Working Paper Series

Assessing the performance of targeting mechanisms

B. Essama-Nssah

ECINEQ WP 2018 - 457

ECINEQ 2018 - 457
February 2018

www.ecineq.org

Assessing the performance of targeting mechanisms*

B. Essama-Nssah†

Abstract

A policy intervention entails a distribution of social resources to a target population in order to solve a particular problem they face. In that context, targeting is designed to assign treatment according to some eligibility criterion. While effective targeting is not an end in itself, it is central to both fairness and cost-effectiveness in policymaking. Decision makers and other stakeholders would therefore like to know the extent to which targeting respects eligibility. Answering this key question is essentially an exercise in associational inference. This paper proposes an evaluative framework for assessing targeting outcomes. The framework focuses on measuring, judging and explaining targeting performance. Among indicators of performance considered, the conditional probability of assignment emerges as a local measure which is more informative than global indicators. The paper advocates for the use of chance-corrected measures of interobserver agreement to judge the extent of agreement between assignment mechanisms, and demonstrates the use of bivariate probit regression analysis to identify proximate determinants of targeting outcomes. The proposed framework is applied to the evaluation of the targeting of cash transfers in the context of the Social Safety Nets Pilot Project (SSNPP) in Northern Cameroon.

Keywords: Targeting outcomes, community-based targeting (CBT), proxy means testing (PMT), self-targeting, interobserver agreement, assignment probability, biprobit model, marginal effects, kernel regression, cash transfers, Cameroon.

*The author is grateful to Quentin Stoeffler and Carlo del Ninno for sharing their targeting assessment for the Social Safety Nets Pilot Project in Cameroon (Stoeffler et al. 2016), and the data they used for that analysis. He is also grateful to Aly Sanoh and Thierry Hounsa for help with data preparation. In addition, Aly Sanoh provided useful comments on an earlier version of this paper.

†**Contact details:** B. Essama-Nssah is independent consultant and former World Bank staff; email: essama.nssah@gmail.com.

1. Introduction

A policy intervention can be viewed as a means-ends relationship wherein resources are transformed into activities designed to solve some socioeconomic problem (the *target problem*) by effectively addressing the conditions or needs of the *target population.* Depending on the nature of the problem, the target population is either the population at risk or the population in need[1]. Ultimately, policy implementation entails a distribution of resources to eligible units (e.g. individuals, households, firms or districts). Such a distribution requires an allocation rule which Young (1994) defines as a process or method of dividing goods or burdens among a group of potential claimants on the basis of salient *characteristics* of those claimants, and the nature and amount of goods or burdens involved. The relative strength of each person's claim stems from relevant and observable characteristics of that person which reflect need, risk or a combination of factors defining desert.

Targeting is essentially a *distributive (or allocation) mechanism* used to identify eligible units for an intervention and screen out the ineligible from a given population (Devereux et al. 2017). It is commonly used in the context of social transfer programs aimed at reducing poverty or protecting vulnerable groups such as people with disabilities, orphans and older persons. In the context of such programs, eligibility is based on a relevant measure of poverty or vulnerability at the individual, household or community level.

One can argue for targeting on the basis of *fairness*, *effectiveness* and *efficiency*. Perception of fairness in the distribution of social resources contributes to social peace. Most targeting mechanisms use the *priority method* to determine who gets the intervention and who goes without. The priority principle requires that resources go to those with the greatest claim. When priority is based on individual (or household) characteristics, one can group applicants by *type* so that all individuals (or households) of the same type have the same characteristics, and create *a standard of comparison* to guide the identification of beneficiaries and the distribution of resources using the priority method[2]. Young (1994) explains that allocation methods based on *priority* are the only methods that respect the two basic principles of equity, namely: *impartiality* and *consistency*[3]. In the context of pro-poor interventions, priority is based on the relative poverty status of individuals or households. In particular, relevant units are ranked on the basis of an indicator of the living standard

[1] The population at risk consists of units that have a significant probability of having or developing the condition the intervention is designed to address while the population in need is the group of units facing the problem that the program is supposed to solve.

[2] A standard of comparison is a list of all types *ordered* from highest to lowest priority (Young 1994).

[3] A distributive mechanism is *impartial* if the associated allocations depend only on the relevant characteristics of the claimants and the total amount to be distributed. In other words, an impartial mechanism pays no attention to personal attributes that are irrelevant to the problem at hand, according to the chosen ethical standard. An allocation mechanism is *consistent* if the allocations within any sub-group of claimants do not depend on the presence of other claimants. Let, for instance, two claimants share a single unit of an indivisible good in such a way that it goes to the person with higher priority among the two claimants. That allocation rule is independent of the other claimants present, and is said to be pairwise consistent (Young 1994).

from the worst to best off. A cutoff point is then imposed on the implied priority list to separate eligible from non-eligible units.

An intervention is relevant if it is the right thing to do for the target population, given the problem and the circumstances they face. It is commonly accepted that when contextual factors are favorable, the likelihood of success is higher when the intervention is appropriate, adequate and *targeted* to the right segment of the population who will make use of it as prescribed. Thus, targeting remains an important element of policymaking despite the difficulties associated with the design of targeting mechanisms. If program services do not reach the population in need or at risk, then it would not have an impact on them (Grosh et al. 2008). In that case the intervention will be considered *ineffective*. Similarly, if the intervention does reach not only the intended beneficiaries, but many non-eligible persons, then it would most likely incur a relatively high operational cost. Such an outcome would be *inefficient*. These considerations suggest that a good targeting assessment can provide evidence that explains, at least in part, the success or failure of an intervention.

Targeting mechanisms are therefore designed to minimize errors of exclusion and errors of inclusion. An *error of exclusion* occurs when an eligible person is excluded from the program. An *error of inclusion* is committed whenever a non-eligible is included in the program. These errors constitute a violation of impartiality, a key principle of equity. Impartiality implies that claimants of the same type should be treated equally. This is the same as saying that unequal treatment should reflect only differences in type. In other words, impartiality implies both *horizontal* and *vertical equity*. The imposition of a cutoff point on the priority list induced by a targeting mechanism creates two broad types of applicants: the eligible and non-eligible. It is clear that errors of exclusion violate horizontal equity which demands equal treatment of equal. Similarly, errors of inclusion violate vertical equity which demands that differences in type be appropriately taken into account when assigning eligibility.

The design of targeting mechanisms is complicated by the fact that the objects of targeting are not simply patients to whom things are done, but socioeconomic agents whose choices and actions are critical factors affecting the *operation* of targeting mechanisms (Sen 1995). Agents' reactions to eligibility rules may even distort targeting arrangements and amplify selection errors. Difficulties in observing individual characteristics create incentive for claimants to game the system by misrepresenting their type. This is an informational distortion that is bound to limit the effectiveness of targeting. In other situations, eligibility criteria may create incentives for socioeconomic agents to adjust their behavior in order to meet eligibility criteria. For instance, an individual may choose to work less to bring his or her earned income below the threshold required to qualify for a subsidy.

Beyond informational and incentive distortions, stigma, administrative costs and political feasibility are all considerations that affect the design and hence the implementation of targeting arrangements. Eligibility conditions for some antipoverty programs may have, in addition to incentive effects, effects on participants' self-respect as well as on the respect they get from others. Such *stigmatization* can discourage participation. There may be substantial administrative costs associated with the need to gather information to identify potential beneficiaries. A byproduct of this information gathering is the loss of privacy (Sen 1995). Means testing, for instance, requires detailed disclosure of individual circumstances. Finally, political feasibility relates to the fact that an antipoverty program is likely to lose political support unless a significant amount of benefits spill over to non-poor. Such leakage would increase errors of inclusion.

There are six mechanisms that are commonly used to identify and select program beneficiaries (Devereux et al. 2017). They include: means testing, proxy means testing (PMT), categorical targeting, geographic targeting, community-based targeting (CBT), and self-targeting.

Means testing requires a direct assessment of income, assets or wealth of applicants to determine whether they deserve social support. *Proxy means testing* determines eligibility on the basis of a score which is a weighted combination of attributes or characteristics that are believed to be highly correlated with the condition of interest (e.g. poverty or vulnerability). *Categorical targeting* bases the selection of beneficiaries on characteristics that are of interest to policymakers. Such characteristics may or may not be correlated with the target condition. In the case of *geographic targeting*, location or area of residence determines eligibility. For instance, an antipoverty program could base eligibility on the level of poverty incidence in the district of residence. *Community-based targeting* means that eligibility is assessed by members of the community who are presumed to have better local knowledge of individual circumstances than the central government. Finally, *self-targeting* usually relies on the power of incentives offered by the program to attract eligible participants while deterring participation of non-eligible. Self-targeting is commonly used in public work programs which set work conditions and the wage rate in such a way that they can appeal only to the needy. A single program can also use a combination of some of these targeting mechanisms to select participants. Presumably, such a combination would produce better targeting than a single method (Grosh et al. 2008).

While effective targeting is not an end in itself, it ensures that the intervention reaches the right segment of the population. This is an important intermediate outcome and a necessary condition for impact. Evidence-based decision making creates the need to know the extent to which targeting mechanisms work as intended.

This paper therefore proposes an evaluative framework for assessing the performance of targeting mechanisms. It is based on the idea that evaluation is meant to generate credible evidence to answer key policy or programmatic questions which decision makers and other stakeholders care about (Gertler et al. 2016). Such questions focus the evaluation and determine the logical pathway to credible answers[4]. There is a core question that undergirds virtually all *ex post* policy evaluations, namely: *Did what was supposed to*

[4] See Essama-Nssah (2013) for a discussion of key questions that evaluations are supposed to answer.

happen, in fact, happen? Answering this question entails measuring what happened, comparing it with what was supposed to happen, and explaining any significant discrepancies between the two states of affairs (Frechtling 2007). This is the logic underlying the evaluative framework proposed here.

Measuring targeting performance quantifies the extent to which targeting respects eligibility. It turns out that all common measures of targeting performance quantify in some sense the association between treatment receipt and eligibility. In general, these measures capture the difference between actual outcomes and counterfactual outcomes that would result from a *neutral* mechanism. A neutral targeting mechanism is blind to eligibility to the extent that it enforces equal treatment by assigning to all claimants on the priority list either equal shares of resources or equal chance of selection into treatment.

The computation of a measure of a targeting outcome is a purely descriptive exercise which produces a statistic indicating the level of performance of the underlying mechanism. A key consideration relates to the strength of this evidence. In other words, there is a need to determine whether the observed level of performance is *significant* both statistically and substantively. Statistical significance limits the likelihood that the observed level of performance is due to chance to the chosen level of significance. Substantive significance or *meaningfulness* is a matter of value judgments. Indeed, a statistically significant measure of performance would be useless in decision-making unless it can be contextualized relative to some frame of reference. Such a frame of reference may take the form of a normative scale for judging the extent to which targeting respects eligibility, or the extent of agreement between two targeting mechanisms.

The evaluation framework presented in this paper also proposes a statistical model of association between targeting outcomes and individual or household characteristics to help identify proximate determinants of observed outcomes. In particular, it is assumed that the contingency table representing targeting outcomes is a manifestation of a *latent bivariate probit process* driven by observable and non-observable individual or household characteristics. Within that framework, one can identify the proximate determinants of targeting outcomes on the basis the *marginal effects* of these characteristics on the probability of the outcomes of interest.

The outline of the paper is as follows. Section 2 presents the evaluation framework. It focuses on *measures* of targeting outcomes and discusses the assessment of the strength of evidence based on such measures. It also explains the use of a biprobit model to identify the drivers of targeting outcomes. Section 3 discusses an empirical illustration using data on a pilot cash transfer program in Northern Cameroon. Section 4 contains concluding remarks.

2. Evaluation Framework

Any effective evaluation begins with well-posed evaluation questions which focus the evaluation and govern its design (Gertler et al. 2016, Ravallion 2009a). The fundamental question driving any targeting assessment may be stated as follows: *To what extent does the* *targeting mechanism respect eligibility?* The search for evidence to answer this question requires that we measure targeting outcomes, compare them to the intended outcomes and provide an explanation for significant discrepancies between the two states of affairs. In other words, there is a need to *measure*, *judge* and *explain* targeting performance. That is an exercise in *associational inference* given that most commonly used indicators of targeting performance quantify the association between targeting outcomes and eligibility. Furthermore, one can identify proximate determinants of targeting performance on the basis of a statistical model of association between outcomes and household (or individual) characteristics.

Measuring Targeting Performance

Measurement is a *quantification* process designed to translate constructs or concepts into their concrete manifestations. Quantification enables decision-makers to set targets and assess progress in achieving them (Allin and Hand 2014). In particular, accurate measurement makes it possible to reliably compare what happened with what was supposed to happen. Measuring targeting performance entails a quantification of the agreement between the targeting outcomes and the intended outcomes based on the eligibility criterion. In other words, measurement in this context quantifies *the extent to which the targeting mechanism respects the eligibility criterion*.

All measures of targeting performance commonly found in the literature are constructed from answers to the following two key questions: (1) Who benefits from the operation of a targeting mechanism? (2) What is the size of the benefit received? One can answer these questions using data from a representative household survey that provides information on the following: (i) eligibility variables, (ii) whether or not the household benefits from the intervention; and (iii) the level of benefits. Whether or not a given household is eligible is represented by a binary variable which takes the value of one if the household is eligible (given the criterion), and zero otherwise. Similarly, benefit receipt is a binary variable (measured on a nominal scale) while the size of the benefit received is measured on a ratio scale.

Given that targeting seeks to focus benefits on the eligible, measures of targeting performance quantify the extent to which targeting outcomes (i.e. answers to the above two key questions) depend on the standard of comparison induced by eligibility. In essence, indicators of targeting performance measure the association between eligibility and benefit receipt. There are two basic categories of measures of targeting performance depending on the information required for their computation. We refer to the first category as *concentration-based indicators*. The second category consists of *frequency-based measures*.

Concentration-based indicators use information on both benefit receipt and the size of the benefit received. Members of this class of measures thus quantify the extent to which a given mechanism concentrates benefits “in the hands” of eligible individuals (or households). Ravallion (2009b) offers an excellent description of these measures which include the concentration curve, the share of benefits going to the eligible, the normalized share as well as the concentration index.

Assume that individuals (or households) are ranked according to some criterion of social desert so that they are ordered from highest to lowest priority. Let p stand for the relative rank of an individual in the priority parade (p is the proportion of individuals with the highest priority). Let $C(p)$ represent the share of benefits going to p. The concentration curve maps[5] $C(p)$ (on the vertical axis) against p (on the horizontal axis) where both p and $C(p)$ vary from 0 to 1. Consider, for instance, a cash transfer intervention targeting the poor. The priority list entails ranking individuals or households from poorest to richest according to some indicator of the living standard (e.g. *per capita* consumption expenditure). In that case, $C(p)$ represents the share of total transfers going to the poorest proportion, p, of the population.

Let $p = h$ for a particular eligibility cutoff point. The *share going to the eligible* is read off the concentration curve as $S(h) = C(h)$. The *normalized share* is equal to $NS(h) = \frac{C(h)}{h}$. This expression says that the normalized share, also known as targeting ratio (or TR) is equal to the percentage of program benefits going to the target group divided by the percentage of the population found in the target group. If the size of the benefit is uniform over the priority list, then $C(p) = p$. This describes the 45 degree line also referred to as the line of equality. The concentration curve shows actual outcomes induced by the targeting mechanism under consideration, while the line of equality represents counterfactual outcomes associated with a neutral targeting mechanism. The concentration index, CI, is defined in terms of the area of concentration, which is the area between the concentration curve and the line of equality. In particular, the concentration index is equal twice the area of concentration[6]. This measure is most conveniently computed in terms of the covariance between benefits and the rank of the recipients in the priority list. The formal expression is: $CI = \frac{2cov(b,p)}{\mu_b}$. This is clearly a measure of association between benefits and eligibility.

When everybody receives the same amount of benefits regardless of her level of priority, the concentration index equals zero[7]. The sign of the concentration index indicates whether or not targeting agrees with eligibility while the magnitude indicate the extent of agreement (or disagreement). A negative CI indicates an inverse relationship between benefits received (b) and the rank (p) in the priority list. Since high priority units have lower rank relative to low priority units, a negative concentration index reflects agreement

[5] This mapping clearly shows that the concentration curve describes the association between eligibility and benefit receipt. Formally, the concentration curve is defined as: $C(p) = \frac{1}{\mu_b}\int_0^p b(t)dt$, where $b(t)$ is the amount of benefit received at quantile t, and μ_b stands for the average benefit.

[6] Let AC stand for the area of concentration. Formally, we have: $AC = \int_0^1 [p - C(p)]\, dp$. Therefore $CI = 2AC$.

[7] The covariance between a variable and a constant is equal to zero. The computational expression of the CI also implies that $AC = \frac{cov(b,p)}{\mu_b}$. It can be shown that $AC = -\frac{cov[b,(1-p)]}{\mu_b}$. Therefore, the concentration index can also be expressed as: $CI = -\frac{2cov[b,(1-p)]}{\mu_b}$.

between targeting and eligibility. Similarly, a positive sign indicates disagreement. The concentration index takes values between -1 and 1.

Frequency-based measures of targeting performance quantify the association between targeting outcomes and eligibility on the basis of frequency counts of outcomes associated with two binary variables representing eligibility and treatment status respectively. Relevant information is usually organized in a contingency table which entails cross-classification of the relevant population by eligibility and treatment receipt. The rest of the paper focuses on this category of indicators.

Table 2.1: Frequency Distribution of Targeting Outcomes

	Participant	Nonparticipant	Total
Eligible	n_{11}	n_{12}	$n_{11} + n_{12}$
Non-Eligible	n_{21}	n_{22}	$n_{21} + n_{22}$
Total	$n_{11} + n_{21}$	$n_{12} + n_{22}$	n

Table 2.1 shows the classification of n subjects by their *eligibility* status and *participation* in the intervention. The table is structured in such a way that rows represent eligibility status which is determined on the basis of policy relevant characteristics of applicants. The columns represent participation which is an outcome of the effective operation of the targeting mechanism. As such, this frequency table is an indicator of targeting outcomes. Its contents are the basic ingredients for other measures of targeting performance. Elements along the main diagonal (i.e. n_{11} and n_{22}) represent cases where the results of the targeting mechanism agree with the eligibility criterion. Indeed, n_{11} participants are eligible and n_{22} nonparticipants are non-eligible. This is exactly what an effective targeting mechanism is expected to do: ensure that eligible participate in the intervention and that non-eligible do not. Off diagonal elements (n_{12} and n_{21}) represent targeting errors. In particular, n_{12}, the number of eligible who do not receive the intervention, is an indicator of *exclusion errors*. Similarly, n_{21} indicates *inclusion errors*. It is the number of non-eligible who get the intervention. The row-totals represent the total number of eligible and the total number of non-eligible respectively. Similarly the column-totals are the total number of participants and nonparticipants respectively.

Table 2.2: Relative Frequency of Targeting Outcomes

	Participant	Nonparticipant	Total
Eligible	p_{11}	p_{12}	$p_{11} + p_{12}$
Non-Eligible	p_{21}	p_{22}	$p_{21} + p_{22}$
Total	$p_{11} + p_{21}$	$p_{12} + p_{22}$	1

Dividing each cell in table 2.1 by n turns the contents into relative frequencies which can be interpreted as probabilities. Also, both tables 2.1 and 2.2 are subject to an adding-up constraint to the extent that all four cells of table 2.1 must add-up to n while those of table

2.2 must add up to 1. This constraint has an interesting implication. In terms of frequencies presented in table 2.1, if the total number of eligible is equal to the total number of participants then the number of eligible excluded (n_{12}) from the program must equal the number of non-eligible included (n_{21}). Similarly for table 2.2, if the proportion of eligible is equal that of participants then the proportion of eligible who are excluded (p_{12}) must equal the proportion of non-eligible included (p_{21})[8].

A variety of measures of targeting performance are defined in terms of the information contained in table 2.1 or in table 2.2. Sumarto and Suryahadi (2001) use many of these measures in their analysis of the targeting outcomes of a group of Indonesian Social Safety Net programs. The targeting success rate is defined as the proportion of cases where targeting outcomes agree with the eligibility criterion. This indicator can be expressed as: $TSR = \frac{(n_{11}+n_{22})}{n} = p_{11} + p_{22}$. Similarly, the overall targeting error rate is the proportion of cases where the results of targeting disagree with the eligibility criterion. Formally, we write: $TER = \frac{(n_{12}+n_{21})}{n} = p_{12} + p_{21}$.

Other widely used indicators include the coverage rate, under-coverage rate, the leakage rate and the targeting differential. The coverage rate (CR) is the rate of participation of the eligible, that is: $CR = \frac{n_{11}}{n_{11}+n_{12}}$. By definition, the under-coverage rate (UR) is equal to one minus the coverage rate. The leakage rate (LR) is the proportion of participants who are not eligible. In other words, $LR = \frac{n_{21}}{n_{11}+n_{21}}$. This is a manifestation of errors of inclusion. The under-coverage rate and the leakage rate are equivalent to what Brown et al. (2016) call the *Exclusion Error Rate* (EER) and the *Inclusion Error Rate* (IER) respectively. Finally, the targeting differential is defined as the difference between the coverage rate and the leakage rate (Ravallion 2009b), that is : $TD = CR - LR$.

To fix ideas, we illustrate the computation of some of these indicators using the targeting outcomes for PROGRESA[9] presented in table 2.3. Application of the formulae defined above shows that the targeting success rate is equal to 72.5 percent, while the overall targeting error rate is about 27.5 percent. The Exclusion Error Rate is equal to 26.45 percent. The Inclusion Error Rate is about 26.44 percent, which is basically the same as the EER. As noted above, this is a consequence of the fact that the total number of eligible households is almost equal to that of participating households.

[8] Brown et al. (2016) explain this point in the context of their assessment of the performance of proxy means testing (PMT) in identifying the poor. Kidd and Wylde (2011) make the same point intuitively. They explain that when the program size (i.e. the number of participants) is equal to the total target population, if k% of eligible are excluded, then the other k% included (besides the $(100-k)$% of eligible who remain in the program) must not have been eligible. Thus both exclusion and inclusion error rates are equal to k%.

[9] PROGRESA (Programa de Educación, Salud, y Alimentación) in Mexico. This program, based on an integrated approach to poverty reduction (involving education, health and nutrition), started in 1997 focusing initially on rural poor. It was extended to urban poor in 2002 when its name changed to *Oportunidades*.

Table 2.3: Targeting Outcomes for PROGRESA (households)

	Participant	Nonparticipant	Total
Eligible	9209	3312	12521
Non-Eligible	3310	8246	11556
Total	12519	11558	24077

Source: Skoufias et al. (1999)

The values of the above frequency-based indicators depend critically on the extent to which the targeting mechanism respects the eligibility criterion. For instance, when the targeting mechanism is in total agreement with eligibility, only eligible persons receive the intervention and ineligible do not. In that case there would be no targeting errors. The TSR would be equal to one while the TER would be equal to zero. Similarly, when there is total disagreement between the targeting mechanism and eligibility, the TER would be equal to 1 and the TSR would be zero. These indicators are therefore manifestation of the association between targeting outcomes and eligibility.

In the context of a contingency table, one can measure the association between the classificatory variables on the basis of some statistics linked to the Chi-square statistic which entails a comparison of the observed frequency in each cell with the frequency that would be expected under some null hypothesis, e.g. that the relevant classificatory variables are independent. In particular, the observed frequencies are the actual outcomes produced by the prevailing targeting mechanism while the expected frequencies are the counterfactual outcomes that would result from a neutral mechanism (i.e. random assignment).

The standard Pearson's Chi-square statistic is defined as follows.

$$\chi^2 = \sum_i \sum_j \frac{\left(n_{ij}-\mu_{ij}\right)^2}{\mu_{ij}} \qquad (2.1)$$

where n_{ij} is the observed frequency in the cell defined by row i and column j, and μ_{ij} is the corresponding expected cell frequency under the null hypothesis of independence. This test statistic asymptotically follows a Chi-square distribution with the number of degrees of freedom equal to $df = (r-1) \times (c-1)$, where r is the number of rows and c the number of columns in the table. Clearly, there is only one degree of freedom associated with a 2x2 table.

The *phi coefficient* (φ) is a well-known measure of association for two-way contingency tables. This coefficient is defined as the squared root of the normalized Chi-square. That is the Chi-square statistic divided by n, where n is interpreted as the maximum value of Chi-square (Acock, 2016). In other words, $\varphi = \pm\sqrt{\frac{\chi^2}{n}}$. Cramer's V is the appropriate measure of association for larger tables. In that case, the maximum value achievable by Chi-square is equal to n times the smaller of $(r-1)$ or $(c-1)$. Thus, Cramer's V is equal to $V = \pm\sqrt{\frac{\chi^2}{n \times min(r-1,c-1)}}$. For 2x2 tables, Cramer's V and the phi coefficient are equivalent.

All measures of targeting performance discussed above are summary statistics that characterize performance on the basis of a single number. Their aggregative nature may hide some of the information that can help one get a fuller appreciation of the performance of the mechanisms involved. A possibly more informative way to describe targeting performance is to consider how a mechanism assigns treatment across the priority list based on the distribution of the variable defining eligibility. For instance, when eligibility is based on an indicator of the living standard such as *per capita* consumption expenditure, we can describe the performance of a targeting mechanism in terms of the probability that the mechanism selects a household given that household's standard of living. This probability is essentially a *propensity score*[10].

Let y stand for the outcome of the assignment by a targeting mechanism. It is a binary variable that is equal to 1 if the unit (e.g. household) is selected by the mechanism, and 0 otherwise. Let the variable x be the basis of eligibility (e.g. *per capita* consumption expenditure). The idea is to analyze targeting performance on the basis of the relationship between the assignment outcome variable, y, and the eligibility variable x. Angrist and Pischke (2009) explain that the conditional expectation function (CEF) is the best predictor of an outcome variable, say y, given x. Because the assignment outcome is a binary variable, the CEF is a conditional probability of assignment, hence a propensity score which we interpret as a *local indicator of targeting performance*.[11]

We propose to estimate this propensity score locally using kernel smoothing methods[12]. Let $p(x) = \Pr(y = 1|x)$ be the *assignment probability* conditional on x. This is the conditional expectation function of y given x. Given a sample of n observations, we can generally write a local estimator of that regression at a specific point x as follows.

$$\hat{p}(x) = \sum_i^n w_i(x, x_i) y_i \tag{2.2}$$

where $w_i(x, x_i)$ is some weighing function such as a kernel function. In the context of kernel smoothing used in this paper, the performance indicator defined by equation (2.2) is a weighted average of unit-level assignment outcomes (y_i) for all observations within the smoothing window around the focal value x[13]. Kernel averaging can be interpreted as

[10] In the context of treatment effect analysis, the propensity score for a unit is the conditional probability of assignment to treatment. In other words, it is the probability of selecting into treatment given a set of covariates which measure the observable characteristics of the unit under consideration.
[11] The indicator is local because it depends on specific values of the eligibility variable. Furthermore, the underlying CEF will be locally estimated using nonparametric kernel regression methods.
[12] In many situations, global parametric models of association between variables do not fit the data very well. In such cases, nonparametric smoothing methods offer a better alternative to the extent that they allow the fitted curve to take a more general form that might better fit the data at hand (Loader 2005). In other words, smoothing methods allow the data "*to speak for themselves*" as much as possible.
[13] The estimation method used in this paper relies on kernel averaging, which involves sliding a "*window*" or band across the data along the x-axis and taking some average of the outcome variable (y) for all observations in the window. The result is an estimate of the value of the outcome variable associated with the focal point of the window. One may define a simple smoothing window as the open interval $(x - h, x + h)$ where h is the bandwidth. The choice of the size of the *bandwidth* determines how close to the focal point x observations have

running a weighted regression of an outcome variable on a constant with the weights given by a kernel function[14]. The kernel estimator defined by (2.2) is known as the Nadaraya-Watson estimator (Ahamada and Flachaire 2010).

The propensity scores defined by (2.2) are computed over the entire distribution of the eligibility variable, x. We can therefore plot these scores against x (or a monotonic transformation of x) and obtain a curve which we call the *assignment probability curve*. It is a mapping of the probability of selection (by the mechanism) onto the range of variation of the eligibility variable. The curve describes the response of the targeting mechanism to variations in the eligibility threshold. Like any other measure of targeting performance discussed so far, this one quantifies the association between targeting outcomes and eligibility.

To see how this curve can inform targeting assessment, let x_0 be an eligibility cutoff such that any unit (e.g. household) is considered eligible if $x \leq x_0$, otherwise it is not eligible. When the targeting mechanism is in perfect agreement with the eligibility criterion, the assignment probability curve will be a step function such that $\hat{p}(x) = 1$ for all eligible units (i.e. with $x \leq x_0$), and 0 for all other units (i.e. units with $x > x_0$). This frame of reference suggests that there will be errors of exclusion whenever the assignment probability curve is less than one in the eligibility region ($x \leq x_0$). We also refer to the eligibility region as the *inclusion region*. There will be errors of inclusion whenever the curve is greater than zero outside the eligibility region ($x > x_0$), which we refer to as the *exclusion region.* The shape of the assignment probability curve therefore provides valuable information on the local behavior of the underlying targeting mechanism. One can compare targeting mechanisms by plotting the corresponding assignment probability curves on the same graph. We demonstrate this in the empirical section.

Considering the Strength of Evidence

A targeting assessment generates evidence on the extent to which a targeting mechanism respects eligibility. The reliability of conclusions and recommendations emerging from such an assessment depends on the strength of that evidence. This creates the need to examine the strength of evidence based on observed measures of performance. The computation of a measure of targeting performance produces a statistic which is the sample evidence about the relationship between the targeting mechanism and eligibility in the population of interest. Assuming that the underlying sample data are valid and reliable, the observed value of a measure of targeting performance is determined by sampling variation and the true strength of the association between targeting and eligibility. Sampling

to be in order to contribute to the computation of the average outcome at that point. In principle, observations that are far from the focal point receive smaller weights relative to observations that are near it (Deaton 1989).

[14] A kernel function satisfies the following conditions: (i) it is positive, (ii) integrates to unity over the smoothing window, (iii) it is symmetric around zero so that the points below the focal point get the same weight as those located an equal distance above that point, and (iv) it is decreasing in the absolute value of its argument. The two most commonly used kernel functions in applied work are the Epanechnikov and the Gaussian. Our empirical results are based on the Epanechnikov kernel.

variation therefore introduces uncertainty about the *existence* and the *strength* of the association of interest. There is indeed the risk of concluding that a relationship exists between targeting and eligibility while in fact none exists. Similarly, one may erroneously conclude that there is no such a relationship while in fact there is one.

Assessing the *strength of evidence* in this context entails answering two basic questions: (i) Is there a relationship between targeting and eligibility? (ii) If the relationship appears to exist, how meaningful is it? To answer the first question, one needs to determine the extent to which actual targeting outcomes differ from counterfactual outcomes associated with a neutral mechanism. This may take the form of testing the null hypothesis that there is no difference between actual and counterfactual outcomes, meaning there is no relationship between targeting and eligibility. The alternative hypothesis would be that such a relationship exists. This is essentially what the Chi-square test of independence does in the context of contingency tables. Failure to reject the null hypothesis at the chosen level of significance implies that the targeting mechanism does not respect eligibility at all. Rejecting the null hypothesis means that, there is a relationship between targeting and eligibility. In other words, there is a statistically significant relationship between targeting and eligibility. Further specification is needed to determine whether the relationship reflects agreement or not.

A statistically significant relationship is not necessarily meaningful. Statistical significance relates to the probability that the observed relationship is due to chance while meaningfulness is a matter of value judgments or substantive significance. This is where the second question comes in. Assessing the strength of evidence therefore entails both a statistical test of significance and a normative scale for determining meaningfulness. The meaningfulness of a targeting mechanism could also be assessed in terms of its contribution to fairness, effectiveness and efficiency.

In the case of *concentration-based measures* of targeting performance, considering the strength of the association between targeting and eligibility entails establishing the statistical significance of the observed measure relative to a *distribution-neutral* allocation of benefits, and placing the relevant indicator on a normative scale to assess its strength. We know that when the concentration curve is equal to the line of equality, the area of concentration and the concentration index are both equal to zero, and the normalized share is equal to one. This would indicate that the targeting mechanism is blind to eligibility. The farther the concentration curve lies above the line of equality the better the targeting. The benefits are more concentrated among individuals or households with the highest priority. The concentration index is negative in this region. When the concentration curve lies below the line of equality, the concentration index is positive meaning that targeting disagree with eligibility. O'Donnell et al. (2008) show how to test statistically dominance relations among concentration curves. They also demonstrate how to estimate concentration indices and test their statistical significance.

Coady, Grosh and Hoddinott (2004a) conducted a meta-analysis of targeting performance across 122 targeted antipoverty interventions implemented in 1985-2000 in 48 countries. Their analysis is based on the normalized share as an indicator of targeting performance. They found that targeting performance, as measured by this indicator, ranged from 0.26 to 4.00 with a median value of 1.25. The top ten performers have scores in the range of 2.02 to 4.00. This distribution of targeting ratios provides a scale for assessing the meaningfulness (or practical importance) of evidence based on the observed targeting ratio in a particular situation. We refer to this scale as the Coady-Grosh-Hoddinott or CGH scale.

Sabates-Wheeler et al. (2015) use this scale to assess the targeting performance of the Hunger Safety Net Program (HSNP) in Northern Kenya. In that context, observed performance based on the TR varied from 1.07 when targeting on the basis of an indicator of food insecurity to 1.12 when targeting on the basis of consumption poverty. These authors conclude that is poor performance because it falls below the median value on the CGH scale. In particular, they concluded HSNP targeting is *mildly* pro-poor.

When it comes to *frequency-based indicators*, assessing the strength of evidence is analogous to the approach followed in the case of concentration-based measures. It entails comparing observed targeting outcomes to those generated by a random assignment mechanism, and referring to some scale to determine substantive significance. In particular, one needs to check whether or not the targeting mechanism and eligibility are independent processes. If so, any observed agreement between the two would be due to chance. If the two processes are *contingent* (i.e. not independent), the second step is to place the appropriate indicator on a scale to determine the value of the evidence provided by the observed association between the two processes.

Based on information provided in a *contingency table* such as table 2.1, we start with a Chi-square test of independence to determine whether there is any statistically significant association between the two categorical variables representing the eligibility and targeting processes. Failure to reject the null hypothesis that the two variables are independent leads to the conclusion that the targeting mechanism is ineffective. Rejection of the hypothesis means that the mechanism is somewhat effective[15]. In the second step, we ascertain the degree of effectiveness (or the meaningfulness of the association between the two variables) on the basis of the value of some measure of association (e.g. φ) and a normative scale.

Acock (2016) proposes a scale for judging the substantive significance of the association between two binary variables. According to that scale, a φ with an absolute value between 0.0 and 0.19 reflects a *weak* association. The association is considered *moderate* for values between 0.2 and 0.49. Any value above 0.50 represents a *stron*g association. In the case of targeting outcomes for PROGRESA presented in table 2.3 the Chi-square test indicates a very high level of statistical significance for the association between eligibility and participation. The null hypothesis of independence is rejected at a level below 0.0001.

[15] See Essama-Nssah and Sarr (2015) for an application of this idea in the context of their targeting assessment for a secondary school stipend program in Nepal.

However, on the basis of Acok's scale that association must be considered as moderate because $\varphi = 0.45$.

Another way to assess the extent to which targeting respects eligibility is to frame the evaluation within the logic of *interobserver agreement* and pass judgment on the basis of the Cohen's Kappa, and the scale proposed by Landis and Koch (1977). The issue of interobserver agreement arises in situations where two or more observers classify the same set of subjects (or objects) in two or more categories based on an examination of relevant characteristics. An example would be two doctors diagnosing a given disease for the same group of patients based on lab results. For each patient, each doctor renders a verdict on whether or not the patient has the disease in question. It is of interest to quantify and interpret the extent to which both doctors (observers) agree about the diagnoses they produce. This leads to the notion of an agreement coefficient indicating the extent to which the observers agree.

It is desirable that an agreement coefficient account for the fact that there may be cases where the observers agree by chance. Such cases must be factored out because agreement by chance is not informative about the validity of the screening process. The *Kappa coefficient* is a chance-corrected agreement coefficient that removes from the observed agreement the amount due to chance and standardizes the result by dividing it by the extent of agreement which is not expected to occur by chance.

To see clearly what is involves, we use table 2.1 to assess the extent of agreement between two methods for determining program eligibility. For instance, column-outcomes could indicate eligibility based on information provided by applicants on administrative forms while row-outcomes could be based on direct observation through a household survey. The same logic applies to the comparison of a targeting outcomes with eligibility. Let p_a stand for the percent agreement between the two processes. On the basis of table 2.1, this is equal to TSR (targeting success rate). In other words: $p_a = \frac{(n_{11}+n_{22})}{n}$. *If participation is independent of eligibility, their joint probability is equal to the product of the marginal probabilities*. This fact implies that the expected percent agreement, p_e, is equal to the following expression.

$$p_e = \left[\left(\frac{n_{c1}}{n}\right) * \left(\frac{n_{r1}}{n}\right) + \left(\frac{n_{c2}}{n}\right) * \left(\frac{n_{r2}}{n}\right)\right] \tag{2.3}$$

where $n_{ri} = (n_{i1} + n_{i2})$ for $i = 1, 2$, and $n_{cj} = (n_{1j} + n_{2j})$ for $j = 1,2$.

Chance-agreement correction entails subtracting this value from p_a. Furthermore, $(1 - p_e)$ indicates the amount of agreement which is not expected to occur by chance. Hence, the Kappa coefficient is defined as follows.

$$K = \frac{p_a - p_e}{1 - p_e} \tag{2.4}$$

Table 2.4: Landis and Koch (1977) Interpretation of the Kappa Statistic

Kappa	**Quality of Agreement**
Under 0.20	Poor
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost Perfect

Source: Adapted from Viera and Garrett (2005)

Referring back to the case of PROGRESA reported in table 2.3, we find that participation agrees with eligibility 72.5 percent of the time. Agreement by chance would occur about 50 percent of the time. The corresponding Kappa coefficient is therefore equal to 0.45. *How good is such a performance?* The Kappa statistic can vary from -1 to 1. The value of 1 indicates perfect agreement while 0 is the expected value when agreement is purely by chance. A negative value indicates systematic disagreement between the observers (Viera and Garrett 2005). The quality of agreement indicated by the Kappa statistic can be determined on the basis the Landis-Koch (LK) scale presented in table 2.4. According to this scale, there is a *moderate* agreement between eligibility and the targeting mechanism used to select participants in PROGRESA. Given the possible range of values for the *phi coefficient*, one could also use the LK scale to value φ. On the basis of that scale, a phi coefficient of 0.45 also represents a moderate association between eligibility and participation.

Understanding Targeting Outcomes

So far, the formulation of the evaluative framework has focused on *measuring* and *judging* targeting performance in terms of the association between eligibility and participation. This association ultimately depends on the characteristics of individuals or households (depending on the unit of analysis). We may therefore base our understanding of targeting outcomes on a model of association linking such outcomes to individual (or household) characteristics. According to Pearl (2009), association is any relationship that can be defined on the basis of a *joint probability distribution* of observed variables. Thus statistical association is quantified by features of joint distributions such as *conditional probabilities* and *conditional averages* also known as regressions. We frame the identification of the proximate determinants of targeting performance within the logic of the bivariate probit model. As noted earlier, we assume targeting outcomes are a manifestation of a *latent bivariate probit process* driven by observable and non-observable individual or household characteristics.

Let y_1 and y_2 be two binary indicators of eligibility and participation (or treatment receipt based on the targeting mechanism used to select beneficiaries) respectively. These indicator variables take on the values of 0 or 1 depending on whether or not some latent variables y_1^* and y_2^* cross some thresholds. In particular, for an eligible individual, we have $y_1 = 1$ if $y_1^* > 0$, otherwise the person is ineligible and $y_1 = 0$. Similarly for participation,

$y_2 = 1$ if $y_2^* > 0$, and 0 otherwise. To tie these outcomes to observable individual characteristics x, we assume that each latent variable can be expressed as a function of the observable characteristics and some random disturbance, ε. In particular, we write the system of equations for the latent variables as follows.

$$y_1^* = x_1\beta_1 + \varepsilon_1 \quad \text{and} \quad y_2^* = x_2\beta_2 + \varepsilon_2 \tag{2.5}$$

where, ε_j is a standard normal variate[16] for $j = 1, 2$. The degree to which the error terms are associated with each other is measured by the correlation coefficient, ρ, which Greene (2008) refers to as the conditional *tetrachoric correlation*. It is the correlation that would be measured between the underlying latent variables if they could be observed.

These assumptions imply that ε_1 and ε_2 follow a bivariate normal distribution, the density function[17] of which is commonly denoted by $\phi_2(\cdot)$. The corresponding cumulative distribution function or CDF is denoted by: $\Phi_2(\cdot)$. We will write the density function and the cumulative distribution function for the univariate normal respectively as ϕ and Φ.

There are four possible outcomes associated with the bivariate probit model, namely: $(y_1, y_2) = (1,1)$; $(y_1, y_2) = (1,0)$; $(y_1, y_2) = (0,1)$; and $(y_1, y_2) = (0,0)$. When these are interpreted as targeting outcomes, the first and the last outcomes represent successful targeting corresponding to the elements of the main diagonal of table 2.1 or table 2.2. The second and third outcomes are targeting errors. In particular, $(y_1, y_2) = (1,0)$ is an error of exclusion while $(y_1, y_2) = (0,1)$ represents an error of inclusion. The probability of each of these possible outcomes can be inferred from the structure of the bivariate probit model in a manner analogous to the univariate probit model. For instance, the probability of an error of exclusion, that is $(y_1, y_2) = (1, 0)$, is given by the following joint probability: $P_{10} = Pr(\varepsilon_1 > -x_1\beta_1, \varepsilon_2 \leq -x_2\beta_2)$. The other joint probabilities are similarly defined.

Table 2.5: Bivariate Probit Model of Targeting Outcomes

	Participation ($y_2 = 1$)	**Nonparticipation ($y_2 = 0$)**	**Marginal**
Eligibility ($y_1 = 1$)	$\Phi_2(x_1\beta_1, x_2\beta_2, \rho)$	$\Phi_2(x_1\beta_1, -x_2\beta_2, -\rho)$	$\Phi(x_1\beta_1)$
Ineligibility ($y_1 = 0$)	$\Phi_2(-x_1\beta_1, x_2\beta_2, -\rho)$	$\Phi_2(-x_1\beta_1, -x_2\beta_2, \rho)$	$1 - \Phi(x_1\beta_1)$
Marginal	$\Phi(x_2\beta_2)$	$1 - \Phi(x_2\beta_2)$	1

Greene (2008) proposes a compact representation of these probabilities in terms of the CDF, $\Phi_2(\cdot)$. Let $q_j = 2y_j - 1, j = 1, 2$. This implies that $q_j = -1$ when $y_j = 0$, and $q_j = 1$ when $y_j = 1$. Furthermore, let $\rho^* = q_1 q_2 \rho$. These transformations ensure that each argument of the CDF carry the right sign consistent with the observed value of the

[16] In other words, ε_j follows a normal distribution with mean 0 and variance equal to 1.

[17] It can be shown that if z_1 and z_2 are standard normal variates with a correlation coefficient equal to ρ, then their joint density function can be expressed as: $\phi_2(z_1, z_2, \rho) = \frac{1}{2\pi\sqrt{(1-\rho^2)}} exp\left(-\frac{1}{2}\left(\frac{z_1^2 + z_2^2 \_ 2\rho z_1 z_2}{1-\rho^2}\right)\right)$.

corresponding binary variable[18]. Furthermore, let $v_1 = q_1 x_1 \beta_1$ and $v_2 = q_2 x_2 \beta_2$. The probabilities of the four possible outcomes can be derived from the following expression for $s = 0, 1$ and $t = 0,1$.

$$P_{st} = \Pr(y_1 = s, y_2 = t) = \Phi_2(v_1, v_2, \rho^*) \tag{2.6}$$

These probabilities are specified in table 2.5, which is considered a statistical model underlying the relative frequencies presented in table 2.2. Measures of targeting performance based on contingency tables are functions of some elements of table 2.5. One can identify the effects of covariates on targeting outcomes by studying the effects of such covariates on the relevant probabilities presented in table 2.5. In particular, the *marginal effect* associated with a given covariate, say x_{jk}, $(j = 1, 2)$ is an estimate of the change in the relevant probability induced by a change in x_{jk} holding all other variables constant.

The marginal effect of x_{jk} on the joint probability defined in equation (2.6) is given by the following expression.

$$\frac{\partial P_{st}}{\partial x_{jk}} = \frac{\partial \Phi_2(v_1, v_2, \rho^*)}{\partial x_{jk}} = \frac{\partial \Phi_2}{\partial v_1}\frac{\partial v_1}{\partial x_{jk}} + \frac{\partial \Phi_2}{\partial v_2}\frac{\partial v_2}{\partial x_{jk}} \tag{2.7}$$

where $\frac{\partial \Phi_2}{\partial v_1} = \phi(v_1)\Phi\left(\frac{v_2 - \rho^* v_1}{\sqrt{1-\rho^{*2}}}\right)$, $\frac{\partial \Phi_2}{\partial v_2} = \phi(v_2)\Phi\left(\frac{v_1 - \rho^* v_2}{\sqrt{1-\rho^{*2}}}\right)$ and that $\frac{\partial v_j}{\partial x_{jk}} = q_j \beta_{jk}$ for $j = 1, 2$. To simplify the notation, we let $\omega_j = \frac{\partial \Phi_2}{\partial v_j}$ for $j = 1, 2$. Equation (2.7) can therefore be rewritten compactly as[19]:

$$\frac{\partial \Phi_2(v_1, v_2, \rho^*)}{\partial x_{jk}} = (q_1 \omega_1)\beta_{1k} + (q_2 \omega_2)\beta_{2k} \tag{2.8}$$

Note that, the coefficient β_{jk} will be zero if the variablex_{jk} is not part of x_j, for $j = 1, 2$.

Equation (2.8) describes a general expression from which one can derive all marginal effects for the 4 joint probabilities presented in table 2.5. Specific expressions are obtained by using the values of q_1 and q_2 corresponding to the outcome of interest. For instance, if one is interested in the probability of committing an error of exclusion, then expression (2.8) must be evaluated at $(q_1, q_2) = (1, -1)$ since these values correspond to the outcome: $(y_1, y_2) = (1,0)$. Similarly, for the probability of committing an error of inclusion, the expression should be evaluated at $(q_1, q_2) = (-1, 1)$ because the underlying outcome is: $(y_1, y_2) = (0,1)$.

Within this framework, the exclusion error rate and the inclusion error rate are conditional probabilities. By definition, $EER = P\,r(y_2 = 0|y_1 = 1) = \frac{Pr(y_1=1, y_2=0)}{Pr(y_1=1)}$.

[18] The following is the mathematical expression of the CDF: $\Phi_2(v_1, v_2, \rho^*) = \int_{-\infty}^{v_1}\int_{-\infty}^{v_2} \phi_2\,(z_1, z_2, \rho^*) dz_2 dz_1$

[19] The derivation of the marginal effects for the joint probabilities associated with the biprobit model relies on the following: (i) the chain rule of differentiation, (ii) total differentiation, (iii) the rule for differentiation under an integral, and (iv) the factorization of the joint density function as: $\phi_2(z_1, z_2, \rho) = \phi(z_1)\phi(z_2|z_1) = \phi(z_2)\phi(z_1|z_2)$.

Similarly, $IER = Pr(y_1 = 0|y_2 = 1) = \frac{Pr(y_1=0,y_2=1)}{Pr(y_2=1)}$. In terms of the biprobit model, we have the following expressions: $EER = \frac{\Phi_2(v_1,v_2,\rho^*)}{\Phi(v_1)}$ for $(q_1, q_2) = (1, -1)$, and $IER = \frac{\Phi_2(v_1,v_2,\rho^*)}{\Phi(v_2)}$ for $(q_1, q_2) = (-1, 1)$.

The marginal effect of x_{jk} on the *Exclusion Error Rate* is defined by the following partial derivative.

$$\frac{\partial EER}{\partial x_{jk}} = \frac{\frac{\partial \Phi_2(v_1,v_2,\rho^*)}{\partial x_{jk}}\Phi(v_1) - \frac{\partial \Phi(v_1)}{\partial x_{jk}}\Phi_2(v_1,v_2,\rho^*)}{[\Phi(v_1)]^2} \tag{2.9}$$

Building on equation (2.8), we can show that:

$$\frac{\partial EER}{\partial x_{jk}} = \left[\frac{\omega_1}{\Phi(v_1)} - \frac{\phi(v_1)\Phi_2(v_1,v_2,\rho^*)}{[\Phi(v_1)]^2}\right](q_1\beta_{1k}) + \frac{\omega_2}{\Phi(v_1)}(q_2\beta_{2k}) \tag{2.10}$$

A specific expression of this marginal effect is obtained by letting $(q_1, q_2) = (1, -1)$.

Similarly, the marginal effect of the same covariate (x_{jk}) on the *Inclusion Error Rate* is equal to:

$$\frac{\partial IER}{\partial x_{jk}} = \frac{\omega_1}{\Phi(v_2)}(q_1\beta_{1k}) + \left[\frac{\omega_2}{\Phi(v_2)} - \frac{\phi(v_2)\Phi_2(v_1,v_2,\rho^*)}{[\Phi(v_2)]^2}\right](q_2\beta_{2k}) \tag{2.11}$$

In this case, the expression should be evaluated at $(q_1, q_2) = (-1, 1)$.

Our frame of analysis of targeting outcomes is analogous to the one used by Stoeffler et al. (2016). The main difference between the two frameworks lies in the statistical models used to identify the drivers of targeting outcomes. Stoeffler et al. (2016) rely on *univariate probit regressions* to model the probabilities of targeting errors as functions of household characteristics. For a given household, they define the exclusion error as a dummy variable that is equal to one if the household is poor (i.e. eligible) and is not selected by the mechanism under consideration, and zero otherwise. They then run the corresponding univariate probit only on the sub-sample of poor households. Similarly, the inclusion error is a dummy variable that is equal to one if a non-poor is selected for treatment, and zero otherwise. The relevant model is also run only on the sub-sample of non-poor (i.e. non-eligible) households.

Finally, to analyze the extent of agreement between two mechanisms such as CBT and PMT, those authors use a multinomial logit model to analyze the four possible outcomes, $(y_1, y_2) = (s, t)$ for $s = 0{,}1$, and $t = 0{,}1$. Within that framework, they select $(y_1, y_2) = (1{,}1)$ as the base or reference category. Compared to that approach, the biprobit model provides a unifying framework for the analysis of targeting outcomes since the model is a straightforward interpretation of a contingency table.

3. Empirical Considerations

In this section, we apply the evaluation framework described above to assess the performance of some targeting mechanisms in the context of the Social Safety Nets Pilot

Project (SSNPP) in Cameroon. This pilot program of cash transfers targeted to the poor was launched in December 2013 (Stoeffler et al. 2016). The project area, selected on the basis of geographic targeting, covers the poorest 15 villages in *Soulédé-Roua* (an *arrondissement* in Northern Cameroon[20]). The project used a *hybrid method*[21] combining community based targeting (CBT) and proxy means testing (PMT) to select beneficiary households who receive a monthly transfer of 15,000 CFA Francs. This amount of money represents about 20 percent of the average poor household consumption expenditure. There is also information about a version of *self-targeting* whereby households determine their own poverty status. We first assess the performance of these mechanisms, and then attempt to identify the proximate determinants of some of the observed outcomes.

Targeting Performance

A targeting mechanism is designed to assign eligibility to treatment (i.e. the intervention). The following question is therefore a key consideration in a targeting assessment: *To what extent does the targeting mechanism under evaluation respect the eligibility criterion*? Given that the program is targeted to the poor, eligibility is based on household *per capita* consumption expenditure relative to the poverty line. Thus eligible households are those with *per capita* expenditure less than or equal to the poverty line. Following Stoeffler et al. (2016), we consider two different cases. First, we set the poverty line in local currency at 96,880 CFA Francs. This implies that 67.5 percent of the households in the sample are poor. Let H stand for the poverty rate among households. We characterize the first scenario as follows: $H = 67.5\%$. The second poverty line, which implies a poverty rate of 35 percent among sample households ($H = 35\%$), is equal to 63,934 CFA Francs. This defines the second scenario.

Table 3.1: Targeting Outcomes under Scenario 1: H=67.5%

Measure	Self-Targeting	CBT	PMT
Targeting Success Rate	0.50	0.54	0.67
Cohen's Kappa	-0.01	-0.05	0.25
Cramer's V	-0.01	-0.05	0.25
Exclusion Error Rate	0.49	0.34	0.25
Inclusion Error Rate	0.33	0.34	0.25

Source: Author's calculations

Table 3.1 presents some aggregate measures of the targeting performance for three mechanisms under scenario one: self-targeting, CBT and PMT. Each of the measures indicates, in some sense, the extent to which each mechanism agrees with eligibility. The

[20] Cameroon is divided into 10 regions administered by a governor and a regional council. Each region is subdivided into administrative units known as *départements* (in French). Each *département* is further subdivided into *arrondissements* each of which may also be divided into *districts*.

[21] The method worked as follows. A survey among households deemed eligible by the community produced necessary data to compute PMT scores. A cutoff score was chosen to determine a list of potential beneficiaries that was submitted to the community for validation. Transfers were then distributed to the validated beneficiaries. See Stoeffler et al. (2015) for details.

targeting success rate ranges from 50 percent for self-targeting to 67 percent for the PMT mechanism. It is equal to 54 percent for community based targeting. The agreement between these mechanisms and eligibility seems substantial when looking at the targeting success rate in the absolute. However, both chance-corrected measures (Cohen's Kappa and Caramer's V) indicate that observed agreement is no different from a chance-agreement, particularly in the cases of self-targeting and CBT. The underlying results, not shown here, reveal that the expected agreement due to chance is equal to 50.56 percent for self-targeting and 56 percent for CBT. The corresponding Kappa coefficients are not statistically significant. According to the Landis-Koch scale, the agreement between these two mechanisms and eligibility is poor. Both chance-corrected measures are equal to 25 percent for PMT. This represents only a *fair agreement* between PMT and eligibility on the basis of the Landis-Koch scale. This level of agreement would be considered moderate on Acock's scale.

The targeting error rates presented in table 3.1 are computed at the household level[22] (i.e. the household is our unit of analysis). The exclusion error rate is equal to the inclusion error rate for both CBT and PMT. Again, this is a consequence of the fact that both mechanisms select the same number of eligible households subject to the adding-up constraint that a 2x2 contingency table must respect. Indeed, the scenario is constructed in such a way that the CBT and PMT mechanisms select the same number of households as poor. That is, 1163 out of 1723 (i.e. 67.5 percent of the households). This is also the same number of households considered poor on the basis of the poverty line defining the scenario. Self-targeting selects only 51.6 percent of the households. As a result, the exclusion error rate differs from the inclusion error rate in that case.

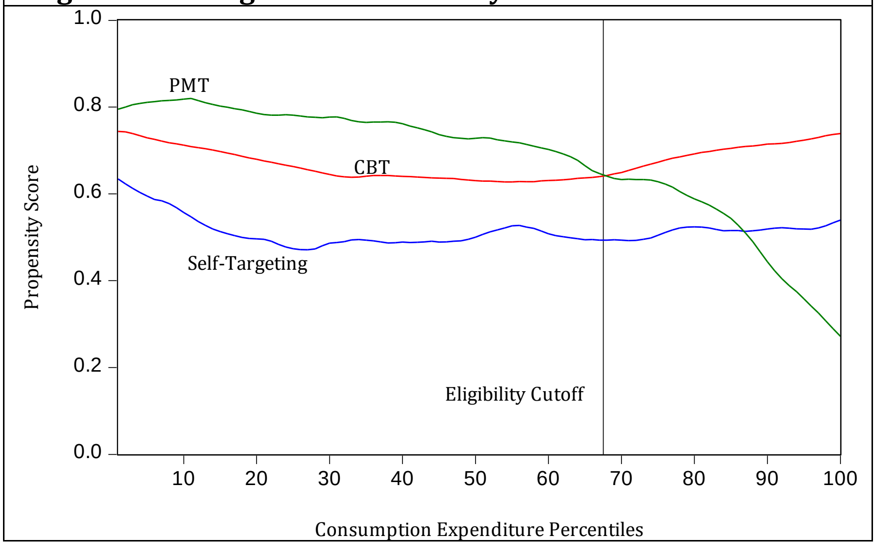
Overall, PMT outperforms the other two mechanisms for all measures presented in table 3.1. On the basis of the same information, CBT performs better than self-targeting, particularly in terms of the exclusion errors. The exclusion error rate associated with CBT is much lower (34 percent) than the one associated with self-targeting (49 percent). The inclusion error rate is about the same for both mechanisms.

Figure 3.1 shows assignment probability curves for the three mechanisms (PMT, CBT and Self-Targeting). The vertical line marks the eligibility cutoff point defining scenario 1 on the basis of the underlying poverty line. The eligibility region or *inclusion region* lies to the left of the cutoff point. Households to the right of the eligibility cutoff are not eligible. They fall in the *exclusion region*. For each mechanism, the assignment probability shows how the probability of declaring a household eligible for a cash transfer depends on the rank of the household in the distribution of per capita consumption expenditure. The configuration of

[22] Stoeffler et al. (2016), using the same dataset, report targeting errors for the CBT and PMT mechanisms that are different from ours because they use an inflation factor to obtain estimates in terms of individuals, not households. Thus, under scenario 1, they report the following targeting error rates. For CBT the exclusion error rate is equal to 0.47 while the inclusion error rate is 0.26. For PMT these error rates are respectively 0.17 and 0.21. However, no loss of generality is involved by choosing the household as the unit of analysis. In fact, we end up with the same ranking of the mechanisms studied as those authors.

the three curves confirms the superior performance of PMT under this scenario. In the inclusion region, the PMT mechanism has the highest probability of selection than any of the other two mechanisms. Within that region, moving from low to high consumption expenditure (i.e. from highest to lowest priority households), the assignment probability of the PMT mechanism falls from a maximum of 0.82 to minimum of 0.65. The CBT mechanism is the second best performer in that same region. Its assignment probability falls from 0.74 to 0.63, while that of self-targeting falls from 0.63 to 0.47.

Figure 3.1: Assignment Probability Curves under Scenario 1



Source: Author's calculations

In the exclusion region, as we move from low to high consumption expenditure, the assignment probability increases for both CBT and self-targeting, while that of PMT keeps on falling. In that region, the assignment probability for the PMT mechanism falls from 0.64 to 0.27 while that of CBT increases from 0.64 to 0.74. The propensity score for self-targeting also increases from 0.49 to 0.54. This pattern clearly shows that, under scenario 1, CBT and self-targeting are more prone to inclusion errors than the PMT mechanism. In particular, the available information indicates that CBT is more likely to include both the poorest households and those located at the top end of the distribution of consumption expenditure. This may be a sign of elite capture.

Table 3.2: Targeting Outcomes under Scenario 2: H=35%

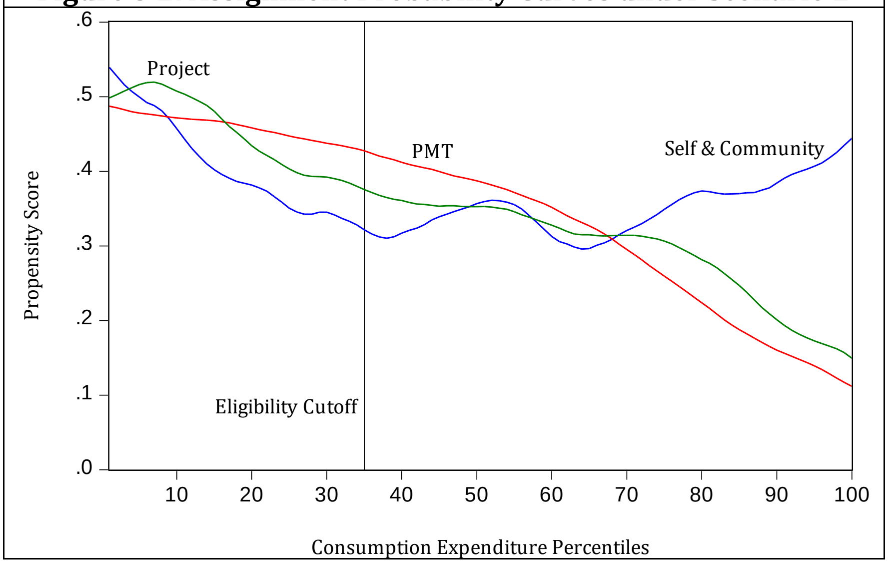
Measure	Self-and-Community	PMT	Project
Targeting Success Rate	0.56	0.63	0.62
Cohen's Kappa	0.05	0.17	0.16
Cramer's V	0.05	0.17	0.16
Exclusion Error Rate	0.60	0.54	0.55
Inclusion Error Rate	0.62	0.54	0.55

Source: Author's calculations

The pattern of the assignment probability curves presented in Figure 3.1 explains the fact that the targeting error rates associated with the PMT mechanism are lower than those induced by the other two mechanisms. This outcome is due mainly to the fact that the assignment probability curve for the PMT mechanism dominates those of the other two mechanisms everywhere in the inclusion region. Furthermore, the PMT assignment probability curve falls throughout the exclusion region while those for the other two mechanisms (CBT and self-targeting) keep on increasing.

Table 3.2 contains aggregate targeting outcomes for three mechanisms under scenario 2 when the target population includes only the 35 percent poorest households. Besides PMT, the other two mechanisms are hybrids. The column labeled "Self-and-Community" presents results for a mechanism that grants eligibility only to households selected by both self-targeting and CBT. That mechanism selects about 37 percent of the households (or 643 households). The mechanism labeled "project" is also a hybrid combining PMT and CBT. Among those households considered eligible by CBT, the project mechanism selects the 35 percent poorest according to PMT. This yields 598 households or about 35 percent of the sample households. Note that, in this scenario, the cutoff point for the mechanism based on PMT alone is also adjusted to pick up from the whole sample the 35 percent poorest households on the basis of the PMT score.

Figure 3.2: Assignment Probability Curves under Scenario 2

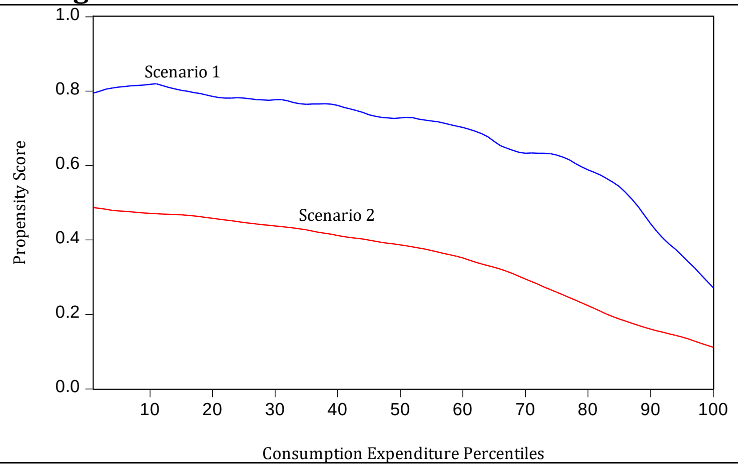


Source: Author's calculations

As in scenario 1 the observed agreement between the eligibility criterion and the three mechanisms is above 50 percent. It is just about the same for PMT and the project mechanism: 63 and 62 percent respectively. The agreement for self-and-community is the lowest, 56 percent. However, once we correct for chance-agreement and place the results on the Landis-Koch scale, we discover that there is poor agreement between eligibility and each of these three targeting mechanisms. The ranking of these three mechanisms does change when we consider targeting error rates. These error rates are about the same for PMT and the mechanism used by the project (54 and 55 percent respectively). They are higher for the combined self-targeting and CBT. For this hybrid mechanism, the exclusion error rate is 60 percent while the inclusion error rate is 62 percent.

The aggregate indicators presented in table 3.2 for scenario 2 suggest that there is no significant difference in performance between PMT and the project mechanism. However the corresponding assignment probability curves presented in figure 3.2 paint a clearer picture that reveals interesting differences between these two mechanisms depending on the location of households on the distribution of *per capita* consumption expenditure (the standard of comparison). Within the inclusion region, the project mechanism outperforms PMT in selecting the 16 percent poorest households. In the interval running from zero to the 16th percentile of the distribution of *per capita* consumption expenditure, the project assignment probability curve lies above the PMT curve. When it comes to selecting households located between the 16th and 35th percentile PMT performs better than the project mechanism. The corresponding assignment probability curves reverse positions. That of PMT dominates the one for the project.

Figure 3.3: Change in the Performance of PMT from Scenario 1 to Scenario 2



Source: Author's calculations

In the exclusion region, the project mechanism dominates PMT in the interval running from the 35th to the 68th percentile. The relative position of the assignment probability curves in that interval indicates that inclusion errors are more likely under PMT than under the project mechanism. Beyond that interval, the reverse is true. PMT performs better than the project mechanism. Given that the project mechanism is a combination of PMT and CBT, the fact that PMT performs better than the project mechanism beyond the 68th percentile is explained by the finding under scenario 1 that beyond the cutoff point (67.5%) CBT has a greater propensity to include non-poor than PMT alone (see figure 3.1).

The interpretation of the configuration of the assignment probability curves presented in figures 3.1 and 3.2 suggests that the coverage rate or the selection of the eligibility cutoff affects the performance of targeting mechanisms. We focus here on the case of PMT which is the only targeting mechanism common to both scenarios 1 and 2. A comparison of the performance of PMT based on the indicators presented in tables 3.1 and

3.2 clearly indicates a deterioration in performance from scenario 1 to scenario 2. In particular, targeting errors increase by more than twofold. Also, as noted above, the assignment probability curves associated with PMT and the project mechanism indicate a poor performance of PMT relative to the project mechanism for the 16 percent poorest households. Given that the project mechanism is a combination of CBT and PMT, and PMT dominates CBT under scenario 1, we conclude that the ability of PMT to select households located at the bottom of the distribution of the eligibility variable must deteriorate from scenario 1 to scenario 2. That conclusion is supported by figure 3.3 showing the assignment probability curves for PMT under scenarios 1 and 2. It is also consistent with the finding by Kidd and Wylde (2010) that the PMT mechanism performs worse at the bottom tail of the distribution of the eligibility variable (e.g. a living standard indicator). Thus targeting errors for PMT increase the lower the eligibility cutoff. This is due mainly to the fact that regression analysis underpinning the PMT method tends to explain at most 50 percent of the variation in the eligibility variable, and eligibility predictions based on such analysis are correct only on average (Coady, Grosh and Hoddinott 2004b).

Proximate Determinants of Targeting Outcomes under Scenario 2

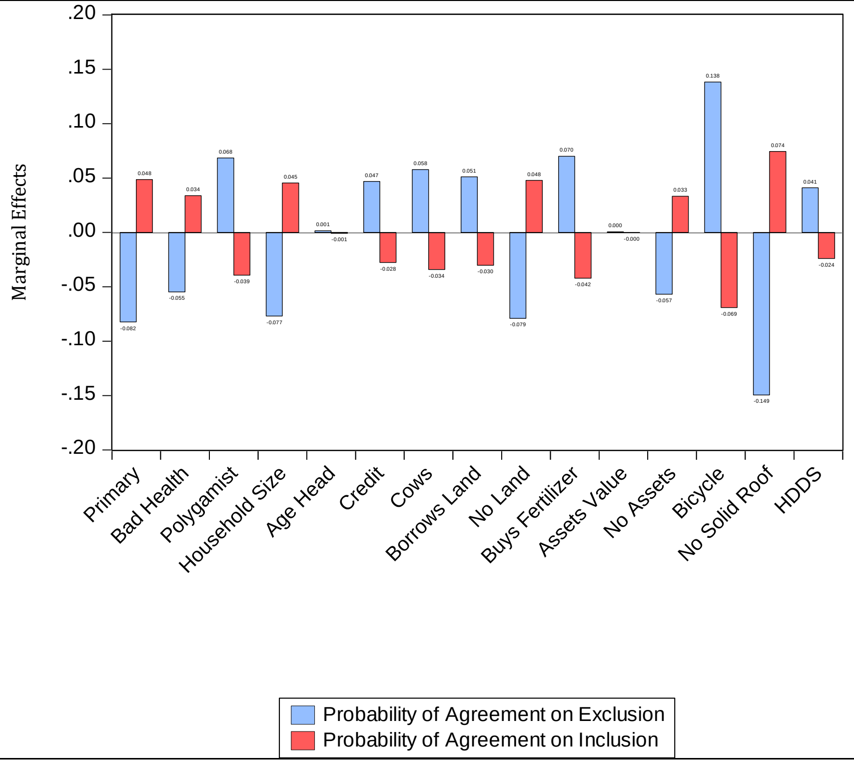
We now illustrate the use of the biprobit model to identify proximate determinants of targeting outcomes. We focus on scenario 2 which involves the actual mechanism used to select households into the intervention. We refer to this mechanism here as the project mechanism. First, we estimate a bivariate probit model of targeting outcomes associated with the project mechanism. We use the same set of covariates as Stoeffler et al. (2016). Table A1 in the Appendix presents these covariates along with their population and sample means. However, we base the identification of the proximate determinants of targeting performance on the marginal effects of the covariates on the probability of the relevant outcomes as described in table 2.5. Second, we apply the same methodology to the analysis of the extent of agreement between the PMT and project mechanisms. The marginal effects for both cases are presented in the appendix, tables A2 and A3.

There is a statistically significant positive correlation between eligibility and the project mechanism. The conditional tetrachoric correlation is about 15 percent while the unconditional correlation between the two binary variables is estimated at about 25 percent[23]. As argued earlier, the fact that this correlation is statistically significant suggests that the project mechanism respects eligibility to some extent. We now consider the drivers of this correlation.

Figure 3.4 shows the proximate determinants of the agreement between the eligibility criterion (i.e. household must be among the 35 percent poorest) and the project assignment mechanism. These are covariates with statistically significant marginal effects on the probability of agreement either on *exclusion* (P_{00}) or on *inclusion* (P_{11}). The associated level of significance is 5 percent or less.

[23] This is obtained by running the bivariate probit regression without the covariates.

Figure 3.4: Drivers of Agreement between Eligibility and the Project Mechanism



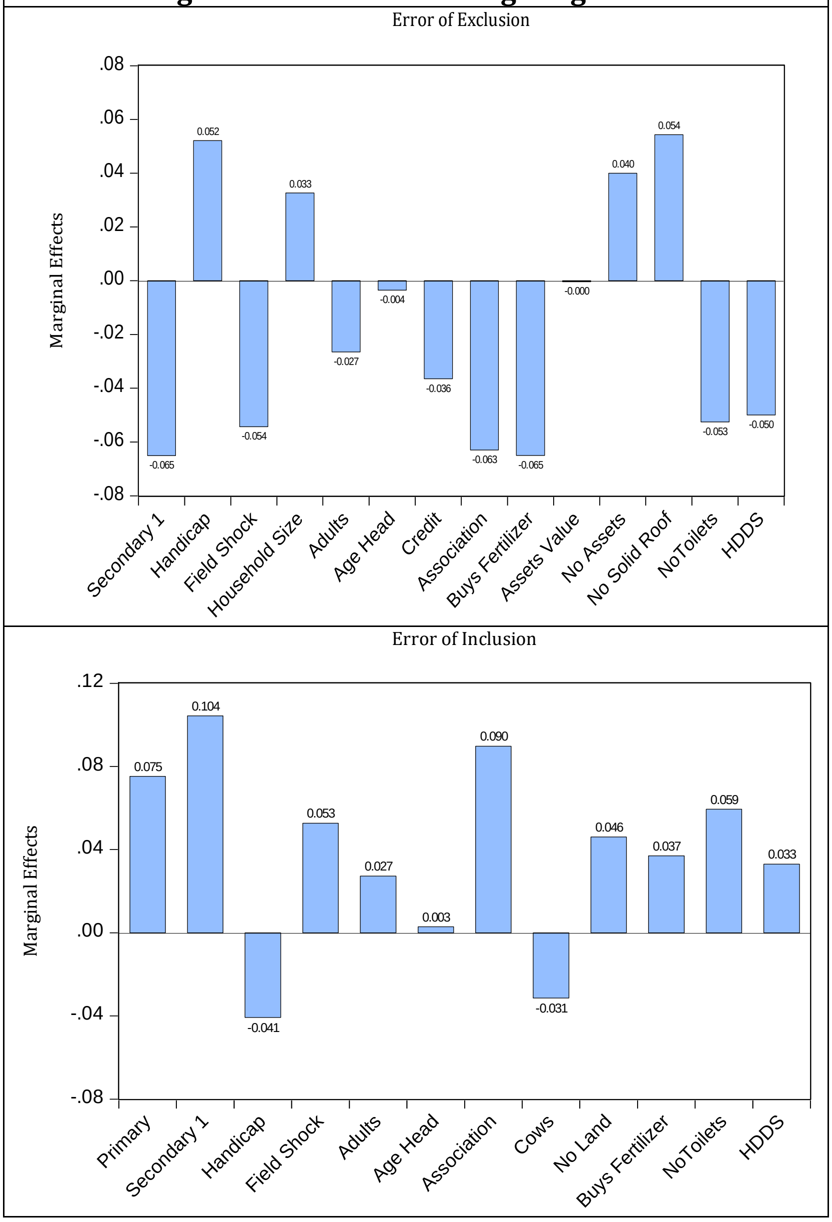
Source: Author's calculations

Six covariates have negative marginal effects on the probability of agreement on exclusion. These include: primary education, poor health, household size, owning no land, owning no assets and having a house with no solid roof. It is less likely that the eligibility criterion and the project mechanism will agree to exclude households with those characteristics. In particular, they are less likely to agree to exclude large households than small ones, or landless households compared to those owning land. The interpretation is similar for the other characteristics. The sign pattern for the marginal effects presented in figure 3.4 implies that any factor that reduces the probability of agreement on exclusion has the opposite effect on the probability of agreement on inclusion. Thus, the same six covariates with a negative effect on the probability of agreement on exclusion increase the probability of agreement on inclusion.

There are nine factors which increase the probability of agreement on exclusion and hence reduce the probability of agreement on inclusion. These are: polygamy, age, access to credit, the number of cows, ability to borrow land, being able to purchase fertilizer, the value of assets, owning at least one bicycle, and the household dietary diversity score (HDDS).

In total, there are about 15 core household characteristics (out of 34 included in the model) which drive the likelihood of agreement between the eligibility criterion and the assignment mechanism used in the project. While the marginal effects associated with these covariate are all statistically significant, they have different magnitudes. For instance, the effects of age and the value of assets are negligible. Among the factors which increase the probability of agreement on exclusion, owning at least one bicycle is the one with the greatest marginal effect, about 0.14. This is twice the effect associated with the ability to buy fertilizer, which is the factor with the next greatest positive effect on the probability of agreement on exclusion. Having a house with no solid roof is the characteristics with the greatest effect on reducing the probability of agreement on exclusion. This characteristic is also the one that increases the probability of agreement on inclusion the most. Owning at least one bicycle reduces the probability of agreement on inclusion the most among all factors presented in figure 3.4.

Figure 3.5: Drivers of Targeting Errors



Source: Author's calculations

Figure 3.5 shows covariates that are significantly associated with targeting errors in a statistical sense. Some of the variables affect both the exclusion and inclusion errors, while the remaining covariates have a significant effect on only one type of error. The covariates affecting both types of errors include: Secondary education level 1, having an handicap, having experienced an agricultural shock, the number of adults in the household, the age of the household head, membership in an association, ability to buy fertilizer, no toilets in the house and the household dietary diversity score (HDDS). Among these nine covariates, those with a positive effect on the probability of exclusion error have a negative effect on the probability of inclusion error and vice versa. Thus having a handicap increases the probability of errors of exclusion. Having first level secondary education, having experienced an agricultural shock, the number of adults in the households, the age of the household head, membership in an association , ability to buy fertilizer, no toilets and the HDDS all reduce the probability of errors of exclusion, while increasing that of errors of inclusion.

The covariates that have a significant statistical effect only on the probability of errors of exclusion are: household size, having access to credit, the value of assets, owning no assets, and no solid roof. Household size, owning no assets, and having no solid roof increase the probability of errors of exclusion. The remaining two variables have the opposite effect. Similarly, primary education, the number of cows and landlessness affect only the probability of errors of inclusion. Primary education and landlessness increase that probability while the number of cows reduces it.

As in the case of the probability of agreement between eligibility and the project mechanism, the effect of the age of the household head and that of the value of assets on the probability of targeting errors are negligible, though statistically significant.

Table 3.3: Frequency of Agreement between PMT and Project (scenario 2)

PMT	Project		
	Participant	Nonparticipant	Total
Eligible	357	241	598
Non-Eligible	241	884	1125
Total	598	1125	1723

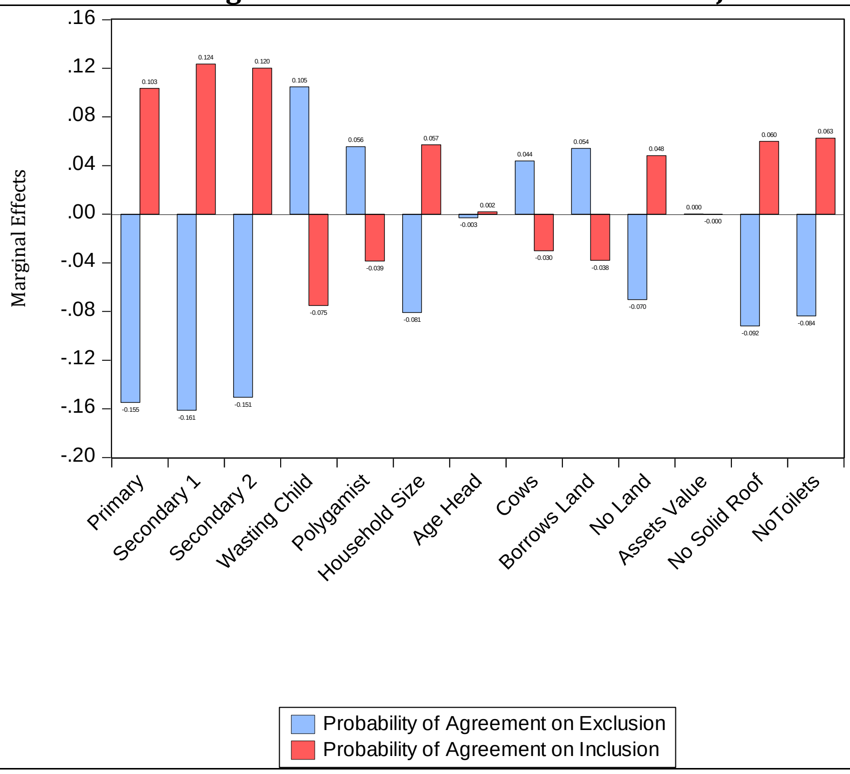
Source: Author's calculations

We also apply the binary probit framework to the analysis of the extent of agreement between the project mechanism and PMT under scenario 2. To facilitate the interpretation of our results, we base eligibility on the PMT score so that the 35 percent poorest households according to PMT are considered eligible. We then analyze the extent to which the project mechanism respects this eligibility criterion. Table 3.3 reports the frequency of agreement between PMT and project. The unadjusted rate of agreement between the two mechanisms is quite high, about 72 percent. Furthermore, the conditional tetrachoric correlation is equal 44 percent while the unconditional correlation is about 58 percent.

All these measures of association between the PMT and project mechanisms are higher than any of the ones we observed earlier. Even the Cohen's Kappa coefficient, which is equal to 0.38, is higher than those reported in tables 3.1 and 3.2. The strong association between the two mechanisms is understandable given that the project mechanism is a hybrid combining PMT and CBT. However, despite this strong association, the observed agreement between the PMT and project mechanisms is at best moderate based on Acock's scale. This level of agreement would be judged as fair (not at all substantial) on the basis of the LK scale.

In this particular context, we define *Type 1 disagreement* by analogy to the error of exclusion. This type of disagreement occurs when the PMT mechanism considers a household eligible but the project mechanism disagrees. Similarly, *Type 2 disagreement* is analogous to an error of inclusion whereby the project mechanism assigns a cash transfer to a household deemed ineligible on the basis of its PMT score. Under scenario 2, the results presented in table 3.3 imply an exclusion error rate of about 40 percent, and an inclusion error rate of the same size. These targeting error rates are 15 percentage points lower than those reported in table 3.2 when eligibility is based on *per capita* consumption expenditure. Given the structure of the project mechanism, these targeting errors are the result of the influence of the CBT component of the project mechanism. In other words, the disagreement between the PMT and project mechanisms is driven by the CBT component of the latter.

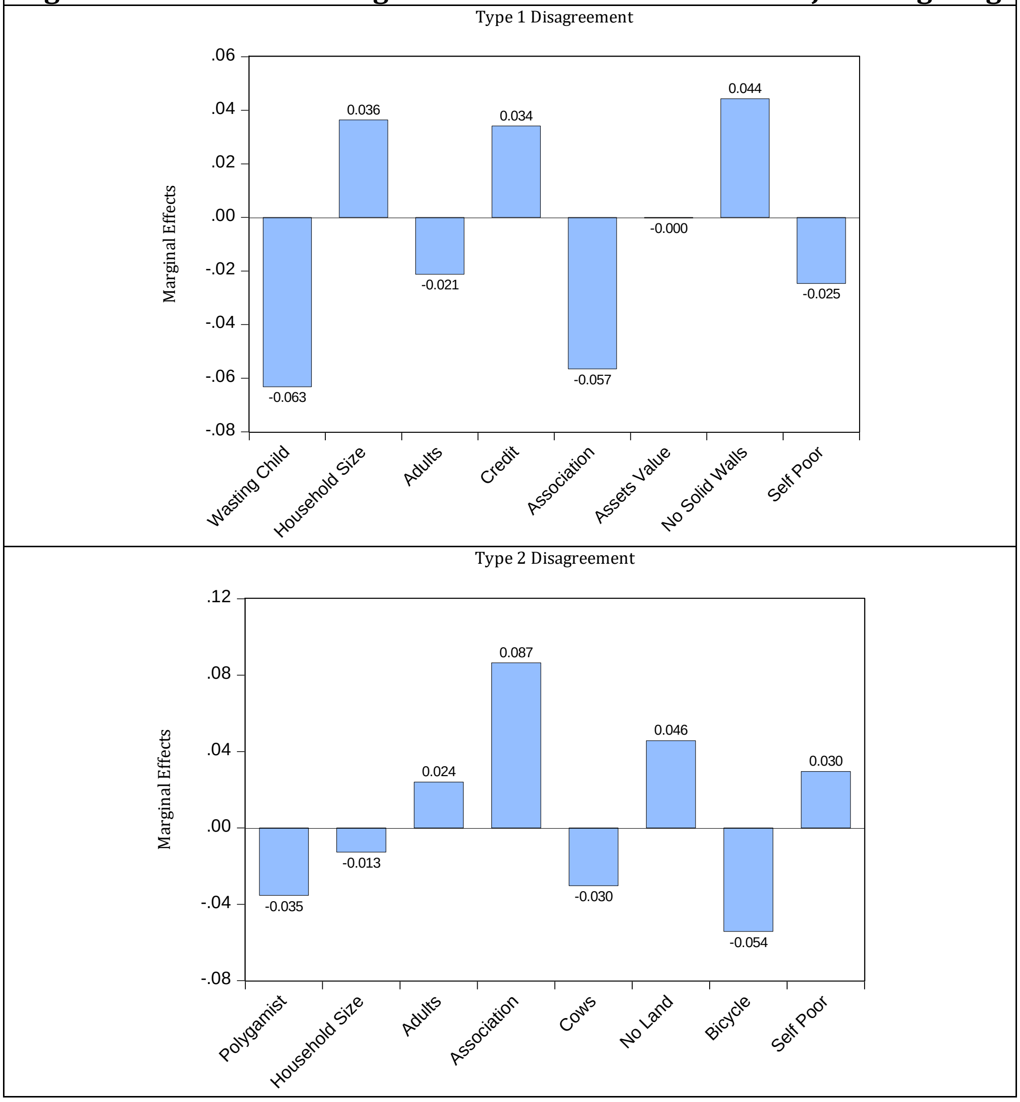
Figure 3.6: Drivers of Agreement between the PMT and Project Mechanisms



Source: Author's calculations

Figure 3.6, which is analogous to figure 3.4, shows the proximate determinants of agreement between the PMT and project mechanisms. There are four covariates which increase the probability of agreement on exclusion: having a wasting child in the household, being a polygamist, the number of cows, and the ability to borrow land. As expected, these covariates have a negative effect on the probability of agreement on inclusion. Similarly the eight covariates which reduce the probability of agreement on inclusion increase the probability of agreement on exclusion. These include: primary education, secondary education (both level 1 and level 2), household size, age of the head of household (negligible), owning no land, no solid roof and no toilets.

Figure 3.7: Drivers of Disagreement between PMT and Project Targeting



Source: Author's calculations

Figure 3.7 shows the marginal effects of covariates affecting Type 1 and Type 2 disagreement between the PMT and project mechanisms in a statistically significant way. Four of these covariates, namely household size, the number of adults, membership in an association and thinking of one's household as poor, affect both Type 1 and Type 2 disagreement. Household size increases the probability of Type 1 disagreement while decreasing that of Type 2. The other three covariates reduce the probability of Type 1 disagreement while increasing the probability of Type 2 disagreement.

Three covariates have a significant effect only on the probability of Type 1 disagreement. These include: having a wasting child in the household, access to credit and having a house with no solid walls. The effect associated with a wasting child is negative, while the other two effects are positive. Having a wasting child and membership in an association are the two covariates with the largest negative effects on the probability of Type 1 disagreement between PMT and the project mechanism.

4. Concluding Remarks

The important role of targeting in policymaking and the demand for evidence-based decision making create the need for evaluating the performance of targeting mechanisms. This paper proposes an *evaluative framework* that can guide the production of credible evidence to answer key questions about targeting performance that decision makers and other stakeholders care about. The key question for any targeting assessment is: *To what extent does targeting respect eligibility*? Answering this key question is an exercise in benefit incidence analysis which entails a study of the possible *association* between treatment receipt and eligibility.

In the context of a performance assessment, one needs to *measure* what happens, *compare* it with what is supposed to happen and *explain* any significant discrepancies between the two states of the world. The proposed framework thus focuses on measuring, judging and explaining the association between targeting and eligibility. All common indicators of targeting performance, be they concentration- or frequency-based, quantify in some sense the association between targeting and eligibility. To judge this association, one needs first to establish that such an association exists by comparing the observed targeting outcomes with counterfactual outcomes from a neutral assignment mechanism which enforces equal treatment of all potential beneficiaries. If the two sets of outcomes are significantly different in a statistical sense, then one places the observed outcomes on an appropriate scale to determine their substantive significance. Finally, to understand targeting outcomes, it is assumed that the contingency table characterizing the relationship between targeting and eligibility is a manifestation of *a latent bivariate probit process* driven by observable and non-observable individual or household characteristics. One can therefore identify the proximate determinants of targeting outcomes on the basis of the marginal effects of the relevant covariates on the probability of the outcomes of interest.

The application of this evaluation framework to the targeting of cash transfers in the context the Social Safety Nets Pilot Project (SSNPP) in Northern Cameroon led to the following findings. When the target population is the 67.5 percent poorest households, the three mechanisms, CBT, PMT and self-targeting, all agree with the eligibility criterion to an extent that is statistically significant. However, the agreement between eligibility and CBT and self-targeting is no better than a chance agreement. The agreement between eligibility and PMT is assessed as fair to moderate, depending on the chosen scale. All aggregate indicators and *the assignment probability curves* show that PMT outperforms the other two mechanisms.

When the target population is the 35 percent poorest households, there is poor agreement between eligibility and the three mechanisms considered: the project mechanism, PMT and a combination of CBT and self-targeting. While aggregate measures of performance imply no difference between PMT and the project mechanism, assignment probability curves show that, in the *inclusion region*, there is an interval where the project mechanism dominates PMT, while latter dominates the former in the rest of the region. A similar observation is made in the *exclusion region*. The identification of the proximate determinants of these targeting outcomes reveals that the lack of assets (human and physical) increases the likelihood of agreement over inclusion between eligibility and the project mechanism. Asset ownership has the opposite effect on the likelihood of this agreement. The likelihood of targeting errors is driven mostly by sociodemographic characteristics of the household.

A direct comparison of PMT and the project mechanism shows a stronger association between the two mechanisms due to the fact that PMT is a component of the selection mechanism used by the project. Socioeconomic factors are the main variables which increase the probability of agreement between the two mechanisms over exclusion. They have the opposite effect on the probability of agreement over inclusion. Similarly, the level of education and the quality of housing are the key factors which increase the probability of agreement on inclusion, with an opposite effect on the probability of agreement on exclusion. Disagreement between the two mechanisms is driven mostly by socioeconomic factors, and not by assets.

Effective policy making requires an understanding of the observed outcomes. In principle, such an understanding must come from a causal explanation which accounts for causal pathways and implementation processes underlying the observed performance. Because our results are based on a purely *statistical model of association*, they can shed some light only on proximate determinants of targeting performance. As noted earlier, objects of targeting are socioeconomic agents whose choices and actions affect the operation of targeting mechanisms. A deeper understanding of targeting performance would therefore require that we view a targeting mechanism as a social arrangement and use *structural modeling* to understand targeting performance as a result of individual behavior and social interaction.

Appendix

Table A1. Population and Sample Means of Explanatory Variables

Variable Name	Population Level Mean	Sample Mean
Primary Education	0.60	0.57
Secondary Education Level 1	0.24	0.20
Secondary Education Level 2	0.07	0.06
A Wasting Child in the Household	0.07	0.05
Christian	0.42	0.40
Animist	0.41	0.42
Handicap	0.20	0.20
Poor Health (Self Evaluation)	0.20	0.23
Agricultural Shock	0.42	0.40
Female Head of Household	0.20	0.27
Polygamist	0.38	0.29
Widow	0.06	0.11
Household Size	7.46	6.02
Household Members between the age of 15 and 59	2.87	2.40
Age of Household Head	45.52	46.73
Access to Credit	0.44	0.41
Member of an Association	0.12	0.11
Number of Cows	0.30	0.21
Value of Livestock Sales	1.21	0.94
Household Borrows Land	0.51	0.48
Household Owns No Land	0.33	0.36
Household Buys Fertilizer	0.43	0.38
Household Has No Agricultural Tools	0.05	0.06
Value of Agricultural Sales	75.39	70.65
Household Owns a Micro-Enterprise	0.26	0.22
Value of Assets	46.89	46.18
Household Owns No Assets	0.71	0.75
Household Owns at least one Bicycle	0.06	0.05
House Has No Solid Walls	0.87	0.86
House Has No Solid Roof	0.91	0.93
House Has No Toilets	0.07	0.10
Household Dietary Diversity Score	6.33	6.21
Household Considers itself very Poor	0.47	0.52
Needs to Go into Debt	0.54	0.53

Source: Author's Calculations

Table A2. Marginal Effects of Household Characteristics on the Probability of Targeting Outcomes for the Project Mechanism under Scenario 2

Variable	P_{00}	P_{11}	P_{10}	P_{01}
Primary Education	**-0.0824**	**0.0485**	-0.0414	**0.0753**
Secondary Education Level 1	**-0.0830**	0.4375	**-0.0651**	**0.1043**
Secondary Education Level 2	-0.0928	0.0522	-0.0498	0.0904
Wasting Child in the Household	0.0235	-0.0148	0.0212	-0.0298
Christian	0.0371	-0.0217	-0.0071	-0.0083
Animist	0.0026	-0.0020	0.0314	-0.0320
Handicap	-0.0209	0.0095	**0.0521**	**-0.0408**
Poor Health (Self Evaluation)	**-0.0548**	**0.0338**	0.0156	0.0054
Agricultural Shock	0.0062	-0.0046	**-0.0543**	**0.0527**
Female Head of Household	-0.0086	0.0046	0.0228	-0.0188
Polygamist	**0.0684**	**-0.0395**	0.0038	-0.0327
Widow	0.0592	-0.0327	-0.0198	-0.0067
Household Size	**-0.0769**	**0.0454**	**0.0327**	-0.0012
Number of Adults	-0.0021	0.0014	**-0.0265**	**0.0273**
Age of Head of Household	**0.0015**	**-0.0009**	**-0.0035**	**0.0029**
Access to Credit	**0.0469**	**-0.0277**	**-0.0364**	0.0172
Member of an Association	-0.0450	0.0182	**-0.0629**	**0.0897**
Number of Cows	**0.0576**	**-0.0341**	0.0079	**-0.0314**
Value of Livestock Sales	-0.0001	0.0001	0.0005	-0.0004
Household Borrows Land	**0.0511**	**-0.0303**	-0.0039	-0.0169
Household Owns No Land	**-0.0791**	**0.0478**	-0.0147	**0.0461**
Household Buys Fertilizer	**0.0700**	**-0.0421**	**-0.0650**	**0.0370**
Household Has No Agricultural Tools	0.0473	-0.0267	-0.0253	0.0047
Value of Agricultural Sales	-0.0001	0.0001	-0.0001	0.0001
Household Owns a Micro Enterprise	0.0269	-0.0157	0.0022	-0.0134
Assets Value	**0.0005**	**-0.0003**	**-0.0004**	0.0002
Household Owns No Assets	**-0.0570**	**0.0332**	**0.0400**	-0.0162
Household Owns at least one Bicycle	**0.1382**	**-0.0692**	-0.0512	-0.0178
House Has No Solid Walls	0.0222	-0.0132	-0.0162	0.0072
House Has No Solid Roof	**-0.1494**	**0.0745**	**0.0543**	0.0207
House Has No Toilets	-0.0044	-0.0025	**-0.0525**	**0.0594**
Household Dietary Diversity Score	**0.0409**	**-0.0240**	**-0.0499**	**0.0330**
Household Considers itself Very Poor	-0.0214	0.0128	-0.0169	0.0255
Household Needs to Go into Debt	0.0160	-0.0094	-0.0107	0.0041

Source: Author’s calculations. Results in bold are statistically significant at a level equal to 5 percent or less.

Table A3. Marginal Effects of Household Characteristics on the Probability of Agreement or Disagreement between the PMT and Project Mechanisms under Scenario 2

Variable	P_{00}	P_{11}	P_{10}	P_{01}
Primary Education	**-0.1548**	**0.1034**	0.0306	0.0208
Secondary Education Level 1	**-0.1613**	**0.1236**	0.0113	0.0265
Secondary Education Level 2	**-0.1506**	**0.1202**	0.0059	0.0245
Wasting Child in the Household	**0.1047**	**-0.0750**	**-0.0632**	0.0335
Christian	0.0289	-0.0201	0.0003	-0.0091
Animist	0.0154	-0.0105	0.0169	-0.0218
Handicap	0.0158	-0.0111	0.0151	-0.0198
Poor Health (Self Evaluation)	-0.0165	0.0095	-0.0237	0.0307
Agricultural Shock	-0.0329	0.0223	-0.0143	0.0248
Female Head of Household	0.0430	-0.0312	-0.0307	0.0189
Polygamist	0.0556	**-0.0385**	0.0184	**-0.0354**
Widow	0.0451	-0.0302	-0.0043	-0.0106
Household Size	**-0.0809**	**0.0572**	**0.0364**	**-0.0127**
Number of Adults	-0.0071	0.0044	**-0.0212**	**0.0240**
Age of Head of Household	**-0.0029**	**0.0020**	0.0008	0.0001
Access to Credit	-0.0233	0.0167	**0.0341**	-0.0275
Member of an Association	-0.0541	0.0241	**-0.0565**	**0.0865**
Number of Cows	**0.0439**	**-0.0301**	0.0165	**-0.0303**
Value of Livestock Sales	0.0011	-0.0008	-0.0008	0.0005
Household Borrows Land	**0.0542**	**-0.0380**	-0.0066	-0.0096
Household Owns No Land	**-0.0701**	**0.0482**	-0.0238	**0.0457**
Household Buys Fertilizer	-0.0050	0.0037	0.0078	-0.0065
Household Has No Agricultural Tools	0.0393	-0.0272	-0.0175	0.0054
Value of Agricultural Sales	-0.0001	0.0001	-0.0000	0.0000
Household Owns a Micro Enterprise	0.0260	-0.0179	0.0052	-0.0133
Assets Value	**0.0002**	**-0.0002**	**-0.0002**	0.0001
Household Owns No Assets	-0.0310	0.0218	0.0149	-0.0057
Household Owns at least one Bicycle	0.0392	-0.0333	0.0484	**-0.0542**
House Has No Solid Walls	-0.0407	0.0327	**0.0443**	-0.0363
House Has No Solid Roof	**-0.0919**	**0.0599**	-0.0019	0.0339
House Has No Toilets	**-0.0836**	**0.0626**	0.0267	-0.0057
Household Dietary Diversity Score	-0.0060	0.0041	-0.0027	0.0046
Household Considers itself Very Poor	-0.0142	0.0093	**-0.0246**	**0.0295**
Household Needs to Go into Debt	0.0130	-0.0092	-0.0057	0.0018

Source: Author's calculations. Results in bold are statistically significant at a level equal to 5 percent or less.

References

Acock Alan C. 2016. A Gentle Introduction to Stata. College Station (Texas): Stata Press.

Ahamada Ibrahim and Flachaire Emmanuel. 2010. Non-Parametric Econometrics. Oxford: Oxford University Press.

Allin Paul and Hand David J. 2014. The Wellbeing of Nations: Meaning, Motive and Measurement. Chichester (UK): Wiley & Sons, Ltd.

Angrist Joshua D. and Pischke Jorn-Steffen. 2009. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton: Princeton University Press.

Brown Caitlin, Ravallion Martin and van de Walle Dominique. 2016. A Poor Means Test? Econometric Targeting in Africa. World Bank Policy Research Working Paper No. 7915. Washington, DC: The World Bank Group.

Coady David, Grosh Margaret and Hoddinott John. 2004a. Targeting Outcomes Redux. The World Bank Research Observer, Vol. 19, No. 1: 61-85.

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

Cameron Lisa A. 2000. An Analysis of the Role of Social Safety Net Scholarships in Reducing School Drop-Out During the Indonesian Economic Crisis. Innocenti Working Paper No. 82. Florence: UNICEF Innocenti Research Center.

Deaton Angus. 1989. Rice Prices and Income Distribution in Thailand: A Non-Parametric Analysis. The Economic Journal, 99: 1-37.

Devereux Stephen, Masset Edoardo, Sabates-Wheeler Rachel, Samson Michael, Rivas Althea Maria and te Lintelo Dolf. 2017. The Targeting Effectiveness of Social Transfers. Journal of Development Effectiveness 9 (2): 162–211.

Essama-Nssah B., 2013. "What Questions Do Evaluations Answer?" PREM Note 25, Special Series on the Nuts and Bolts of M&E Systems, Washington, DC: The World Bank Group.

Essama-Nssah B. and Sarr Leopold R. 2015. Who Benefits from the Pro-Poor Targeted Secondary School Stipend Program in Nepal? Evidence from a Targeting Assessment. World Bank South Asia Region Education Global Practice, Discussion Paper No. 81.

Frechtling Joy A. 2007. Logic Modeling Methods in Program Evaluation. San Francisco: Jossey Bass.

Gertler Paul J., Martinez Sebastian, Premand Patrick, Rawlings Laura B., and Vermeersch Christel M. J.. 2016. Impact Evaluation in Practice: Second Edition. Washington, DC: The World Bank Group.

Greene William H. 2008. Econometric Analysis. Upper Saddle River, NJ: Pearson Education, Inc.

Grosh Margaret, Carlo del Ninno, Emil Tesliuc, and Azedine Ouerghi. 2008. For Protection and Promotion: The Design and Implementation of Effective Safety Nets. Washington, DC: The World Bank Group.

Handa Sudhanshu, Huang Caolyn, Hypher Nicola, Teixeira Clarissa, Soares Fabio V. and Davis Benjamin. 2012. Targeting Effectiveness of Social Cash Transfer Programmes in Three African Countries. Journal of Development Effectiveness, Vol.4, No. 1: 78-108.

Kidd Stephen and Wylde Emily. 2011. Targeting the Poorest: An Assessment of the Proxy Means Test Methodology. AusAID Research Paper. Canberra (Australia): Australian Agency for International Development.

Landis J. Richard and Koch Gary G. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics*, Vol. 33, No.1:159–174.

Loader Catherine. 2005. Regression, Likelihood, Smoothing. Cleveland: Department of Statistics, Case Western Reserve University (Presentation at the Neuroinformatics Workshop, Wood Hole, MA, August 17, 2005).

O'Donnell Owen, van Doorslaer Eddy, Wagstaff Adam and Lindelow Magnus. 2008. Analyzing Health Equity Using Household Survey Data: A Guide to Techniques and Their Implementation. Washington, D.C.: The World Bank Group.

Pearl Judea. 2009. Causal Inference in Statistics: An Overview. Statistics Surveys, Vol. 3: 96-146.

Ravallion, Martin. 2009a. "Evaluating Three Stylised Interventions." Journal of Development Effectiveness 1 (3): 227–36.

Ravallion Martin. 2009b. How Relevant Is Targeting to the Success of an Antipoverty Program? The World Bank Research Observer, Vol. 24, No. 2: 205-231.

Sabates-Wheeler, Hurrell Alex and Devereux Stephen. 2015. Targeting Social Transfer Programmes: Comparing Design and Implementation Errors across Alternative Mechanisms. Journal of International Development, **27**: 1521-1545.

Sen Amartya. 1995. The Political Economy of Targeting. In Dominique van de Walle and Kimberly Nead (eds) "Public Spending and the Poor: Theory and Evidence". Baltimore: The Johns Hopkins University Press.

Skoufias Emmanuel, Davis Benjamin and Behrman Jere R. 1999. An Evaluation of the Selection of Beneficiary Households in the Education, Health, and Nutrition Program (PROGRESA) of Mexico. Washington, DC: International Food Policy Research Institute (IFPRI)

Stoeffler Quentin, Mills Bradford and del Ninno Carlo. 2016. Reaching the Poor: Cash Transfer Program Targeting in Cameroon. World Development, Vol.83:244-263.

Stoeffler Quentin, Nguetse-Tegoum Pierre and Mills Bradford. 2015. Generating a System for Targeting Unconditional Cash Transfers in Cameroon. In Carl del Ninno and Mills Bradford (eds), Effective Targeting Mechanisms for the Poor and Vulneravle in Africa. Washington D.C.: The World Bank Group.

Sumarto Sudarno, and Suryahadi Asep. 2001. Principles and Approaches to Targeting: With Reference to the Indonesian Social Safety Net Programs. SMERU Working Paper, The SMERU Research Institute, Jakarta.

Viera Anthony J. and Garrett Joanne M. 2005. Understanding Interobserver Agreement: The Kappa Statistic. Family Medicine, Vol. 37, No.5: 360-363.

Young Peyton H. 1994. Equity: In Theory and Practice. Princeton: Princeton University Press.

Wednesday, February 7, 2018