TI 2020-086/V
Tinbergen Institute Discussion Paper

Welfare Measurement and Poverty Targeting Based on Participatory Wealth Rankings

Martin Wiegand[1]

[1] Vrije Universiteit Amsterdam

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and Vrije Universiteit Amsterdam.

Contact: discussionpapers@tinbergen.nl

More TI discussion papers can be downloaded at https://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900

Welfare Measurement and Poverty Targeting Based on Participatory Wealth Rankings

Martin Wiegand
Department of Economics, Vrije Universiteit Amsterdam

December 23, 2020

Abstract

Participatory wealth rankings (PWRs) present an inclusive and inexpensive targeting method to identify poor households. They tend to be well received by participants but point to a systematically different understanding of welfare than implied by consumption-based rankings. This suggests that PWRs could be used as the basis for alternative welfare measures that aim to reflect local perceptions of poverty. This paper demonstrates how such a measure can be constructed, using data from a field experiment on poverty targeting in Indonesia. It then explores the potential impact of using this welfare measure as targeting goal on participants' and village leaders' satisfaction. I find that higher targeting accuracy—using the PWR-based measure as benchmark—increases satisfaction with the program. However, after controlling for targeting accuracy, the PWR does not lead to discernibly higher satisfaction than a proxy means targeting mechanism. The PWRs thus seem to be appreciated for their resulting allocations rather than valued intrinsically. I also find that targeting accuracy explains satisfaction outcomes better when it is measured against PWR-based welfare rather than predicted consumption. This holds true even for communities where no actual PWRs had been conducted. The results suggest that the information contained in PWRs can be used as a meaningful basis for targeting and poverty measurement.

Keywords: poverty, targeting preferences, welfare measures, participatory wealth ranking.

JEL Classification: C8, D63, I32, I38, O1.

1 Introduction

Social security programs that target the poor are staple policies in many developing countries. Where tax bases are shallow and incidences of poverty high, selecting beneficiaries is often preferred to universal coverage. The problem of how best to choose who should be eligible is subject of ongoing discourse. A substantial branch of literature is concerned with the question how well different targeting methods, e.g. proxy means testing, geographical targeting, participatory methods, or self-selection, align with certain targeting objectives, usually poverty status as measured by consumption or income (Coady, Grosh and Hoddinott, 2004; Zeller, Feulefack and Neef, 2006; Banerjee et al., 2009; Coady and Parker, 2009; Yusuf, 2010; Alatas et al., 2012; Alatas et al. (2016); Alatas et al., 2019; Bah et al., 2019; Karlan and Thuysbaert, 2019). A related yet less explored question is how the method of targeting affects acceptance and satisfaction with the programs in question.[1] How well a targeted intervention is being received may depend on a number of components: the degree to which the allocation meets shared intuitions of justice, the amount of self-determination allowed to communities in the targeting process, as well as the extent to which the specific needs of different localities are met.

This paper is an attempt to formalize these factors, and to identify how they impact program satisfaction in the context of a field experiment on targeting in Indonesia, conducted by Alatas et al. (2012). To understand what drives satisfaction with antipoverty programs, it is essential to take into account local views of what constitutes poverty. One way of doing so is through participatory approaches, where data collection and targeting come with an active involvement of the local population. The prime example of such approaches is the participatory wealth ranking (PWR): representatives of a community rank all households according to their wealth. Such rankings can help to understand how poverty is perceived by locals, and how different household characteristics are weighted in the assessment. An important insight from PWRs is that lo-

[1] The significance of dissatisfaction due to targeting and program implementation can hardly be underestimated. The Indonesian Direct Cash Assistance (Bantuan Langsung Tunai, or BLT) programs launched in 2005 and 2008 illustrate this point. For the implementation of 2005, Widjaja (2012) reports nationwide protests, threats to staff of the Central Statistics Bureau, and cases of vandalism against government facilities as a result of deficient program implementation. Cameron and Shah (2014) find for the same program that an increase in crime and a decrease in people's participation in community groups are associated with mistargeting. Alatas et al. (2012) remark that for the 2008 implementation of the BLT program, dissatisfaction with beneficiary lists was so immense that more than 2000 village officials refused to participate in the program.

cal perceptions of welfare differ systematically from the traditional assessments based on income or consumption (Shaffer, 2013). Furthermore, evidence from field experiments suggests that participants are generally satisfied with the results of interventions that use PWRs as the targeting method (Alatas et al., 2012; Schüring, 2014). Taken together, these findings suggest that PWRs are not only pragmatic ways to allocate benefits at the local level, but that they may also be used as the basis for alternative welfare measures.

The first goal of the paper is to demonstrate how welfare measures grounded in local concepts of material well-being can be constructed from PWRs. The idea is to estimate the relationship between rankings and household characteristics, and to predict scores based on this model. The resulting welfare measure has a number of desirable properties: it does not depend on preselected dimensions of wealth or deprivation, or on predefined weights[2]; it does not rely on subjective categories of welfare; it can be constructed for localities where no actual PWRs have been conducted; and it can be used to relate households from different communities to each other, and thus overcome the principal incomparability of ranking outcomes between villages. The new welfare scores can in turn be used as targeting goals, or as benchmarks to assess targeting performance and to measure local poverty.

A number of arguments can be made why these scores may be more appropriate welfare indicators than consumption. Figures of consumption are usually constructed on the basis of assets and expenses within the days before data collection. This means that prospects for future consumption, income volatility, and the ability to smooth out shocks, are not fully reflected in this measure—though they might be visible to other locals and get incorporated in the PWRs. Furthermore, an antipoverty program should arguably not distribute benefits to households with the lowest consumption, but rather to those with the highest marginal utility thereof. Or the aim might be to facilitate yet another welfare goal, such as equality of basic capabilities (Sen, 1980). If villagers share these intuitions for distributive justice, the PWRs may lead to more favorable outcomes than rankings based on consumption. Another attempt to assigning normative validity to the outcomes of community rankings is offered by Kanbur and Shaffer (2007). They view participatory approaches in the light of discourse ethics, according to which norms receive validity through practical discourse—an ideal communicative exchange in which participants engage in

[2]This distinguishes it from other concepts of multi-dimensional poverty, notably the one by Alkire and Foster (2011).

rational argumentation, and which allow the fair involvement of everyone.[3]

The second goal of the paper is to estimate the impact that targeting method, targeting accuracy, and amount of provided benefits have on program satisfaction. I use a dataset from a field experiment in Indonesia (Alatas et al., 2012), in which multiple targeting methods were compared to distribute a one-time lump-sum payout. The experiment contained three treatments regarding the within-village allocation of payments to households: a proxy means test (i.e., a predicted consumption score), a PWR, and a hybrid method between the first two. The authors of the study find that people in the PWR villages were more satisfied than those who were targeted based on a proxy means test. It is not clear, though, whether this difference in satisfaction is a result of PWRs leading to an allocation more in line with people's preferences, or because they grant more agency in the allocation process. After defining a measure of targeting accuracy, the experimental setup allows me to separate the impact of the participatory process and of the degree to which the resulting allocation is aligned with the welfare measure. I find that targeting accuracy based on perceived welfare has clearly positive impacts on various outcomes related to program satisfaction. At the same time, there is hardly any evidence that the ranking exercises themselves impact satisfaction, when controlling for targeting accuracy. The effects on satisfaction of (mis)allocating benefits within and between villages can also be measured separately. The results suggest that, on average, reallocating benefits within a village has a stronger effect on satisfaction than providing additional benefits to the village.

The satisfaction outcomes—besides being of inherent interest—also represent a neutral yardstick against which to evaluate the adequacy of various welfare measures for targeting. In addition to the measures based on local perceptions, I construct targeting accuracy and local poverty based on per capita consumption and examine how they compare in explaining satisfaction. It turns out that the measures based on perceived welfare have a significantly stronger impact than the ones based on consumption. This holds up even when consid-

[3]In how far this ideal of inclusiveness and equal treatment is being met depends on the design and implementation of the PWR, as well as the cultural context. For the wealth rankings from Indonesia considered in this paper, Alatas et al. (2012) report that when all households of a community were invited, almost half of them participated. The facilitators who moderated the meetings reported only in 15% of the meetings that a few individuals dominated the discussion about the rankings. This indicates that in the case of this particular field experiment, the resulting rankings were indeed the product of a relatively fair and democratic discourse. It is worthwhile to note, however, that the willingness to participate in communal projects and to contribute to public goods in general may be uniquely high in Indonesia for historical reasons (Mansuri and Rao, 2013, chapter 2).

ering only those villages that never conducted a PWR.

The results confirm that local perceptions of welfare are different from consumption-based welfare, and show that this difference is large enough that choosing one over the other as targeting goal translates into noticeable differences in satisfaction. Furthermore, while understanding local perceptions of welfare is thus important for successful targeting, the participatory process itself seems to matter little, if at all. This is a useful insight especially for contexts in which PWRs may not be feasible.

The paper continues as follows. Section 2 summarizes the field experiment and the corresponding dataset of Alatas et al. (2012). In section 3, I outline the baseline model of satisfaction and the construction of the different welfare measures, targeting accuracy, and local poverty rates. The welfare prediction models and impact estimations on satisfaction are discussed in section 4, and their results are presented in section 5. Section 6 concludes.

2 Summary of the Field Experiment and Data Description

The paper by Alatas et al. (2012), which serves as the starting point and primary data source for this study, describes a field experiment conducted in 640 villages in three provinces of Indonesia: North Sumatra, South Sulawesi, and Central Java. The sample of villages is randomly divided into three treatment groups, in which the beneficiary households of an unconditional one-off cash transfer of 30,000 Indonesian Rupees (around 3 US$) are determined in different ways.

First, in the *PMT* group, household indicators for consumption were collected by the Central Statistics Bureau (BPS) and composed into a proxy means test (PMT) score, using a key that the government had determined through survey data. Households in each village were then ranked according to the PMT score, and the lowest ranked households would receive the benefit. The number of benefits available for each village was determined using an existing poverty map and the Village Potential Statistics (PODES) dataset of 2008, and based on a PPP2$ per-day poverty line. Second, in the *Community* group, representatives of households in the village were invited to participate in a ranking exercise, in which households were ranked from poorest to wealthiest. The poorest households would then receive the benefit. A number of sub-treatments were conducted to elicit whether the composition of participants led to any differences. Third, in the *Hybrid* group, the same ranking was conducted as in the Community group, then the BPS collected data to calculate PMT scores for the lowest ranked households, with the cutoff being 1.5 times the number of available benefits for the village. The households with the lowest PMT scores among them received the benefit.

The authors conducted a survey in all the participating villages and constructed a detailed figure of per capita consumption for a sample of nine households per village. They find that the rank-correlation between consumption and the rankings produced by the treatment is highest for the PMT treatment and lowest for the Community treatment—unsurprisingly, given the PMT score was meant to predict consumption. At the same time, satisfaction with the program is higher for the Community treatment than for the two other treatments. Furthermore, the authors do not find any evidence that different subgroups (local elites, women) ranked households differently, that ethnic or religious minorities were discriminated against, or that local community leaders or their relatives were favored. The authors attribute the differences in rank correlation between the treatments to a local understanding of poverty that differs from the consumption metric. Villagers seem to weight especially those factors that are not decisive for consumption but consumption capacity and the ability to smooth shocks. Households whose head was less educated, widowed, disabled, seriously ill, or spent lots of money on tobacco or alcohol were rated relatively lower conditional on consumption, while those with connections to local elites were ranked relatively higher.

For this paper, I use a variety of components from the dataset of Alatas et al. (2012).[4] The first component is a baseline survey, which contains detailed household information of nine households—the village head and eight households selected at random—from each village, including the results of the ranking exercise and per capita consumption. The baseline survey contains a total of 5,755 observed households. The second component is the data collected by the BPS to obtain PMT scores. It was obtained for every household in the PMT group and for about 47% of households in the Hybrid group. It contains basic information about household demographics, education, occupation, and housing characteristics. On top of that, information about some easily observable assets was collected, including household appliances, electronic devices, livestock, vehicles, productive machinery, and agricultural land. The BPS

[4] Available online under http://dx.doi.org/10.1257/aer.102.4.1206.

data includes 10,718 households for the PMT group and 5,129 households for the Hybrid group. Table A.1 in the appendix gives a complete list of all the characteristics and assets that were available in both data sets and could be successfully matched and meaningfully used. The baseline survey and the BPS data are used to estimate the relationship between household characteristics and welfare. In addition, the BPS data is used to construct measures of targeting accuracy and poverty headcount for each village.

The third data components is an endline survey, conducted only in Central Java province, with five out of the eight randomly selected households in each village. The fourth component is another endline survey, conducted with the village leaders from all the sampled villages. The two endline surveys are used to evaluate the impact of the treatments and of targeting performance on satisfaction. They include several questions revolving around satisfaction with the program, which are used to construct the set of outcome variables. Households are being asked whether they are satisfied with the program (on a scale of 1 to 4), and village heads are asked whether they think the people in the village are satisfied with the program. Furthermore, village heads and households are asked whether there are any poor households not covered by the program, whether there are any households on the list of beneficiaries who do not belong there, whether the targeting method is correct, whether targeting is worse, equal, or better than the method formerly used for the Direct Cash Assistance (Bantuan Langsung Tunai, or BLT) programs, and if there were too few, enough, or too many benefits given out in the village. For all these questions, I treat answers such as "don't know" or "no opinion" as missing. In addition, village heads were asked how many complaints about the list of beneficiaries they received. Lastly, there was also a letter box for anonymous complaints, from which the number of complaints is documented.[5] The household and village level outcome variables are summarized in Tables 2.1 and 2.2, respectively. These tables display a clear pattern: the Community treatment group has significantly better[6]

[5] The numbers of complaints are divided by the number of households in the village to make them comparable. The resulting complaints-per-household variables are highly skewed, with very few villages registering a lot of complaints. Therefore, in order to avoid results being driven by a few extreme values, I use a log-transformation of the complaint variables in the regressions. More precisely, in order to deal with zeros among the complaint variables, 0.5 times the smallest non-zero value is added before taking the logarithm.

[6] *Better* here means: higher satisfaction, more agreement with the list of beneficiaries, fewer instances of poor households not on the list or non-poor households on the list, the targeting method being more correct and comparing favorably to the BLT method, the number of available benefits per village being closer to correct, and fewer complaints to the village head and in the complaint box.

Table 2.1: Outcomes from the household endline survey

Treatment group	PMT	Hybrid	Community
Are you satisfied with the targeting activities in this village in general? (1 = worst, 4 = best)	3.042	3.080	3.280
	(0.0367)	(0.0367)	(0.0326)
Do you agree with the households on the list of targeted households? (0 = no, 1 = yes)	0.669	0.778	0.878
	(0.0221)	(0.0192)	(0.0149)
Are there any poor households that should be added to the list? (0 = no, 1 = yes)	0.579	0.600	0.389
	(0.0231)	(0.0225)	(0.0223)
Are there any non-poor households that should be subtracted from the list? (0 = no, 1 = yes)	0.456	0.357	0.220
	(0.0233)	(0.0220)	(0.0189)
Is the method applied to determine the targeted households appropriate? (1 = worst, 4 = best)	3.243	3.228	3.388
	(0.0400)	(0.0418)	(0.0348)
How does this method compare to other methods (like BLT) in targeting households?	2.546	2.520	2.615
(1 = worse, 3 = better)	(0.0337)	(0.0362)	(0.0315)
Is the number of households on the list too small, correct, or too large?	1.611	1.595	1.716
(1 = too small, 2 = correct, 3 = too large)	(0.0283)	(0.0286)	(0.0268)
Number of interviewed households	465	480	490

The table shows group means, with standard errors in parentheses.

Table 2.2: Outcomes from the village head endline survey and the complaints box

Treatment group	PMT	Hybrid	Community
In your opinion, are villagers satisfied with the targeting activities in this village in general?	2.456	2.986	3.389
(1 = worst, 4 = best)	(0.0610)	(0.0525)	(0.0451)
Are there any poor households that should be added to the list? (0 = no, 1 = yes)	0.732	0.673	0.565
	(0.0307)	(0.0319)	(0.0340)
Are there any non-poor households that should be subtracted from the list? (0 = no, 1 = yes)	0.0574	0.0369	0.0467
	(0.0161)	(0.0128)	(0.0145)
Is the method applied to determine the targeted households appropriate? (0 = no, 1 = yes)	0.565	0.753	0.939
	(0.0344)	(0.0295)	(0.0165)
How does this method compare to other methods (like BLT) in targeting households?	2.236	2.610	2.821
(1 = worse, 3 = better)	(0.0561)	(0.0448)	(0.0318)
How many households complained about the list of targeted households? (per household)	0.0849	0.0539	0.0337
	(0.00947)	(0.00621)	(0.00529)
Number of complaints in the complaint box (per household)	0.0392	0.0287	0.0145
	(0.00748)	(0.00637)	(0.00258)
Is the number of households on the list too small, correct, or too large?	1.309	1.358	1.466
(1 = too small, 2 = correct, 3 = too large)	(0.0404)	(0.0377)	(0.0408)
Number of villages	209	217	214

The table shows group means, with standard errors in parentheses.

outcomes than the Hybrid and the PMT group, with only a few of the outcomes being statistically indistinguishable. Similarly, the Hybrid group has either significantly better outcomes than the PMT group (mostly for the outcomes at the village level) or is statistically indistinguishable (mostly for the outcomes at the household level).

3 Empirical Model

In this section, I first introduce a baseline model of satisfaction. Two of its components—targeting accuracy and local poverty—depend on the choice of a welfare measure. The construction of the different welfare measures is discussed further below.

3.1 Model of Satisfaction

Consider the model

$$y_{ij} = \beta_0 + \beta_1 I_H + \beta_2 I_C + \beta_3 t_j^m + \beta_4 b_j + \beta_5 h_j^m + \beta_6 x_{ij} + \varepsilon_{ij}. \tag{3.1}$$

y_{ij} stands for any of the satisfaction outcomes in Tables 2.1 and 2.2 for household i in village j. For village level outcomes, subscript i becomes obsolete. I_H and I_C are indicators for the Hybrid and the Community treatment group, respectively, with the PMT group being the excluded category. t_j^m refers to within-village targeting accuracy, i.e., the share of correctly targeted households based on welfare measure m—also counting those households mistargeted as a result of too many or too few available benefits for the village. b_j denotes the benefit ratio, i.e., the number of benefits over the number of households n_j. h_j^m denotes the local poverty headcount ratio, which depends on the underlying welfare measure m. x_{ij} is a vector of household and village characteristics that may potentially affect satisfaction as well as targeting accuracy or benefit ratio. These characteristics include regional dummies (for all combinations of the three provinces and urban/rural), log village size, and—only for the outcomes from the endline household survey—dummies for whether the household received the benefit and whether household members felt entitled to it, as well as the interaction of the two.

Coefficients β_1 and β_2 indicate what difference the treatment group makes for satisfaction, after controlling for how well the different treatments work at distributing funds to the poor. In other words, they should tell how much participation is intrinsically valued. β_3 and β_4 reveal the relative importance of (mis)allocation within villages and the total amounts to be distributed to each village. Lastly, using different measures m can show how well different concepts of welfare are able to explain the variation in satisfaction.

3.2 Within-Village Targeting Accuracy and Welfare Measures

Within-village targeting accuracy shows how well the benefits available for a given village are distributed. Let b_{ij} be an indicator that household i in village j with n_j households received the benefit, and let p_{ij}^m be an indicator that household i is among the poorest $\sum_{k=1}^{n_j} b_{kj}$ households in village j according to welfare measure m. Within-village targeting accuracy[7] is defined as

$$t_j^m := n_j^{-1} \sum_{k=1}^{n_j} \left(b_{kj} p_{kj}^m + (1 - b_{kj}) \left(1 - p_{kj}^m \right) \right). \tag{3.2}$$

While the recipient households are fixed, the within-village rankings of welfare—and thus p_{ij}^m—depend on how welfare m is being defined. (In what follows, superscripts m are omitted where it serves readability.) One way to do so implicitly is to assume that welfare is perfectly observed and reported by the participants of the PWR, which is sufficient to define t_j. I call this approach *rank-consistent welfare*. It implies that for the Community treatment group, targeting accuracy equals 1 for each village, as benefits were given to the $\sum_k b_{kj}$ lowest ranked households. For the Hybrid treatment group, targeting accuracy would be lower than 1 on average, as the allocation among the lowest ranked $1.5 \cdot n_j b_j$ households was determined by a PMT score ranking instead of the PWR. For the PMT group, targeting accuracy is not observed, as the villages in this group did not conduct a PWR.

An alternative way to compute targeting accuracy using PWRs, which does allow to include households that were not ranked, is to use predicted values based on a latent welfare model with household characteristics. I call this approach

[7] There are other measures of targeting quality than accuracy, that assign different weights to poor non-beneficiaries (type 1 errors) and non-poor beneficiaries (type 2 errors), or factor in the severity of household poverty (see e.g. Ravallion, 2009). These are, however, largely incompatible with the ordinal nature of a rank-based welfare measure, and with the fact that for within-village targeting, any type 1 error automatically also results in a type 2 error.

rank-score welfare (the word *score* indicating that this is an estimated rather than observed ranking). The predicted ranking outcomes will inevitably differ to some degree from the order of the actual PWRs. This may be seen as a disadvantage, as the true PWRs may reflect some important factors that are not being asked for in the surveys, for instance because they are sensitive or hard to quantify. But it may also be seen as an advantage, as the predicted rankings only take into account the factors that are considered relevant in multiple villages, while idiosyncratic factors such as a household's popularity or connectedness to local elites are left out of the score—a desirable effect.

Lastly, per capita consumption also constitutes a welfare measure. The BPS data does not contain detailed enough information to construct these figures, though, so that only predicted values can be used here as well.[8] To train the consumption model, I use the consumption figures from the baseline survey. In the following paragraphs, I outline how consumption and rank-score welfare are being constructed.

To estimate the relation between welfare w_{ij} of household i in village j and observable household characteristics z_{ij}, I assume a linear model,

$$w_{ij} = z_{ij}\gamma + \xi_{ij}. \tag{3.3}$$

When the welfare measure to be predicted is consumption, 3.3 can be estimated via OLS. For rank-score welfare, I propose to use a rank-ordered logit (ROL) model (Beggs, Cardell and Hausman, 1981), which was originally developed to study consumer preferences. The key assumption that needs to be made is that the random disturbance term ξ_{ij} in 3.3 is iid type I extreme value (EV1) distributed. In accordance with levels of welfare, village j provides a complete ranking over the set of households, R_j. For ease of notation, assume $w_{1j} < \ldots < w_{n_j j}$. The zero-probability case of equal welfare of two households is being ignored here. The probability for any particular ranking to occur given

[8] Importantly, the PMT score, that also aimed at predicting per capita consumption based on the BPS data, does a much poorer job than the prediction model used in this paper: for the PMT group, the correlation coefficient of log per capita consumption with the log PMT score is 0.53, while the correlation coefficient with the predicted per capita consumption estimated here is 0.72. This leads to large variation in within-village targeting accuracy (based on predicted consumption), which is needed to identify its impact on satisfaction.

household characteristics $Z_j = \left(z_{1j}, \ldots, z_{n_j j}\right)'$ is then

$$\begin{aligned}
\Pr\left[R_j | Z_j; \gamma\right] &= \Pr\left[w_{1j} < \ldots < w_{n_j j} | Z_j; \gamma\right] && (3.4)\\
&= \Pr\left[w_{1j} < w_{2j} | Z_j; \gamma\right] \cdot \Pr\left[w_{ij} < w_{3j}\ \forall i = 1, 2 | Z_j; \gamma\right] \cdot \ldots \\
&\quad \ldots \cdot \Pr\left[w_{ij} < w_{n_j j}\ \forall i = 1, \ldots, n_j - 1 | Z_j; \gamma\right] && (3.5)\\
&= \prod_{k=2}^{n_j} \frac{\exp\left(z_{kj}\gamma\right)}{\sum_{l=1}^{k} \exp\left(z_{lj}\gamma\right)}. && (3.6)
\end{aligned}$$

The ROL formula is a product of multinomial logit probabilities. Step 3.5 follows from the assumption that the conditional distribution of the highest ranked household from any subset is independent of the ranking of the other households. This is equivalent to the irrelevance of independent alternatives (IIA) property, which follows from the EV1 specification of ξ_{ij}. The reduction to the above closed-form expression allows estimation via maximum likelihood. The log-likelihood for the sample is

$$L(\gamma) = \sum_{j=1}^{J} \ln\left(\Pr\left[R_j | Z_j; \gamma\right]\right). \tag{3.7}$$

Welfare scores $\hat{w}_{ij}$ can be then be predicted using estimated coefficients $\hat{\gamma}$ just the same way as in the model of consumption, $\hat{w}_{ij} = z_{ij}\hat{\gamma}$. These scores are ordinal, i.e., unlike consumption scores they are not interpretable on their own.

3.3 Local Poverty Rate

An important control variable is the poverty headcount for each village. Constructing this requires—in addition to welfare scores that allow comparisons of households between villages—a global poverty threshold. While for some welfare measures, such as consumption, there are natural poverty thresholds such as the 2$ per day poverty line, there is none such for the rank-score welfare measure. For this paper, I set the overall poverty rate equal to the overall benefit ratio. This has the advantage that the total poverty rate is the same across different different welfare measures. It is also consistent with the poverty rate implied by the 2$ per day consumption poverty line, which the total benefit ratio in the field experiment aimed to meet.

Village poverty rates are computed as follows: welfare scores $\hat{w}_{ij}^{m}$ of households in the entire sample are being ranked. The lowest ranked households are declared poor, $h_{ij}^{m} = 1$, and the remaining ones non-poor, $h_{ij}^{m} = 0$, with the threshold being the total number of benefits allocated, $\sum_l \sum_k b_{kl}$. The local poverty rate of village j is then defined as

$$h_j^m := n_j^{-1} \sum_{k=1}^{n_j} h_{kj}^m. \tag{3.8}$$

3.4 Total Targeting Accuracy

Total targeting accuracy T_j^m is constructed just like within-village targeting accuracy, but using the absolute poverty indicators h_{ij}^{m},

$$T_j^m := n_j^{-1} \sum_{k=1}^{n_j} \left(b_{kj} h_{kj}^m + (1 - b_{kj}) \left(1 - h_{kj}^m\right)\right). \tag{3.9}$$

There are a number of reasons for using within-village targeting accuracy instead of total targeting accuracy in model 3.1. To begin with, it is conceivable that mistargeting within and between villages is perceived differently. Given that the targeting process has two stages, targeting errors have their origins within and outside the village, which may affect satisfaction in different ways. And as views on absolute poverty may vary between villages, violations of the ordering within the village may be perceived stronger than missing the correct number of benefits given the regional poverty line. In addition, for some of the welfare models being used it is impossible to construct total targeting accuracy, while for others it is possible but at the cost of relatively lower precision. Despite these caveats, total targeting accuracy may be a meaningful metric, and it is being used as a robustness check when comparing the explanatory power of different welfare models.

4 Model Selection and Estimation

In this section, I first line out how the welfare scores are being constructed given a set of predictors. I then go on to describe how the model of satisfaction is being estimated given certain limitations in the data.

4.1 Model Selection

To account for regional differences in the factors and weights constituting welfare, each welfare model is being estimated separately by province (North Sumatra, South Sulawesi, and Central Java) and urban/rural areas. In order to achieve high predictive performance, for each model and each sample, I use an alternating forward/backward model selection procedure in order to identify an adequate set of variables from a list of potential predictors. The procedure minimizes the bias-corrected Akaike information criterion, as defined by Hurvich and Tsai (1989), which is asymptotically equivalent to leave-one-out cross-validation. The candidate variables are taken from Table A.1, extended by polynomial terms and logarithms of age, household size, and floor area per capita, as well as by interaction terms of gender and marriage status, of age and education, of education and occupational sector, as well as a number of cluster-level variables and interactions of all individual-level and cluster-level variables.

There are some subtle differences in the respective variable pools for the different models. For the models of consumption, including village- or higher level variables as predictors may help to increase the model fit. On the other hand, this is pointless when using the ranking outcomes, as γ is identified only from comparisons of households within villages. The inability of the ROL model to incorporate village-level variation is not relevant for the construction of t_j, but it may render h_j less reliable when compared to the version based on consumption. To what extent this is true depends on how much of the variation in welfare is captured by household-level variables relative to village-level variables. Apart from that, for the construction of within-village targeting accuracy t_j based on rank-score welfare, it may be helpful to include interactions of household-level variables with village-level variables. These may emphasize local differences in the relative importance of certain household level factors for poverty, and thus improve the estimated rankings. On the other hand, for the absolute poverty indicators h_{ij}—which assume that welfare scores are comparable between villages—, such interactions should not be included. Comparing predictions from a ROL model can only be meaningful if its component factors are being compared during tuning. For this reason, total targeting accuracy T_j based on rank-score welfare is predicted using a smaller set of covariates than within-village targeting accuracy t_j^m, presumably leading to less precise predictions of village rankings.

The prediction models of consumption are estimated using all the households in the baseline survey. The prediction models of rank-score welfare are estimated based on the BPS data of the Hybrid group. To prevent overfitting due to using the same villages both for estimation and prediction, I use a one-village-out cross-fitting procedure: for the predicted ranking in village j, the estimation includes all villages except j.

To construct t_j^m and h_j^m with a predicted ranking as benchmark, the joint distribution of $(b_{kj}, z_{kj})_{k=1,\ldots,n_j}$ is required. b_{ij} is known for every household in the PMT and the Hybrid group, but z_{ij} is only fully observed for the PMT group, while in the Hybrid group it is observed only for the $1.5 \cdot n_j b_j$ lowest-ranked households from the PWR. Therefore, for estimations that involve the Hybrid group, I use imputed values $\hat{t}_j^m$ and $\hat{h}_j^m$. The imputation procedure is outlined in section A.1 of the appendix, together with an assessment of the bias arising from it.

4.2 Estimation

Due to the fact that ranking and consumption data is not available for every household of the experiment, equation (3.1) cannot be estimated directly. In particular, it is not possible to construct targeting accuracy and excess ratio based on rank-score welfare or predicted consumption for the villages in the Community group, J_C, since the kind of data collected by the BPS for every household to construct PMT scores was not collected there. On the other hand, for the Hybrid group J_H and the Community group J_H, it is possible to construct targeting accuracy based on the true rankings—but not for the PMT group J_P, as no PWR was conducted there. To get around this issue, I conduct separate sets of estimations.

The first one uses rank-score welfare and only villages from the Hybrid group and the PMT group,

$$\begin{aligned} y_{ij} = \beta_0 + \beta_1 I_H + \beta_3 \hat{t}_j^{\text{rank-score}} + \beta_4 b_j \\ + \beta_5 \hat{h}_j^{\text{rank-score}} + \beta_6 x_{ij} + \varepsilon_{ij}, \quad j \in J_P \cup J_H. \end{aligned} \tag{4.1}$$

The PMT group is the excluded category. Estimating equation 4.1 shows the difference in intrinsic value between the PMT and the Hybrid ranking method, as well as the relative significance of targeting accuracy.

The next set of estimations uses within-village targeting accuracy based on rank-consistent welfare, and can only be applied to the Hybrid group and the

Community group:

$$y_{ij} = \beta_0 + \beta_2 I_C + \beta_3 t_j^{\text{rank-consistent}} \\ + \beta_4 b_j + \beta_{5'} a_j + \beta_6 x_{ij} + \varepsilon_{ij}, \quad j \in J_H \cup J_C. \tag{4.2}$$

The Hybrid group is the excluded category. For the villages of the Community group, local poverty rates are not available. Instead, I include the village attendance rate for the PWR, a_j, as an additional control. Since the ranking procedure and the allocation mechanism were only explained at the meetings, a_j should be independent of treatment status, but it may proxy affluence and social capital, which could affect both targeting accuracy as well as satisfaction. Estimating equation 4.2 shows the difference in intrinsic value between the hybrid and the community ranking method, and allows to inspect how β_2 varies compared to equation 4.1.

A further set of estimations has the objective to compare the ability of rank-score welfare and predicted consumption to explain satisfaction. This is done by pooling the different measures of targeting accuracy and local poverty in one equation:

$$y_{ij} = \beta_0 + \beta_{31} t_j^{\text{rank-score}} + \beta_{32} t_j^{\text{cons.}} + \beta_4 b_j \\ + \beta_{51} h_j^{\text{rank-score}} + \beta_{52} h_j^{\text{cons.}} + \beta_6 x_{ij} + \varepsilon_{ij}, \quad j \in J_P. \tag{4.3}$$

A significant estimate of, say, coefficient β_{31} means that targeting accuracy based on the rank-score likely helps to explain the respective satisfaction outcome, given the information provided by the targeting accuracy measure that is based on consumption. Equation 4.3 focuses on the PMT group, to avoid having to use imputations for targeting accuracy or local poverty. Leaving out households from the Hybrid group also rules out a possible bias caused by the degree to which the final allocation resembles the PWRs. For instance, if people in the Hybrid group are asked to rank households, but then those rankings are not being adhered to due to the second-stage ranking, people may feel actively ignored, leading to lower satisfaction that would falsely be attributed to targeting accuracy alone. These issues also apply to estimating 4.1, and will be addressed in the next section.

Equations 4.1 to 4.3 are estimated via OLS.[9] For outcomes measured at

[9]Some of the outcomes y_{ij} are binary or ordered categorical. I also ran the corresponding regressions as logit and ordered logit models, respectively. The relative coefficient sizes and

the village level, heteroscedasticity-consistent standard errors are used. For outcomes at the household level, standard errors are clustered at the village level.

5 Results and Discussion

The results of estimating equation 4.1 for household level satisfaction outcomes are shown in Table 5.1. The table also includes the results of a simple regression of the satisfaction outcome on the treatment group, to see how much treatment effects change after including covariates. Within-village targeting accuracy appears to be important and is significant for most outcomes. Higher $\hat{t}_j$ increases satisfaction and agreement with the list of beneficiaries, reduces the chance of households not being on the list or wrong households to be included, and increases the likelihood that participants find the allocation method to be correct. The effect of the Hybrid group, on the other hand, is mostly small and insignificant, and does not show a consistent direction. However, when compared to the simple regression results, one can see that including covariates pulls the effect of the Hybrid group away from the expected direction for almost all outcomes. The effect of the benefit ratio is mostly consistent with the effect of targeting accuracy: more coverage increases satisfaction and reduces the chance of poor households being excluded. While mostly having the same sign as the effect of targeting accuracy, the magnitude of the effect of the benefit ratio is consistently lower. Correctly targeting households within the village seems to weigh more than adding additional benefits. Lastly, as one would expect, the impression of whether there were enough benefits is strongly affected by the benefit ratio but not by within-village targeting accuracy.

The results of the regressions of village level outcomes, reported in Table 5.2, show a slightly different picture. Village heads believed on average that households in the Hybrid group were more satisfied than in the PMT group, and that the targeting method was better, even after controlling for targeting accuracy. However, this notion is not supported by significant effects of the treatment group on list errors, or complaints to the village head or in the complaint box. Targeting accuracy, on the other hand, did not only increase perceived household satisfaction and correctness of the method, but also led to significantly fewer complaints. Lastly, a higher share of beneficiaries led to a

p-values were almost indistinguishable from those of the OLS results.

Table 5.1: Simple treatment effects and estimation results of equation 4.1 – household level outcomes

	Satisfied	Agree with list	Household not on list	Wrong household on list	Targeting method correct	Method better than BLT	Enough benefits
Hybrid group	0.0384	0.110***	0.0202	-0.0992**	-0.0147	-0.0260	-0.0146
(simple regression)	(0.0747)	(0.0402)	(0.0459)	(0.044)	(0.0779)	(0.0727)	(0.0548)
Adj. R^2	-0.001	0.014	-0.001	0.009	-0.001	-0.001	-0.001
Hybrid group	-0.0225	0.0777*	0.0548	-0.0507	-0.0555	-0.0201	-0.0368
	(0.0727)	(0.042)	(0.047)	(0.0443)	(0.0778)	(0.0773)	(0.0557)
Targeting accuracy	1.164***	0.510**	-0.730**	-0.974***	1.009**	0.0681	0.261
	(0.447)	(0.248)	(0.324)	(0.271)	(0.510)	(0.424)	(0.371)
Benefit ratio	0.515*	0.200	-0.293*	0.0383	0.166	-0.0985	0.614***
	(0.281)	(0.138)	(0.166)	(0.151)	(0.305)	(0.248)	(0.190)
Adj. R^2	0.068	0.074	0.028	0.056	0.042	0.021	0.056
Observations	807	924	930	934	728	753	938

* Significant at the 10% level. ** Significant at the 5% level. *** Significant at the 1% level.

Estimations are done by OLS. Standard errors (in parentheses) are clustered at the village level. Households are from the PMT or Hybrid group. The first rows show the results of a simple regression of the outcome on the Hybrid group dummy. Targeting accuracy and local poverty use rank-score welfare as benchmark. The multiple regression specifications control for local poverty, regional dummies, log village size, and dummies indicating whether the household received the benefit and whether household members felt entitled to it, as well as the interaction of those two. The coefficients of these controls are omitted.

Table 5.2: Simple treatment effects and estimation results of equation 4.1 – village level outcomes

	Households satisfied	Household not on list	Wrong household on list	Targeting method correct	Method better than BLT	Complaints to village head (log)	Complaints in complaint box (log)	Enough benefits
Hybrid group	0.530***	-0.0592	-0.0205	0.189***	0.375***	-0.317**	-0.198	0.0493
(simple regression)	(0.0805)	(0.0443)	(0.0206)	(0.0453)	(0.0718)	(0.140)	(0.125)	(0.0553)
Adj. R^2	0.092	0.002	0.000	0.037	0.059	0.010	0.004	0.000
Hybrid group	0.435***	-0.0504	-0.0106	0.143***	0.325***	-0.194	-0.0992	0.0531
	(0.0885)	(0.0448)	(0.0224)	(0.0484)	(0.0785)	(0.147)	(0.108)	(0.0559)
Targeting accuracy	1.502***	-0.233	-0.184	0.912***	0.86	-2.201**	-1.330*	0.00834
	(0.563)	(0.337)	(0.187)	(0.311)	(0.540)	(0.993)	(0.688)	(0.443)
Benefit ratio	0.823***	-0.887***	0.251**	0.187	0.173	-1.502***	-1.082**	1.099***
	(0.311)	(0.167)	(0.113)	(0.187)	(0.300)	(0.543)	(0.438)	(0.234)
Adj. R^2	0.127	0.094	0.092	0.069	0.062	0.080	0.378	0.136
Observations	421	426	426	424	421	426	426	419

* Significant at the 10% level. ** Significant at the 5% level. *** Significant at the 1% level.

Estimations are done by OLS. Standard errors (in parentheses) are heteroscedasticity-consistent. Households are from the PMT or Hybrid group. The first row shows the results of a simple regression of the outcome on the Hybrid group dummy. Targeting accuracy and local poverty use rank-score welfare as benchmark. The multiple regression specifications control for local poverty, regional dummies, and log village size. The coefficients of these controls are omitted.

higher share of village heads to think that households were satisfied, fewer excluded poor households, more non-poor households receiving benefits, and fewer complaints.

As mentioned earlier, two factors could bias the results in Tables 5.1 and 5.2: targeting accuracy and local poverty are imputed, and the targeting measure and allocation method more or less coincide for the Hybrid group. Table A.2 in the appendix shows the results of the same estimations for the PMT group only, and using true targeting accuracy and local poverty. The coefficients of targeting accuracy and benefit ratio are statistically indistinguishable from those in Tables 5.1 and 5.2 and show no evidence of bias in any direction.

Tables 5.3 and 5.4 show the results of estimating equation 4.2. The household level outcomes show fewer significant impacts of targeting accuracy (which is now using actual instead of predicted rankings as benchmark). This is likely due to higher standard errors caused by the lack of variation in targeting accuracy for the Community treatment group. However, the magnitudes in the effects on satisfaction and indicators of correctness of the beneficiary list are comparable to those in Table 5.1. The village level outcomes strongly suggest that targeting accuracy increases household satisfaction and correctness of list and method, and reduces complaints to the village head. At the same time, the effect of the Community group (versus the Hybrid group) is mostly insignificant and fairly inconsistent in sign. This can partly be attributed to large standard errors, stemming from a high correlation between targeting accuracy and treatment group. When comparing the coefficients between the simple and multiple regressions, it becomes evident that including targeting accuracy and other covariates strongly pulls the effect of the treatment group away from its expected direction. Lastly, a higher share of beneficiaries increased satisfaction and perceived correctness of list and method, and reduced complaints. Again, the magnitude of these effects appears lower than that of within-village targeting accuracy.

In summary, household members and village heads seem to notice and appreciate increases in targeting accuracy and the number of benefits. The effects of the treatment group on satisfaction-related outcomes are mostly small and statistically insignificant. The exception is that village heads perceived households to be more satisfied in the Hybrid group than in the PMT group, and generally believed the former method was better than the latter. This notion, however, is not backed up by the respective counterpart variables at the household level, nor by significant effects of more tangible indicators of

Table 5.3: Simple treatment effects and estimation results of equation 4.2 – household level outcomes

	Satisfied	Agree with list	Household not on list	Wrong household on list	Targeting method correct	Method better than BLT	Enough benefits
Community group	0.200***	0.0989***	-0.210***	-0.137***	0.160**	0.094	0.120**
(simple regression)	(0.069)	(0.032)	(0.045)	(0.040)	(0.075)	(0.067)	(0.054)
Adj. R^2	0.018	0.016	0.043	0.022	0.010	0.004	0.009
Community group	0.0443	0.0792	-0.113	-0.00511	0.154	-0.0137	0.136
	(0.125)	(0.0514)	(0.093)	(0.0700)	(0.147)	(0.112)	(0.0939)
Targeting accuracy	0.931	0.0957	-0.54	-0.779**	0.147	0.739	-0.159
	(0.583)	(0.250)	(0.488)	(0.355)	(0.646)	(0.525)	(0.462)
Benefit ratio	0.516**	0.0465	-0.185	0.268**	0.117	0.255	0.545***
	(0.227)	(0.0887)	(0.146)	(0.122)	(0.264)	(0.201)	(0.181)
Adj. R^2	0.094	0.057	0.075	0.060	0.105	0.054	0.085
Observations	831	955	952	958	747	767	967

* Significant at the 10% level. ** Significant at the 5% level. *** Significant at the 1% level.

Estimations are done by OLS. Standard errors (in parentheses) are clustered at the village level. Households are from the Hybrid or Community group. The first row shows the results of a simple regression of the outcome on the Community group dummy. Targeting accuracy uses rank-consistent welfare as benchmark. The multiple regression specifications control for attendance rates of the PWR, regional dummies, log village size, and dummies indicating whether the household received the benefit and whether household members felt entitled to it, as well as the interaction of those two. The coefficients of these controls are omitted.

Table 5.4: Simple treatment effects and estimation results of equation 4.2 – village level outcomes

	Households satisfied	Household not on list	Wrong household on list	Targeting method correct	Method better than BLT	Complaints to village head (log)	Complaints in complaint box (log)	Enough benefits
Community group	0.403***	-0.107**	0.010	0.185***	0.210***	-0.410***	-0.116	0.108*
(simple regression)	(0.069)	(0.047)	(0.019)	(0.034)	(0.055)	(0.125)	(0.108)	(0.056)
Adj. R^2	0.072	0.010	-0.002	0.063	0.031	0.022	0.000	0.007
Community group	0.0544	0.0846	0.0715**	0.0669	-0.0451	0.113	-0.16	-0.0035
	(0.126)	(0.0839)	(0.0282)	(0.0573)	(0.0933)	(0.214)	(0.186)	(0.0946)
Targeting accuracy	2.152***	-1.129***	-0.385**	0.733**	1.523***	-3.059***	0.208	0.671
	(0.668)	(0.421)	(0.195)	(0.332)	(0.556)	(1.132)	(0.944)	(0.519)
Benefit ratio	1.038***	-0.811***	0.0291	0.059	0.359**	-1.340***	-0.278	0.781***
	(0.200)	(0.149)	(0.0544)	(0.104)	(0.141)	(0.354)	(0.336)	(0.184)
Adj. R^2	0.137	0.081	0.053	0.072	0.055	0.097	0.294	0.076
Observations	423	431	431	427	425	431	431	423

* Significant at the 10% level. ** Significant at the 5% level. *** Significant at the 1% level.

Estimations are done by OLS. Standard errors (in parentheses) are heteroscedasticity-consistent. Households are from the Hybrid or Community group. The first row shows the results of a simple regression of the outcome on the Community group dummy. Targeting accuracy uses rank-consistent welfare as benchmark. The multiple regression specifications control for attendance rates of the PWR, regional dummies, and log village size. The coefficients of these controls are omitted.

dissatisfaction such as complaints received. It is thus possible that, while the targeting method itself has little intrinsic value to most people, village heads have a preference for an allocation by ranking. This might be due to a sense of importance derived from being involved in the execution of the PWRs, or because PWRs were perceived to be more different than the PMT method from the allocation method of the BLT, which was deemed highly inadequate.

Tables 5.5 and 5.6 show the results of estimating equation 4.3. On the household level, the results indicate that targeting accuracy based on rank-score welfare is significantly predictive for most satisfaction outcomes—even conditional on targeting accuracy measured by predicted consumption. The reverse is not true for any outcome. On the village level, the same conclusion holds for the number of complaints to the village head and in the complaint box. Taken together, these results suggest that targeting accuracy explains satisfaction better when it is based on predicted ranks rather than predicted consumption.

One reason for this might be that the consumption model simply has a lower predictive performance than the model of ranks. After all, the two models are estimated based on different samples: the consumption model uses the baseline survey of 9 households per village, whereas the rank model uses all the households from the Hybrid group in the BPS data. To rule out this potential element of unfairness in the comparison, I create an alternative set of welfare scores based on a ROL model that uses the baseline survey instead. Tables A.3 and A.4 in the appendix show estimations of equation 4.3 with targeting accuracy and local poverty based on these new welfare scores. The results remain qualitatively the same. Tables A.5 and A.6 in the appendix display estimation results of estimating equation 4.3 using total targeting accuracy (and therefore leaving out the benefit ratio). The results confirm the above findings, and show significant effects of targeting accuracy based on rank-score welfare even for most of the village level outcomes. Lastly, to confirm that the results are not a coincidental product of focusing on the PMT group only, I also ran the same regressions including the Hybrid group, using imputed targeting accuracy and local poverty and including a treatment group dummy. The results, reported in Tables A.7 and A.8 in the appendix, again remain qualitatively unchanged.

The way in which local poverty would impact program satisfaction is not clear a priori. However, one would expect that, controlling for the benefit ratio, villages with higher poverty incidence register more omitted poor households and fewer targeted non-poor households, and a comparatively less sufficient

Table 5.5: Estimation results of equation 4.3 – household level outcomes

	Satisfied	Agree with list	Household not on list	Wrong household on list	Targeting method correct	Method better than BLT	Enough benefits
Targeting accuracy	1.278***	0.811**	-0.31	-1.212***	1.125**	0.438	0.917**
(rank-score welfare)	(0.473)	(0.334)	(0.377)	(0.323)	(0.552)	(0.504)	(0.427)
Targeting accuracy	-0.649	0.062	-0.0641	-0.23	-0.596	-1.120*	-0.737
(predicted cons.)	(0.695)	(0.384)	(0.421)	(0.383)	(0.684)	(0.643)	(0.518)
Local poverty rate	-0.135	0.316*	0.409*	-0.489**	-0.162	0.0475	-0.463*
(rank-score welfare)	(0.252)	(0.180)	(0.238)	(0.197)	(0.337)	(0.330)	(0.266)
Local poverty rate	-0.420*	-0.205	-0.0298	0.411**	-0.219	-0.598***	-0.0183
(predicted cons.)	(0.231)	(0.172)	(0.166)	(0.172)	(0.290)	(0.202)	(0.201)
Benefit ratio	0.406	0.282*	-0.0383	-0.243	-0.175	-0.0763	0.244
	(0.337)	(0.154)	(0.216)	(0.176)	(0.389)	(0.292)	(0.236)
Adj. R^2	0.088	0.080	0.012	0.088	0.041	0.024	0.062
Observations	383	453	456	458	342	357	460

* Significant at the 10% level. ** Significant at the 5% level. *** Significant at the 1% level.

Estimations are done by OLS. Standard errors (in parentheses) are clustered at the village level. Households are from the PMT group only. All specifications control for regional dummies, log village size, and dummies indicating whether the household received the benefit and whether household members felt entitled to it, as well as the interaction of those two. The coefficients of these controls are omitted.

Table 5.6: Estimation results of equation 4.3 – village level outcomes

	Households satisfied	Household not on list	Wrong household on list	Targeting method correct	Method better than BLT	Complaints to village head (log)	Complaints in complaint box (log)	Enough benefits
Targeting accuracy	0.744	0.0456	-0.0729	0.467	1.019	-4.341***	-1.720**	-0.146
(rank-score welfare)	(0.603)	(0.330)	(0.188)	(0.393)	(0.619)	(1.057)	(0.689)	(0.435)
Targeting accuracy	1.14	-0.034	0.16	0.348	-0.43	1.876	1.275	0.0421
(predicted cons.)	(1.020)	(0.402)	(0.205)	(0.517)	(0.883)	(1.407)	(1.113)	(0.527)
Local poverty rate	0.474	0.344	-0.223*	0.468*	0.700*	-1.240*	0.252	-0.644**
(rank-score welfare)	(0.451)	(0.232)	(0.135)	(0.246)	(0.405)	(0.684)	(0.640)	(0.324)
Local poverty rate	-0.0593	0.207	0.028	-0.163	-0.0872	0.952	0.314	-0.171
(predicted cons.)	(0.371)	(0.184)	(0.121)	(0.207)	(0.354)	(0.581)	(0.466)	(0.269)
Benefit ratio	1.292***	-0.944***	0.336**	0.256	0.166	-1.499**	-1.051*	1.181***
	(0.483)	(0.228)	(0.166)	(0.261)	(0.434)	(0.750)	(0.612)	(0.347)
Adj. R^2	0.027	0.114	0.066	0.003	-0.023	0.136	0.435	0.126
Observations	206	209	209	209	208	209	209	204

* Significant at the 10% level. ** Significant at the 5% level. *** Significant at the 1% level.

Estimations are done by OLS. Standard errors (in parentheses) are heteroscedasticity-consistent. Households are from the PMT group only. All specifications control for regional dummies and log village size. The coefficients of these controls are omitted.

amount of benefits. The results in the tables above and the appendix confirm this pattern for local poverty measured against rank-score welfare, but not for consumption. It is noteworthy that the poverty rate based on the ROL model—which does not take into account village level predictors and does not compare households between villages—is able to pick up these relationships better than the poverty rate based on a consumption regression model that does include village level predictors.

6 Conclusion

The findings in this paper suggest that PWRs are popular largely due to their outcome, not the procedure. Furthermore, allocations coinciding with scores from PWRs were better-received than those coinciding with consumption scores. These insights could lead to improvements in targeting: similar to PMT scores of consumption, one could use welfare scores, trained on PWRs for a representative sample of villages. Then, the households with the lowest scores below a predefined threshold are targeted. The targeting threshold could come from a poverty map or be set according to budgetary restrictions. It could also come from the PWR itself, for instance if it takes the form of a poverty classification exercise instead of a continuous welfare ranking.

Using PWR-based welfare scores has advantages compared to conducting actual PWRs in all locations, as the latter may not be feasible or desirable in some contexts. This is the case if people are not expected to know most of the households in their neighborhood, or if poverty is too shameful a topic to be discussed publicly. Further worries might include selfish ranking behavior, discrimination, elite capture, or coercion. Using scores based on training PWRs without rank-dependent payouts should mitigate such concerns. The targeting cost should not differ much between the two methods.[10]

Constructing welfare scores based on PWRs is much less unambiguous than constructing predictions of consumption. In this paper, PWR results were treated as correct representations of welfare. But by doing so, systematic biases regarding minorities or others might creep into the score. Going forward, it will

[10] For the field experiment by Alatas et al. (2012), the cost per village of administering the asset surveys was slightly higher than administering the PWR meetings (153$ vs. 110$). However, the survey questionnaire included roughly 100 questions (depending on household size), while for the rank score predictions in this paper, the questions needed per household ranged between 11 and 25 (depending on the province and urban/rural). The actual implementation should thus be comparatively less costly.

be useful to better understand the conditions leading to such biases. Beyond that, it may be worth it to apply debiasing techniques to the prediction algorithm. Exploring the effects of bias corrections on the resulting allocations and on satisfaction outcomes may be a worthwhile task for future research.

References

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A Olken, and Julia Tobias. 2012. "Targeting the Poor: Evidence from a Field Experiment in Indonesia." *American Economic Review*, 102(4): 1206–1240.

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A. Olken, Ririn Purnamasari, and Matthew Wai-Poi. 2019. "Does Elite Capture Matter? Local Elites and Targeted Welfare Programs in Indonesia." *AEA Papers and Proceedings*, 109: 334–339.

Alatas, Vivi, Ririn Purnamasari, Matthew Wai-Poi, Abhijit Banerjee, Benjamin A Olken, and Rema Hanna. 2016. "Self-targeting: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy*, 124(2): 371–427.

Alkire, Sabina, and James Foster. 2011. "Counting and Multidimensional Poverty Measurement." *Journal of Public Economics*, 95(7): 476–487.

Bah, Adama, Samuel Bazzi, Sudarno Sumarto, and Julia Tobias. 2019. "Finding the Poor vs. Measuring Their Poverty: Exploring the Drivers of Targeting Effectiveness in Indonesia." *The World Bank Economic Review*, 33(3): 573–597.

Banerjee, Abhijit, Esther Duflo, Raghabendra Chattopadhyay, and Jeremy Shapiro. 2009. "Targeting Efficiency: How Well Can We Identify the Poorest of the Poor?" *Institute for Financial Management and Research Centre for Micro Finance Working Paper*, 21.

Beggs, Steven, Scott Cardell, and Jerry A. Hausman. 1981. "Assessing the Potential Demand for Electric Cars." *Journal of Econometrics*, 17(1): 1–19.

Cameron, Lisa, and Manisha Shah. 2014. "Can Mistargeting Destroy Social Capital and Stimulate Crime? Evidence from a Cash Transfer Program in Indonesia." *Economic Development and Cultural Change*, 62(2): 381–415.

Coady, David, Margaret Grosh, and John Hoddinott. 2004. "Targeting Outcomes Redux." *The World Bank Research Observer*, 19(1): 61–85.

Coady, David P., and Susan W. Parker. 2009. "Targeting Performance under Self-selection and Administrative Targeting Methods." *Economic Development and Cultural Change*, 57(3): 559–587.

Hurvich, Clifford M., and Chih-Ling Tsai. 1989. "Regression and Time Series Model Selection in Small Samples." *Biometrika*, 76(2): 297–307.

Kanbur, Ravi, and Paul Shaffer. 2007. "Epistemology, Normative Theory and Poverty Analysis: Implications for Q-Squared in Practice." *World Development*, 35(2): 183–196. Experiences of Combining Qualitative and Quantitative Approaches in Poverty Analysis.

Karlan, Dean, and Bram Thuysbaert. 2019. "Targeting Ultra-Poor Households in Honduras and Peru." *The World Bank Economic Review*, 33(1): 63–94.

Mansuri, Ghazala, and Vijayendra Rao. 2013. *Localizing Development: Does Participation Work?* Washington, D.C.:The World Bank.

Ravallion, Martin. 2009. "How Relevant Is Targeting to the Success of an Antipoverty Program?" *The World Bank Research Observer*, 24(2): 205–231.

Schüring, Esther. 2014. "Preferences for Community-based Targeting – Field Experimental Evidence from Zambia." *World Development*, 54: 360–373.

Sen, Amartya. 1980. "Equality of What?" *The Tanner Lecture on Human Values*, I: 197–220.

Shaffer, Paul. 2013. "Ten Years of "Q-Squared": Implications for Understanding and Explaining Poverty." *World Development*, 45: 269–285.

Widjaja, Muliadi. 2012. "An Economic and Social Review on Indonesia's Direct Cash Transfer Program to Poor Families in 2005." *Economics and Finance in Indonesia*, 60: 197–212.

Yusuf, Moeed. 2010. *Community Targeting for Poverty Reduction: Lessons from Developing Countries.* Boston University.

Zeller, Manfred, Joseph Feulefack, and Andreas Neef. 2006. "How Accurate is Participatory Wealth Ranking (PWR) in Targeting the Poor? A Case Study from Bangladesh." International Association of Agricultural Economists 2006 Annual Meeting, August 12-18, 2006, Queensland, Australia.

A Appendix

A.1 Imputation Procedure

In this section, I address the issue that the better-off households (according to the PWR) in the Hybrid group do not appear in the BPS data. I subsume the Hybrid group households represented in the BPS dataset under H_1^{Hybrid} and the other ones under H_2^{Hybrid}. If we use actual rankings as benchmark, the fact that the households in H_2^{Hybrid} are unobserved is not a problem, as they are all counted as non-poor and did not receive the benefit, and thus all count as correctly targeted. However when using predicted welfare as benchmark, it cannot be ruled out that the model would have classified some of the unobserved households as poor. Thus, in assuming that all of them were targeted correctly we would overstate targeting accuracy.

I propose to look at the households within H_2^{Hybrid} that were visited for the baseline survey, $H_2^{\text{Hybrid, baseline}}$, to get an average targeting accuracy for the unobserved part in each village. To do so, I go through the following steps.

1. For the PMT group, I establish which households would most likely have been included in the BPS survey, had they been in the Hybrid group instead. This is to make predictions in the PMT group just as in the Hybrid group. With the households in the BPS survey, I estimate a logit model of treatment status on household characteristics, and assign propensity scores to the households of the PMT group. The lower these are, the less likely it is to find a similar household in the Hybrid group, meaning that its equivalent in the Hybrid group would less likely have been included in the BPS data. I then rank households in each village j of the PMT group by propensity score and subsume the $\min\{1.5 \cdot n_j b_j, n_j\}$ highest scoring households under the set H_1^{PMT}, and the remaining ones under H_2^{PMT}, with subset $H_2^{\text{PMT, baseline}}$ of households observed in the baseline survey.

2. I predict welfare for the households in $H_1^{\text{PMT}} \cup H_2^{\text{PMT}}$ and rank all of them according to their score. I define the absolute poverty line as the welfare score of the $B^{\text{PMT}th}$ poorest household, with B^{PMT} the number of benefits made available to all households in the PMT group. Furthermore, for each village j in the PMT group I compute t_j and h_j.

3. I predict welfare for the households in $H_1^{\text{Hybrid}} \cup H_2^{\text{Hybrid, baseline}}$. Then for each village j, both in the PMT group and the Hybrid group, I com-

pute the *reduced* targeting accuracy $t_{1,j}$, that pretends the households in $H_1^{\text{PMT}} \cup H_1^{\text{Hybrid}}$ make up the entire village.

4. Using the households in $H_2^{\text{PMT, baseline}}$ and $H_2^{\text{Hybrid, baseline}}$, respectively, I calculate t_2^{PMT} and t_2^{Hybrid}—the overall fractions of households below the local poverty lines used to calculate $t_{1,j}$ in the step above. These averages are used as a proxy for mistargeting among the households unobserved by the BPS. Since the number of households per village in $H_2^{\text{PMT, baseline}}$ and $H_2^{\text{Hybrid, baseline}}$ is very low—0 in some cases—it is not feasible to use village-specific averages instead.

5. Eventually, the two separate estimates are combined in the following way. I construct two factors, $z_{1,j} = s_j t_{1,j}$ and $z_{2,j} = (1 - s_j)\, t_{2,g}$, where s_j is the fraction of households from $H_1^{\text{PMT}} \cup H_1^{\text{Hybrid}}$ in village j. I then regress t_j on $z_{1,j}$ and $z_{2,j}$ for the PMT villages and construct linear predictions, $\hat{t}_j$, for both the PMT group and the Hybrid group.

The reason for not simply adding up $z_{1,j}$ and $z_{2,j}$ is that both $t_{1,j}$ and $t_{2,g}$ are not unbiased estimates of their respective shares of targeting accuracy: $t_{1,j}$ does not take into account that the poorer households in $H_2^{\text{PMT}} \cup H_2^{\text{Hybrid}}$ may render some of the households declared poor in $H_1^{\text{PMT}} \cup H_1^{\text{Hybrid}}$ non-poor, and $t_{2,g}$ does not take into account that this is likely to change the local poverty line.

Just as targeting accuracy, the local poverty rate h_j also needs to be estimated, since the total number of poor households is not observed for the Hybrid group. I take a slightly different approach than for targeting accuracy, though, as the treatment group should have no influence on the total amount of benefits (which was determined from a government census, independently of the treatment group assignment). Therefore, predictions $\hat{h}_j$ of the poverty headcount for the PMT and the Hybrid group are made with the PMT group as training data. As candidate predictors I use dummies for region, (log) village size, the fraction of observed households s_j, and the poverty headcount only considering the households in $H_1^{\text{PMT}} \cup H_1^{\text{Hybrid}}$, as well as various transformations and interactions of these variables. Just as in the welfare models, I use a stepwise model selection procedure as well as cross-fitting to prevent overfitting.

Comparing the effects of t_j and $\hat{t}_j$ as well as h_j and $\hat{h}_j$ for rank-score welfare in the PMT group shows that the imputation does not significantly change measured impacts. The results of this comparison are available on request.

A.2 Tables

Table A.1: List of household and local characteristics

Variable description
Household demographics
Household size
Age of household head
Gender of household head
Marriage status of household head
Number of children (between 0 and 4, going to primary school, going to junior high school)
Dependency ratio
Village head lives in household
Education
Household head's education (graduated from no school, primary school, junior high school, high school or higher)
Household member with highest qualification (graduated from no school, primary school, junior high school, high school or higher)
Occupation
Household head works in agricultural sector (including mining / quarrying)
Household head works in industrial sector
Household head works in service sector
Housing characteristics
Privately owned house
Per capita floor area
Type of floor, walls, and roof
Private toilet
Clean drinking water
Electricity source
Cooking fuel
Assets
Kitchen appliances (gas burner, fridge/freezer, rice cooker, mixer/blender)
Electronic devices (air conditioning, fan, radio, TV, DVD/VCD player, laptop/PC, dish antenna, cell phone)
Livestock (poultry, pig, goat, cow/buffalo, horse)
Means of transport (bike, motorbike, car)

Table A.1: List of household and local characteristics

Variable description
Productive machinery (sewing machine, electric pump)
Jewelry/gold
Household ever received credit
Village characteristics
Number of households
Schools (primary school, junior high school)
Medical facilities (doctor, midwife, neighborhood medical center, medical center, clinic)
Semi/permanent market place
Credit facility
Road type
Mean agricultural land area
Subdistrict characteristics
Ratios of household heads working in agricultural / industrial / service sector
Ratios of household heads graduated from no school / primary school / junior high school / high school or higher
Mean per capita floor area
Mean agricultural land area
Ratio of households with clean drinking water

Table A.2: Estimation results of equation 4.1 using only the PMT group and no imputations

Dependent variable (household level)	Satisfied	Agree with list	Household not on list	Wrong household on list	Targeting method correct	Method better than BLT	Enough benefits
Targeting accuracy	1.226**	0.834**	-0.316	-1.276***	1.076*	0.345	0.822**
	(0.478)	(0.330)	(0.374)	(0.310)	(0.589)	(0.539)	(0.4060)
Benefit ratio	0.474	0.234	-0.0317	-0.128	-0.100	0.045	0.376*
	(0.321)	(0.157)	(0.203)	(0.196)	(0.347)	(0.2900)	(0.219)
Adj. R^2	0.080	0.078	0.016	0.066	0.042	-0.006	0.059
Observations	383	453	456	458	342	357	460

Dependent variable (village level)	Households satisfied	Household not on list	Wrong household on list	Targeting method correct	Method better than BLT	Complaints to village head (log)	Complaints in complaint box (log)	Enough benefits
Targeting accuracy	0.919	0.0679	-0.0431	0.502	0.936	-3.908***	-1.471**	-0.159
	(0.567)	(0.312)	(0.191)	(0.376)	(0.605)	(1.005)	(0.634)	(0.423)
Benefit ratio	1.121**	-0.906***	0.319**	0.182	0.211	-1.604**	-1.176*	1.147***
	(0.451)	(0.205)	(0.157)	(0.251)	(0.418)	(0.724)	(0.627)	(0.310)
Adj. R^2	0.029	0.117	0.073	0.007	-0.014	0.128	0.435	0.133
Observations	206	209	209	209	208	209	209	204

* Significant at the 10% level. ** Significant at the 5% level. *** Significant at the 1% level.

Estimations are done by OLS. For outcomes at the household level, standard errors (in parentheses) are clustered at the village level. For outcomes at the village level, standard errors are heteroscedasticity-consistent. Households are from the PMT group. Targeting accuracy and local poverty use rank-score welfare as benchmark. All specifications control for local poverty, regional dummies, and log village size. In addition, the specifications with household-level outcomes control for dummies indicating whether the household received the benefit and whether household members felt entitled to it, as well as the interaction of those two. The coefficients of these controls are omitted.

Table A.3: Estimation results of equation 4.3 using survey welfare model – household level outcomes

	Satisfied	Agree with list	Household not on list	Wrong household on list	Targeting method correct	Method better than BLT	Enough benefits
Targeting accuracy	0.754	0.680**	-0.650*	-0.668**	1.589***	-0.00168	1.105***
(rank-score welfare)	(0.504)	(0.290)	(0.373)	(0.327)	(0.543)	(0.495)	(0.387)
Targeting accuracy	-0.563	0.153	-0.00948	-0.374	-0.623	-1.059	-0.744
(predicted cons.)	(0.706)	(0.392)	(0.413)	(0.372)	(0.669)	(0.643)	(0.517)
Local poverty rate	0.200	0.294	0.159	-0.565***	0.211	0.0683	-0.340
(rank-score welfare)	(0.346)	(0.189)	(0.250)	(0.211)	(0.396)	(0.402)	(0.276)
Local poverty rate	-0.602**	-0.26	0.0158	0.556***	-0.441	-0.626**	0.00364
(predicted cons.)	(0.285)	(0.184)	(0.201)	(0.176)	(0.341)	(0.271)	(0.236)
Benefit ratio	0.380	0.314**	-0.0841	-0.216	-0.106	-0.098	0.322
	(0.346)	(0.155)	(0.201)	(0.201)	(0.379)	(0.291)	(0.213)
Adj. R^2	0.087	0.076	0.002	0.086	0.040	0.024	0.061
Observations	383	453	456	458	342	357	460

* Significant at the 10% level. ** Significant at the 5% level. *** Significant at the 1% level.

Estimations are done by OLS. Standard errors (in parentheses) are clustered at the village level. Households are from the PMT group only. All specifications control for regional dummies, log village size, and dummies indicating whether the household received the benefit and whether household members felt entitled to it, as well as the interaction of those two. The coefficients of these controls are omitted.

Table A.4: Estimation results of equation 4.3 using survey welfare model – village level outcomes

	Households satisfied	Household not on list	Wrong household on list	Targeting method correct	Method better than BLT	Complaints to village head (log)	Complaints in complaint box (log)	Enough benefits
Targeting accuracy	0.835	-0.00471	-0.0454	1.047***	0.636	-0.919	-3.380***	0.0991
(rank-score welfare)	(0.660)	(0.374)	(0.183)	(0.377)	(0.705)	(1.248)	(0.781)	(0.482)
Targeting accuracy	1.136	-0.0444	0.163	0.233	-0.304	0.770	1.515	0.0195
(predicted cons.)	(1.010)	(0.396)	(0.212)	(0.509)	(0.878)	(1.432)	(1.058)	(0.530)
Local poverty rate	0.531	0.414*	-0.270*	0.746***	0.754*	-0.416	-0.928	-0.736**
(rank-score welfare)	(0.448)	(0.230)	(0.149)	(0.258)	(0.450)	(0.823)	(0.608)	(0.329)
Local poverty rate	-0.186	0.104	0.0963	-0.419*	-0.242	0.646	0.959*	-0.00243
(predicted cons.)	(0.409)	(0.211)	(0.144)	(0.231)	(0.412)	(0.669)	(0.503)	(0.315)
Benefit ratio	1.277***	-0.975***	0.355**	0.306	0.0727	-1.030	-1.185**	1.256***
	(0.484)	(0.232)	(0.168)	(0.267)	(0.444)	(0.814)	(0.598)	(0.349)
Adj. R^2	0.028	0.118	0.072	0.017	-0.022	0.126	0.439	0.132
Observations	206	209	209	209	208	209	209	204

* Significant at the 10% level. ** Significant at the 5% level. *** Significant at the 1% level.

Estimations are done by OLS. Standard errors (in parentheses) are heteroscedasticity-consistent. Households are from the PMT group only. All specifications control for regional dummies and log village size. The coefficients of these controls are omitted.

Table A.5: Estimation results of equation 4.3 using total targeting accuracy – household level outcomes

	Satisfied	Agree with list	Household not on list	Wrong household on list	Targeting method correct	Method better than BLT	Enough benefits
Targeting accuracy	1.309**	0.554	-0.427	-1.175***	1.349**	0.519	0.626
(rank-score welfare)	(0.560)	(0.388)	(0.418)	(0.340)	(0.614)	(0.595)	(0.466)
Targeting accuracy	0.789	0.833*	-0.098	-0.439	-0.515	-1.646	-0.982
(predicted cons.)	(1.013)	(0.469)	(0.648)	(0.516)	(1.138)	(1.015)	(0.725)
Local poverty rate	-0.243	0.177	0.575**	-0.181	-0.417	-0.183	-0.670***
(rank-score welfare)	(0.255)	(0.214)	(0.229)	(0.232)	(0.385)	(0.347)	(0.240)
Local poverty rate	-0.496**	-0.0463	-0.0438	0.241	-0.243	-0.762***	-0.0601
(predicted cons.)	(0.226)	(0.179)	(0.176)	(0.179)	(0.274)	(0.230)	(0.210)
Benefit ratio	0.921**	0.419**	-0.171	-0.388*	0.154	0.0366	0.335
(predicted cons.)	(0.369)	(0.205)	(0.253)	(0.226)	(0.462)	(0.392)	(0.252)
Adj. R^2	0.088	0.067	0.020	0.053	0.046	0.054	0.059
Observations	383	453	456	458	342	357	460

* Significant at the 10% level. ** Significant at the 5% level. *** Significant at the 1% level.

Estimations are done by OLS. Standard errors (in parentheses) are clustered at the village level. Households are from the PMT group only. All specifications control for regional dummies, log village size, and dummies indicating whether the household received the benefit and whether household members felt entitled to it, as well as the interaction of those two. The coefficients of these controls are omitted.

Table A.6: Estimation results of equation 4.3 using total targeting accuracy – village level outcomes

	Households satisfied	Household not on list	Wrong household on list	Targeting method correct	Method better than BLT	Complaints to village head (log)	Complaints in complaint box (log)	Enough benefits
Targeting accuracy	0.926	0.556	-0.38	0.978*	1.067	-2.396	-1.682	-0.986
(rank-score welfare)	(0.816)	(0.486)	(0.351)	(0.513)	(0.909)	(1.535)	(1.036)	(0.724)
Targeting accuracy	-0.111	0.328	0.549	-0.879	-2.419*	2.976	0.867	0.125
(predicted cons.)	(1.390)	(0.667)	(0.429)	(0.751)	(1.351)	(2.146)	(1.747)	(0.969)
Local poverty rate	0.794*	0.539**	-0.301**	0.393	0.58	-0.471	-0.13	-0.920***
(rank-score welfare)	(0.443)	(0.253)	(0.134)	(0.271)	(0.448)	(0.719)	(0.585)	(0.335)
Local poverty rate	0.101	0.347*	-0.0772	-0.0257	0.184	0.878	0.293	-0.406
(predicted cons.)	(0.375)	(0.186)	(0.128)	(0.219)	(0.346)	(0.669)	(0.500)	(0.272)
Benefit ratio	0.968*	-0.878***	0.396*	0.128	-0.244	-1.076	-1.052	1.177***
(predicted cons.)	(0.549)	(0.271)	(0.219)	(0.311)	(0.527)	(0.922)	(0.672)	(0.450)
Adj. R^2	0.024	0.142	0.100	0.007	-0.014	0.083	0.423	0.165
Observations	206	209	209	209	208	209	209	204

* Significant at the 10% level. ** Significant at the 5% level. *** Significant at the 1% level.

Estimations are done by OLS. Standard errors (in parentheses) are heteroscedasticity-consistent. Households are from the PMT group only. All specifications control for regional dummies and log village size. The coefficients of these controls are omitted.

Table A.7: Estimation results of equation 4.3 including PMT and Hybrid group – household level outcomes

	Satisfied	Agree with list	Household not on list	Wrong household on list	Targeting method correct	Method better than BLT	Enough benefits
Total targeting acc.	1.264***	0.435	-0.229	-0.313	1.393**	1.092**	0.407
(rank-score welfare)	(0.414)	(0.309)	(0.327)	(0.316)	(0.539)	(0.497)	(0.420)
Total targeting acc.	-0.326	-0.0698	-0.0248	0.0493	-0.575	-0.596*	-0.338
(predicted cons.)	(0.369)	(0.234)	(0.294)	(0.257)	(0.441)	(0.332)	(0.357)
Local poverty rate	0.0802	0.383**	0.377	-0.516**	0.0207	0.193	-0.432
(rank-score welfare)	(0.255)	(0.192)	(0.251)	(0.226)	(0.385)	(0.352)	(0.285)
Local poverty rate	-0.366*	-0.199	-0.0279	0.443**	-0.253	-0.555***	0.0279
(predicted cons.)	(0.213)	(0.166)	(0.174)	(0.180)	(0.267)	(0.197)	(0.223)
Adj. R^2	0.092	0.068	0.014	0.053	0.063	0.041	0.046
Observations	383	453	456	458	342	357	460

* Significant at the 10% level. ** Significant at the 5% level. *** Significant at the 1% level.

Estimations are done by OLS. Standard errors (in parentheses) are clustered at the village level. Households are from the PMT or Hybrid group. All specifications control for treatment group, regional dummies, log village size, and dummies indicating whether the household received the benefit and whether household members felt entitled to it, as well as the interaction of those two. The coefficients of these controls are omitted.

Table A.8: Estimation results of equation 4.3 including PMT and Hybrid group – village level outcomes

	Households satisfied	Household not on list	Wrong household on list	Targeting method correct	Method better than BLT	Complaints to village head (log)	Complaints in complaint box (log)	Enough benefits
Total targeting acc.	1.727**	0.611*	-0.513**	0.772**	0.916	-1.060	-2.891***	-1.065**
(rank-score welfare)	(0.682)	(0.330)	(0.211)	(0.391)	(0.656)	(1.220)	(0.947)	(0.469)
Total targeting acc.	-0.299	0.0649	0.102	0.151	-0.412	0.932	0.247	-0.0806
(predicted cons.)	(0.620)	(0.294)	(0.160)	(0.309)	(0.500)	(0.957)	(0.705)	(0.394)
Local poverty rate	0.970**	0.368	-0.301**	0.657**	0.834*	-1.195	-0.479	-0.764**
(rank-score welfare)	(0.484)	(0.256)	(0.136)	(0.256)	(0.436)	(0.735)	(0.564)	(0.347)
Local poverty rate	0.0637	0.162	0.0415	-0.0709	-0.0494	0.704	-0.0463	-0.136
(predicted cons.)	(0.391)	(0.212)	(0.137)	(0.211)	(0.365)	(0.632)	(0.487)	(0.312)
Adj. R^2	0.021	0.036	0.066	0.021	-0.020	0.067	0.452	0.067
Observations	206	209	209	209	208	209	209	204

* Significant at the 10% level. ** Significant at the 5% level. *** Significant at the 1% level.

Estimations are done by OLS. Standard errors (in parentheses) are heteroscedasticity-consistent. Households are from the PMT or Hybrid group. All specifications control for treatment group, regional dummies, and log village size. The coefficients of these controls are omitted.