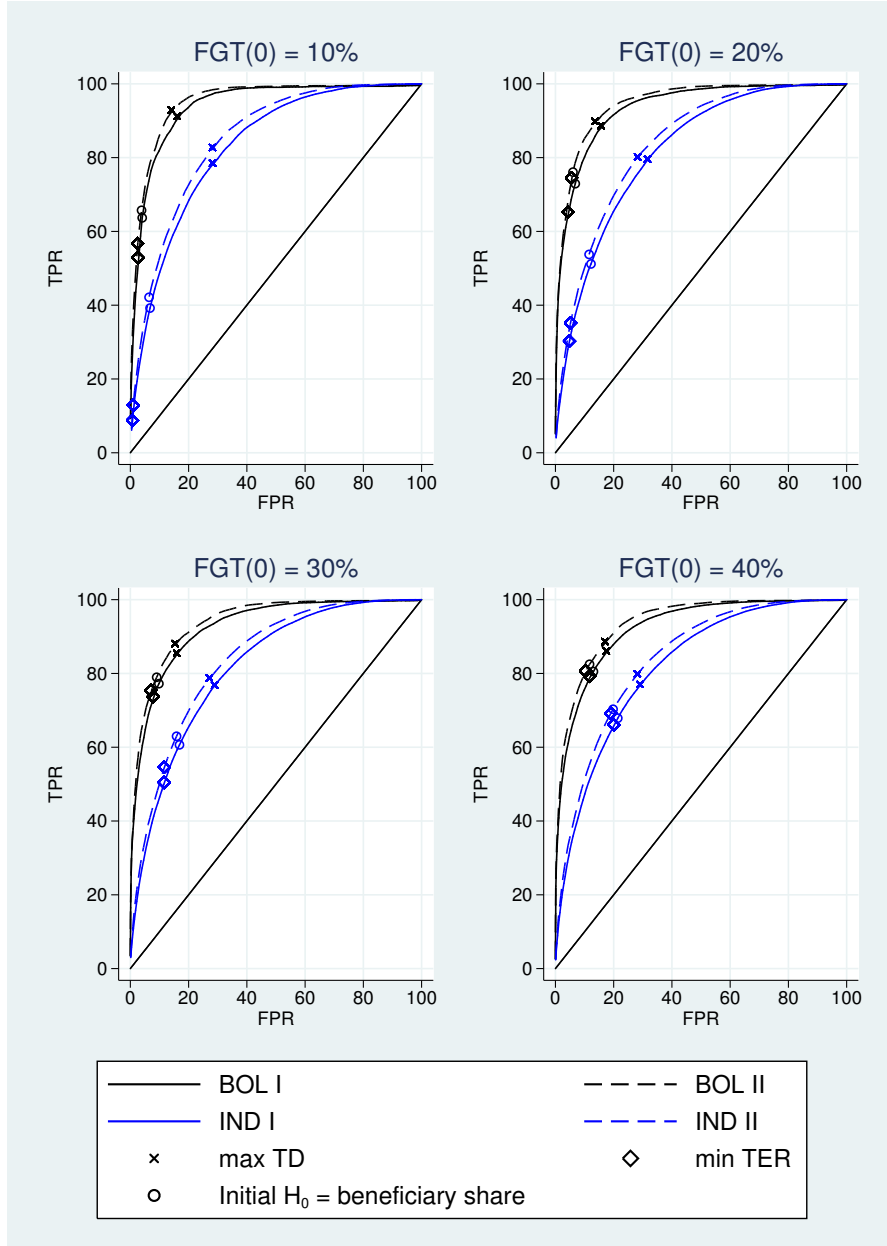


FIGURE 2: ROC-curves for both PMTs and four poverty rates of ten, 20, 30 and 40 percent. The black (blue) lines represent ROC-curves obtained by averaging TPRs and FPRs over 20 random sample splits for Bolivia (Indonesia). The solid (dashed) lines indicate use of PMTs M1 (M2). We also indicate in the graphs the points at which TD and TER are optimized and the point at which the participation rate equals the poverty rate. We omit confidence bands as these turned out to be very narrow.

TABLE 5: Optimal beneficiary share, $TPRs$, and $FPRs$ by country, initial headcount, and PMT.

		min TER		max TD		$H_0 = r$	
	H_0 (%)	M1	M2	M1	M2	M1	M2
<i>Panel A: Optimal beneficiary share (in %)</i>							
Bolivia	10	7.4 (1.8)	7.8 (1.5)	23.6 (2.6)	22.0 (2.0)	10.0 (0.0)	10.0 (0.0)
	20	16.2 (2.1)	19.2 (2.1)	30.2 (2.1)	28.9 (3.2)	20.0 (0.0)	20.0 (0.0)
	30	27.5 (1.9)	27.5 (2.0)	36.8 (2.9)	37.2 (1.8)	30.0 (0.0)	30.0 (0.0)
	40	38.7 (2.0)	38.4 (2.1)	44.9 (2.8)	45.6 (2.8)	40.0 (0.0)	40.0 (0.0)
Indonesia	10	1.5 (0.7)	2.2 (0.8)	33.2 (3.2)	33.7 (2.1)	10.0 (0.0)	10.0 (0.0)
	20	9.9 (0.9)	11.1 (1.1)	41.2 (2.0)	38.6 (1.3)	20.0 (0.0)	20.0 (0.0)
	30	23.2 (2.2)	24.5 (1.8)	43.3 (1.3)	42.6 (2.2)	30.0 (0.0)	30.0 (0.0)
	40	38.5 (2.8)	39.0 (1.2)	48.3 (2.0)	48.8 (2.0)	40.0 (0.0)	40.0 (0.0)
<i>Panel B: True positive rate (in %)</i>							
Bolivia	10	52.1 (9.6)	56.7 (8.1)	91.2 (2.6)	92.9 (2.6)	63.7 (2.0)	65.7 (1.8)
	20	64.6 (6.1)	74.5 (5.6)	88.7 (2.1)	89.9 (3.0)	72.9 (1.4)	76.1 (1.1)
	30	73.7 (3.0)	75.4 (3.4)	85.5 (3.0)	88.1 (1.7)	77.2 (0.8)	79.0 (0.9)
	40	79.3 (2.7)	80.7 (2.8)	86.1 (2.3)	88.7 (2.8)	80.5 (0.9)	82.5 (0.7)
Indonesia	10	8.8 (3.8)	12.9 (4.4)	78.5 (3.4)	82.8 (1.9)	39.2 (1.2)	42.2 (1.2)
	20	30.3 (2.4)	35.2 (2.9)	79.6 (2.0)	80.2 (1.5)	51.2 (0.6)	53.8 (0.8)
	30	50.5 (3.6)	54.6 (3.0)	76.9 (1.2)	78.7 (2.3)	60.7 (0.6)	63.0 (0.4)
	40	66.1 (3.6)	69.2 (1.6)	77.1 (1.9)	79.9 (1.9)	67.9 (0.5)	70.3 (0.4)
<i>Panel C: False positive rate (in %)</i>							
Bolivia	10	2.5 (0.9)	2.4 (0.8)	16.1 (2.6)	14.1 (2.0)	4.1 (0.2)	3.8 (0.2)
	20	4.1 (1.2)	5.5 (1.2)	15.7 (2.2)	13.6 (3.3)	6.8 (0.4)	6.0 (0.3)
	30	7.8 (1.4)	7.0 (1.5)	15.9 (3.0)	15.4 (1.9)	9.8 (0.3)	9.1 (0.4)
	40	11.7 (1.7)	10.3 (1.7)	17.5 (3.1)	17.0 (3.0)	13.1 (0.6)	11.7 (0.5)
Indonesia	10	0.7 (0.3)	1.0 (0.4)	28.2 (3.2)	28.2 (2.1)	6.8 (0.1)	6.4 (0.1)
	20	4.8 (0.5)	5.1 (0.7)	31.6 (2.0)	28.2 (1.3)	12.2 (0.2)	11.6 (0.2)
	30	11.6 (1.6)	11.6 (1.3)	28.9 (1.3)	27.1 (2.2)	16.9 (0.2)	15.9 (0.2)
	40	20.0 (2.3)	19.0 (1.0)	29.1 (2.2)	28.1 (2.2)	21.4 (0.3)	19.8 (0.3)

Means and associated standard deviations (in parentheses) based on 20-fold cross validation. Results that indicate that the beneficiary share should be greater than the initial poverty headcount are highlighted.

panel A of table 5. In fact, our results suggest that the number of beneficiaries should always exceed the number of poor when the aim is to maximize TD and be lower than the number of poor when the aim is to minimize TER . Since the relative weight the TER attaches to the TPR is increasing in H_0 , the differences is more pronounced for low initial poverty rates. Both criteria would converge in terms of the optimal beneficiary share for $H_0 = 0.5$. This translates into low $FPRs$ and low $TPRs$ if policy-makers minimize the TER and high $FPRs$ and high $TPRs$ if they maximize TD with the third strategy providing a compromise between the two.

Our results here also suggest that the use of alternative PMTs will improve targeting. Figure 2 indicates that a higher TPR can be had for every FPR if one moves from the more parsimonious PMT to the more sophisticated one. This is also evident from panels B and C of table 5: conditional on country and initial poverty headcount, $TPRs$ tend to be higher and $FPRs$ lower as one moves from M1 to M2. However, these differences are hardly very important.

Indeed, the more important differences are between countries as shown in both the ROC-

curves as well as table 5. *TPRs* are much higher and *FPRs* much lower in Bolivia and this often tends to be more pronounced when the poverty rate is low. This mirrors our findings in the previous section that indicated that it is much easier to distinguish the poor from the non-poor in Bolivia: since the set of variables we use is very much comparable and the out-of-sample R-squared is considerably higher in Bolivia (section 3), these differences are like driven by greater inequality and, hence, a stronger signal-to-noise ratio in the Bolivian data.

One interesting exception occurs when targeting is based on the *TER* and the initial poverty headcount is ten percent. Under these circumstances, the slope of the ROC-curve for Indonesia declines sufficiently fast and the weight on *TPRs* is sufficiently low to result in very narrow targeting of only one to three percent of the population in Indonesia. This, in turn, results in an *FPR* that is lower than the one for Bolivia in the same situation and under the same targeting strategy. But it achieves this by running a tiny program—only one percent of the population would be included—that is likely to have hardly any poverty effect, a subject to which we now turn.

5 Poverty simulations

5.1 Empirical strategy and set-up

While discussions about the extent of targeting often evolve around errors of targeting, ultimately, it is unclear what this implies for effects on measures of poverty.¹¹ In this section we run poverty simulations for transfer schemes and gauge to what extent beneficiary shares that result from optimizing targeting measures—the targeting differential and the total error rate—will coincide with optimal impacts on poverty. We will also compare these strategies to a policy-maker who simply targets the same proportion that is also poor.

The results will clearly depend on our assumptions about budgets. In terms of poverty effects, larger budgets will tend to favor broad targeting and, keeping the poverty headcount fixed, broad targeting will also have a greater effect on poverty if the available budget increases with the share of beneficiaries. Below, we make alternative assumptions about the size of the budget and, in addition, about the relationship between the budget and the beneficiary share.

5.1.1 Poverty measures

We will assume that policy-makers are concerned with the impact that targeted transfers have on poverty measures of the Foster-Greer-Thorbecke-class (Foster et al., 1984). These can be written as

$$FGT(\alpha) = \frac{1}{N} \sum_{y_i < z} \left(\frac{z - y_i}{z} \right)^\alpha,$$

¹¹Poverty measures attach a zero weight to levels of living above the poverty line and are thus sometimes interpreted as *exclusive* measures of social welfare (e.g. Ravallion, 1994).

where N denotes total population, y_i denotes real per capita expenditure, z denotes the poverty line in the space of real per capita expenditure, and α is the ‘poverty aversion parameter’, the weight attached to the severity of poverty. We obtain the poverty headcount, the poverty gap index, and the squared poverty gap (or ‘poverty severity’) index for $\alpha = 0$, $\alpha = 1$, and $\alpha = 2$, respectively. While used widely, both the poverty headcount and the poverty gap index have severe shortcomings in that they violate Dalton’s (1920) transfer principle that states that a small transfer from a rich to a poor person should always result in less poverty. We therefore focus on the squared poverty gap index in our simulations. Below, we also report results from a robustness check that treats the poverty gap index as the indicator of interest.

5.1.2 Budget and transfers

A common assumption in poverty simulations is that the budget is fixed in which case halving the number of beneficiaries will result in a doubling of the available per capita transfer; put differently, the elasticity of the budget with respect to the beneficiary share is zero. This is equivalent to assuming that the elasticity of average per capita transfers among beneficiary with respect to the beneficiary share is -1 . This assumption ignores key findings from the literature on the political economy of targeting that suggest that increasingly narrow programs will lack political support and will end up underfunded (De Donder and Hindriks, 1998; Gelbach and Pritchett, 2001, 2002; Moene and Wallerstein, 2001). These theoretical studies suggests that the elasticity will be somewhere between -1 and zero: as more voters benefit from the program, the politically viable budget will increase yet this will not completely compensate the diluting effect of broadening programs. In appendix A, we obtain plausible values for the latter elasticity by analyzing data from China’s urban *Dibao*-transfer program. While there are some caveats, our findings suggest that an elasticity as high as $-1/3$ is consistent with the data.

We therefore vary the elasticity and the budget size systematically: we assume that the elasticity of average per capita transfers with respect to the beneficiary share, which we will denote by β , is either -1 or $-1/3$, *viz.* the first assumption will result in a fixed budget while the latter will result in a total budget that is an increasing function of the beneficiary share.

We also investigate the sensitivity of our results to variation in the size of the budget: we assume that, irrespective of β , the budget available for a program that includes as many individuals as are poor is a fraction s of the aggregate poverty shortfall, i.e. the sum over the shortfalls from the poverty line over all household.¹² We assume that s is either $1/3$ or 1 .¹³ The budget b as a function of the participation rate r , the budget fraction s , and the elasticity β can thus be written as

$$b(r, s, \beta) = s \times \left\{ \sum_{y_i < z} (z - y_i) \right\} \times \exp\{(1 + \beta) \times \ln(r/H_0)\},$$

¹²This is done for both countries and all validation samples separately.

¹³Budgets for single programs that are greater than the aggregate shortfall from the poverty line are hardly plausible and rare in practice. Skoufias and Coady (2007) use a budget of roughly one-half of the aggregate gap in their simulations while Grosh and Baker (1995) fix the budget at about one percent of total expenditure in their samples.

where H_0 denotes the poverty headcount ratio. It is easy to verify that β is indeed the elasticity of the average transfer among beneficiaries with respect to the beneficiary share and that $b = s \times \sum_{y_i < z} (z - y_i)$ for $H_0 = r$.

A simplifying assumption we make is that the transfers received do not vary across beneficiaries, i.e. conditional on the beneficiary share, per capita transfers are fixed. While this is quite common in the literature (e.g. [Grosh and Baker, 1995](#); [Skoufias and Coady, 2007](#)) and in practice, it is not in general efficient if the goal is to minimize poverty (and if behavioral responses are assumed away as in our simulations). There have been some attempts in the past to provide poverty-minimizing algorithms that allow for variation in transfers across beneficiaries ([Ravallion and Chao, 1989](#); [Glewwe, 1990](#)). Moreover, programs that aim to establish a minimum livelihood guarantee are also sometimes found in practice, particularly in developed countries ([Chen et al., 2008](#); [Umapathi et al., 2013](#)). Note that in what follows, ‘poverty-minimizing’ or ‘optimal’ targeting (or beneficiary share) will refer to the extent of targeting that will minimize poverty *conditional on the restriction of constant per capita transfers across beneficiaries*.

We illustrate the implications of our assumptions for budgets and transfers in figure 3 in which we depict budgets and per capita transfer for the first random validation sample for Bolivia. Panel (a) plots the resulting budget as a share of total expenditure against the share of beneficiaries and panel (b) plots the per capita transfer as a share of average expenditure (on a log scale) against the share of beneficiaries when the poverty headcount is 30 percent. The apolitical simulations, depicted by the black lines, assume a fixed budget that is between one and four percent of total expenditure. The corresponding figures for Indonesia (not shown) are somewhat lower, a consequence of lower poverty gaps in the data.

Two points are worth noting about the assumption of elastic budgets: on the one hand, the assumption of a fixed budget translates into very large per capita transfers for narrowly targeted programs to the point that they are unrealistic ultimately. For instance, when only two percent of the population are targeted and the budget would suffice to fill in the entire poverty shortfall, the per capita transfer would be two times average per capita expenditure. In contrast, if the budget is increasing in the beneficiary share at an elasticity of two thirds (i.e. $\beta = -1/3$), the corresponding figure is only 33 percent. On the other hand, budgets become very large for programs that are universal if we allow for political constraints: when we assume that the relevant elasticity is the same as the one we find in data for China’s *Dibao*-program, it comes close to nine percent of total expenditure. Below, we will be concerned primarily with situations in which the budget is modest and β is either -1 or $-1/3$ or the budget is large and the elasticity is -1 .

5.1.3 Simulation steps and outcomes

As previously, the simulation is carried out by changing the beneficiary share from one percent of the population to one hundred percent in increments of one percentage point. Under the above scenarios and based on our PMTs introduced in section 3.3, we can calculate for all PMTs \mathcal{X} the

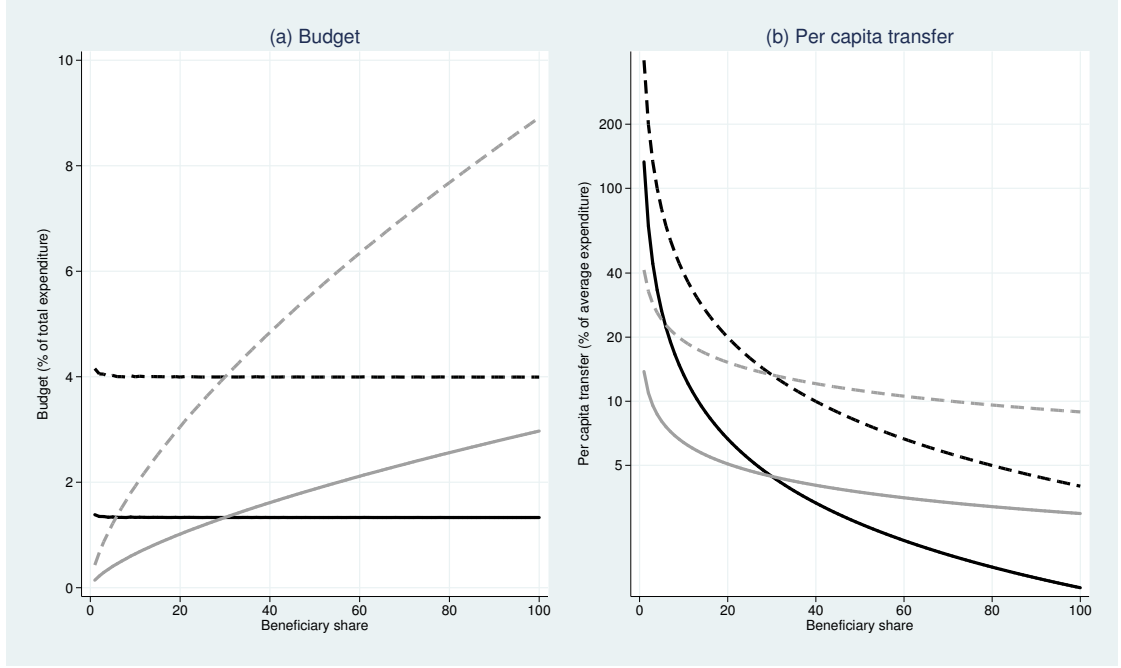


FIGURE 3: The figure reports budgets and transfers (on a log scale) for Bolivia when 30 percent of the population are poor. This is done for the first random subsample. *Panel (a)*: Budget as a share of total expenditure against population share targeted under alternative scenarios. *Panel (b)*: Per capita transfers as a share of average per capita expenditure (on a log scale) against population share targeted. The black (gray) lines denote elasticities of per capita transfers with respect to the targeting rate of -1 ($-1/3$), respectively, and the solid (dashed) lines indicate budgets that are 33 (100) percent of the aggregate poverty gap. The respective figures for Indonesia are similar qualitatively yet budgets are somewhat smaller in relative terms, a consequence of a greater average relative poverty gap.

‘optimal’ beneficiary shares that result from using a targeting measure T , where $T \in \{TD, TER\}$, and from using instead the poverty measure $FGT(\alpha)$, where $\alpha \in \{1, 2\}$. Note that both depend on the PMT and the initial poverty headcount yet only the latter depends on assumptions about the budget.

5.2 Results

A first set of results from our simulations are depicted in figure 4, where the change in the squared poverty gap is plotted against the beneficiary share. In this first depiction we fixed the budget at one-third of the aggregate poverty gap at the point in which $H_0 = r$. Note that this is the point in which the two graphs intersect. Panels (a) and (b) depict results for Bolivia and Indonesia, respective, and an initial poverty gap of ten percent while panels (c) and (d) depict results if the initial poverty gap is 40 percent. Poverty effects are depicted with 90-percent confidence bands that are based on 20-fold cross validation. Indicated are also the beneficiary shares that would

result from maximizing TER and TD , respectively.

Three general results are worth noting that are line with our expectations: first, potential poverty effects are more pronounced in Bolivia. Under the assumption of a budget that is fixed at one-third of the aggregate poverty gap and using M1 as the classifier, panel (a) [(c)] shows that the scenario captured here by the solid line, the greatest poverty effect that is attainable (under the restrictions we impose on transfers) is 35 [50] percent of the squared poverty gap if the initial poverty headcount is ten [40] percent. For Indonesia, the corresponding figure is only about 20 [36] percent (panel (b) [(d)]). Recall that the initial poverty gap in Bolivia is larger and that PMTs are significantly better at distinguishing between the poor and the non-poor. Hence, there is more potential for poverty reduction. Second, while all graphs exhibit only one local minimum, these tend to come out more sharply if the budget is assumed to be fixed (solid lines). This will have implications for targeting: if we assume that the budget increases by two-thirds of a percent for each one-percent increase in the beneficiary share (dashed lines), the question of how narrow policy-makers concerned with poverty-removal will want to target becomes less salient. This is to be expected as a larger beneficiary share not only affects targeting but also increases the size of the program itself. Third, it is clear from figure 4 that if budgets are elastic, broad targeting is more advisable. This reflects policy-recommendations from the literature on the political economy of targeting.

Our primary interest is in whether beneficiary shares that result from optimizing targeting measures are in line with targeting that also has a large poverty-removal effect. Results depicted in figure 4 suggest that this is not generally the case. Recall that optimizing TER will tend to result in narrow targeting while optimizing TD will tend to result in broad targeting. If the initial headcount is below 50 percent—as is the case in all our simulations and in most policy applications—the former puts a larger negative weight on leakage while the latter tends to emphasize coverage among the poor. Consider scenarios in which the budget is fixed and the poverty headcount is ten percent. In these situations, optimizing TER will result in targeting that is too broad in Bolivia yet too narrow in Indonesia (panels (a) and (b), respectively). And the poverty-removal effect one would forego in the latter case is pronounced. If, however, the initial headcount stands at 40 percent, optimizing TER will result in targeting that is too broad in both Bolivia and Indonesia. Optimizing TD , on the other hand, will always tend to result in targeting that is too broad when budgets are fixed.

What if budgets are elastic? We noted above that poverty effects in this situation are generally less sensitive to the beneficiary share. The exception is very narrow targeting. Therefore, optimizing TER , particularly when the initial headcount is low and it is difficult to identify poor households (as is the case in Indonesia), results in large losses in terms of poverty removal (see panel (b)).

Part of the information that can be had from figure 4 is also reported in panel A of table 6, in which we report poverty effects that would result from optimizing TER and TD , respectively, relative to the largest effect that would result from choosing the beneficiary share r to minimize

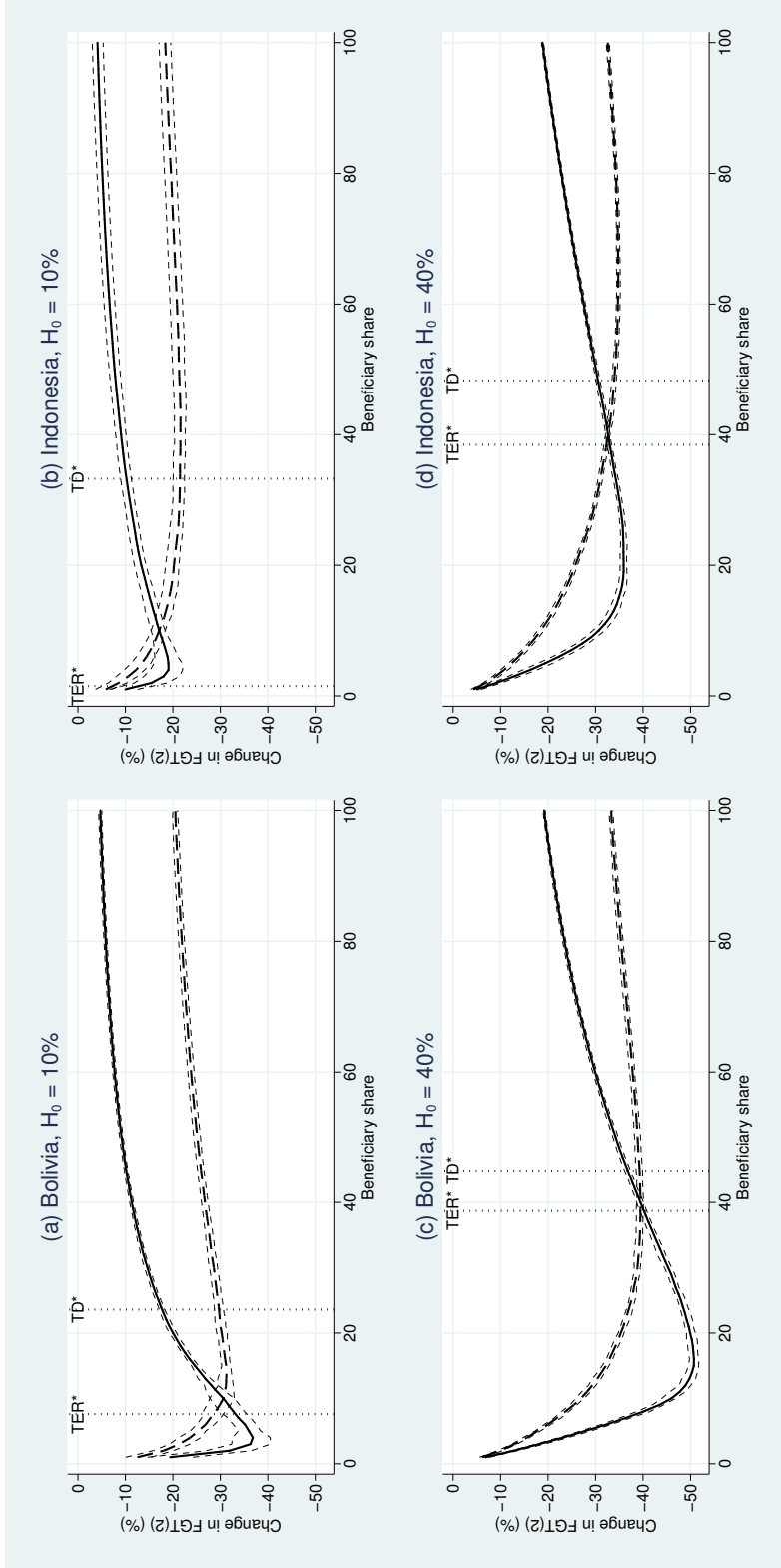


FIGURE 4: Effects on the squared poverty gap (in %) against population proportion targeted by country and initial poverty headcount. The solid lines assume that the budget is fixed while the dashed lines assume that the elasticity of available budgets with respect to the beneficiary share is two-thirds. We always assume here that the budget is one-third of the aggregate poverty gap if $r = H_0$. 90-percent-confidence bands for poverty effects are depicted as short-dashed lines and are based on 20 random sample splits. We also indicate the beneficiary share that would result from optimizing TER and TD , respectively, through dotted, vertical lines.

TABLE 6: Relative poverty effects by targeting strategy, country, initial headcount.

	H_0 (%)	Reduction in poverty (%)	Reduction in poverty relative to optimum (%)		
		min $FGT(\alpha)$	min TER	max TD	$H_0 = r$
<i>Panel A:</i> Fixed budget, $s = 1/3$, M1, $\alpha = 2$.					
Bolivia	10	-37.5 (2.2)	90.4 (6.8)	48.1 (5.2)	81.9 (4.6)
	40	-50.9 (0.6)	79.0 (2.1)	72.5 (3.1)	77.5 (1.1)
Indonesia	10	-19.4 (1.6)	64.5 (20.2)	53.3 (5.7)	88.4 (4.0)
	40	-36.0 (0.4)	91.9 (2.1)	84.9 (1.8)	90.6 (0.8)
<i>Panel B:</i> Fixed budget, $s = 1$, M1, $\alpha = 2$.					
Bolivia	10	-66.7 (2.5)	96.6 (4.2)	69.7 (5.4)	98.8 (1.5)
	40	-84.1 (0.7)	98.5 (0.9)	95.0 (1.9)	97.9 (0.3)
Indonesia	10	-37.7 (1.8)	36.4 (15.3)	74.4 (5.0)	98.2 (1.4)
	40	-67.6 (0.6)	99.1 (0.9)	99.2 (0.5)	99.4 (0.4)
<i>Panel C:</i> $\beta = -1/3$, $s = 1/3$, M1, $\alpha = 2$.					
Bolivia	10	-31.7 (0.8)	91.1 (6.2)	94.4 (2.2)	96.7 (3.1)
	40	-39.6 (0.4)	99.8 (0.3)	99.3 (0.4)	99.7 (0.2)
Indonesia	10	-21.7 (0.7)	33.4 (11.5)	99.4 (0.5)	78.8 (3.2)
	40	-34.8 (0.2)	93.0 (2.1)	97.6 (0.8)	93.8 (0.7)
<i>Panel D:</i> Fixed budget, $s = 1/3$, M2, $\alpha = 2$.					
Bolivia	10	-40.7 (2.4)	87.5 (6.8)	47.8 (4.1)	78.3 (5.2)
	40	-52.0 (0.5)	79.0 (2.4)	71.1 (2.9)	77.0 (1.0)
Indonesia	10	-21.5 (1.5)	80.7 (16.4)	49.6 (3.0)	88.4 (3.1)
	40	-37.8 (0.5)	90.4 (1.5)	82.7 (1.5)	89.6 (0.8)
<i>Panel E:</i> Fixed budget, $s = 1/3$, M1, $\alpha = 1$.					
Bolivia	10	-24.7 (0.9)	89.6 (4.6)	51.1 (5.6)	82.5 (4.5)
	40	-32.8 (0.5)	80.3 (2.3)	75.0 (3.0)	78.9 (1.3)
Indonesia	10	-14.2 (0.7)	71.0 (18.8)	57.8 (4.5)	90.8 (3.7)
	40	-25.3 (0.3)	90.3 (2.0)	84.1 (1.8)	89.2 (1.1)

Means and associated standard deviations (in parentheses) based on 20-fold cross validation. Results associated with strategies that result in the largest relative effect on poverty are highlighted.

$FGT(\alpha)$. In addition, we report in the last column relative poverty effects from simply setting r equal to the beneficiary share. Remaining panels in this table report robustness checks. In panel B we consider a much larger budget that equals the size of the aggregate poverty gap. In panel C we allow a smaller budget to be elastic with respect to the beneficiary share. Panel D goes back to a smaller, inelastic budget but relies on the more sophisticated PMTs M2. Panel E considers effects on poverty as measured by the poverty gap indicator $FGT(1)$. Standard errors for poverty effects based on 20-fold cross validation are reported in parentheses. We also highlight the strategy that results in the largest relative effect on poverty.

If budgets are fixed, targeting should be narrow—more narrow than targeting just whoever is poor. This is particularly the case in countries like Bolivia in which the initial poverty gap is large. Optimizing TER will therefore often come closer to the optimal poverty effect than alternative strategies. This is reflected in panel A: optimizing TER will achieve relative poverty effects that are about 79–90 percent of the optimal poverty effect in Bolivia and 65–92 percent

in Indonesia. As was also evident from figure 4, targeting would tend to be too narrow in Indonesia when the initial poverty headcount stands at ten percent. While the estimate of the poverty effect comes with a large standard error, on average, targeting just whoever is poor would yield a larger poverty effect.

Note that these results are neither affected by using a ‘better’ PMT (panel D) nor by focusing on effect on the poverty gap indicator (panel E). These results change, however, if a larger budget is available (panel B) or if the budget increases with the beneficiary share (panel C). In the former case, we find that setting $r = H_0$ results in very good approximations of the poverty effects that are attainable while in the latter case, optimizing TD —which results in broad targeting—results in very large relative poverty effects.

It is striking, however, that the losses in terms of poverty effects from simply targeting the same proportion of the population that one deems poor are only modest in the worst case yet optimizing either TER or TD will sometimes result in poverty effects that are less than half the size they could be. As demonstrated in section 4, this strategy has the virtue of striking a compromise between targeting based on optimization of TER and TD . Results here suggest in particular that an emphasis on leakage—as captured by the false positive rate—is not warranted in situations in which the initial poverty gap is small and it is challenging to distinguish the poor from the non-poor. Conversely, an emphasis on coverage—as captured by the true positive rate—is not in the interest of the poor when the poverty gap is large and the poor are readily discernible.

6 Conclusion

In this paper we consider the usefulness of targeting measures in program design. We first show that two targeting measures popular in practice can be rationalized in terms of a social welfare function defined over errors of targeting. We then present simulation evidence that suggests that relying on targeting measures in order to decide about the extent of targeting will in many situations result in programs that have an effect on poverty that is considerably lower than what could be had. Optimizing the targeting differential, the difference between true positive rate and false positive rate, will result in too much emphasis on coverage among the poor except when the budget is highly responsive to the number of beneficiaries. Optimizing the total error rate, on the other hand, will in many situations imply an unwarranted emphasis on the avoidance of leakage to the non-poor. As a result, targeting will often be too narrow for the effect on poverty to be close to the maximum. While not optimal in general, setting the targeting rate equal to the initial poverty headcount, i.e., setting the number of beneficiaries equal to the number of poor, will often strike a warranted compromise in terms of targeting and result in poverty reductions that are never very different from the maximum reductions that are viable. Our results are robust to alternative poverty measures and models to distinguish the poor from the non-poor.

This suggests that using targeting measures that combine information about coverage among

the poor and leakage to the non-poor should be used cautiously if at all. Whether they are good proxy objectives depends on the context. And while it may be that targeting errors are by themselves important for the success of an anti-poverty program (for instance, because the public is concerned primarily with the unfairness that leakage of benefits to the non-poor entails), their relationship to the actual effect of programs on poverty seems tenuous at best.

While our simulations consider the potential effects of politico-economic constraints in targeting, there are some caveats. First, we have not considered behavioral responses, variation in administrative costs, and hidden penalties such as welfare stigma. Second, we have obtained rough estimates of the elasticity of average per capita transfers with respect to the participation rate from published data on one particular transfer scheme. We used this as a lower bound yet while our estimates are not implausible, econometric concerns remain.

On a side note, our results also suggest that country-context (i.e. the incidence and depth of poverty as well as politico-economic constraints) and the extent of targeting will often carry more significance than the use of the particular classifier.

References

- Ahmed, A. U. and H. E. Bouis (2002). Weighing What’s Practical: Proxy Means Tests for Targeting Food Subsidies in Egypt. *IFPRI Food Consumption and Nutrition Division Discussion Paper No. 132*.
- Akerlof, G. A. (1978). The Economics of “Tagging” as Applied to the Optimal Income Tax, Welfare Programs and Manpower Planning. *American Economic Review* 68(1), 8–19.
- Alatas, V., A. Banerjee, R. Hanna, B. A. Olken, R. Purnamasari, and M. Wai-Poi (2013). Ordeal Mechanisms and Targeting: Theory and Evidence from a Field Experiment in Indonesia. *NBER Working Paper 19127*.
- Alatas, V., A. Banerjee, R. Hanna, B. A. Olken, and J. Tobias (2012). Targeting the Poor: Evidence from a Field Experiment in Indonesia. *American Economic Review* 102(4), 1206–1240.
- Appleton, S., J. Knight, L. Song, and Q. Xia (2002). Labor Retrenchment in China: Determinants and Consequences. *China Economic Review* 13(2–3), 257–275.
- Atkinson, A. B. (1995). On Targeting Social Security: Theory and Western Experience with Family Benefits. In D. van de Walle and K. Nead (Eds.), *Public Spending and the Poor: Theory and Evidence*, pp. 25–68. John Hopkins University Press.
- Bardhan, P. and D. Mookherjee (2005). Decentralizing Antipoverty Program Delivery in Developing Countries. *Journal of Public Economics* 89, 675–704.
- Besley, T. and S. Coate (1992). Workfare versus Welfare: Incentive Argument for Work Requirements in Poverty-Alleviation Programs. *The American Economic Review* 82(1), 249–261.
- Besley, T. and R. Kanbur (1988). Food-Subsidies and Poverty Alleviation. *The Economic Journal* 98(392), 701–719.
- Besley, T. and R. Kanbur (1993). The Principles of Targeting. In M. Lipton and J. van der Gaag (Eds.), *Including the Poor*. Washington, D.C.: World Bank.
- Bigman, D. and H. Fofack (2000). Geographic Targeting for Poverty Alleviation: An Introduction to the Special Issue. *World Bank Economic Review* 14(1), 129–146.
- Brady, D. and A. Bostic (2015). Paradoxes of Social Policy: Welfare Transfers, Relative Poverty, and Redistribution Preferences. *American Sociological Review* 80(2), 268–298.
- Camacho, A. and E. Conover (2011). Manipulation of Social Program Eligibility. *American Economic Journal: Economic Policy* 3, 41–65.

- Canstañeda, T. (2005). Targeting Social Spending to the Poor with Proxy-Means Testing: Colombia's SISBEN System. *World Bank Human Development Network Social Protection Unit Discussion paper 0529*.
- Chen, S., M. Ravallion, and Y. Wang (2008). Does the Dibao Program Guarantee a Minimum Income in China's Cities? In J. Lou and S. Wang (Eds.), *Public Finance in China: Reform and Growth for a Harmonious Society*. Washington DC: World Bank.
- Coady, D., M. Grosh, and J. Hoddinott (2004). Targeting Outcomes Redux. *The World Bank Research Observer* 19(1), 61–85.
- Conning, J. and M. Kevane (2002). Community-Based Targeting Mechanisms for Social Safety Nets: A Critical Review. *World Development* 30(3), 375–394.
- Cornia, G. A. and F. Stewart (1995). Two Errors of Targeting. In D. van de Walle and K. Nead (Eds.), *Public Spending and the Poor: Theory and Evidence*. John Hopkins University Press.
- Dalton, H. (1920). The Measurement of the Inequality of Incomes. *Economic Journal* 55, 749–764.
- Datt, G. and M. Ravallion (1994). Transfer Benefits From Public-Works Employment: Evidence for Rural India. *Economic Journal* 104(427), 1346–1369.
- De Donder, P. and J. Hindriks (1998). The Political Economy of Targeting. *Public Choice* 95(1–2), 177–200.
- Deaton, A. (1997). *The Analysis of Household Surveys: Microeconometric Analysis for Development Policy*. Baltimore: John Hopkins University Press.
- Drèze, J. and P. Srinivasan (1997). Widowhood and Poverty in Rural India: Some Inferences from Households Survey Data. *Journal of Development Economics* 54, 217–234.
- Dutta, P., R. Murgai, M. Ravallion, and D. van de Walle (2012). Does India's Employment Guarantee Scheme Guarantee Employment? *World Bank Policy Research Working Paper 6003*.
- Foster, J., J. Greer, and E. Thorbecke (1984). A Class of Decomposable Poverty Measures. *Econometrica* 52(3), 761–766.
- Fujii, T. (2008). How Well Can We Target Aid with Rapidly Collected Data? Empirical Results for Poverty Mapping from Cambodia. *World Development* 36(10), 1830–1842.
- Galasso, E. and M. Ravallion (2005). Decentralized Targeting of an Antipoverty Program. *Journal of Public Economics* 89, 705–727.
- Gao, Q., I. Garfinkel, and F. Zhai (2009). Anti-poverty Effectiveness of the Minimum Living Standard Assistance Policy in Urban China. *Review of Income and Wealth* 55, 630–655.

- Gelbach, J. and L. Pritchett (2001). Indicator Targeting in a Political Economy: Leakier can be Better. *Journal of Policy Reform* 4(2), 113–145.
- Gelbach, J. and L. Pritchett (2002). Is More for the Poor Less for the Poor? The Politics of Means-Tested Targeting. *The B.E. Journal of Economic Analysis & Policy* 2(1).
- Glewwe, P. (1990). Efficient Allocation of Transfers to the Poor: The Problem of Income. *LSMS Working Paper #70*.
- Grosh, M., C. del Ninno, E. Tesliue, and A. Ouerghi (2008). *For Protection & Promotion: The Design and Implementation of Effective Safety Nets*. Washington, D.C.: The World Bank.
- Grosh, M. E. and J. L. Baker (1995). Proxy Means Tests for Targeting Social Programs: Simulations and Speculation. *LSMS Working Paper No. 118*.
- Gustafson, B. A. and D. Quheng (2011). Di Bao Receipt and Its Importance for Combating Poverty in Urban China. *Poverty and Public Policy* 3(1).
- INE (2015). Bolivia—Encuesta de Hogares 2011. Technical report, Instituto Nacional de Estadísticas.
- Jalan, J. and M. Ravallion (1998). Are There Dynamic Gains from a Poor-area Development Program? *Journal of Public Economics* 67(1), 65–85.
- Johannsen, J. (2008). *Operational Assessment of Monetary Poverty by Proxy Means Tests: The Example of Peru*. Number 65 in Development Economics and Policy. Peter Lang Verlag.
- Kidd, S. and E. Wylde (2011). Targeting the Poorest: An Assessment of the Proxy Means Test Methodology. Published by the Australian Agency for International Development (AUSAid).
- Landau, K., S. Klasen, and W. Zucchini (2012). Measuring Vulnerability to Poverty Using Long-Term Panel Data. *Courant Research Centre Discussion Paper No. 18, Göttingen University*.
- Lindert, K., E. Skoufias, and J. Shapiro (2006). Measuring Vulnerability to Poverty Using Long-Term Panel Data. *World Bank Social Protection Discussion Paper No. 0605*.
- Mares, I. and M. E. Carnes (2009). Social Policy in Developing Countries. *Annual Review of Political Science* 12, 93–113.
- Moene, K. O. and M. Wallerstein (2001). Targeting and the Political Support for Welfare Spending. *Economics of Governance* 2(1), 3–24.
- Moffitt, R. (1983). An Economic Model of Welfare Stigma. *American Economic Review* 73(5), 1023–1035.
- Niehaus, P., A. Atanassova, M. Bertrand, and S. Mullainathan (2013). Targeting with Agents. *American Economic Journal: Economic Policy* 5(1), 206–238.

- Ravallion, M. (1994). Measuring Social Welfare With and Without Poverty Lines. *American Economic Review Papers and Proceedings* 84(2), 359–364.
- Ravallion, M. (2003). Targeted Transfers in Poor Countries: Revisiting the Trade-Offs and Policy Options. *Social Safety Net Primer Series*.
- Ravallion, M. (2009a). Decentralizing Eligibility for a Federal Antipoverty Program: A Case Study for China. *World Bank Economic Review* 23(1), 1–30.
- Ravallion, M. (2009b). How Relevant is Targeting to the Success of an Antipoverty Program? *World Bank Research Observer* 24(2), 205–231.
- 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 S. Chen (2015). Benefit Incidence with Incentive Effects, Measurement Errors and Latent Heterogeneity: A Case Study for China. *Journal of Public Economics* 128, 124–132.
- Schady, N. R. (2002). Picking the Poor: Indicators for Geographic Targeting in Peru. *Review of Income and Wealth* 48(3), 417–433.
- Sen, A. (1995). The Political Economy of Targeting. In D. van de Walle and K. Nead (Eds.), *Public Spending and the Poor: Theory and Evidence*, pp. 11–24. John Hopkins University Press.
- Skoufias, E. and D. P. Coady (2007). Are the Welfare Losses from Imperfect Targeting Important? *Economica* 74, 756–776.
- Skoufias, E., B. Davis, and S. de la Vega (2001). Targeting the Poor in Mexico: An Evaluation of The Selection of Household for PROGRESA. *IFPRI Food Consumption and Nutrition Division Discussion Paper No. 103*.
- Surbakti, P. (1995). Indonesia’s National Socio-Economic Survey: A Continual Data Source for Analysis on Welfare Development. Technical report, Central Bureau of Statistics.
- Tang, J. and Z. Ren (2002). Policy Suggestions for Improving China’s Social Assistance System. *Economic Reference* 76, 20–30.
- Thompson, M. L. and W. Zucchini (1989). On the Statistical Analysis of ROC Curves. *Statistics in Medicine* 8(10), 1277–1290.
- Umapathi, N., D. Wang, and P. O’Keefe (2013). Eligibility Thresholds for Minimum Livelihood Guarantee Programs. *World Bank Social Protection & Labor Discussion Paper No. 1307*.
- van de Walle, D. (1998). Targeting Revisited. *World Bank Research Observer* 13(2), 231–248.
- Wodon, Q. (1997). Targeting the Poor Using ROC Curves. *World Development* 25(12), 2083–2092.

A Gauging political constraints

A.1 Empirical framework

In this section we aim to gauge the importance of politico-economic constraints by estimating the relationship between per capita transfers and the beneficiary share. Assuming no politico-economic constraints is equivalent to assuming that the budget is fixed and one would expect to find that the elasticity of average per capita transfers with respect to the beneficiary share is minus one. To see this, denote by $t_{ij} \geq 0$ the transfer to individual j in constituency i . Let n_i be the total number of individuals in that constituency and n_i^b the number of beneficiaries. The average transfer to beneficiaries in i is $t_i = (\sum_j^{n_i} t_{ij})/n_i^b = b_i/n_i^b$, where b_i denotes the budget allocated to the program. Taking logs on both sides and expanding yields

$$\ln t_i = -\ln(n_i^b/n_i) + \ln b_i - \ln n_i. \quad (\text{A.1})$$

(A.1) is, of course, an identity, not an empirical model. If resources available for a program do not depend on how narrowly the program is targeted and if these resources and the total population are kept constant, a one-percent increase in the beneficiary share will invariably result in a one-percent decrease in the average transfer payment to beneficiaries. In order to obtain an empirical model that can be used to test whether the budget is exogenous, we proxy the capacity to fund programs through GDP, denoted y_i . Adding a constant and a white-noise-error term, we obtain

$$\ln t_i = \beta_0 + \beta_1 \ln(n_i^b/n_i) + \beta_2 \ln y_i + \beta_3 \ln n_i + \epsilon_i. \quad (\text{A.2})$$

In what follows we estimate variants of (A.2) based on data from China's urban *Dibao*-program. Note again that β_1 in this model is interpreted as the elasticity of average transfers among beneficiaries with respect to the beneficiary share. Given (A.1), β_2 and β_3 are expected to be close to one and minus one, respectively. However, if, for instance, GDP conditional on population is correlated with capacity to administer a program, it may be that $\beta_2 > 1$. If, on the other hand, the center provides co-funding to cities with low GDP, we may find that $\beta_2 < 1$. If larger populations are associated with economies of scale in administering the program or congestion effects, β_2 would be greater or smaller than -1 , respectively. Measurement error in the participation rate and reverse causality are concerns that we aim to address in what follows.

A.2 China's *Dibao*-program

China's *Minimum Livelihood Guarantee* program, popularly known as the *Dibao* (DB)-program, is a cash transfer program that aims to bring individuals' incomes up to the poverty line. The DB program with its urban and rural arm is by now probably the largest cash transfer program in the world with around 52 million people covered.¹⁴ It has been one key element in China's

¹⁴See *China Daily*, November 9th, 2015, available at http://www.chinadaily.com.cn/kindle/2015-11/09/content_22408338.htm. Accessed May 13th, 2016.

overhaul of its social policies that aimed to address major changes in the country’s urban labor market that became apparent in the mid-1990s (Appleton et al., 2002; Gustafson and Quheng, 2011).

On paper, the program aims to remove all poverty shortfalls. This would result in a marginal tax rate of 100 percent for beneficiaries: each increase in income should result in a equal reduction in benefits as long as income remains below the poverty line. In practice, however, this rate seems to be much lower at around 12–14 percent (Ravallion and Chen, 2015). In line with this finding, targeting as assessed by widely-used indicators is exceptional yet still far from perfect. Past research found that leakage is very limited but so is coverage among the poor (Chen et al., 2008; Gao et al., 2009).

The implementation of the *Dibao*-program was mostly decentralized with the role of the the national and provincial governments restricted to providing guidelines and co-financing, particularly in recent years (Tang and Ren, 2002; Umapathi et al., 2013). For instance, claimants must first apply to a local office in order to receive benefits.¹⁵ In terms of threshold levels, the center provides some guidance on how these are to be set. Guidelines mention the requirement that basic needs be met and that prevailing prices be taken into account. Yet they also take into account local fiscal constraints and, ultimately, individual municipalities are responsible for the selection of beneficiaries. Resulting differences in implementation of threshold levels seem to be driven by local economic conditions and fiscal capacities: local governments in large, coastal cities such as Beijing, Shanghai, and Jiangsu tend to have higher threshold levels while governments in medium and small cities often opted for lower thresholds (Tang and Ren, 2002; Umapathi et al., 2013).

We assemble a cross-city dataset from the early 2000s collected for the program’s urban arm. Data on average transfers, beneficiary shares, threshold levels (i.e., *Dibao* poverty lines), and pre-transfer headcounts for 35 large cities are taken from published work by Ravallion (2009a). The original data come from the *Urban Household Short Survey* for 2003–04, a light survey-instrument with large sample sizes that allows it to be representative at the level of these cities. These are supplemented by data on population and GDP for the year 2002 from the UN¹⁶ and the *China City Statistical Yearbook 2002*, respectively.

While this is a small dataset, the main advantage for our purpose is that it provides variation both in beneficiary shares and average transfers. The former roughly range from one to about 27 percent while average transfers range from between 1.1 to 4.4 percent of average incomes. Figure 5 shows that beneficiary shares are usually lower than headcounts implied by the *Dibao*-thresholds but with considerable variation. At the same time, the variation in potential confounding factors conditional on GDP and population size will likely be less pronounced than in cross-country settings.

There are still econometric concerns in estimating equation (A.2) that need to be addressed.

¹⁵There is also a community vetting process: the names of proposed participants are displayed on notice boards and community members are encouraged to identify any undeserving applicants (Ravallion, 2009a).

¹⁶The data were taken from (<http://data.un.org/Data.aspx?d=POP&f=tableCode%3A240>)

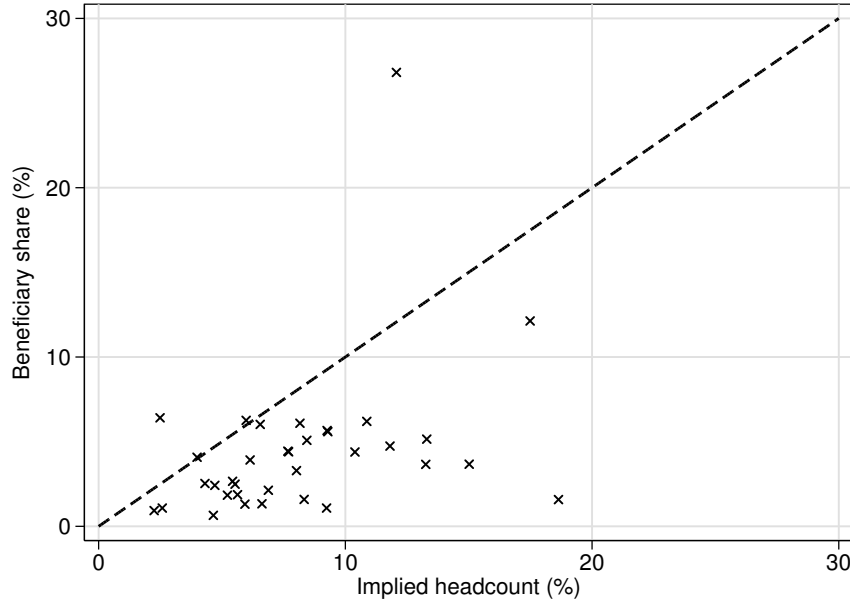


FIGURE 5: Scatter plot of participation rates against headcounts implied by *Dibao*-thresholds. The dashed line is a 45-degree-line.

One is that reverse causality may tend to upward-bias our estimate of β_1 , the coefficient of interest. This will be the case if, say, higher transfers induce more eligible beneficiaries to sign on to the program, increasing the beneficiary share. A second concern is measurement error in that very share. This would also tend to upward-bias our estimates (i.e. attenuate β_1 towards zero). One way of addressing these concerns is through using an instrumental variable that predicts the average transfer yet is uncorrelated to the error term in (A.2). We propose that the poverty headcount implied by *Dibao*-eligibility may be such an instrument: it is clearly expected to predict the beneficiary share yet may be uncorrelated to the error term once we condition on both GDP and population size. While this is far from perfect, our results below are in line with the notion that OLS estimates are upward biased.

A.3 Results

Results from estimating (A.2) are reported in table 7. White-standard errors are reported in parentheses. Column (1) reports results from regressing log average transfers only on log participation rates, whereas column (2) reports results from the fully specified model. We then add additional controls, a dummy for cities in coastal provinces (column (3)) and a the Gini coefficient¹⁷ (column(4)), and both controls jointly (column (5)).

¹⁷The Gini coefficient is calculated based on the assumption that incomes are log-normally distributed. In this case, the statistic can be inferred from the poverty rate, the mean income, and the poverty line.

TABLE 7: Estimates of the elasticity of per capita transfers with respect to the beneficiary share in the China's urban *Dibao*-program.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	OLS	OLS	OLS	OLS	OLS	2SLS
Constant	4.682*** (0.248)	3.787 (2.587)	4.094 (2.876)	2.650 (1.945)	2.989 (2.349)	3.724 (2.805)	1.498 (2.247)
Log beneficiary share	-0.238*** (0.079)	-0.209*** (0.061)	-0.214*** (0.066)	-0.194*** (0.056)	-0.200*** (0.063)	-0.213** (0.083)	-0.333*** (0.133)
Log GDP in 2002		0.353** (0.170)	0.393 (0.251)	0.336** (0.140)	0.381 (0.239)	0.350* (0.178)	0.318 (0.212)
Log population in 2002		-0.084 (0.231)	-0.121 (0.281)	-0.050 (0.184)	-0.092 (0.246)	-0.079 (0.244)	0.011 (0.220)
Coastal city			-0.052 (0.203)		-0.059 (0.217)		-0.110 (0.208)
Gini coefficient				0.020* (0.011)	0.020* (0.011)		0.019* (0.011)
R-squared	0.168	0.554	0.555	0.598	0.600	0.540	0.561
Observations	35	35	35	35	35	34	35
p -value ($H_0 : \beta_1 = -1$)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Log headcount (in first stage-regression)							0.689*** (0.209)
First-stage F -test for excluded instrument							8.300

White-standard errors in parentheses. *, **, and *** denote significance at the ten-, five-, and one-percent level, respectively. Based on data from Ravallion (2009a), China City Statistical Yearbook 2002, and the UN.

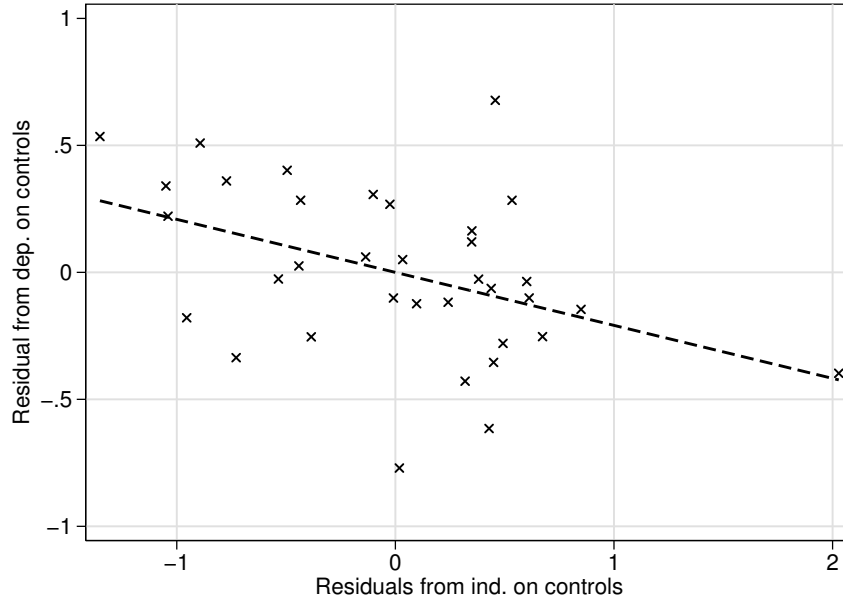


FIGURE 6: Partial regression plot for the log beneficiary share pertaining to results reported in column (2) of table 7.

Results suggest that the elasticity of the average transfer with respect to the beneficiary share is around -0.2 . t -tests indicate that all estimates are always significantly different from both zero and minus one. In other words, there is some evidence here that budgets are not fixed but increasing in the the proportion that benefits.

The final two columns implement two robustness checks. First, in column (6), we omit one outlier in terms of the beneficiary share, the city of Kunming in which almost 27 percent of residents receive transfers. This is five standard deviations above the mean of that variable. We find that excluding this observation has no effect on our estimates. This is also apparent from figure 6, a partial regression plot for the log beneficiary share pertaining to column (2) of table 7).

Second, we address concerns about reverse causality and measurement error as noted above: it may be the case that a large per capita transfer induces many households to sign up for the program in which case there would be positive feedback from the dependent back to the regressor of interest. There will also be classical measurement error in the variable of interest as the data are estimated from an underlying survey. Both would result in upward-biased estimates of the elasticity.

In column (7), we use the log of the implied poverty headcount as an instrument for the log beneficiary rate. The program aims to include all individuals identified as poor. It should therefore be the case that the two variables are positively correlated in the first stage regression.

However, cities have some say in how the *Dibao*-poverty lines are set in that they can decide from a number of methods to do so (Umapathi et al., 2013). This is far from perfect as it may still be that some unobserved, city-specific heterogeneity affects both the way in the threshold was set and is present to some extent in the error term. Conditioning on the coastal city-dummy, GDP and population should attenuate the problem somewhat. But the results should still be interpreted only carefully.

The results are nevertheless very much in line with our expectations. The first stage results suggest that that a one percent-increase in the poverty headcount increases the beneficiary share by almost 0.7 percent. The second stage-results suggest an elasticity of the average transfer with respect to the beneficiary share that is somewhat lower: around -0.33 . This is the value we use as a lower bound in our simulations in section 5.

Overall, our results here suggest that it may be important to take into account politico-economic considerations in the design of targeted transfers. One concern with these results may be whether we can actually extrapolate from findings in China to programs implemented elsewhere. In particular, one may argue that budgets are more responsive to the population proportion that benefits from the program in democracies in which voters can voice support or opposition to particular programs by casting their votes. But note that even though many developing countries are non-democratic, “they have social policies, their public has preferences, and public support matters” (Brady and Bostic, 2015). In fact, as noted by Mares and Carnes (2009), many social insurance programs in now developed countries were initially adopted by non-democratic governments.