Self-Targeting: Evidence from a Field Experiment in Indonesia

Vivi Alatas

World Bank

Abhijit Banerjee

Massachusetts Institute of Technology

Rema Hanna

Harvard University

Benjamin A. Olken

Massachusetts Institute of Technology

Ririn Purnamasari

World Bank

Matthew Wai-Poi

World Bank

This paper shows that adding a small application cost to a transfer program can substantially improve targeting through self-selection. Our village-level experiment in Indonesia finds that requiring beneficiaries

This project was a collaboration involving many people. We thank Jie Bai, Talitha Chairunissa, Amri Ilmma, Donghee Jo, Chaeruddin Kodir, Gabriel Kreindler, He Yang, Ariel Zucker, and Gabriel Zucker for their excellent research assistance and Raj Chetty, Esther Duflo, Amy Finkelstein, and numerous seminar participants for helpful comments. We thank Mitra Samya, the Indonesian Central Bureau of Statistics, the Indonesian Na-

Electronically published March 7, 2016
[*Journal of Political Economy*, 2016, vol. 124, no. 2]

to apply for benefits results in substantially poorer beneficiaries than automatic enrollment using the same asset test. Marginally increasing application costs on an experimental basis does not further improve targeting. Estimating a model of the application decision implies that the results are largely driven by the nonpoor, who make up the bulk of the population, forecasting that they are unlikely to pass the asset test and therefore not bothering to apply.

I. Introduction

In designing targeted aid programs, a perennial problem is how to separate the poor from the rich. One solution is to impose requirements that are more costly for the rich than for the poor (Nichols, Smolensky, and Tideman 1971; Nichols and Zeckhauser 1982; Ravallion 1991; Besley and Coate 1992). These self-selection mechanisms are common: welfare programs, from the Works Progress Administration in the United States during the Great Depression to the National Rural Employment Guarantee Act (right-to-work) scheme in India today, often have manual labor requirements to receive aid. Similarly, subsidized food schemes often provide low-quality food, which leads those who can afford better food to choose not to purchase subsidized products.

The challenge with these self-selection mechanisms is that they may be quite inefficient: in order to dissuade the rich from participating, the poor are forced to incur substantial utility costs in order to receive transfers, whether by toiling in the hot sun or eating unappetizing food. In this paper, we ask whether much smaller costs can still achieve substantial self-selection. In particular, we show that when applying for benefits has a cost and there is a good but not perfect procedure for screening out unsuitable applicants, those who are not supposed to get benefits will correctly foresee that they face only a very small chance of slipping through the screening procedure and therefore will not bother to apply. The resulting reductions in inclusion error may substantially improve the degree to which the program is targeted to the poor.

We conduct a randomized experiment in the context of Indonesia's Conditional Cash Transfer program, known as PKH. Conditional cash transfer programs have spread rapidly throughout the developing world and are present in over 30 countries today. In Indonesia, PKH provides

tional Team for the Acceleration of Poverty Reduction (TNP2K, particularly Sudarno Sumarto and Bambang Widianto), the Indonesian Social Affairs Department (DepSos), and SurveyMeter for their cooperation implementing the project. Most of all, we thank Jurist Tan for her truly exceptional work leading the field implementation. This project was financially supported by the World Bank, AusAID, and 3ie, and analysis was supported by the National Institutes of Health under grant P01 HD061315. All views expressed are those of the authors and do not necessarily reflect the views of the World Bank, TNP2K, Mitra Samya, DepSos, or the Indonesian Central Bureau of Statistics.

beneficiaries with US$130 per year for 6 years and is one of the country's largest social assistance programs, covering about 2.4 million households. The program is aimed at the poorest 5–10 percent of the population, with eligibility determined on the basis of a weighted sum of about 30 easy-to-observe assets (e.g., size of house, materials used to construct household roof, motorbike ownership).

Working with the Indonesian government, we experimentally varied the enrollment process for PKH across 400 villages, comparing a process that required households to apply for the program with the procedure that the government implements in other areas, in which the government statistical system conducts the asset test for potential beneficiaries (chosen through prior asset surveys and consultations with village leadership) at their home and automatically enrolled those that passed. In both cases, eligibility was determined on the basis of an asset screen known as a proxy means test (PMT), so the key difference we studied was whether households had to actively apply to be screened for eligibility or instead were automatically screened on the basis of the results of a government survey administered to a subset of the population. These two approaches to targeted social assistance programs—automatic screening based on a top-down survey or enrollment limited to those who actively apply—are the two most common ways of determining beneficiary lists for targeted transfer programs in the developing world (Grosh et al. 2008; Kidd and Wylde 2011).[1]

In villages randomized to receive the application process ("self-targeting" villages), interested households were required to go to a central registration site to take an asset test administered by the statistics office. This entailed both traveling up to a few kilometers to the application site and waiting in line to apply. Within these areas, we randomly varied the application costs by varying the distance to the application site. However, the program was intentionally set up such that even the highest level of application costs—the sum total of about half a day's missed work, a few kilometers of travel, and a few hours of waiting—pales in comparison to the benefits on offer, which amount to $130 per year for 6 years.

In control areas (automatic screening villages), the usual government procedure was followed: the statistics office, working with local government officials, drew up a list of potential beneficiaries, interviewed every-

[1] Examples of automatic screening PMTs include the Mexican Progresa program, the Colombian social assistance programs, the Indian Below Poverty Line card, and the Indonesian cash transfer programs; examples of self-selection-based PMTs include the expansion of Progresa under the name Oportunidades to urban areas in Mexico, the Chilean social assistance system, the Costa Rican Sistema Información de la Población Objetivo system, and the Mongolian Child Money Program (see, e.g., Castaneda and Lindert 2005; Hodges et al. 2007; Coady and Parker 2009; Martinelli and Parker 2009). Brazil's Bolsa Familia program uses a combination of the two methods (Lindert et al. 2007).

one at their homes, and then automatically enrolled those who passed using the same asset test that was used in self-targeting.

We begin with a description of the experiment and the data. We then ask what we would expect from such an experiment on purely a priori grounds. Specifically, we adapt the classical theory of self-selection into social programs developed by Nichols et al. (1971), Nichols and Zeckhauser (1982), Besley and Coate (1992), and others to a context in which, after selecting into applying, one receives the program stochastically, with the probability of receiving the program declining with income. The fact that the likelihood of receiving benefits is stochastic but declining with income captures the fact that most screening mechanisms (including but not limited to PMTs) differentiate between rich and poor but not perfectly, so that people cannot exactly forecast before applying whether they will turn out to be eligible. The standard Nichols and Zeckhauser self-selection idea depends on a single-crossing property, where the ordeal is more costly for the rich than for the poor. Time-based ordeals are the canonical example, since the rich presumably have a higher opportunity cost of time than the poor. In this context, we discuss a number of reasons why requiring people to spend time traveling to and applying at the application site does not necessarily generate single crossing: for example, the poor and rich may have different means of travel, which might make distance less costly at the margin for the rich than for the poor. On the other hand, we argue that the fact that the probability of receiving benefits slopes downward in income provides a very straightforward reason why self-selection might improve targeting: since the rich have only a small chance of passing through the PMT if they apply, they may not bother, even if the costs of applying are relatively small.

Our empirical analysis then proceeds in four stages. First, we examine who selects to apply for the PKH program in the 200 villages where the application-based process was administered. To do so, we utilize the data on households' per capita consumption that we collected before the program was announced or targeting began. We find that the probability of applying is decreasing in a household's per capita consumption. Decomposing consumption into that which is potentially observable to the government (i.e., the part that can be predicted on the basis of observable assets) and the unobservable residual, we show that those who apply are poorer on both observables and unobservables than those who choose not to. This implies not only that self-selection can potentially save resources (since many who would fail the asset test, i.e., have high observables, are no longer tested) but that it also has the potential to improve targeting even over a universally administered asset test (since those who apply are poorer on unobservables than the population at large). However, we also find evidence for the view that inviting all the poor to apply is not enough to ensure that every deserving candidate gets benefits: for example, only about 60 percent of the very poor apply under self-targeting.

The question for most governments, however, is not necessarily how self-targeting would perform relative to a counterfactual of perfect targeting, but rather how it would compare against the next-best alternative targeting strategy. To this end, the second step of our empirical analysis is to use the experiment to compare self-targeting with the usual government automatic screening procedure. In comparison with this real alternative, we find that per capita consumption was 21 percent lower for beneficiaries in the self-targeting villages. Moreover, exclusion error was actually less of a problem in self-targeting than in the usual government procedure: the very poorest households were twice as likely to receive benefits in self-targeting as in control areas.

We further show that these findings are not driven by the government ineptly choosing whom to interview under the usual government procedure. Using asset data that we independently collected at baseline, we find that the beneficiaries under self-targeting would still be, on average, poorer than those under a "hypothetical," universal automatic targeting system in which everyone is interviewed for the asset test. Intuitively, this is possible because—as we showed above—self-selection includes selection on unobservables. That is, conditional on passing the asset test, those that self-select into applying have lower consumption than the average person in the population.

The third step in our empirical analysis is to consider whether a marginal increase in the severity of the ordeal further increases targeting performance. We examine the results from experimentally varying the distance to the registration site (i.e., increasing travel costs) and find no evidence that the marginal increase in application costs further improves selection. It reduces overall take-up but does not differentially discriminate between rich and poor.

The final step of our empirical analysis uses generalized method of moments to estimate a parametric version of our model. Simulations from the estimated model suggest that the key driver of selection is the fact that rich households forecast that they have a very small likelihood of receiving benefits conditional on applying and therefore do not bother to apply if there is any cost of applying. This helps explain both why small costs can produce substantial selection and why marginally increasing the intensity of the costs reduces overall application rates without substantially improving targeting.

We conclude by considering the net impacts of different techniques on poverty reduction. We find that, even taking into account the fact that self-targeting imposes higher costs on households, including many households that do not receive benefits, a program using self-targeting leads to between 29 and 41 percent more reduction in the poverty gap than a program with the identical budget targeted using automatic screening. Increasing distance to the application center or wait times under self-targeting has no distinguishable additional effect on the poverty gap.

This paper builds on several theoretical and empirical papers on the decision to apply for programs and how that affects program take-up and targeting. Parsons (1991) sets up a model of the decision to apply for disability insurance similar to the one used in this paper but focuses on selection due to differences in the time discount rate. He then shows, in the context of a natural experiment, that increasing beneficiary selection delays leads to an applicant pool with higher (subsequent) mortality. Heckman and Smith (2004) discuss how unequal participation (selection) may occur at different stages of a prototypical social program: eligibility, awareness, application, acceptance, and enrollment. Kleven and Kopczuk (2011) study complexity in screening for a social program; in their model, the government chooses a parameter that controls the precision of the screening process, which improves selection but imposes a cost on households. By comparison, in this paper we model explicitly several mechanisms through which self-targeting may affect selection but note that, a priori, it is not clear whether it will improve selection.[2]

The remainder of the paper is organized as follows. Section II discusses the setting, experimental design, and data. Section III introduces our model, which revisits the standard screening model with nonlinear costs, idiosyncratic shocks, and differences in sophistication. Section IV examines the self-targeting data to ask who chooses to apply for the program. Section V uses the experiment to compare self-targeting with the usual government approach. Section VI examines the marginal effect of targeting when the ordeal is changed experimentally. Section VII estimates the model to help shed light on which of the possible theoretical mechanisms that we outline best explains the results and discusses the impact of different approaches on the poverty gap. Section VIII presents conclusions.

[2] In addition, another strand of papers looks at determinants of program take-up but does not focus as much on how this differentially affects the beneficiary pool, and hence targeting of the program, as we do here. For example, Thornton et al. (2010) found that reducing distance to enrollment for health insurance increased enrollments substantially. Currie and Grogger (2001) use variation in recertification requirements for the food stamps program across US states and across time and find that longer recertification periods are associated with higher program take-up. Bhargava and Manoli (2015) measure benefit take-up in an experiment in which they vary the information contained in a mailing advertising a negative income tax benefit program (Earned Income Tax Credit), the complexity of the attached application form, and the perceived level of stigma of applying. They find that the complexity of the application form significantly affects the decision to apply. See also the review of this literature in Currie (2006).

II. Setting and Experimental Design

A. Setting: The PKH Program

This project explores self-targeting mechanisms within Program Keluarga Harapan (PKH), a conditional cash transfer project administered by the Ministry of Social Affairs (DepSos) in Indonesia. The program targets households that have per capita consumption below 80 percent of the poverty line (approximately the poorest 5 percent of the study population) and that meet the demographic requirements of having a pregnant woman, a child between the ages of 0 and 5, or children below 18 years old who have not finished 9 years of compulsory education. Program beneficiaries receive direct cash assistance ranging from Rp. 600,000 to Rp. 2.2 million (USD$67–$250) per year—about 3.5–13 percent of the average yearly consumption by poor households in our sample—depending on their family composition, school attendance, pre-/postnatal checkups, and completed vaccinations.[3] The payments are disbursed quarterly for up to 6 years. In 2013, approximately 2.4 million households were enrolled in the program.

Determining whether households fall below the consumption requirement ("targeting") is difficult because per capita consumption, while the intended target of the program, is not easily observed by the government. Instead, PKH uses a proxy means test approach with automatic screening for households that meet the demographic requirements. Specifically, every 3 years, enumerators from the Central Statistical Bureau (BPS) conduct a survey of households nationwide that are potentially eligible for antipoverty programs, including but not limited to PKH. They survey all households that were included in previous surveys (regardless of whether they previously qualified or not) and supplement this list with recommendations from local leaders and their own observations of the kinds of houses that the households inhabit. After passing an initial five-question filter, each household is asked a series of about 30 questions, including attributes of their home (e.g., wall type, roof type), ownership of specific assets (e.g., motorcycle, refrigerator), household composition, and the education and occupation of the household head. These measures are combined with location-based indicators, such as population density, distance to the district capital, and access to education. Using independent survey data, the government then estimates the relationship between these variables and the household per capita consumption to generate a district-level formula for predicting consumption lev-

[3] Note, however, that although PKH is formally a conditional cash transfer program, with transfers dependent on health take-up and school enrollment, these conditions are typically not enforced in practice, so this can be thought of as closer to a "labeled" cash grant, as in Benhassine et al. (2013).

els on the basis of the responses to the survey. Individuals with predicted consumption levels below each district's very poor line are eligible for the program.

Panel *A* in figure 1 shows the probability of passing the government asset test and being determined eligible for the program as a function of log per capita consumption, as estimated from our baseline data. Note that the particular function used to map assets to eligibility is estimated by the government separately for each district and for urban and rural areas, which is why several different downward-sloping curves are visible in the figure. Several key points are worth observing about this function. First, it is strongly downward sloping: the poor are much more likely to receive benefits than the rich. Second, there is substantial noise in the process, driven by how hard it is to accurately estimate consumption from assets. PKH targets approximately the bottom 5 percent of the population, but even the very poorest rarely have more than a 40 percent chance of receiving benefits, and even those with incomes more than twice the target threshold (i.e., about 13 log points, as opposed to the cutoff of about 12.3 log points) still have as much as a 5–10 percent chance of receiving them.[4]

Panel *B* in figure 1 shows the probability of passing the government asset test as a function of the PMT score calculated from our baseline data. The relationship is downward sloping and significantly steeper than in panel *A*, yet it is not deterministic. This suggests that assets are also measured with noise, and thus even a household that knew the PMT formula exactly would still be uncertain of whether it would receive benefits. The two functions in figure 1 are the main building blocks of the model that we set up in Section III.

B. *Sample Selection*

This project was carried out during the 2011 expansion of PKH to new areas that had never had PKH before. We chose six districts (two each in the provinces of Lampung, South Sumatra, and Central Java) from the expansion areas to include a wide variety of cultural and economic environments. Within these districts, we randomly selected a total of 400 villages, stratified such that the final sample consisted of approximately 30 percent urban and 70 percent rural. Within each village, we

[4] The PMT formulas were determined using household survey data from the National Socio-Economic Household Survey (SUSENAS, 2010) and village survey data from Village Potential Statistics (PODES, 2008). On average, these regressions had an R^2 of .52. The questions chosen for the PMT survey were those that the government was considering for the next nationwide targeting survey (the Data Collection on Social Protection Programme [PPLS, 2011]).

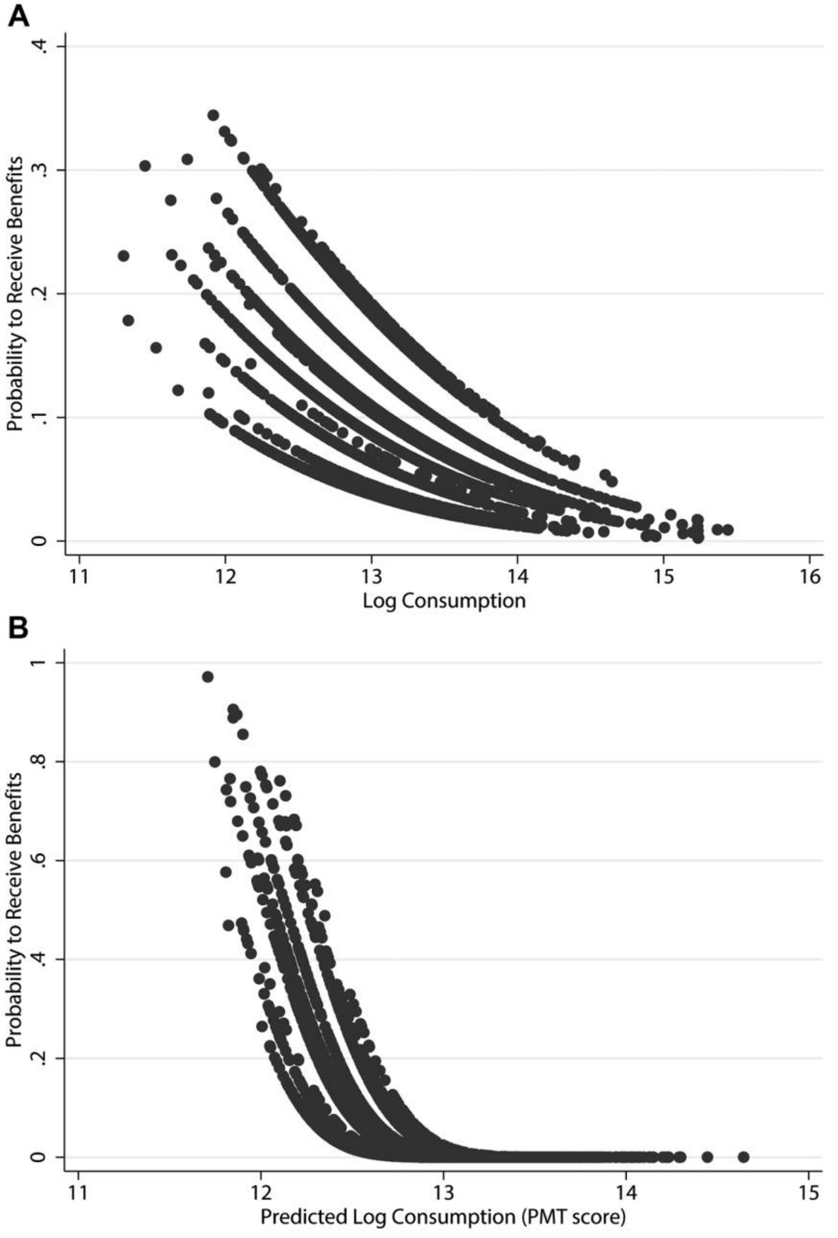

FIG. 1.—Probability of obtaining benefits versus log per capita consumption and PMT score. *A*, Probability of obtaining benefits versus log per capita consumption. *B*, Probability of obtaining benefits versus PMT score. Panel *A* shows the predicted probability of receiving the benefit, conditional on applying, from a probit model of receiving the benefit as a function of log per capita consumption. Panel *B* repeats the same exercise replacing log per capita consumption by the predicted values from the PMT using baseline survey asset data. The predicted values from panel *B* are the $\mu(y_i^o)$ that we use in the model. We include urban/rural interacted with district fixed effects in the probit equations in both panels, since the PMT cutoff for inclusion varies slightly for each urban/rural times district cell.

randomly selected one hamlet to be surveyed.[5] These hamlets are best thought of as neighborhoods that consist of about 150 households and that each have their own administrative head, whom we refer to as the hamlet head.

C. *Experimental Design*

We randomly allocated each of the 400 villages to one of two targeting methodologies: self-targeting or an automatic screening system, that is, the usual government procedure in place in other areas.[6]

1. Automatic Screening Treatment

In Indonesia, the automatic screening treatment is the usual government procedure, and the procedure discussed in Section II.A was followed.[7] For each hamlet in this treatment, the government BPS enumerators were given a preprinted list of households from the last targeting survey (PPLS, 2008). When they arrived at a village, the enumerators showed the list to village leaders and asked them to add any households that they thought were inappropriately excluded. The enumerators also had the option of adding households to the list of interviewees if they observed that a household was likely to be quite poor. For each interviewed household, a computer-generated poverty score was generated using the district-specific PMT formulas. A list of beneficiaries was generated by selecting all households with a score below the score cutoff for their district.[8]

[5] Villages in both rural and urban areas are administratively divided by neighborhood into subvillages, which we henceforth refer to as "hamlets." In rural areas, each hamlet ranges from about 30 to 330 households, while in urban areas, they each range from 70 to 410 households.

[6] We also randomly assigned an additional 200 villages to a "hybrid treatment," which is similar to the community treatment in Alatas et al. (2012). As this is largely an extension of previous work on community-based targeting and was done for policy purposes, we do not include it in this paper.

[7] Owing to cost considerations, for this treatment, the automatic screening was conducted only in the one randomly selected hamlet per village that we also surveyed in the baseline. To select beneficiaries in the other hamlets, the government used the 2008 automatic screening survey.

[8] The only difference between the PMT formula used in the automatic screening treatment and the self-targeting treatment was that, in automatic screening, for each potential interviewee, the enumerator conducted an initial five-question filter; only those households that passed this filter were given the full PMT survey. The filter consists of five questions: Is the household's average income per month in the past 3 months more than Rp. 1 million (US$110)? Was the average transfer received per month in the past 3 months more than Rp. 1 million (US$110)? Did they own a television or refrigerator that cost more than Rp. 1 million (US$110)? Was the value of their livestock, productive building, and large agricultural tools owned more than Rp. 1.5 million (US$167)? Did they own a motor

2. Self-Targeting Treatment

The PMT to determine eligibility under the self-targeting mechanism was the same as in automatic screening, but households in the self-targeting treatment were required to apply and take this test at a central registration station.

To publicize the application process, a community facilitator from a local nongovernmental organization (Mitra Samya) met with village leaders to inform them about the program, brainstorm with them about the best indicators of local poverty, and set a date for a series of hamlet-level meetings that were aimed at the poor. In these hamlet-level meetings, the facilitators described the PKH program and explained the registration process. In particular, they stressed that the program was geared toward the very poor. They listed examples of questions that would be asked during the interview (e.g., type of house, motorbike), informed households that there would be a verification stage after the interview, and highlighted a set of local poverty criteria (the criteria that locals would typically use to characterize very poor households) to help villagers understand how the PMT screening would operate. Though they did not convey the exact criteria used in the PMT, the goal was to ensure that the households generally understood that their chances of obtaining PKH conditional on showing up to be interviewed would be much higher for the poor than for the rich. Since these meetings helped households understand how the government would make selections, they should be considered an integral part of the self-targeting treatment.

Registration days for each area were scheduled in advance on the basis of the number of predicted applicants and their relative proportion within the hamlet. During the registration days, the BPS enumerators were present at the registration station from 8:00 a.m. to 5:00 p.m. Households that wanted to apply were required to come to the registration site. Once they arrived, they were signed in and given a number in the queue. When their number was called, BPS conducted the asset interview for the PMT.

vehicle, and did they own jewelry worth more than Rp. 1 million (US$110)? Households that answered yes on four or more of the questions were instantly disqualified. Of the 6,406 households on the potential interviewee list, 16 percent were eliminated on the basis of the initial filter, and 5,383 households (or about 37.8 percent of each hamlet) were given the full PMT survey of 28 questions. The idea is that these thresholds are so high that any household answering yes to a majority of these questions would likely have been eliminated anyway by the PMT. We reran the main experimental analysis (e.g., tables 5 and 6 below) dropping any household in either treatment that would have failed this filter, using answers to the same questions in the baseline survey, so that in this sample the PMTs used in automatic screening and self-targeting were exactly identical. The results are virtually unchanged.

Households that applied were subsequently categorized by eligibility on the basis of the PMT regression formula and the district-specific very poor line, using the same PMT formula and questions as in the automatic screening treatment. Any household that both was classified as very poor on the basis of assets disclosed in their interview and had also been visited by government enumerators in the previous 2008 poverty census and found to be very poor (about 37 percent that passed the interview at the registration site) was selected as a recipient. All other households that were classified as very poor on the basis of their interview were subjected to a verification process: the BPS enumerators visited their homes to redo the asset test. The results of this home-based survey were used, with the same PMT regression formula and poverty lines, to determine the final beneficiary list. About 68 percent of those who got to the verification stage were ultimately considered eligible after the verification.

Within self-targeting treatment villages, we experimentally varied the costs of registration by varying the distance to the registration site.[9] The idea was to vary the time and travel costs required to sign up, while ensuring that all locations could still potentially be reached by walking, so as not to impose substantial financial transportation costs on poor households. In urban areas, we randomly allocated villages to have the registration site at the subdistrict office (far location) or the village office (close location). Distances are greater in rural areas than in urban ones, so we randomly allocated rural villages to have the registration site at the village office (far location) or in the subvillage (close location).[10]

On net, these application costs are small relative to the potential benefits received. We can compute the costs of applying from the household survey (described in more detail in Sec. II.E below) by adding up reported time and monetary costs to travel to the location where the interview would take place (which we obtain in the baseline survey for all households, even before they know about the targeting program), as well as the average time people spent waiting multiplied by an estimate of the household's likely wage rate. (See Sec. VII for more details on this calculation.) On average, the total time and monetary cost of applying is about Rp. 17,000 (US$1.70) per household, with costs being higher for wealthier households with higher implied wage rates. By contrast, the per-household benefits average Rp. 1.3 million (US$130) per year

[9] In addition, we also attempted to vary the opportunity cost of signing up by varying whether any household member could sign up or whether both household members were required. In practice, this treatment had little force as anyone who could not bring a spouse could easily obtain an exemption letter. This is discussed in online app. B.

[10] The distance subtreatment was violated in four villages as a result of village leader objections. All analysis reports intent-to-treat effects in which these four villages are categorized on the basis of the randomization result, not actual implementation.

TABLE 1
Experimental Design

	Number of Villages (Households)
Automatic screening	200 (1,998)
Self-targeting:	
Close subtreatment	100 (1,000)
Far subtreatment	100 (1,000)
Total	200 (2,000)

Notes.—This table provides the number of villages in each treatment cell. The number of households in each cell is shown in parentheses.

for 6 years. For households with very low probabilities of receiving benefits conditional on applying, it may not make sense to apply, but for those with high probabilities of receiving benefits, the expected return from doing so appears substantial.[11]

D. *Randomization Design and Timing*

We randomly assigned each of the 400 villages to the treatments (see table 1), stratifying by 58 geographic strata, where each stratum consisted of all the villages from one or more subdistricts and was entirely located in a single district. We then randomly and independently allocated each self-targeting village to the subtreatments, with each of these two subtreatment randomizations stratified by the previously defined strata.

From December 2010 to March 2011, an independent survey firm (SurveyMeter) collected the baseline data from one randomly selected hamlet in each village. After surveying was completed in each subdistrict, the government conducted the targeting treatments. The targeting treatments thus occurred from January through April 2011. SurveyMeter conducted a midline survey in early August 2011, after the targeting was complete but before the beneficiary lists were announced to the villages. Fund distribution occurred starting in late August 2011. Finally, we conducted an endline survey from January 2012 to March 2012, after two

[11] More specifically, we can compute the break-even discount factor that makes a household with a given consumption level indifferent between applying and not applying. We assume that risk-neutral households weight the costs (described above) and the benefit that pays out yearly for 6 years, starting 1 year after the application, multiplied by the estimated probability of receiving the benefits. For example, a household with the median level of consumption has only a 3.6 percent chance of receiving benefits if it applies, and this number decreases further for wealthier households. Online app. fig. C.1 plots the break-even discount factor. Households with per capita consumption above 13.6 log points have an implied break-even discount factor above one, implying that they should never apply. If we assume households have a yearly discount factor of 0.5, then any household with more than 13.1 log consumption points (roughly 44 percent of the population) should never apply.

fund distributions had occurred. Table 2 presents the time line of the experiment.

E. Data, Summary Statistics, and Balance Test

1. Data Collection

We collected three main sources of data.

Baseline data.—The baseline survey was completed in each subdistrict before any targeting occurred, and there was no mention of the experiment in the villages until SurveyMeter had completed the baseline survey in the entire subdistrict. The mean time elapsed between the baseline survey and the commencement of targeting activities was 22 days. Within each village, we randomly selected one hamlet, and within that hamlet, we randomly sampled nine households from the set of those that met the demographic eligibility requirements for PKH, as well as the subvillage head, for a total of 3,998 households across the 400 villages. The survey included detailed questions on the household's consumption level and demographics. We also collected data for all of the variables that enter the PMT formula so that we could calculate PMT scores for each surveyed household.

Targeting data.—We obtained all of the targeting data from the government, including who was interviewed, all data from the interview (at the interview site, at home, or both), each household's predicted consumption score, and whether the household qualified to receive PKH. For the self-targeting villages, we additionally asked the government to record data on each step of the process (e.g., where and when the regis-

TABLE 2
TIME LINE OF THE EXPERIMENT

	Self-Targeting Villages	Automatic Screening Villages
December 2010 to March 2011	Baseline survey	
January to April 2011	Application process publicized. Registration days: Households that showed up to apply received the PMT interview at the registration site. Verification process: A subset of households received home visits and received another PMT interview.	Prescreen list: Households suggested by village leaders or BPS enumerators were added to the prescreen list. PMT interviews: BPS enumerators conducted home visits and PMT interviews with all prescreened households.
Early August 2011	Midline survey	
Late August 2011	Beneficiary lists were announced to the villages. First round of PKH benefits distributed.	
January to March 2012	Endline survey	

tration meetings occurred, how the socialization was done in each village).

Midline and endline surveys.—We administered a midline survey and an endline survey, both of which were conducted by SurveyMeter. The first occurred in August 2011, prior to announcements of the beneficiary lists. We surveyed up to three beneficiary households per village and revisited one household from the baseline survey per village in 97 randomly chosen automatic screening villages and 193 self-targeting villages, for a total sample of 1,045 households.[12] In this survey, we collected detailed data on each household's consumption level, as well as the respondents' experience and satisfaction with the targeting process (e.g., whether they applied, how long they waited to be interviewed). We conducted the endline from January 2012 to March 2012, after two rounds of PKH fund distribution. In this survey, we revisited all 10 of the baseline households, collecting consumption data, as well as data on satisfaction with PKH.

2. Summary Statistics and Experimental Validity

Table 3 shows the flow of households through the experiment. Column 1 shows the total number of households in the baseline survey in each of the two primary treatments. The next columns show the number of households that applied to be interviewed for self-targeting (754 out of 2,000, or 38 percent) or were interviewed as part of the automatic screening treatment (706 out of 1,998, or 35 percent). Column 3 shows the number of baseline households that were ultimately chosen as beneficiaries (73 out of 2,000, or 3.65 percent, in self-targeting; 86 out of 1,998, or 4.3 percent, in automatic screening).

Online appendix table C.1 presents summary statistics and a check on the experimental validity using data from the baseline survey and a village census. We chose all of the variables for this table prior to analyzing the data. Column 3 shows the difference between villages in automatic screening and self-targeting (with associated standard errors), and column 4 shows this difference after controlling for stratum fixed effects. Only one of the 20 differences presented is statistically significant (at the 10 percent level), confirming balance at the baseline. In the final row, we also provide the p-value from a joint test of the treatment across all baseline characteristics that we consider. The p-values of .99 and .67, respectively, confirm that the groups are balanced in the baseline.

[12] Owing to safety and travel concerns that were independent of the project, the survey company asked that we not return to 10 villages in the midline and 13 villages in the endline. These were spread among treatment and control villages.

TABLE 3

Descriptive Statistics for Households Surveyed in the Baseline

	Total Number of Households (1)	Number of Households Interviewed (2)	Number of Beneficiaries (3)	Households Interviewed (%) (4)	Interviewed Households That Received Benefits (%) (5)	Total Households That Received Benefits (%) (6)
Automatic screening	1,998	706	86	35.34	12.18	4.30
Self-targeting	2,000	754	73	37.70	9.68	3.65

Note.—This table provides information on the flow of surveyed households through the experiment.

III. Model

A. *Model Setup*

In this section, we reexamine self-selection into a welfare program based on the expected benefits and costs of applying. We assume that households live for two periods and have linear per-period utility in current consumption.[13] We assume that households do not save or borrow and have no other assets and therefore that their consumption is equal to their flow income, including both labor income and any transfers, less any "travel" costs (this is the only cost we will consider and includes both the money cost of traveling and the time cost of both travel and any associated wait time).

Households vary in their per-period labor income, denoted by y, but for a given household this is the same number in both periods. At the beginning of the first period, which is before the household makes the choice of whether to apply for PKH, and obviously therefore before transfers have been allocated, its consumption is also y. We assume that this is what we measure in our baseline survey and also what the government is trying to target. However, the government observes only a part of this consumption: we denote the portion that is observable to the government by y^o and the portion that is unobservable to the government by y^u, so $y = y^o + y^u$.[14] The application cost is denoted by $c(l, y)$, where l is the distance to the registration site. The inclusion of y in this function captures the key idea that the opportunity cost of time for the household is related to its income/consumption; but as we will see below, y could also affect the money cost of travel.

Conditional on applying, households have a probability $\mu(y^o)$ of passing the asset-based test and actually qualifying for the program ($\mu'(y^o) \leq 0$). Note that this is a function of the observable portion of consumption. If the observable consumption was perfectly measured by the government, this would be a step function, with households receiving the transfer if the observable portion of consumption was less than a cutoff value, that is, if $y^o < y^*$. In practice, even the observable portion of consumption is measured by the government with noise, so household beliefs about their eligibility for the program take the probit form, with

[13] The assumption of linearity rules out the possibility that for some of the poor, given their very high marginal utility of consumption, it does not make sense to bear the cost of applying (missed work, travel costs), which has to be paid before they get the benefits. This could lead to the wrong kind of selection. In a previous version of this paper (Alatas et al. 2013), we showed that the concavity of the utility function played no role in explaining the observed variation in show-up. Thus, for conciseness, we therefore assume linearity from the start and point the interested reader to that version.

[14] In practice, the government predicts y^o from a range of observable characteristics, and thus y^o and y^u are statistically independent.

$\mu(y^o) = \text{Prob}(y^o + \pi < y^*)$, where π is an independent and identically distributed normal noise term.

We assume there are two types of households in the population. Sophisticated households understand how the government computes $\mu(y^o)$. Unsophisticated households, however, do not know what the government observes and what it does not. These households know the actual empirical probability that someone with their consumption level receives the program conditional on applying, $\lambda(y)$. While $\lambda'(y) \leq 0$, intuitively μ will be closer to a step function than λ, since the sophisticated households know more about the true rule the government uses.

If the household qualifies for the program, it receives an additional income b in the future period (for simplicity, we assume there is just one future period). Otherwise, it receives no additional income.

We think of period 1 in the model as the period during which the application process takes place and period 2 as the period when the chosen beneficiaries get to enjoy the net present value (NPV) of program benefits, where δ is the weight given to the second period.

To complete the description of the model, assume that each person receives a random utility shock, ε, that encourages (or discourages) him to go to apply, and $F(\varepsilon)$ is the cumulative distribution function (CDF) of ε. These utility shocks could capture, for example, psychological costs of applying. These idiosyncratic shock terms will be important in Section VII below when we estimate the model explicitly.

Taken together, the sophisticated household's expected utility upon applying is

$$y - c(l, y) + \mu(y^o)\delta(y + b) + [1 - \mu(y^o)]\delta y + \varepsilon, \tag{1}$$

and the unsophisticated household's expected utility upon applying is

$$y - c(l, y) + \lambda(y)\delta(y + b) + [1 - \lambda(y)]\delta y + \varepsilon. \tag{2}$$

If the household does not apply, expected utility is

$$y + \delta y. \tag{3}$$

The expected gain from applying is the difference, that is,

$$- c(l, y) + \mu(y^o)\delta b + \varepsilon \tag{4}$$

for sophisticated households and

$$-c(l, y) + \lambda(y)\delta b + \varepsilon \tag{5}$$

for unsophisticated households. It will turn out to be convenient to define

$$g(y^o, y, l) = -c(l, y) + \mu(y^o)\delta b, \tag{6}$$

$$h(y, l) = -c(l, y) + \lambda(y)\delta b \tag{7}$$

to denote the net gains for sophisticated and unsophisticated households, respectively. Define

$$A_s(y^o, y, l) = \text{Prob}[g(y^o, y, l) > \varepsilon] = 1 - F(-g(y^o, y, l))$$

to be the probability that sophisticated households apply and

$$A_u(y, l) = \text{Prob}[h(y, l) > \varepsilon] = 1 - F(-h(y, l))$$

to be the probability that unsophisticated households apply.

To close the model we assume that unsophisticated households' beliefs about their probability of receiving benefits conditional on applying, $\lambda(y)$, are internally consistent in the sense that if these households show up on the basis of $\lambda(y)$ and the sophisticated households show up on the basis of $\mu(y^o)$, the average probability that someone with income y gets benefits is indeed $\lambda(y)$. Formally this is captured by the following condition, which says that $\lambda(y)$ equals the function $\lambda_{\text{induced}}(y)$, which is a weighted average of $\mu(y^o)$ among all the people with income y who apply. If α denotes the share of sophisticated, this is given by

$$\lambda(y) = \lambda_{\text{induced}}(y) \equiv \frac{\begin{array}{c}\alpha \iint \mu(y^o) A_s(y^o, y, l)\vartheta(y^o, l|y)\, dl dy^o + \\ (1-\alpha)\iint \mu(y^o) A_u(y, l)\vartheta(y^o, l|y)\, dl dy^o\end{array}}{\begin{array}{c}\alpha \iint A_s(y^o, y, l)\vartheta(y^o, l|y)\, dl dy^o + \\ (1-\alpha)\iint A_u(y, l)\vartheta(y^o, l|y)\, dl dy^o\end{array}} \tag{8}$$

where $\vartheta(y^o, l|y)$ is the conditional distribution of y^o and l given y. This condition simply states that at any consumption level y, the probability of receiving benefits is given by the underlying government decision rule $\mu(y^o)$, integrated over the distribution of the set of people (i.e., with observable incomes y^o and costs l) that apply at a given full income y. Note that equation (8) is a fixed-point condition, because $A_u(y, l)$ is a function of $\lambda(y)$. It provides an additional moment restriction that helps us identify the unobserved parameters of the model.

B. Analysis

We start with the most basic model and add elements to the model one by one in order to understand how each affects selection.

1. The Benchmark Case

Suppose that the time cost of applying is linear in distance τl and that all households are unsophisticated and do not know the difference between observable and unobservable components of consumption. For someone who earns a wage w, this imposes a monetary cost of $\tau l w$. Assuming that wages are proportional to income/consumption, $w = \phi y$, then the monetary application cost is $\tau l \phi y$. Assume also that there are no shocks ($\varepsilon \equiv 0$). Thus, $h(y)$ can be written such that a household applies if

$$-\tau l \phi y + \delta \lambda(y) b \geq 0. \tag{9}$$

Since the left-hand side of this expression is decreasing in y, this expression defines a cutoff value y^* such that those with incomes less than y^* apply and those with incomes greater than y^* do not. Moreover, an inspection of equation (9) shows that $\partial y^*/\partial l < 0$; that is, making the ordeal more onerous increases the degree of selection and implies that the applicants will be poorer. This expression captures the basic intuition for using ordeal mechanisms for selection that is captured by Nichols and Zeckhauser (1982).

2. Adding Shocks

Consider what happens if we reintroduce the utility shock term. A household applies iff

$$\tau l \phi y - \delta \lambda(y) b \leq \varepsilon. \tag{10}$$

Consider two levels of income, y_1 and $y_2 > y_1$, and assume that the cutoff value of ε in both cases is interior to the support of its distribution. The ratio of their show-up rates is

$$\frac{1 - F(\tau l \phi y_1 - \delta \lambda(y_1) b)}{1 - F(\tau l \phi y_2 - \delta \lambda(y_2) b)}. \tag{11}$$

This ratio is always greater than one because the rich are less likely to sign up as their costs are higher and their probability of getting the benefit is lower. Note that this ratio is a measure of how well targeted the application process is: the higher the ratio, the higher the fraction of the poor in the applicant population. Making the ordeal tougher reduces the number of poor applicants and imposes deadweight costs on all applicants, which are both undesirable. Therefore, the only reason to do so is that it improves the ratio of poor to rich, which may reduce the government's program costs per eligible beneficiary.

Taking the derivative with respect to l, the distance to the registration site, tells us that targeting efficiency measured by this ratio improves when l increases if and only if

$$\frac{f(\tau l\phi y_2 - \delta\lambda(y_2)b)}{1 - F(\tau l\phi y_2 - \delta\lambda(y_2)b)}\tau\phi y_2 - \frac{f(\tau l\phi y_1 - \delta\lambda(y_1)b)}{1 - F(\tau l\phi y_1 - \delta\lambda(y_1)b)}\tau\phi y_1 > 0. \tag{12}$$

When costs, l, are marginally increased by a small amount, the share of people who are lost is proportional to the density of people right on the margin—given by the probability density function (PDF) $f(y)$—to the number of people who are inframarginal, given by the $1 - F(y)$ term.

Thus, a sufficient condition for targeting efficiency to be improving as l increases is that the hazard rate,

$$\frac{f(\tau l\phi y - \delta\lambda(y)b)}{1 - F(\tau l\phi y - \delta\lambda(y)b)}, \tag{13}$$

is weakly increasing with y, since if this is true, then clearly

$$\frac{f(\tau l\phi y - \delta\lambda(y)b)}{1 - F(\tau l\phi y - \delta\lambda(y)b)}\tau\phi y$$

is increasing in y. This property holds if $F(\varepsilon)$ represents a uniform, logistic, exponential, or normal distribution, but not in other relevant cases such as the Pareto distribution and other "thick-tailed" distributions. The log-logistic distribution function $F(\varepsilon) = \varepsilon^\beta/(c^\beta + \varepsilon^\beta)$, where c and β are two known positive parameters and $\varepsilon \geq 0$, exhibits declining hazard rates as long as $\beta \leq 1$, but not otherwise.

What this discussion illustrates is that single crossing in the classical screening sense is not sufficient for increasing ordeals to increase targeting effectiveness. Instead, one also needs to consider the density of people who are near the threshold and, hence, who will be affected by any marginal change in ordeals.

3. Nonlinearities in the Application Cost

We now model a nonlinearity in the cost of applying, $c(l, y)$. This nonlinearity may be more realistic because there are different transportation modes: one can either walk or take a bus. Buses are faster, but they cost money. Given that l is the distance to the registration site, walkers face a calorie cost γl and a time cost τlw, where w is their wage rate and τl

is defined to include the waiting time. Taking a bus requires a fixed bus fare, ν, plus a time cost, λ/w, where $\lambda < \tau$. Again, λl includes waiting time. Assuming that the wage is proportional to income/consumption, $w = \phi y$, the decision rule is

$$D = \begin{cases} \text{bus} & \text{if } \nu + \lambda l \phi y < \gamma l + \tau l \phi y \\ \text{walk} & \text{if } \nu + \lambda l \phi y \geq \gamma l + \tau l \phi y. \end{cases} \tag{14}$$

Applying is optimal if and only if

$$-\min\{\gamma l + \tau l \phi y, \nu + \lambda l \phi y\} + \delta \lambda(y) b \geq \varepsilon. \tag{15}$$

The expression on the left-hand side is declining in y. Therefore, richer people always apply less.

To explore the effect of increasing l, consider two income levels y_1 and y_2, such that at y_1, an individual just prefers to walk if he applies, and at y_2, he just prefers to take a bus, so that y_1 and y_2 are separated by some small distance ψ. For those with income y_1, the cost of travel is $\gamma l + \tau l \phi y_1$. For those at y_2, it is $\nu + \lambda l \phi y_2$. The fall in utility due to an increase in distance of Δl will be greater at y_1 than y_2: $(\gamma + \tau \phi y_1)\Delta l > (\lambda \phi y_2)\Delta l$. Therefore, an increase in distance can increase travel costs more for the poor than for the rich.

4. Sophisticated versus Unsophisticated Households

The cases that we considered thus far were all based on unsophisticated households, with a single $\lambda(y)$ function. We now reintroduce the distinction between sophisticated and unsophisticated households.

Observe from equations (1) and (2) that the only difference between sophisticated and unsophisticated households is that the sophisticated ones understand that their probability of obtaining benefits is based on the observable portion of their income $\mu(y^o)$, whereas the unsophisticated use the coarser rule $\lambda(y)$.

This difference is important because it affects the ways in which selection on unobservables may occur. For sophisticated households, the unobservable portion of income (y^u) affects the show-up decision only through the cost of applying $(c(l, y))$. For unsophisticated households, the unobserved portion of consumption (y^u) also affects their beliefs about the probability of receiving the benefits, $\lambda(y)$. Unsophisticated households should therefore exhibit more selection on unobservables than sophisticated ones.

To the extent that there are errors in the government's targeting formula (e.g., because there is a substantial portion of income that is unobserved or poorly measured), selection on unobservables could substantially improve targeting efficiency. Having these unsophisticated households with an intermediate amount of knowledge (i.e., they understand that $\lambda(y)$ is downward sloping, but not what aspects of income figure into the μ function) could therefore improve targeting relative to either having all households be sophisticated or having all households lack information about the government decision rule and instead believing that $\lambda(y)$ was a constant rather than downward sloping. Having only sophisticated households would be bad, since sophisticated households that know that they have a low y^o (and hence pass the screen) may choose to sign up, meaning that the government gets precisely those rich households that are likely to slip through the asset screen.

IV. Who Self-Selects?

We begin by examining whether richer or poorer households were more likely to apply for PKH in the 200 villages where the government implemented the self-targeting treatment. Specifically, we plot a nonparametric Fan (1992) regression of the probability of applying against baseline log per capita consumption (fig. 2). This corresponds to total consumption, y, in the model. Note that the consumption data were collected be-

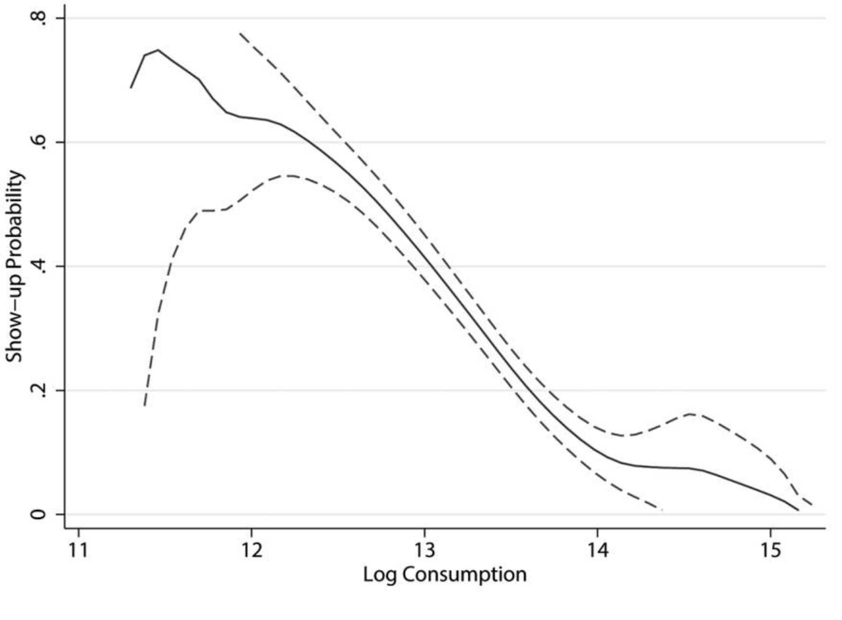

FIG. 2.—Show-up rates versus log per capita consumption. This figure provides a nonparametric Fan regression of the probability of applying for PKH against baseline log per capita consumption in the 200 self-targeting villages. Bootstrapped pointwise 95 percent confidence intervals, clustered at the village level, are shown in dashes.

fore any mention of targeting occurred. Bootstrapped pointwise 95 percent confidence intervals, clustered at the village level, are shown in dashes.

Across all expenditure ranges, figure 2 shows that the poor are more likely to apply than the rich. This is evident as the probability of applying falls monotonically with per capita consumption. At the very bottom of the expenditure distribution, a majority of households apply: 61 percent of households at the 5th percentile of the consumption distribution do so. The share that apply falls rapidly as consumption increases: at the middle of the expenditure distribution, only 39 percent of households apply, and by the 75th percentile, only 21 percent do so. At the 95th percentile of per capita expenditure, only 10 percent of households apply.

As described in the model, from the perspective of the government, self-selection could affect targeting along two distinct dimensions. First, there could be selection on characteristics that are observable to the government (i.e., y^o); that is, households that have more assets, and are therefore less likely to pass the PMT, may be less likely to show up. This type of selection could potentially reduce the government's administrative costs since it would reduce the number of interviews that it would have to conduct for those who are likely to fail the PMT anyway, but it would not necessarily change the poverty profile of beneficiaries compared to automatic screening. Second, there could be selection on the unobservable component of consumption (i.e., y^u); that is, conditional on a household's PMT score, households with higher unobservable consumption might be less likely to attend. This could arise if there is self-selection based on the opportunity cost of time or if households do not perfectly understand the construction of the PMT score. If this type of selection on unobservables is occurring, then introducing self-selection has the potential to lead to a poorer distribution of beneficiaries than automatic screening.

As in the model, we can decompose household consumption into the components that are observable and unobservable to the government:

$$y_i = y_i^o + y_i^u, \tag{16}$$

where y_i is the household's log per capita consumption, y_i^o is the projection of y_i on the predicted PMT score for the household based on the observable characteristics that enter the PMT formula, and y_i^u is the residual from the regression of y_i on the PMT score, or the unobserved component of consumption. We then examine the relationship between the probability of applying and both the observable component, y_i^o, and the unobservable component, y_i^u.

We first examine these relationships graphically, presenting nonparametric Fan regressions of the probability of showing up as a function of the observable (fig. 3, panel *A*) and unobservable (panel *B*) components

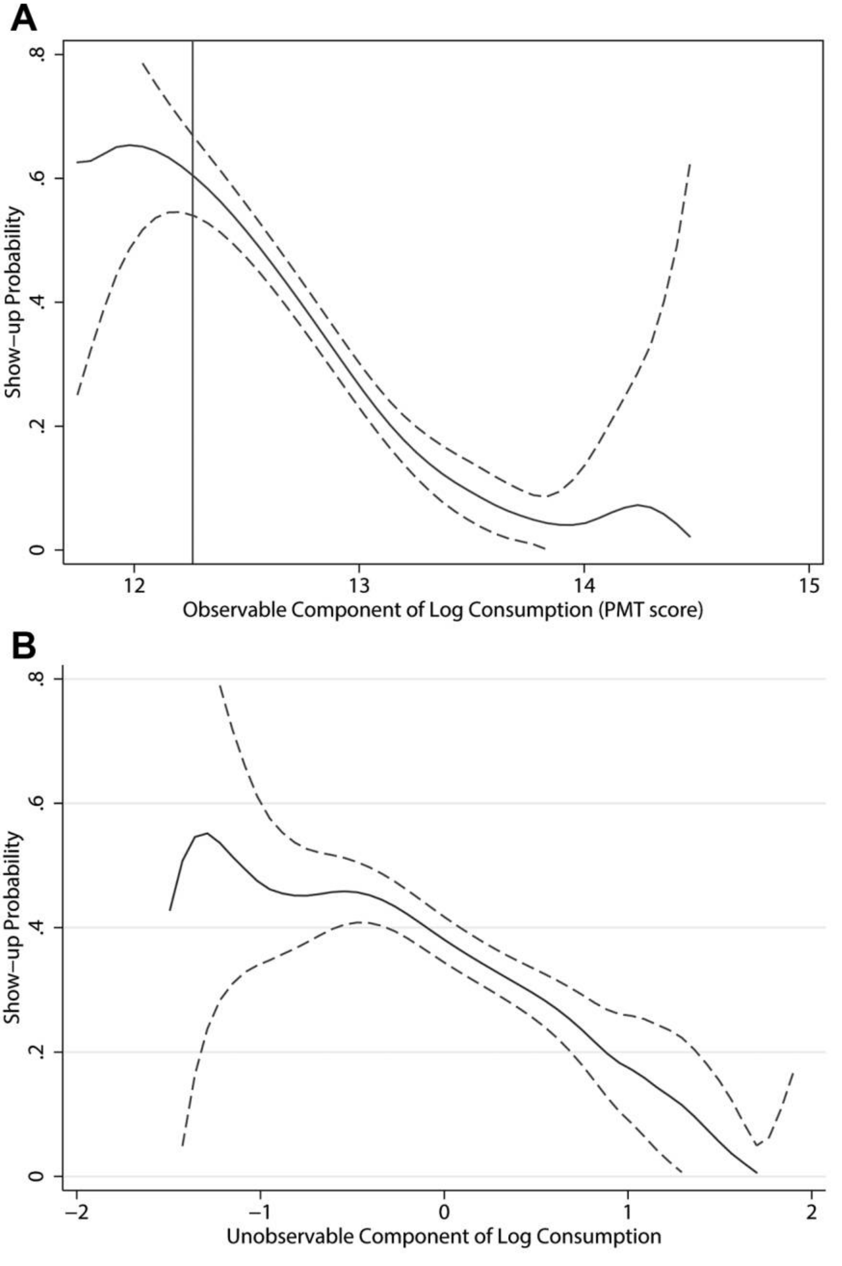

FIG. 3.—Show-up rates versus observable and unobservable components of log per capita consumption. *A*, Show-up as a function of observable consumption (y_i^o). *B*, Show-up as a function of unobservable consumption (y_i^u). Figures provide nonparametric Fan regressions of the probability of applying for PKH against the observable and unobservable components of baseline log per capita consumption in the 200 self-targeting villages. The scales for the *x*-axis are both in logs and so are comparable. Bootstrapped pointwise 95 percent confidence intervals, clustered at the village level, are shown in dashes.

of log per capita consumption. Bootstrapped pointwise 95 percent confidence intervals (clustered at the village level) are shown in dashes, and the vertical line in the top panel shows the average eligibility cutoff for receiving benefits, calculated using baseline data. Strikingly, the probability of applying is decreasing in both the observable and unobservable components of consumption.

We now formally examine these relationships in a regression framework. Table 4 provides the results from estimating the following logit equation:

$$\text{Prob}(\text{show-up}_i = 1) = \frac{\exp\{\alpha + \gamma y_i^o + \psi y_i^u\}}{1 + \exp\{\alpha + \gamma y_i^o + \psi y_i^u\}}, \qquad (17)$$

where y_i^o and y_i^u are as defined in equation (16). We use logit specifications since baseline show-up rates will differ substantially once we start to examine different samples, and therefore, in these settings the logit model is easier to interpret. We show in online appendix table C.2 that the results are qualitatively similar if we use linear probability models instead. All standard errors are clustered by village.

Table 4 confirms the graphical analysis and shows that there is self-selection along both margins. Column 1 provides the coefficient estimates for the full sample. Both the observable and unobservable components of consumption significantly predict applying at the 1 percent level. The relative magnitudes suggest that the observed component of

TABLE 4
PROBABILITY OF SHOWING UP AS A FUNCTION OF THE OBSERVED AND UNOBSERVED COMPONENTS OF BASELINE LOG PER CAPITA CONSUMPTION

	SHOWED UP		
	All (1)	Very Poor (2)	Not Very Poor (3)
Observable consumption (y_i^o)	−2.217***	−.325	−2.310***
	(.201)	(1.785)	(.208)
Unobservable consumption (y_i^u)	−.907***	−.775	−.908***
	(.136)	(.581)	(.138)
Stratum fixed effects	No	No	No
Observations	2,000	114	1,886
Mean of dependent variable	.377	.658	.360

NOTE.—Each column reports the coefficients from a logit regression of the show-up dummy on the observable and the unobservable components of log consumption. Very poor is defined as being eligible for the program on the basis of the PMT score calculated using the baseline asset data (see fn. 19). Robust standard errors, clustered at the village level, are shown in parentheses

* $p < .1$.
** $p < .05$.
*** $p < .01$.

consumption has about 2.5 times the impact of the unobserved component, but both are large: a doubling of the PMT score (i.e., predicted log consumption based on assets) reduces the log odds ratio of showing up by about 1.5; a doubling of the unobserved component of consumption reduces the log odds ratio of showing up by about 0.6. In columns 2 and 3, we split the sample on the basis of whether the household would have been eligible had it chosen to apply. Among the poorest 4 percent of households in our sample, the coefficient on unobservable consumption is negative and has a magnitude similar to that in the first column, implying that those who are poorer on unobservables are more likely to apply, though this result (estimated on a sample of 114 households) is not statistically significant. Overall, the strong selection on unobservables suggests that self-selection has the potential to result in a dramatically poorer distribution of beneficiaries than other methods.[15]

V. Comparing Self-Selection and Automatic Screening

The self-targeting treatment generated considerable self-selection, and yet only about 60 percent of the poorest group showed up, suggesting that there was significant exclusion error. However, it is not clear that we should be comparing self-targeting to the theoretical ideal of no error because, in reality, it is very costly for the government to collect consumption data for each and every household. Therefore, in Section V.A, we compare self-targeting against the real government procedure, which consists of enrolling those who pass a PMT among those selected to be interviewed by the government and local communities.

How self-targeting compares against the automatic screening will depend in part on how well the government selects the set of households to be interviewed. As we show in online appendix table C.3 (discussed in more detail below), prescreening by the government is reasonable given both budgetary and administrative constraints, and therefore, it is a realistic comparison of the true policy options available for many developing countries. However, to better understand how self-targeting operates, in Section V.B, we also compare self-targeting against a hypothetical exercise in which we use the data that we have collected independently to

[15] In online app. table C.4, we add additional variables to eq. (17) to investigate other factors that influence show-up rates, both for the entire sample and for the eligible subset. We find that households' subjective perceptions of their own wealth influence show-up and that those households that have received previous government programs (Raskin [rice for the poor], Askeskin [health insurance for the poor], and Bantuan Langsung Tunai [direct cash assistance for the poor]) are also more likely to show up. Both of these results imply that households may be basing their show-up decisions in part on their perceived likelihood of receiving programs conditional on applying (i.e., their perceptions of $\lambda(y)$).

predict selection if the PMT was implemented universally (i.e., everyone was interviewed).

A. *Experimental Comparison of Self-Targeting with Usual Government Targeting Procedure*

In this section, we test whether the types of individuals selected under self-targeting and automatic screening (the current usual procedure of the Indonesian government) differ. We compare the distribution of beneficiaries in the 200 villages randomized to receive the self-targeting treatment with the 200 villages randomized to receive the automatic screening treatment. Given the randomization, the distribution of beneficiaries and the probability of receiving benefits should be identical in the two sets of villages in the absence of the difference in targeting, so we can ascribe the differences that we observe to the differences in targeting methodologies.

We begin with a graphical analysis in which we compare the distribution of beneficiaries under the self-targeting and automatic screening treatments (fig. 4). In panel *A*, we plot the CDF of log per capita consumption of the final PKH beneficiaries in both sets of villages. The beneficiaries under self-selection appear substantially poorer: the CDF of beneficiaries' consumption under automatic screening first-order stochastically dominates that under selection. A Kolmogorov-Smirnov test of equality of distributions using randomization inference methods that account for clustering at the village level yields a p-value of .103.

While the results in panel *A* imply that the distribution of beneficiaries is poorer under self-selection, it does not tell us whether this is due to the inclusion of more poor households, the exclusion of rich households, or some combination of both. For this reason, we next present nonparametric Fan regressions of the probability of obtaining benefits as a function of log per capita consumption in panel *B* of figure 4. Bootstrapped pointwise 95 percent confidence intervals, clustered at the village level, are shown as dotted lines. The figure shows that the probability of receiving aid is substantially higher for the very poorest households in the self-targeting treatment. For those with log per capita consumption in the bottom 5 percent, that is, those with log per capita consumption below about 12.33, the probability of receiving benefits is more than doubled by the self-targeting treatment: 16 percent receive benefits in the self-targeting treatment as compared with just 7 percent in the automatic screening treatment. This difference is statistically significant at the 5 percent level. While exclusion error is still very high—even in self-targeting, only 16 percent of these very poor households received benefits, meaning that 84 percent were excluded—the rate of receiving benefits is four times higher than the overall rate of 4 percent of households in the sample

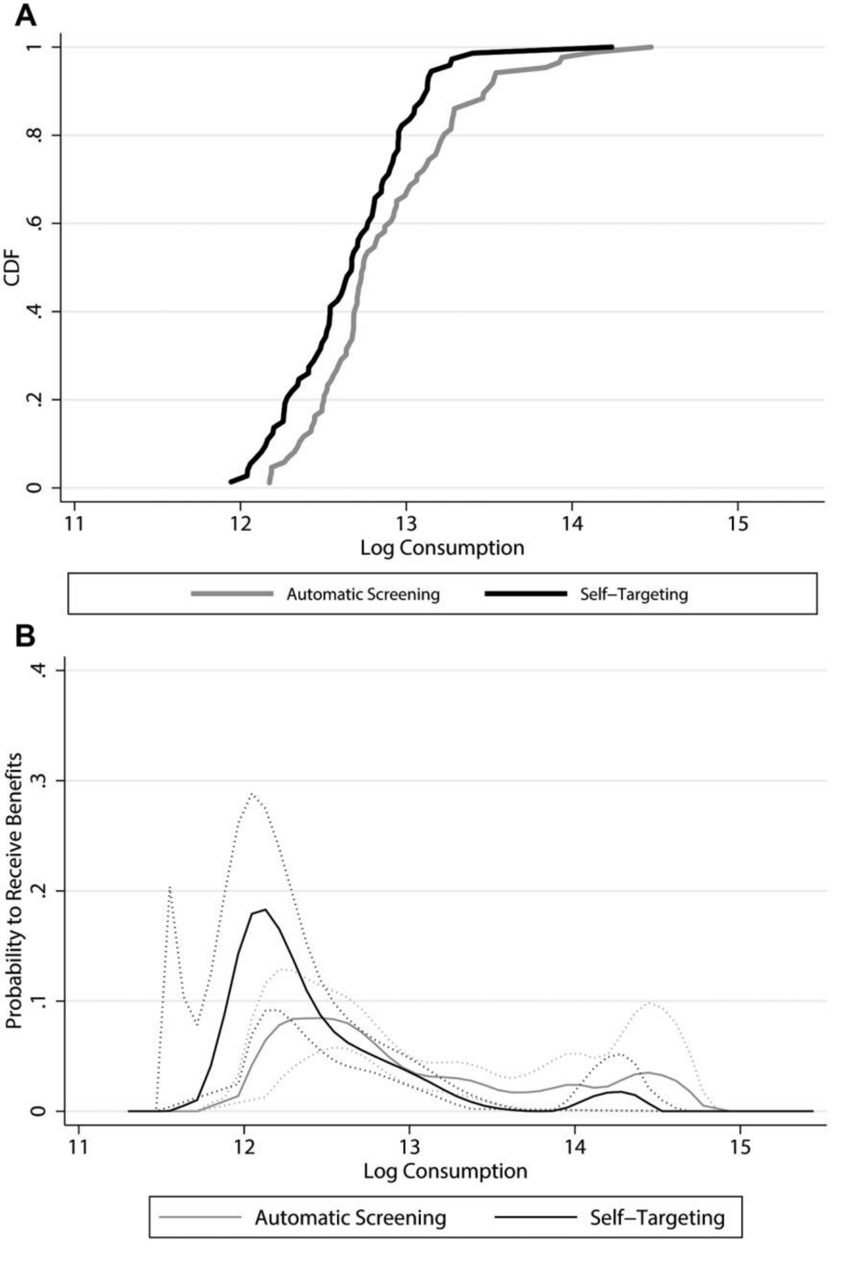

FIG. 4.—Experimental comparison of self-targeting and automatic screening treatments. *A*, CDF of log per capita consumption of beneficiaries. *B*, Receiving benefit as a function of log per capita consumption. Panel *A* shows the CDFs of log per capita consumption of beneficiaries in the self-targeting and automatic screening treatments. A Kolmogorov-Smirnov test of equality yields a p-value of .10. Panel *B* presents nonparametric Fan regressions of benefit receipt on log per capita consumption in the two treatments. Bootstrapped pointwise 95 percent confidence intervals, clustered at the village level, are shown in dashes.

that receive benefits and double what it is in the usual procedure automatic screening villages.

Conversely, households at higher consumption levels are substantially more likely to receive benefits in the automatic screening treatment. Households in the top 50 percent of the per capita expenditure distribution—none of which should be receiving benefits—are more than twice as likely to receive benefits in automatic screening as in the self-targeting treatment: 2.5 percent of such households receive benefits in automatic screening compared with 1 percent of such households in self-targeting (statistically significant at the 5 percent level). One explanation is that there are always errors in the PMT formula that allow some fraction of ineligible households to slip through the PMT, but many of these households would not apply (and therefore not slip through) with self-targeting. In sum, figure 4, panel *B*, suggests that self-targeting both increased the probability that the poor received benefits and decreased the probability that richer households did so, relative to the current usual procedure.

We now quantify these effects using regression analysis, the results of which are presented in table 5. In column 1, we compare average log per capita consumption of the beneficiary populations (y_{vi}) in the two treatments by estimating by ordinary least squares (OLS):

$$y_{vi} = \alpha + \beta \mathrm{SELF}_v + \vartheta_{vi}, \tag{18}$$

where SELF_v is a dummy for village v being in the self-targeting treatment and ϑ_{vi} is the error term. Standard errors are clustered by village. We estimate this model directly (panel A) and with stratum fixed effects (panel B). As suggested by the figures above, the regression analysis confirms that beneficiaries are substantially poorer under self-selection: Column 1 of panel A reports that per capita consumption of beneficiaries is 21 percent lower in self-targeting as compared to automatic screening (significant at the 1 percent level). Including stratum fixed effects (panel B), the difference becomes 11 percent, and the *p*-value increases to .14.

To increase our precision of the difference in consumption levels of beneficiaries, we conducted a midline survey after the targeting was complete but before program beneficiary status had been announced or benefits had begun. In column 2, we compare log per capita consumption of beneficiaries in the two treatments, including both the 159 beneficiaries from our baseline sample and the additional 745 beneficiaries whom we oversampled at midline. Since the average level of consumption may be different in these two survey rounds (e.g., because of seasonality), we include a dummy variable for the survey round. The results in column 2 are similar in magnitude but more precisely estimated: self-targeting selects beneficiaries who are 18–19 percent poorer than those selected by the automatic screening treatment (statistically significant at the 1 percent level).

In column 3 of table 5, we examine the probability of getting benefits—Prob($\text{benefit}_{vi} = 1$)—across the treatments for different groups by estimating the following logit model:

$$\text{Prob}(\text{benefit}_{vi} = 1) = \frac{\exp\{\alpha + \beta \text{SELF}_v + \gamma y_{vi} + \eta \text{SELF}_v \times y_{vi}\}}{1 + \exp\{\alpha + \beta \text{SELF}_v + \gamma y_{vi} + \eta \text{SELF}_v \times y_{vi}\}}. \quad (19)$$

The coefficient of interest is the coefficient η on $\text{SELF}_v \times y_{vi}$, which captures the degree to which there is differential targeting in the self-targeting treatment as compared with automatic screening (the omitted category). The coefficient on η is negative, large in magnitude, and statistically significant. The magnitude suggests that self-targeting is twice as strong in targeting as the automatic screening: the estimates in panel A imply that doubling consumption decreases the log odds of receiving benefits by 0.70 under automatic screening, whereas it decreases the log odds of receiving benefits by 1.37 under self-targeting.

In columns 4–6, we examine alternative dependent variables to quantify the types of inclusion and exclusion error shown in panel *B* of figure 4. Governments may place different weight on these types of errors, that is, between exclusion error (failing to give benefits to a very poor household) and inclusion error (giving benefits to a non–very poor household), an issue we return to in Section VII.B below. In column 4 we define the overall error rate as a dummy that is equal to one if either exclusion error or inclusion error occurs. We find that the log odds ratio of making an error is about 0.2 lower under self-targeting (p-values of .08 without stratum fixed effects and .11 with stratum fixed effects). Column 5 examines exclusion error, defined as a dummy for very poor households failing to receive benefits. The results in the table suggest that the log-odds of such households being excluded are between 0.55 and 0.71 lower in self-selection, though these results are not statistically significant (p-values of .18 and .15, respectively). Likewise, inclusion error, defined as a non–very poor household that does receive benefits, is lower in self-targeting and statistically significant in the specification with stratum fixed effects (col. 6; p-values .14 and .08, respectively).[16]

[16] In online app. table C.7, we add additional variables and their interactions with the self-targeting treatment to eq. (19) to explore other factors that differentially influence the probability of receiving benefits. We find that none of the household characteristics differentially influence whether the household receives the benefit. The only robust finding is that the unobserved component of consumption (y^u) is a stronger predictor of receiving the benefit in the self-targeting treatment.

TABLE 5
Experimental Comparison of Targeting under Self-Targeting and Automatic Screening Treatments

	Log Consumption (Beneficiaries; Baseline; OLS) (1)	Log Consumption (Beneficiaries; Baseline + Midline; OLS) (2)	Receives Benefits (Logit) (3)	Error (Logit) (4)	Exclusion Error (Logit) (5)	Inclusion Error (Logit) (6)
	A. No Stratum Fixed Effects					
Self-targeting	−.208***	−.193***	12.142**	−.190	−.506	−.311
	(.076)	(.060)	(4.894)	(.126)	(.402)	(.210)
Log consumption			−1.016***			
			(.280)			
Log consumption × self-targeting			−.964**			
			(.383)			
Observations	159	904	3,996	3,998	249	3,749
Mean of dependent variable	12.78	13.61	.0398	.0870	.880	.0344

	B. With Stratum Fixed Effects					
Self-targeting	−.114	−.175***	15.180***	−.209	−.649	−.331*
	(.077)	(.058)	(5.295)	(.140)	(.441)	(.192)
Log consumption			−1.042***			
			(.283)			
Log consumption × self-targeting			−1.202***			
			(.416)			
Observations	159	904	3,489	3,938	113	3,130
Mean of dependent variable	12.78	13.61	.0456	.0884	.761	.0412

NOTE.—In each panel, each column reports the coefficients from a logit or OLS regression with dependent variable indicated in the column header. The samples in cols. 1 and 2 consist of beneficiary households included in the baseline survey, and in the baseline and midline surveys, respectively. The sample in cols. 3 and 4 consists of all households; in col. 5, it consists of very poor households (those with baseline consumption below 80 percent of the poverty line); in col. 6, it consists of households that are not very poor. (Smaller sample sizes are due to two households that do not have consumption data in col. 3 and due to dropped strata because of a lack of within-stratum variation in panel B.) Exclusion error is defined to be one if a household is very poor and does not receive PKH and zero if the household is very poor and receives PKH. Inclusion error is defined to be one if a not very poor household does receive PKH and zero if a not very poor household does not receive PKH. Error is defined to be one if either exclusion or inclusion error is equal to one, and zero otherwise. In panel A and cols. 1 and 2 of panel B, robust standard errors, clustered at the village level, are shown in parentheses. In panel B, cols. 3–6, robust standard errors are clustered at the stratum level.

* $p < .1$.
** $p < .05$.
*** $p < .01$.

On net, the nonparametric and parametric results combine to paint a clear picture: self-targeting leads to a poorer distribution of beneficiaries, both because the poor are more likely to receive benefits and because richer households are less likely to receive benefits.

B. *Comparing Self-Targeting to a Hypothetical Universal Automatic Targeting Treatment*

In the automatic screening procedure, not all households were considered for enrollment. Instead, as discussed in Section II.C.1, households received the full PMT interview only if they passed an initial set of screens. These prescreening criteria were designed to save the government the cost of having to conduct a complete long-form census of all households in the country every time it wanted to select beneficiaries. On net, as shown in table 3, about 35 percent of households in the village received the full PMT interview, which is roughly comparable to the share of households that self-selected to be interviewed in the self-targeting treatment. While the prescreening does select a set of households that are poorer than the average household, it is possible that some eligible households are excluded from the prescreen.[17] This could be the case, for example, if many of the very poorest rarely come in contact with government officials; so officials do not realize they are present, and hence they are missed from the survey list.

Comparing self-targeting against the current procedure is interesting because it provides information on the different methods that are realistically within a government's choice set. However, it is also interesting to ask how self-targeting performs relative to a PMT procedure that does not have the prescreening. While this is less realistic (i.e., it is too costly to actually be conducted by the government), it provides us with a greater understanding of the margins through which self-selection occurs. Thus, in this section, we assume, hypothetically, that the government had conducted the full PMT interview on everyone in the community. Recalling the decomposition of who selects to apply in the self-targeting treatment in Section IV into selection on observables and selection on unobservables, we know a priori that self-targeting will perform worse than univer-

[17] Online app. table C.3 compares who applies in self-targeting with who is prescreened in automatic screening. The first two columns replicate the analysis in table 4 and show that the households that apply for benefits in the self-targeting treatment are poorer in terms of observable and unobservable consumption. The last two columns repeat the same exercise in the automatic screening treatment and show that the prescreen also selects poorer households in terms of observable and unobservable consumption. The estimated coefficients show that the two treatments select equally well in terms of observed consumption, though self-targeting is much better at selecting households to be screened on the basis of the unobservable component.

sal automatic targeting with respect to selection on observables, because by definition, universal automatic targeting picks up 100 percent of households with PMT scores less than the cutoff whereas self-targeting limits the beneficiaries to a subset of those who chose to apply. However, it is still possible that self-selection could outperform universal automatic targeting on net if the selection on unobservables is sufficiently large.

To simulate what would have happened in universal automatic targeting, we use asset data we collected in our baseline data to construct PMT scores for the households that applied in the self-targeting treatment and for all households in the automatic screening treatment. Using the same data source for the PMT scores in both treatments ensures that any difference that we find is due to selection. However, the PMT score using the baseline data is a better predictor of poverty compared to the government's PMT score, both because our baseline data are of higher quality than the government's data and because our consumption data come from the same survey. This effect would tend to underestimate the relative quality of selection under self-targeting, because it reduces a (true) benefit of self-targeting, namely, that those rich people who make it through the actual government PMT screen, which is quite noisy, choose not to apply in self-targeting. To correct for this, we add random noise to make the PMT score from baseline data more similar to the government's PMT score,[18] and we assume that households would have received benefits if their constructed PMT score (with random noise) was below the threshold required to receive the program.[19] We then repeat the same analysis in figure 4 and table 5, but instead of comparing

[18] Specifically, we construct the noisy PMT score from the baseline PMT score by adding a normally distributed random variable with mean zero and standard deviation $\sigma = 0.45$. The standard deviation σ is chosen such that the exclusion error using the noisy PMT score matches the exclusion error using the government's PMT score in the sample of prescreened households in the automatic enrollment treatment. Specifically, we calculate the exclusion error using the noisy PMT score for different values of σ. Panel A of online app. fig. C.2 shows the results. The exclusion error with $\sigma = 0$ is lower than if we use the government's data, and it is increasing in σ. We cannot match the government's inclusion error (panel B) because inclusion error is weakly decreasing in σ; this may occur because the density of households is increasing around the poverty line. So adding noise both makes the density gradient flatter, which tends to increase inclusion error, and also pushes down the poverty line (which is defined as a percentile of the distribution; see fn. 19), which tends to decrease inclusion error. We have also verified that beneficiaries selected using the baseline data and $\sigma = 0$ are poorer than beneficiaries selected using the government's data, and this difference shrinks as σ increases; see panel C.

[19] The threshold to receive benefits is computed using the baseline data, in the same way as the government threshold. First, in each district times urban/rural cell, the consumption percentile corresponding to a value of 80 percent of the 16th percentile is calculated. The PMT threshold is the score corresponding to that percentile in the noisy PMT distribution. We do not directly use the government's PMT threshold because the mean levels of assets in our baseline survey are different and generally larger than those in the government's survey. Using the same threshold as the government leads to broadly similar results (see online app. table C.10), although the number of hypothetical beneficiaries is different.

self-targeting to the actual automatic screening treatment, we compare it to the constructed hypothetical universal automatic targeting procedure.

The results are shown graphically in figure 5 and in regression form in table 6.[20] Panel *A* of figure 5 shows that the distribution of beneficiaries looks significantly poorer in self-selection than in the hypothetical universal automatic targeting, and the difference between the two distributions is statistically significant (the p-value from the Kolmogorov-Smirnov test of equality of distributions, with randomization inference to cluster at village level, is .047). Panel *B* of figure 5 reveals that universal automatic targeting and self-targeting have similar patterns in terms of the probability of being selected at the low end of the spectrum (and the error bands cannot reject equality between them) but that nonpoor households are more likely to receive benefits under the universal automatic targeting than under self-targeting. This is related to selection on unobservables shown in figure 4*B*: in the universal automatic targeting treatment, some higher-consumption people make it through the PMT screen as a result of errors in the PMT, whereas those people do not self-select in the self-targeting treatment.

Looking at the regressions, columns 1 and 2 of table 6 confirm that, even under this hypothetical universal automatic targeting treatment, the beneficiaries are poorer in self-targeting than in universal automatic targeting. Exclusion error is higher in self-targeting, and this result is significant at the 10 percent level in the specification without stratum fixed effects. Inclusion error is substantially lower in self-selection. As a result, the overall error rate in targeting is substantially (and statistically significantly) lower in self-targeting than under this hypothetical universal automatic targeting.

An alternative to using baseline data for everyone would be to use the constructed PMT score from baseline data only to screen households that are not prescreened in the automatic enrollment treatment and use the original government's PMT score where available. Online appendix table C.11 shows the results, which are qualitatively very similar.[21]

[20] Online app. table A.1 summarizes alternative specifications of table 6.

[21] Online app. tables C.12 and C.13 are versions of tables 5 and C.11, where we use the PMT score calculated from baseline data, without adding noise. In both cases, the results are slightly muted in magnitude and statistical significance compared to those in the original tables. As mentioned above, the reason is likely that using the higher-quality baseline data for everyone reduces the relative benefit of self-targeting over the government's PMT of screening out richer households. The one substantive change is that exclusion error is higher under self-targeting. The results in fig. 5 and table 6 are computed for a given random draw of noise. Online app. fig. C.3 and table C.8 present results in which we simultaneously bootstrap the sample and resample the noise each time. The results are qualitatively similar yet have lower levels of statistical significance.

VI. Marginal Effect of a Change in the Ordeal

We next examine the results from experimentally varying the distance to the registration site. This experiment was carefully designed to be within the set of policy instruments that potentially could be considered by the government in its real conditional cash transfer program, under the constraints that the ordeals could not be so onerous that they would either discourage the severely credit-constrained poor from applying or impose large application costs on the poor who might still be incorrectly screened out by the asset test.

In the self-targeting villages, we experimentally chose whether the sign-up location would be situated very close or further away from the potential applicants' households. Moving from the far to close registration sites decreased the distance from 1.83 kilometers (km) to 0.27 km, a reduction of 1.61 km (or 1.69 km controlling for strata fixed effects; see online app. table C.15a).[22]

Table 7 explores the impact of the close treatment on targeting outcomes by estimating the following logit equation in the sample of self-targeting villages:

$$\text{Prob}(\text{show-up}_i = 1) = \frac{\exp\{\alpha + \beta\text{Close}_v + \gamma y_{vi} + \eta\text{Close}_v \times y_{vi}\}}{1 + \exp\{\alpha + \beta\text{Close}_v + \gamma y_{vi} + \eta\text{Close}_v \times y_{vi}\}}, \tag{20}$$

where Close_v is a dummy for the close treatment in village v, y_{vi} is household i's log per capita consumption, and $\text{Close}_v \times y_{vi}$ is the interaction between them. Columns 1–3 show results without stratum fixed effects, and columns 4–6 show results with stratum fixed effects.

Increasing distance reduces the number of applicants but does not differentially affect who applies. We first show the results from estimating equation (20) including only the Close_v variable. The results show that the close treatment increases the log odds of applying by between 0.21 (col. 1, no stratum fixed effects, p-value .16) and 0.28 (col. 4, with stratum fixed effects, p-value .10).[23] This means that moving from far to close increases the percentage of households that apply by 15 percent

[22] Given differences in geography, the nature of the variation in distance was not the same across rural and urban locations. In rural areas, the sign-up station in the close treatment was located in each hamlet of the village (essentially zero distance from people's houses), whereas in the far treatment it was in the village office (an average of 1.2 km from people's houses; see online app. table C.15b). In urban areas, the sign-up station in the close treatment was located in the village office (an average of 0.8 km from people's houses), whereas in the far treatment it was in the subdistrict office (an average of 3.1 km from people's houses; see table C.15c).

[23] The OLS version of this coefficient, which is clustered at the village level rather than the stratum level, is statistically significant at the 5 percent level (p-value .024). See app. table C.14.

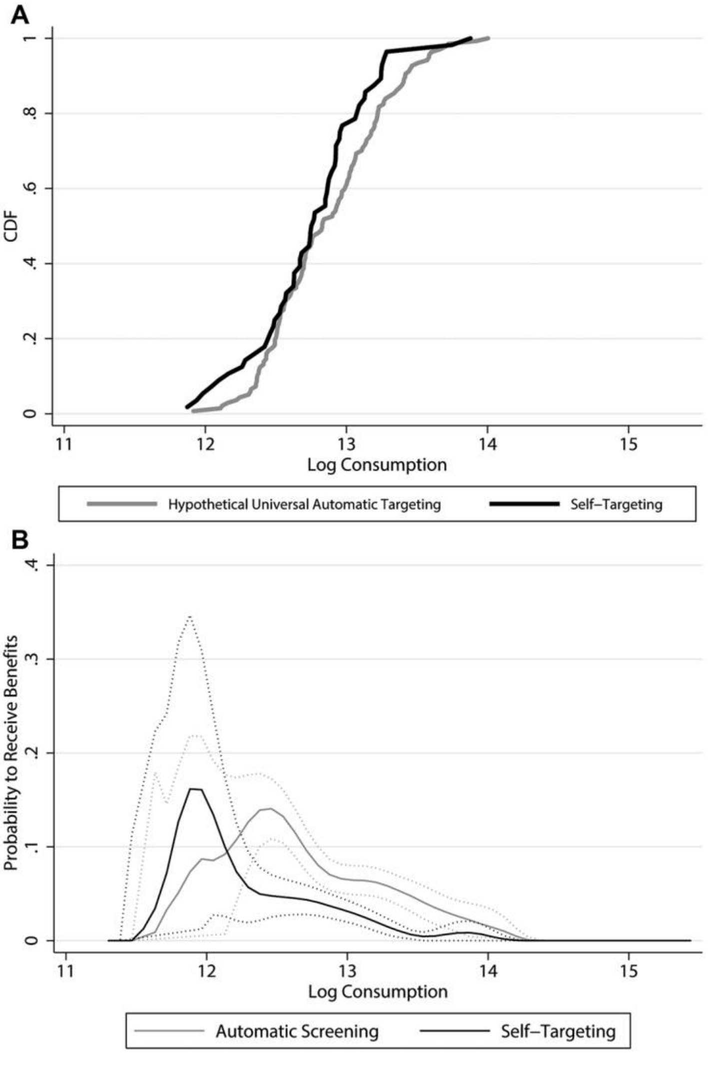

FIG. 5.—Comparison of self-selection and hypothetical universal automatic targeting. *A*, CDF of consumption of beneficiaries. *B*, Getting benefit as a function of log per capita consumption. Panel *A* shows the CDFs of log per capita consumption of beneficiaries in the self-targeting and hypothetical universal automatic targeting treatments. Households in self-targeting villages are defined as beneficiaries if they applied for benefits and if their PMT score according to the baseline asset data (with random noise) was below the required threshold. Households in automatic screening villages are defined as beneficiaries if their PMT score according to the baseline asset data (with random noise) was below the required threshold. (The threshold is computed using the baseline data, in the same way as the government threshold; see fn. 19.) A Kolmogorov-Smirnov test of equality yields a *p*-value of .068. Panel *B* presents nonparametric Fan regressions of benefit receipt on log per capita consumption in the two treatments. Bootstrapped pointwise 95 percent confidence intervals, clustered at the village level, are shown in dashes.

TABLE 6
COMPARISON OF TARGETING UNDER SELF-SELECTION AND HYPOTHETICAL UNIVERSAL AUTOMATIC TARGETING USING BASELINE DATA

	Log Consumption (Beneficiaries; OLS) (1)	Receives Benefits (Logit) (2)	Error (Logit) (3)	Exclusion Error (Logit) (4)	Inclusion Error (Logit) (5)
		A. No Stratum Fixed Effects			
Self-targeting	−.116**	4.556	−.403***	.663*	−.982***
	(.059)	(3.359)	(.119)	(.396)	(.182)
Log consumption		−1.083***			
		(.199)			
Log consumption × self-targeting		−.431			
		(.263)			
Observations	193	3,996	3,998	249	3,749
Mean of dependent variable	12.84	.048	.098	.896	.045
		B. With Stratum Fixed Effects			
Self-targeting	−.136**	6.546*	−.425***	.245	−1.002***
	(.062)	(3.544)	(.119)	(.404)	(.180)
Log consumption		−1.048***			
		(.197)			
Log consumption × self-targeting		−.584**			
		(.278)			
Observations	193	3,437	3,938	103	3,180
Mean of dependent variable	12.84	.056	.099	.796	.053

NOTE.—In each panel, each column reports the coefficients from a logit or OLS regression with dependent variable indicated in the column header. The sample in col. 1 consists of beneficiary households included in the baseline survey. The sample in cols. 2 and 3 consists of all households; in col. 4, it consists of very poor households (those with baseline consumption below 80 percent of the poverty line); and in col. 5, it consists of households that are not very poor. See the note to table 5 for variable definitions. Households in self-targeting villages are defined as beneficiaries if they applied for benefits and if their PMT score according to the baseline asset data (with random noise) was below the required threshold. Households in automatic screening villages are defined as beneficiaries if their PMT score according to the baseline asset data (with random noise) was below the required threshold. (The threshold is computed using the baseline data, in the same way as the government threshold; see fn. 19.) In panel A and col. 1 of panel B, robust standard errors, clustered at the village level, are shown in parentheses. In panel B, cols. 2–5, robust standard errors are clustered at the stratum level.

* $p < .1$.
** $p < .05$.
*** $p < .01$.

(5.8 percentage points).[24] When we test for differential selection by consumption (col. 5), we are unable to distinguish the differential effect of the close treatment at different consumption levels from zero. While the

[24] The fact that the marginal change in costs had any effect is in contrast to the one study we know of with this form in the United States. In that study, Ebenstein and Stange (2010)

TABLE 7
EXPERIMENTAL RESULTS: PROBABILITY OF SHOWING UP AS A FUNCTION OF DISTANCE AND LOG PER CAPITA CONSUMPTION

	No Stratum Fixed Effects			With Stratum Fixed Effects		
	(1)	(2)	(3)	(4)	(5)	(6)
Close subtreatment	.205 (.146)	1.345 (2.841)	.185 (.237)	.275 (.168)	.485 (2.920)	.179 (.314)
Log consumption		−1.434*** (.143)			−1.446*** (.144)	
Close subtreatment × log consumption		−.093 (.217)			−.023 (.218)	
Consumption quintile 2			−.312 (.233)			−.326 (.255)
Consumption quintile 3			−.821*** (.229)			−.792*** (.230)
Consumption quintile 4			−1.072*** (.204)			−1.050*** (.231)
Consumption quintile 5			−2.204*** (.253)			−2.276*** (.271)
Close subtreatment × consumption quintile 2			−.243 (.321)			−.246 (.378)
Close subtreatment × consumption quintile 3			.268 (.295)			.330 (.318)
Close subtreatment × consumption quintile 4			−.382 (.298)			−.262 (.313)
Close subtreatment × consumption quintile 5			.189 (.368)			.308 (.386)
Stratum fixed effects	No	No	No	Yes	Yes	Yes
Observations	2,000	2,000	2,000	1,960	1,960	1,960
Mean of dependent variable	.377	.377	.377	.385	.385	.385

NOTE.—Each column reports the coefficients from a logit regression of the show-up dummy on the close subtreatment and other regressors. The sample is all households in self-targeting villages. In cols. 1–3, robust standard errors are clustered at the village level. In cols. 4–6, robust standard errors are clustered at the stratum level.
* $p < .1$.
** $p < .05$.
*** $p < .01$.

standard errors are large (i.e., in col. 5 of table 7, the confidence interval on the interaction of close treatment and log consumption ranges from about −0.45 to 0.40), the range of effects is substantially smaller than the overall difference between self-targeting and automatic screening shown in table 5. Given that the theory implies that there may be nonlinearities

use cross-state variation to examine the impact of a marginal change in ordeal, where those receiving unemployment insurance could recertify their status online instead of in person. They find no effect on overall take-up from the change.

in the effect on the type of individual who applies when we alter the ordeal, we next explore potential nonlinearities in the effect. Specifically, column 6 interacts the close treatment dummy with dummies for quintiles of log per capita consumption, and once again, we find no evidence that moving the targeting closer to the households differentially changes the distribution of who showed up.

VII. Using the Model to Distinguish Theories and Predict Alternative Policies

The results thus far have shown that requiring households to apply for the program substantially improves targeting to the poor compared to automatic screening; yet marginal increases in application costs do not seem to further improve targeting. In this section, we return to the model in Section III, estimate the unknown parameters of the model from the data in the self-targeting sample, and use it to shed light on which theoretical mechanisms are driving the empirical results.

To take the model to the data, we start with equations (4) and (5) and specify a functional form for the shock term ε, which can be viewed as individuals' psychic costs to apply. We assume that the idiosyncratic utility shocks are drawn from a logistic distribution with mean v_ε and standard deviation σ_ε. We parameterize unsophisticated households' beliefs about the equilibrium probability of receiving benefits if they show up as a function of income, that is, $\lambda(y)$, to take the probit form, so that $\lambda(y) = \Phi(\gamma + \pi y)$, where y is log per capita expenditure and Φ is the standard normal CDF. We use the discount rate at which households can borrow from the government-subsidized credit program (Kredit Usaha Rakyat), 22 percent, to compute the NPV of benefits under the program. We also use a 22 percent discount rate for our base case estimate of δ.[25] We focus on fitting five parameters: v_ε, σ_ε, α (the share of households that are sophisticated), and the two parameters of the $\lambda(y)$ distribution, γ and π.

To estimate the model, we exploit both cross-sectional and experimental variation in registration costs and benefits. We define registration costs as the per capita monetary cost, including forgone wages, of traveling to the registration site, waiting in line, and returning home. That is, for each household, we specify

$$c(y_i, l_i) = \text{wage}_i \times (\text{traveltime}_i + \text{waittime}) + \text{travelmoney}_i, \quad (21)$$

[25] We show in online app. table C.16 that the results of the generalized method of moments (GMM) estimation are similar if we use either a much lower discount rate (5 percent), which corresponds to what households receive on savings accounts, or a much higher discount rate (50 percent), which corresponds to what poor households might pay to borrow from informal money lenders.

where traveltime_i and travelmoney_i are the individual's reports of the time and expenditure required to reach the application site, which we observe in the baseline survey for all households, regardless of whether they show up or not. We compute waittime by taking average wait times by treatment group and urban/rural designation calculated from the midline survey. We calculate the household hourly wage rate wage_i by dividing monthly household expenditure by hours worked by the household in a month.

Figure 6 plots a Fan regression of the total costs of applying, $c(y_i, l_i)$, against per capita consumption y_i. The figure shows that the actual total sign-up cost exhibits some mild concavity of the sort we introduced as a possibility in Section III.B.3.[26]

We calculate the level of benefit, b_i, that the household would receive if enrolled in the program on the basis of the number of children and their respective education levels.[27] Since consumption is likely measured with error, we assume that individuals make their decisions on the basis of their true income y^*, whereas we observe $y = y^* e^{\omega}$, where ω is a normally distributed error term. We use the fact that, for a random subset of our sample, we observe per capita consumption measured 3 months apart in the midline survey and the endline survey to calibrate the standard deviation of ω. We assume that the standard deviation of measurement error ω is half that of the total difference in consumption across the two surveys, or 0.275, suggesting that measurement error in consumption is nontrivial in our setting.

We estimate the model by GMM, using the following moments. The mean values of the show-up rates for the five quintiles of the consumption distribution in the far and close treatments generate 10 moment conditions. The mean values of the show-up rates in all combinations of top and bottom terciles of the distribution of observed consumption (PMT score) and top and bottom terciles of the distribution of unobserved consumption generate four moment conditions. The mean values of the show-up rates in the top and bottom quartiles of the distance distribution generate two moment conditions. For each of these show-up moments, we generate show-up rates from the model by integrating over

[26] A regression of $c(y_i, l_i)$ on y_i and y_i^2 shows that the coefficient on the quadratic term is negative and statistically significant at the 5 percent level. This is not driven by the outliers shown in the figure; we obtain a similar result even when we drop the 17 observations with per capita consumption above Rp. 2 million per month.

[27] The benefit is calculated as follows. Beneficiary households each receive a base benefit of Rp. 200,000 per year. This level increases by Rp. 800,000 if they have a child younger than age 3 or are currently expecting, by Rp. 400,000 if they have a child enrolled in primary school, and by Rp. 800,000 if they have a child in middle school. Since all beneficiaries fall into at least one of these categories, the benefit level is therefore between Rp. 600,000 and Rp. 2.2 million per year, with a mean of about Rp. 1.3 million.

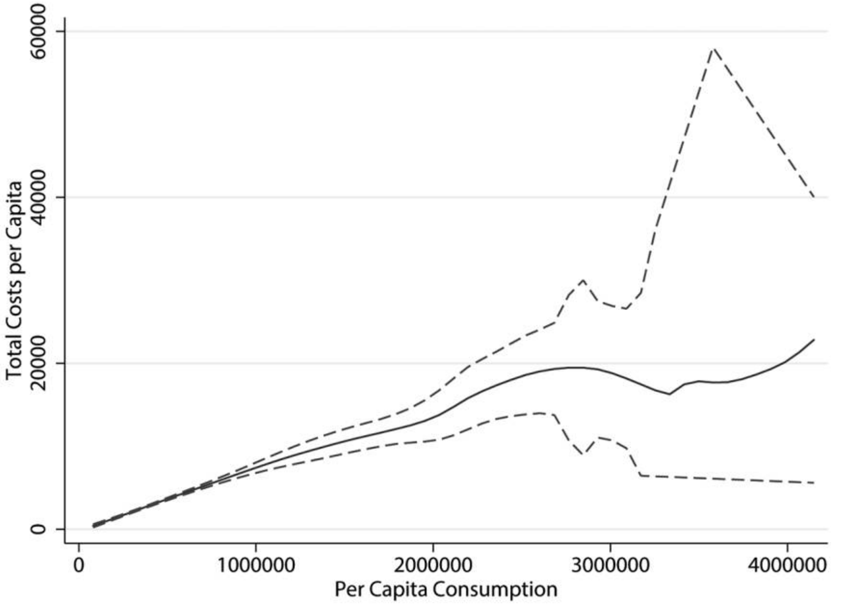

FIG. 6.—Cost of applying by per capita consumption. The figure shows a nonparametric Fan regression of total costs incurred in applying for PKH against per capita consumption. Bootstrapped pointwise 95 percent confidence intervals, clustered at the village level, are shown in dashes. Costs assume that one individual per household goes to a sign-up location, even for households in the opportunity cost subtreatment.

possible unobserved values of the utility shock ε and measurement error ω term as follows:[28]

$$\begin{aligned}\text{Prob}(\text{show-up}_i = 1) = {} & \alpha \int \text{Prob}[\varepsilon > -g(y_i^o e^{\omega}, y_i e^{\omega}, l_i)] df_{\omega} \\ & + (1-\alpha) \int \text{Prob}[\varepsilon > -h(y_i e^{\omega}, l_i)] df_{\omega},\end{aligned} \tag{22}$$

where $g(y^o, y, l)$ and $h(y, l)$ are defined in equations (6) and (7).[29]

[28] To evaluate the integrals we use the explicit formula for the logit CDF and integrate numerically over ω using the trapezoidal method over a grid with 100 points between $-1.1 = 4 \times (-0.275)$ and 1.1.

[29] In contrast to the model in Sec. III, which has only two periods, we assume that utility is defined over monthly consumption. Costs are incurred in the first month, and benefits are distributed monthly for 6 years, starting 1 year after the application date. We assume that households evaluate the benefits using the yearly discount rate (set to 0.22 in our baseline specification) and the NPV of benefits calculated 1 year after the application date.

Finally, we add the four following moments to help pin down the $\lambda(y)$ function. The first moment matches $E[\lambda(y)]$ computed using the λ function using the γ and π model parameters to the mean benefit receipt rate in the data, on the sample of households that applied. The third moment matches $E[\lambda_{\text{induced}}(y)]$ computed using the λ_{induced} function induced if households base their show-up decisions on λ (see eq. [8]) to the mean benefit receipt rate in the data, on the sample of households that applied. The second and fourth moments impose that the errors in the first and third moments are uncorrelated with demeaned y. Specifically, the four moments are[30]

$$E[\Phi(\gamma + \pi y_i) - \text{benefit}_i \mid \text{show-up}_i = 1] = 0,$$
$$E[[\Phi(\gamma + \pi y_i) - \text{benefit}_i](y_i - \bar{y}) \mid \text{show-up}_i = 1] = 0,$$
$$E[\lambda_{\text{induced}}(y_i) - \text{benefit}_i \mid \text{show-up}_i = 1] = 0,$$
$$E[[\lambda_{\text{induced}}(y_i) - \text{benefit}_i](y_i - \bar{y}) \mid \text{show-up}_i = 1] = 0.$$

This gives us a total of 20 moments to estimate five parameters, so we use a standard two-step GMM procedure to compute optimal weights for the 20 moments.

Table 8 shows the estimated parameter values. Specifically, the estimated model parameters are $v_\varepsilon = -79{,}681$, $\sigma_\varepsilon = 59{,}715$, $\alpha = 0.50$, $\gamma = 8.04$, and $\pi = -0.72$. Several observations are worth making about the estimated parameters. The result that $v_\varepsilon < 0$ implies that the idiosyncratic utility shocks, on average, favor not showing up. Since utility is linear, v_ε is interpretable in monetary terms, so the mean ε term is equal to about US$8. The fact that $\alpha = 0.50$ implies that households are equally divided between sophisticated and unsophisticated; that is, roughly half the households appear to self-select on the basis of their total income y, rather than knowing the components of income y^o that feed directly into the PMT. As we will show below, these unsophisticated households further help improve targeting over what would be achieved if all households were sophisticated.[31]

Online appendix table C.22 reports the empirical moments values as well as the simulated moments values using the estimated parameters. Panel A lists the 16 moments that are means of the show-up rate in various subgroups of the population. Panel B lists the four mean λ function

[30] Online app. table C.17 reports the (scaled) sensitivity measure proposed by Gentzkow and Shapiro (2014). The results confirm that indeed the λ parameters are primarily identified by the four λ moments. The test also reveals that the model is approximately equally sensitive to the cross-sectional and experimental moment conditions.

[31] Online app. table C.16 shows the estimated parameter values with alternative values for the 1-year discount factor ($\delta = 0.50$ and $\delta = 0.95$). The results show that changes in the annual discount factor are absorbed by changes in the scale of v_ε and σ_ε.

TABLE 8
Estimated Parameter Values for the Model

v_ε	σ_ε	α	γ	π
−79,681	59,715	.50	8.04	−.72
(6,798)	(11,734)	(.07)	(.63)	(.05)

Note.—This table reports the estimated mean v_ε and standard deviation σ_ε of the utility shock (ε), the fraction of sophisticated households (α), and the constant γ and log consumption coefficient π in the λ function. The parameters are estimated using two-step feasible GMM. For each step, we choose 100 random initial conditions and minimize the objective function using a trust-region-reflective algorithm. Bootstrapped standard errors, calculated using 100 bootstrap iterations, are in parentheses.

moments. The results serve as a goodness-of-fit check for the model and indicate that the estimated moments generally match their empirical counterparts. Online appendix figure C.4 graphs empirical and predicted show-up rates in the close and far subtreatments, by consumption quintile. The top-left graph uses measured show-up rates. The top-center graph (repeated in the middle-left and bottom-left graphs) shows the predicted show-up rates using the estimated model. A comparison of the top-left and top-center graphs offers another view of the fit of the model.

We now use the estimated parameters to predict the application rates under different assumptions on the model. For each possible scenario, we simulate predicted application rates. To summarize what the model predicts, we repeat the same logit regressions we performed in table 7 on the simulated data. We also calculate the predicted show-up rates for close and far subtreatments for those above and below the poverty line.

The results from this exercise are shown in table 9, and the predicted show-up rates by quintile are graphed in figure C.4. For comparison purposes, column 1 of table 9 and the top-left graph of figure C.4 replicate the actual empirical results (e.g., col. 2 of table 7). In addition to the empirical results from the logit model, in panel B we calculate the show-up rates for those above and below the poverty line for both close and far treatments. In panel C, we calculate the ratio of the poor to rich show-up rates (i.e., eq. [11] from the model) for both treatments, as well as the difference in this ratio between the close and far treatments (i.e., eq. [12] from the model). In column 1, the difference is positive but statistically insignificant, indicating no statistically detectable differential targeting induced by moving from close to far in the experiment.[32]

[32] Note that the difference of ratios is positive but insignificant, whereas the interaction term (the estimated coefficient on [Close × log consumption]) in panel A is negative and insignificant. The reason they have different signs is that the logit model in panel A is estimated using the continuous log consumption expenditure variable, whereas the ratios in panel C are based on a dummy variable for poor/nonpoor. If we reestimate the logit model using a dummy variable for rich, we obtain results with the same sign. Note also that the results in this table are based on the actual populations in the close and far subgroups. Since this was randomized, these will be statistically similar, but there may be small-sample differences. Online app. table C.19 replicates the analysis in this table adjusting for these small-sample differences.

TABLE 9
MODELED EFFECTS OF TIME AND DISTANCE COSTS ON SHOW-UP RATES

		PREDICTED SHOW-UP PROBABILITY (Model)				
	SHOW-UP RATE (Experimental) (1)	Baseline Model (2)	$\sigma_\varepsilon = \hat{\sigma}_\varepsilon/2$ (3)	$\sigma_\varepsilon = 0$ (4)	Assuming Same Travel Technology (5)	Constant $\mu(\cdot)$ and $\lambda(\cdot)$ (6)
	A. Logistic Regressions					
Close	1.509	−1.365	−1.825	−1.791	−1.367	−1.742
	(2.972)	(3.098)	(3.472)	(3.765)	(2.967)	(2.18)
Log consumption	−1.423***	−1.630***	−2.181***	−2.456***	−1.631***	−.103
	(.148)	(.163)	(.193)	(.204)	(.166)	(.118)
Close × log consumption	−.105	.105	.141	.138	.106	.136
	(.227)	(.238)	(.268)	(.29)	(.228)	(.166)
Observations	1,971	5,913,000	5,913,000	5,913,000	5,913,000	5,913,000
p-value		.522	.483	.509	.513	.391
	B. Show-Up Rates					
Above poverty line, far	34.09	34.55	30.04	28.12	34.54	45.89
Above poverty line, close	38.99	37.37	33.11	31.17	37.37	47.15
Below poverty line, far	53.23	71.94	72.94	73.83	71.92	46.53
Below poverty line, close	59.32	65.52	65.81	66.25	65.52	43.84

	C. Show-Up Rate Ratios					
Poor to rich ratio, far	1.561	2.082	2.428	2.626	2.082	1.014
	(.213)	(.203)	(.244)	(.262)	(.199)	(.14)
Poor to rich ratio, close	1.522	1.753	1.987	2.126	1.753	.93
	(.169)	(.183)	(.214)	(.221)	(.19)	(.141)
Difference of ratios	.040	.329	.441	.5	.329	.084
	(.268)	(.271)	(.322)	(.34)	(.281)	(.197)
p-value		.448	.338	.288	.456	.893

Note.—This table reports results using measured show-up outcomes in col. 1 and model predicted show-up probabilities in cols. 2–6. The model is simulated using the estimated parameters and actual household application costs in col. 2; i.e., the specification is the same as in col. 1 but using the variation from the model rather than the experimental variation. In cols. 3 and 4, the standard deviation of the utility shocks ε is assumed to be half of the estimated value, and zero, respectively. For col. 5, we regress reported monetary costs and reported travel time on quadratic functions of distance, treating urban and rural populations separately. We then use these predicted average travel costs for all households and recalculate total registration costs $c(y, l)$. In col. 6, the model is simulated assuming that the $\mu(\cdot)$ and $\lambda(\cdot)$ are constant functions, equal to the mean benefit receipt rate in the sample of households that show up. The main sample consists of households in self-targeting villages. Each column in panel A reports the coefficients from a logit regression of the show-up dummy specific to that column, on the close subtreatment dummy, log consumption, and their interaction. In order to run logit regressions using predicted show-up rates, we create 3,000 copies of the data. The copies of each household are assigned to show up or not in proportion to its predicted probability of showing up. In parentheses we report bootstrapped standard errors clustered by village that are designed to be comparable to the standard errors in col. 1. To compute the standard errors, for each bootstrap iteration we sample 2,000 households, clustered at the village level, to make the sample equivalent to that in col. 1. We perform 1,000 bootstrap iterations. The p-value in panel A is the test of whether the coefficient on [Close × log consumption] is equal to the equivalent coefficient in col. 1. Panel B reports show-up rates (probabilities) in all combinations of far/close subtreatments and below the poverty line (poor)/above the poverty line (rich). Panel C reports ratios of the numbers in panel B. The p-value in panel C is the test of whether the difference of ratios is equal to the difference of ratios in col. 1.

* $p < .1$.
** $p < .05$.
*** $p < .01$.

In column 2 of table 9, we begin by estimating the effect on the simulated data of the change in $c(y, l)$ induced by the close treatment; that is, we use the actual costs $c(y_i, l_i)$ for both close and far households calculated using equation (21) and calculate each household's predicted show-up rate using the model with equation (22). We bootstrap the standard errors for the model-generated data in order to make them equivalent to those from the actual data.[33] The results in column 2 thus show what we would have found had the data from our survey been generated by the model.

Comparing the actual empirical estimates in column 1 with the estimates on the model-generated data in column 2, we find similar results of differential targeting between the treatments. In particular, even though the model seems to overpredict show-up rates for the poor on average, the small differential effect between rich and poor show-up ratios moving from close to far in the simulated data is not statistically distinguishable from what we actually observe in the experiment (panel C; *p*-value .448). Consistent with this, the coefficients on the close dummy interacted with log per capita consumption (η in eq. [20]), which is another way of capturing the degree of differential targeting between the close and far treatments, are also statistically indistinguishable between the actual experimental data in column 1 and the simulated data in column 2 (panel A; *p*-value .522).[34]

A. *Distinguishing Alternative Theories*

We now use the structural model to return to the various theories outlined in Section III for why self-targeting may work, by illustrating how the model predicts that show-up rates would change under various parameters. This helps shed light on why, even though there is strong evi-

[33] In order to run the logits using the predicted application rates, we create 3,000 copies of the data. The copies of each individual are assigned to apply or not apply in proportion to that individual's predicted probability of doing so. To make the standard errors comparable to those in the main experiment, we apply a cluster bootstrap approach (clustered at the village level) to this distribution, holding the total number of observations equal to the number of observations in the actual data. These standard errors do not include uncertainty in the estimated parameters.

[34] The one aspect of the model that does not match is that the predicted show-up rates for those below the poverty line are actually higher in the far treatment than in the close treatment (72 percent vs. 66 percent). We have verified that this is not due to the model, but rather due to small-sample differences in the expected benefits from obtaining the program among the poor in these two samples. In particular, the poor in the far group have (statistically insignificantly) more middle schoolers than the poor in the close group, which leads to higher show-up rates. If we simulate the impact of moving from far to close on the exact same group of beneficiaries, we indeed would obtain lower show-up rates in far than in close in both rich and poor samples. See online app. table C.19.

dence of self-selection, both the experiment and the model show no statistically significant marginal increase in the targeting ratio from increasing the severity of the ordeal (i.e., moving from close treatment to far).

One possible explanation is that if the distribution of shocks does not have the monotone hazard rate property, it is possible that targeting could get worse as distance is increased, because the density of poor people induced to drop out by a higher marginal change is higher than the density of the rich (see Sec. III.B.2). The version of the structural model that we estimate and use in column 2 uses logit shocks, which have the monotone hazard rate property, thus suggesting that the distribution of shocks alone is not driving the lack of response to a change in ordeal. However, the magnitude of the shocks may explain why the response is so low. Examining equation (12), which showed the derivative of the show-up ratio with respect to a change in distance l, one can see that increasing the variance of the shocks, which would lower the PDF f at the margin for both rich and poor, would dampen the responsiveness to a marginal increase in ordeals. In column 3 of table 9, we simulate the model in which we cut the standard deviation σ_ε of the shocks ε in half for all households. Doing so increases the point estimate of the impact of moving from close to far on the poor/rich show-up ratio—from 0.33 in the baseline model to 0.44. In column 4 we shut off the shocks entirely: we find that with no shocks, the ratio of poor to rich who show up in the far subtreatment would increase from 2.08 to 2.63. In short, shocks dampen the effect of ordeals.

A second explanation is that the poor and rich use different transportation technologies, so that the marginal monetary cost of distance is smaller for the rich (see Sec. III.B.3).[35] We generate simulated show-up rates under the counterfactual that the poor and the rich use the same travel technology. In column 5 of table 9, we reestimate the logit regressions and calculate the show-up rates for the simulated data using the same predicted costs for all households instead of the actual costs.[36] The results appear similar to the experimental findings, confirming that travel technology does not explain the lack of differential selection in response to an increase in distance.

A third explanation is that most of the selection that we observe in Section IV is being driven by the fact that households anticipate that $\mu(y^o)$ and $\lambda(y)$, the probabilities of receiving benefits conditional on showing

[35] Figure 6 showed that this might be a possible explanation in the data, as the total costs of travel do appear to be concave in per capita consumption.

[36] We model travel costs (time and money) as a function of distance. Treating urban and rural populations separately, we regress reported monetary costs and reported travel time on quadratic functions of distance. We then use these predicted average travel costs for all households and recalculate total registration costs $c(y, l)$.

up as a function of observable and total income, are downward sloping.[37] In column 6 of table 9, we simulate what would happen if all households assume that they will receive benefits with some constant probability $\bar{\mu}$, which we set equal to the average probability of getting benefits in the population of households that apply. The results are dramatic: the coefficient on log per capita expenditure falls from -1.42 and -1.63 (in cols. 1 and 2) to -0.10 (in col. 6) and is no longer significant at the 10 percent level. This suggests that almost the entire selection effect is driven by the fact that the poor and rich applicants have differential beliefs about their probability of receiving benefits. The same result emerges if we compare the change in poor to rich show-up ratios when we move from the baseline model to the model with constant $\bar{\mu}$. This result is consistent with our empirical findings: if most of the selection is coming because households anticipate that they will not receive benefits if they apply, then even small but positive costs can have large selection effects, since people with a low probability of receiving benefits will not sign up, but marginal increases in the costs of the ordeal impose deadweight costs without substantially improving selection.

A final question is how the results would differ if we change the fraction of households that understand the true decision rule $\mu(y^o)$, that is, the sophisticated households, as opposed to those that select only on the basis of total income, that is, using $\lambda(y)$. Here, the key point is who would receive benefits, since the difference is that sophisticated households that are rich but know they have low y^o choose to apply. This appears to be important: if we simulate the model and calculate the difference in average log per capita consumption, we find that moving from all unsophisticated households (i.e., $\alpha = 0$) to all sophisticated ones (i.e., $\alpha = 1$) would increase the beneficiaries' average consumption by 10 percent (see online app. table C.20). Moving from the estimated level of unsophisticated households (i.e., $\alpha = 0.50$) to all sophisticated (i.e., $\alpha = 1$) would increase it by 5 percent. Combined with the results above about constant $\bar{\mu}$, these results suggest that a nontrivial share of the total selection effect of self-targeting comes from the fact that, while households understand that the rich are less likely to obtain benefits, it is important that they do not precisely know the eligibility formula.

[37] Alternatively, it could be that there is a stigma from applying that is increasing with income y; i.e., the rich would feel embarrassed from showing up and applying for an antipoverty program, and the poor would not. Empirically, this will look similar to a downward-sloping $\lambda(y)$ function. Survey responses from the midline survey suggest that stigma is not a first-order issue in our context. The survey, conducted after the application period and before beneficiaries were announced, contained questions on why households did not apply for PKH. Households were given many (nonexclusive) answer options, including two options that measured the stigma of being considered poor by other people. Only one respondent, out of the 237 in this sample who did not choose to sign up, chose either of these stigma-related options.

B. *Impact of Alternative Targeting Approaches on the Poverty Gap*

Self-targeting appears to perform better than automatic screening in identifying the poor, but it also entails costs. There is the cost of the ordeal: households lose valuable time traveling to the interview site and waiting in line to be interviewed and often need to spend money traveling as well. In addition, both self-targeting and automatic screening entail administrative costs: enumerators need to be paid to conduct interviews at self-targeting application sites for self-targeting and to conduct field verification visits to assess PMT scores in both self-targeting and automatic screening. One of the potential benefits of self-targeting is that it reduces the number of surveys that need to be conducted compared to a universal PMT; but if those cost savings to the government were offset by commensurate increases in the waiting and travel costs paid by households, one might not be so sanguine about such a policy.

To help shed light on this issue, we use the poverty gap to integrate the benefits and costs of the program borne by households, and we compare the experimental and simulated targeting policies.[38] To calculate the poverty gap, we assume that households receive their baseline consumption each month. In addition, applicant households incur the application cost in the first month, and benefit recipients receive monthly benefits for 6 years starting at the beginning of the second year (i.e., in months 13–84). We compute the poverty gap in each month and average these measures. This corresponds to a "steady-state" measure of the poverty gap across space if the program was phased in uniformly over time. Finally, we expand or contract the fraction of the country that could be covered by the program in order to hold the government budget constant, and we assume that in the remainder of the country the poverty gap is unchanged.

To calculate the costs of the program, we use the same approach as in the model to calculate the time and money costs to households from applying. For the administrative costs, we note that there could be economies of scale in implementing a national program. For automatic screening, where we indeed know the Indonesian government's costs from implementing the nationwide PMT, we report those "at-scale" costs as well as those from our experiment; for self-targeting, which has yet to be done nationally, we do not have an analogous estimate and therefore use the costs from our experiment.

[38] The advantage of using the poverty gap is that (unlike the poverty head count) it is sensitive to how far below the poverty line households lie. Moreover, given that the intended threshold for PKH benefits is 80 percent of the poverty line used to measure the poverty gap, this measure also captures the effect of the program on the poorest households that are intended nonbeneficiaries.

Table 10 presents the results of this exercise. The first three columns use experimental data from the automatic screening and self-targeting treatments. Column 1 shows the results from the automatic screening treatment using the costs of the national automatic screening program, while column 2 uses the (higher) costs in our sample. Column 3 shows the results for self-targeting, where we have the costs in our sample.[39] Columns 4–8 are based on show-up probabilities derived from the model, under different alternative scenarios for the application costs. Column 9 assumes that the program is perfectly targeted to households under 80 percent of the 16th percentile of the consumption distribution; only these households apply and receive benefits. This scenario serves as a benchmark to evaluate the performance of the preceding experimental and simulated approaches to targeting.

Panel A in table 10 reports the fraction of the population that applies for benefits (or is interviewed, in the case of the usual procedure), the fraction of the population that receives benefits, as well as the components of this number due to very poor and non–very poor households. Panel B reports the average costs of applying borne by households. Panel C reports administrative program costs and the costs of benefits paid. Panel D calculates the poverty gap if the program is implemented under a fixed government budget and the improvement in the poverty gap under each scenario (relative to baseline) as a fraction of the potential improvement were there perfect targeting.

We begin by comparing the usual automatic screening procedure with self-targeting. The key point is that, for a fixed government budget, self-targeting achieves a substantially greater reduction in poverty gap than automatic screening, even taking into account the costs borne by households that spend time applying but do not receive benefits. Specifically, self-targeting achieves 39 percent of the theoretical upper bound of reduction in poverty gap, compared with 27–30 percent for automatic screening. Self-targeting thus achieves between 29 and 41 percent more reduction in the poverty gap than automatic screening. The reason is largely that self-targeting finds much poorer households and it has much lower inclusion error. The lower inclusion error implies that for a fixed budget, the program can be implemented to substantially more locations, covering more poor households.

The results using the model predictions and actual household application costs, presented in column 4, are similar to the results using ex-

[39] Even though the treatment is randomized, the consumption distributions are slightly different in the two treatments because of finite samples, and these differences are amplified when we compute the poverty gap. To correct for this, the consumption distribution in the automatic screening villages is adjusted to be exactly the same as in the self-targeting villages in each district times urban cell. Hence, the comparison between the usual procedure and self-targeting isolates the effect of the different targeting method.

perimental data in self-targeting (38.58 percent of the theoretical upper bound for actual self-targeting in col. 3, compared with 36.38 percent of the theoretical upper bound for simulated self-targeting from the model). Columns 5 and 6 show the results of increasing the distance to the application site for each household in the far subtreatment by 3 and 6 km, respectively. Panel B shows that household costs go up by between 15 percent and 24 percent, and panel A shows that the show-up rate and the fraction of beneficiaries decrease slightly. In columns 7 and 8, we look at the effects of increasing the waiting time by a factor of three and six, respectively. The treatments increase the cost incurred by households that show up by around 75 percent and 171 percent. These changes lead to a decrease in average show-up rates from around 37.9 percent to around 36.8 percent and 35.4 percent, respectively; the average benefit receipt rate falls slightly. In the end, the magnitude of the changes in columns 5–8 relative to column 4 is very small: the poverty gap is essentially unchanged.

On net, the key conclusion is that self-targeting appears to lead to a substantially higher reduction in poverty gap compared with the automatic screening procedure. Increasing the distance or waiting time does not seem to further improve targeting noticeably.

VIII. Conclusion

Using data from a field experiment across 400 villages to examine targeting in Indonesia's conditional cash program (PKH), we showed that introducing application costs meant that the poor are more likely to self-select into applying than the nonpoor. Interestingly, this selection occurred on two types of margins. First, we observe selection on the component of consumption that is observable to governments. This implies that ordeals have the potential to save money by not having to survey rich people who would ultimately fail the asset test. Second, ordeal mechanisms also lead to selection on the unobservable components of consumption, which means that targeting may become more pro-poor by screening out the rich who may get incorrectly screened in by an asset test. On net, introducing self-selection improved targeting as compared with the other targeting mechanisms that we considered, both the current usual government procedure and a universal automatic targeting system.

However, while experimentally increasing the ordeals by increasing the distance to the application site reduced the number of individuals who applied under the self-targeting regime, it did not differentially improve targeting. Put another way, the increase in distance we experimentally induced (a 1.6 km increase in distance) imposed substantial enough costs on households to lower application rates, but these costs did not differentially affect poor and rich households. Estimating our model

TABLE 10
IMPACT OF ALTERNATIVE TARGETING APPROACHES ON THE POVERTY GAP

	Show-Up Rate (Experimental)			Predicted Show-Up Probabilities (Model)					
	Automatic Screening (Scaled) (1)	Automatic Screening (in Sample) (2)	Self-Targeting (3)	Baseline Model (4)	Far Distance + 3 km (5)	Far Distance + 6 km (6)	Far Wait Time × 3 (7)	Far Wait Time × 6 (8)	Perfect Targeting (9)
	A. Program Statistics								
Mean show-up rate (%)	34.62	34.62	37.84	37.93	37.67	37.53	36.79	35.39	5.83
Mean benefit receipt (%)	4.38	4.38	3.64	4.11	4.10	4.09	4.06	4.00	5.83
Mean eligible benefit receipt (%)	.63	.63	.73	.86	.86	.86	.86	.86	5.83
Mean ineligible benefit receipt (%)	3.75	3.75	2.91	3.24	3.23	3.23	3.20	3.14	.00
	B. Average Household Costs for Households That Show Up (Rupees)								
Average cost to households	1,021	1,021	13,674	13,831	15,947	17,218	24,187	37,460	7,621
Average cost to beneficiary households	938	938	12,464	12,797	14,774	16,130	21,987	34,968	7,621
Average cost to nonbeneficiary households	1,033	1,033	13,803	13,957	16,091	17,351	24,459	37,777	. . .
	C. Government Costs and Benefits Paid (Rupees)								
Administrative costs, per household	4,768	31,054	6,764	6,781	6,734	6,710	6,576	6,326	1,042
Expected benefits, per household	332,028	332,028	306,108	353,230	352,305	351,742	349,750	344,826	472,990

	D. Poverty Gap								
Poverty gap under fixed budget (%)	2.736	2.741	2.720	2.724	2.724	2.724	2.724	2.725	2.610
Reduction in poverty gap relative to perfect targeting (%)	29.91	27.42	38.58	36.38	36.71	36.37	36.47	36.26	100.00

Note.—This table reports the estimated effect of the PKH program on poverty gap under different scenarios. Cols. 1–3 use experimentally measured show-up rates, and cols. 4–8 use show-up probabilities predicted using the model. In col. 9 we assume that only households with baseline consumption under 80 percent of the poverty line apply, and all receive the benefit. The sample is the automatic screening villages in cols. 1 and 2 and the self-targeting villages in all other columns. The results in this table are constructed as follows. First, we create 1,000 copies of the data and we smooth the consumption distribution by assigning copied households linearly interpolated values of consumption. The consumption distribution in automatic screening villages is then adjusted to match that in self-targeting villages, and this is done separately in each district times urban cell. The results in automatic screening villages are also weighted such that the relative weights of district times urban cells are the same as in self-targeting, Next, each household (in the enlarged sample) is assigned a show-up outcome; in cols. 1–3 this is whether the original household actually showed up, and in cols. 4–8 it is given by a random draw with probability of show-up predicted by the model. Each household that shows up is randomly assigned whether it receives the benefit or not, with probability predicted from a probit model on log consumption. Panel A reports the average number of households that show up, that receive benefits, and that receive benefits and are very poor (below 80 percent of the 16th percentile of the consumption distribution). Panel B reports per household costs borne by households that show up, that receive the benefits, and th at show up but do not receive the benefits (the show-up and beneficiary outcomes are constructed as explained above). Panel C reports administrative costs per study household (total costs in the study sample divided by the number of households in the sample). Column 1 uses the costs from the national (scaled-up) version of the automatic screening program, and col. 2 uses the costs from the version of the program implemented for this study. For self-targeting, we know only the costs from the study, which are used in col. 3. The first row in panel D reports the steady-state poverty gap when the program is implemented in a fraction of the villages, due to a fixed budget. (We normalize this fraction to 80 percent in self-targeting villages in col. 3.) The steady-state poverty gap is calculated assuming that each applicant household pays the application cost in the first month and receives its baseline consumption level for the following 11 months; then beneficiary households receive the benefit for 6 years starting 1 year after the application. Specifically, the poverty gap is calculated as the average of the poverty gaps in each month, without time discounting. The second row in panel D measures the improvement in the poverty gap under a fixed budget relative to baseline, as a fraction of the improvement under perfect targeting (col. 9).

suggested that the key driver behind the improvement in targeting from application costs was the fact that the rich forecast that they have a low probability of success and hence do not choose to apply.

In short, these types of administrative costs can be a powerful tool to improve targeting relative to automatic screening systems, but making onerous ordeals even more costly may not be the best way to improve targeting further. This suggests that one should not strictly view administrative barriers as a bar to take-up, but instead should carefully consider their power as a screening device. On the other hand, while self-targeting dominates the usual procedure, many of the poor still do not sign up. Understanding how to design screening mechanisms to increase take-up of the poor while still discouraging sign-up of the rich seems a promising direction for future work.

References

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A. Olken, Ririn Purnamasari, and Matthew Wai-Poi. 2012. "Elite Capture or Elite Benevolence? Local Elites and Targeted Welfare Programs in Indonesia." Technical report, Massachusetts Inst. Tech.

———. 2013. "Ordeal Mechanisms in Targeting: Theory and Evidence from a Field Experiment in Indonesia." Working Paper no. 19127, NBER, Cambridge, MA.

Benhassine, Najy, Florencia Devoto, Esther Duflo, Pascaline Dupas, and Victor Pouliquen. 2013. "Turning a Shove into a Nudge? A 'Labeled Cash Transfer' for Education." Working Paper no. 19227, NBER, Cambridge, MA.

Besley, Timothy, and Stephen Coate. 1992. "Workfare versus Welfare: Incentive Arguments for Work Requirements in Poverty-Alleviation Programs." *A.E.R.* 82 (1): 249–61.

Bhargava, Saurabh, and Dayanand Manoli. 2015. "Psychological Frictions and the Incomplete Take-Up of Social Benefits: Evidence from an IRS Field Experiment." *A.E.R.* 105 (11): 3489–3529.

Castaneda, Tarsicio, and Kathy Lindert. 2005. "Designing and Implementing Household Targeting Systems: Lessons from Latin America and the United States." Social Protection Discussion Paper no. 526, World Bank, Washington, DC.

Coady, D., and S. Parker. 2009. "Targeting Social Transfers to the Poor in Mexico." Working Paper no. 9/60, Internat. Monetary Fund, Washington, DC.

Currie, Janet. 2006. "The Take-Up of Social Benefits." In *Poverty, the Distribution of Income, and Public Policy*, edited by Alan Auerbach, David Card, and John Quigley, 80–148. New York: Sage.

Currie, Janet, and Jeffrey Grogger. 2001. "Explaining Recent Declines in Food Stamp Program Participation." *Brookings-Wharton Papers Urban Affairs* 2001: 203–44.

Ebenstein, Avraham, and Kevin Stange. 2010. "Does Inconvenience Explain Low Take-Up? Evidence from Unemployment Insurance." *J. Policy Analysis and Management* 29 (1): 111–36.

Fan, Jianqing. 1992. "Design-Adaptive Nonparametric Regression." *J. American Statis. Assoc.* 87 (420): 998–1004.

Gentzkow, Matthew, and Jesse M. Shapiro. 2014. "Measuring the Sensitivity of Parameter Estimates to Sample Statistics." Working Paper no. 20673, NBER, Cambridge, MA.

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: World Bank.

Heckman, James J., and Jeffrey A. Smith. 2004. "The Determinants of Participation in a Social Program: Evidence from a Prototypical Job Training Program." *J. Labor Econ.* 22 (2): 243–98.

Hodges, Anthony, Anne-Claire Dufar, Khurelmaa Dashdorj, Kang Yun Jong, Tuya Mungan, and Uranchimeg Budragchaa. 2007. "Child Benefits and Poverty Reduction: Evidence from Mongolia's Child Money Programme." Technical report, Maastricht Univ.

Kidd, Stephen, and Emily Wylde. 2011. "Targeting the Poorest: An Assessment of the Proxy Means Test Methodology." Technical report, AusAID, Washington, DC.

Kleven, Henrik Jacobsen, and Wojciech Kopczuk. 2011. "Transfer Program Complexity and the Take-Up of Social Benefits." *American Econ. J.: Econ. Policy* 3:54–90.

Lindert, Kathy, Anja Linder, Jason Hobbs, and Benedicte Briere. 2007. "The Nuts and Bolts of Brazil's Bolsa Familia Program: Implementing Conditional Cash Transfers in a Decentralized Context." Social Protection Discussion Paper no. 709, World Bank, Washington, DC.

Martinelli, Cesar, and Susan W. Parker. 2009. "Deception and Misreporting in a Social Program." *J. European Econ. Assoc.* 7 (4): 886–908.

Nichols, Albert L., and Richard J. Zeckhauser. 1982. "Targeting Transfers through Restrictions on Recipients." *A.E.R.* 72 (2): 372–77.

Nichols, D., E. Smolensky, and T. N. Tideman. 1971. "Discrimination by Waiting Time in Merit Goods." *A.E.R.* 61 (3): 312–23.

Parsons, Donald O. 1991. "Self-Screening in Targeted Public Transfer Programs." *J.P.E.* 99 (4): 859–76.

Ravallion, M. 1991. "Reaching the Rural Poor through Public Employment: Arguments, Evidence, and Lessons from South Asia." *World Bank Res. Observer* 6 (2): 153–75.

Thornton, R. L., L. E. Hatt, E. M. Field, I. Mursaleena, F. S. Diaz, and M. A. Gonzalez. 2010. "Social Security Health Insurance for the Informal Sector in Nicaragua: A Randomized Evaluation." *Health Econ.* 19:181–206.