Ordeal Mechanisms in Targeting: Theory and Evidence from a Field Experiment in Indonesia

Citation

Hanna, Rema, Vivi Alatas, Abhijit Banerjee, Benjamin A. Olken, Ririn Purnamasari, and Matthew Wai-Poi. "Ordeal Mechanisms in Targeting: Theory and Evidence from a Field Experiment in Indonesia." CID Working Paper Series 2012.254, Harvard University, Cambridge, MA, November 2012.

Published Version

https://www.hks.harvard.edu/centers/cid/publications

Permanent link

https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37366278

Terms of Use

Share Your Story

Ordeal Mechanisms in Targeting: Theory and Evidence from a Field Experiment in Indonesia

Rema Hanna, Vivi Alatas, Abhijit Banerjee, Benjamin Olken, Ririn Purnamasari, and Matthew Wai-Poi

Working Papers

Center for International Development
at Harvard University
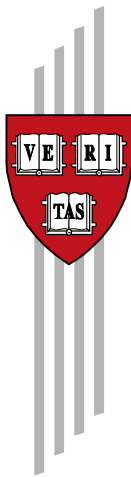
ORDEAL MECHANISMS IN TARGETING:
THEORY AND EVIDENCE FROM A FIELD EXPERIMENT IN INDONESIA

VIVI ALATAS, ABHIJIT BANERJEE, REMA HANNA,
BENJAMIN A. OLKEN, RIRIN PURNAMASARI, AND MATTHEW WAI-POI

ABSTRACT. This paper explores whether ordeal mechanisms can improve the targeting of aid programs to the poor ("self-targeting"). We first show that theoretically the impact of increasing ordeals is ambiguous: for example, time spent applying imposes a higher monetary cost on the rich, but may impose a higher utility cost on the poor. We examine these issues by conducting a 400-village field experiment with Indonesia's Conditional Cash Transfer program (PKH), where eligibility is determined through an asset test. Specifically, we compare targeting outcomes from self-targeting, where villagers came to a central site to apply and take the asset test, against the status quo, an automatic enrollment system among a pool of potential candidates that the village pre-identified. Within self-targeting villages, we find that the poor are more likely to apply, even conditional on whether they would pass the asset test. Exploiting the experimental variation, we find that self-targeting leads to a much poorer group of beneficiaries than the status quo. Self-targeting also outperforms a universal asset-based automatic enrollment system that we construct using our survey data. However, while experimentally increasing the distance to the application site reduces the number of applicants, it does not differentially improve targeting. Estimating our model structurally, we show that there are large unobserved shocks in the decision to apply, and therefore increasing waiting times to 9 hours or more would be required to induce detectable additional selection. In short, ordeal mechanisms can induce self-selection, but marginally increasing the ordeal can impose additional costs on applicants without necessarily improving targeting.

PRELIMINARY AND INCOMPLETE. PLEASE DO NOT CITE WITHOUT PERMISSION.

1. Introduction

In designing targeted aid programs, a perennial problem is how to separate the poor from the rich. One strategy for doing so is to impose program requirements that are differentially costly for the rich and the poor, in order to induce the poor to participate while dissuading the rich from doing so (Nichols, Smolensky and Tideman, 1971; Nichols and Zeckhauser, 1982; Besley and Coate, 1992). These "self-selection" or "ordeal" mechanisms are quite common: welfare programs, from the WPA in the United States during the Great Depression to the NREGA right-to-work scheme in India today, often have manual labor requirements to receive aid; unemployment schemes often require individuals to report to the unemployment office weekly during working hours, which is challenging for the employed; subsidized food schemes often provide lower quality food so that those who can afford tastier food choose not to purchase the subsidized products.

By imposing higher participation costs on the rich, these mechanisms may save governments considerable screening costs and potentially result in better targeted programs. On the other hand, imposing participation costs on the poor, however small, may also dissuade them from partaking. For example, if the poor are very credit constrained or have higher discount rates than the rich, a substantial fraction of them may choose not to apply, leading to a less pro-poor distribution of beneficiaries (see Currie 2006 for a review).

In this paper, we explore the different margins through which self-selection may operate in the context of Indonesia's Conditional Cash Program (PKH), which provides beneficiaries with US\$150 per year for 6 years. The program is aimed at about the poorest 5 percent of the population, with eligibility traditionally being determined based on a weighted sum of about 30 easy-to observe assets (e.g. how large is your house, what material is the roof of your house made of, do you own a motorbike, etc). Working with the Indonesian government, we experimentally varied the eligibility process for PKH across 400 villages. In the treatment villages, households that were interested in the program were required to travel to a central registration site to take an asset-test administered by the statistics office. This entailed both travel costs (time and money) and waiting (more time costs). Within the treatment areas, we randomly varied the application costs along two dimensions: the distance to the application site and whether one or both spouses needed to be present to apply. In control areas, the status quo procedure–automatic enrollment–was followed: the statistics office, working together with local government officials, drew up a list of potential beneficiaries; interviewed everyone at their homes; and then automatically enrolled those who passed the asset test.[1]

We begin with a description of the experiment and the data, and then ask what we would expect from such an experiment on purely a priori grounds. Specifically, we reexamine the classical theory of self-selection into social programs developed by Nichols, Smolensky and Tideman (1971), Nichols and Zeckhauser (1982), Besley and Coate (1992) and others. These papers assume indifference curves that have the property that an ordeal is more costly in utility terms for the rich and that this gap is increasing in the duration of the ordeal. As a result, there is a simple trade off between

---

[1]These two approaches – proxy means tests based on either automatic enrollment or self-selection into being interviewed, followed by verification of assets for those eligible based on the interview results – are quite common. For example, the initial Mexican Progressa program used an automatic enrollment PMT to determine beneficiaries in selected areas, and the subsequent expansion of the program under the name Oportunidades used a self-selection PMT treatment virtually identical to what we investigate in this study. See Coady and Parker (2009).

the dead-weight loss of a longer ordeal and the better self-selection that it generates. We argue that in a more realistic environment, these two properties may not necessarily hold. First, ordeals may be more costly in utility terms for the poor than the rich: given that the poor tend to be savings and credit constrained, the loss of earnings from applying may imply a greater utility cost because their marginal utility of consumption is higher. Moreover, the same ordeal may even cost more in money terms for the poor as the rich may be better able to overcome an ordeal; for example, the rich may perhaps quickly drive to a far-away registration site, while the poor may expend considerable time walking there.[2] Second, the gap may not necessarily increase in the duration of the ordeal. Even if the total *money* cost is higher for the rich, the *marginal* cost of or an ordeal may be lower. For example, if the poor walk to a registration site to apply while the rich take a bus, the cost of traveling a little further may be relatively smaller for the rich (say because the fare is fixed and the bus is fast). Third, traditional selection models assume no idiosyncratic element in the decision to apply, but we show that the effect on self-selection very much depends on the distribution of these idiosyncratic shocks: if we assume that the underlying payoffs are such that the poorest people will always apply even when there is a burdensome ordeal and the rich only do so when it is convenient (i.e. when the utility shocks are very favorable), it may be that those who are not quite the poorest may be most affected by ordeals, and this may have ambiguous effects on targeting.

Our empirical analysis proceeds in four stages. First, we begin by examining who chooses to self-select into applying for the program in the 200 villages where self-targeting was implemented. To do so, we utilize data on households' per capita consumption that we collected before the program was announced or targeting began. We find that the probability of self-selecting to apply for the program is monotonically decreasing in a household's per-capita consumption, i.e. that the poor are always more likely to apply than the rich. Decomposing consumption into that which is potentially observable to the government (i.e. the part that can be predicted based on observable assets) and the unobservable residual, we show that those who apply are poorer on observables and unobservables than those who choose not to. This implies that self-selection can not only potentially save resources (since many who would fail the asset test are no longer tested), but that it also has the potential to improve targeting even over a universally-administered asset test (since those that apply are poorer on unobservables than the population at large). However, we also find evidence for the view that self-targeting may screen out some of the poor: for example, only about 60 percent of the very poorest apply under self-targeting.

The question, though, for most governments is not necessarily how self-targeting would perform relative to a counterfactual of no error, but how would it compare against the next best alternative targeting strategy. The second step of our empirical analysis is use the experiment to compare self-targeting with the current status quo, in which the government conducted the asset-test for all potential beneficiaries (chosen through prior asset surveys and consultations with village leadership) at their homes and automatically enrolled those that passed. Compared against this real alternative, we find that per-capita consumption was 13 percent lower for beneficiaries in the self-targeting villages than those under the status quo. This occurs throughout the entire distribution: per-capita consumption of the beneficiaries in self-targeting areas was first-order stochastically dominated by

---

[2]While the car obviously costs money, most of that is sunk cost from the point of view this intervention.

the per-capita consumption of those under the status quo. Moreover, exclusion error was actually less of a problem in self targeting than in the status quo: the very poorest households were twice as likely to receive benefits in self-targeting than in control areas. Note that these findings are not fully driven by the fact that the government chose who to interview under the status quo: supplementing the government's asset–test data in the automatic enrollment villages with asset data that we independently collected for those not interviewed, we find that the beneficiaries under self-targeting would still be, on average, poorer than those under a "hypothetical" system where 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 marginal increases in the severity of the ordeal further increase targeting performance, which as we discussed above is theoretically ambiguous. We examine the results from experimentally varying the distance to the registration site (increasing travel costs) and the number of household members required to be present at the application site (increasing opportunity costs of time for the family). Note that these experiments were carefully designed to be within a set of policy instruments that could be potentially considered by the government, under the requirements that the ordeals could not be so onerous that they would either discourage the severely credit-constrained poor from applying or that would likely impose large application costs for the poor who might still be incorrectly screened out by the asset-test.

Examining the experimental variation in the extent of the ordeal, we do not observe that increasing ordeals differentially improves targeting. We find that increasing the distance that the applicant has to travel by an average of 1.7 kilometers reduces the overall number of applicants about 17 percent, and thus inclusion error of the rich is reduced. However, there is no detectable differential selection by income groups when we increase distance, and thus increasing distance also additionally screens out a similar fraction of poor households. Similarly, we also do not observe significant differential selection when we increase the opportunity cost of waiting by requiring both spouses to apply in person rather than allowing either spouse to apply alone. In sum, these results show that while ordeal mechanisms can induce self-selection by the poor, increasing the size of the ordeal can impose additional costs on applicants without necessarily improving targeting.

The theoretical model outlined a number of reasons why marginal increases in the extent of ordeals could might not necessarily improve targeting. To understand which factors are in fact empirically relevant, the final step of our empirical analysis uses Generalized Method of Moments to estimate a CRRA utility version of our model with logit shocks. We use as moments the average showup rates in the far distance treatment for each income quintile. Since we estimate the model using only one experimental subtreatment and cross-sectional differences in distances, not the experimental variation, we can check that the model's predictions provide a reasonable approximation to the experimental findings, which indeed they do. We can then use the calibrated model to understand which factors are driving the lack of a differential targeting impact from increasing ordeals.

We use the estimated version of the model to see which of the various mechanisms we outlined in the theoretical section lie behind the fact that marginal increases in the extent of the ordeal do

not seem to differentially improve selection. Simulations from the estimated model suggests that, of the theoretical mechanisms we outline, neither curvature of the utility function nor differential travel technology is driving the results – we show that the data in fact is best explained by a linear utilty function, rather than one with curvature, and we obtain virtually identical predictions when we impose identical travel technologies for both poor and rich. Instead, we show that the estimates imply that there is substantial heterogeneity in utility shocks, which dampens the differential responsiveness to changing ordeals. Although the model as estimated matches the empirical findings, with little differential selection obtained by increasing distance, we show that, if we reduced the standard deviation of the shocks by half, we would have obtained detectable differential selection from our increased distance treatment. From a practical perspective, we find that, with the actual estimated shock distribution, differential selection only occurs in our counterfactuals when we triple wait time at the registration site. This would mean that prospective applicants would need to have waited in line 9 or more hours to be interviewed, which is beyond what appears to be feasible as a policy.[3]

The remainder of the paper is organized as follows. Section 2 discusses the setting, experimental design, and data. Section 3 introduces our model which revisits the standard screening model in light of curvature in the utility function, differential access ways of dealing with costs, and idiosyncratic shocks. Section 4 examines the self-targeting data to ask who chooses to apply for the program. Section 5 uses the experiment to compare self-targeting with the status quo PMT-based approach. Section 6 examines the marginal effect on targeting when the ordeal is changed experimentally, and compares this to what a structurally estimated version of our model would predict. Section 7 concludes.

2. Setting and Experimental Design

2.1. **Setting: The PKH Program.** This project explores self-targeting mechanisms within the context of Program Keluarga Harapan (PKH), a conditional cash transfer project administered by the Ministry of Social Affairs (Depsos) in Indonesia. The program targets households with that have per-capita consumption below 80 percent of the poverty line (approximately the poorest 5 percent of the population we study) and that meet a demographic requirements of having a pregnant women, a child between the ages of 0 to 5, or children below the age of 18 years old that have not finished the nine years of compulsory education. Program beneficiaries receive direct cash assistance ranging from Rp 600,000 to Rp 2.2 million (US$67-US$250) per year—or about 3.5-13 percent of the average yearly consumption of very poor households in our sample—depending on their family composition, school attendance, pre/postnatal check-ups, and completed vaccinations.[4]

---

[3]Interestingly, in our pre-pilot we explicitly piloted treatments aimed at increasing the wait time, with wait times as long as 8 or more hours. Even at wait times well below the level our simulations suggest would be necessary to induce substantial targeting effects, villagers endogenously organized themselves to reduce wait times (e.g. by pre-assigning scheduled times for people to come back to be interviewed). This suggests that actually implementing a policy that requires waiting in line for more than 8 hours may be quite difficult to implement practically.

[4]Note, however, that although PKH is formally a conditional cash transfer program, which conditions transfers on health takeup and school enrollment, these conditions are typically not enforced in practice.

The payments are disbursed quarterly for up to six years. Currently, around 1.12 million households are enrolled in the program.[5]

Determining whether households fall below the consumption requirement ("targeting") is difficult since per-capita consumption is not easily observed by the government. Instead, PKH uses a proxy means-test (PMT) approach with automatic enrollment for all households that meet the demographic requirements and that pass a proxy means test. Specifically, every three years, enumerators from the Central Statistical Bureau (BPS) conduct a survey of households nationwide who are potentially eligible for anti-poverty programs, including but not limited to PKH. They survey all households that were included on 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 the households inhabit. After passing an initial pre-screening, each household is asked a series of about 30 questions, including attributes of their home (wall type, roof type, etc.), ownership of specific assets (motorcycle, refrigerator, etc.), household composition, and the household head's education and occupation. 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 levels based on the responses to the survey. Individuals with predicted consumption levels below each district's very poor line were eligible for the program.

2.2. **Sample Selection.** This project was carried out during the 2011 expansion of PKH to new areas. We chose 6 districts (2 each in the provinces of Lampung, South Sumatra, and Central Java) from the expansion areas to include a wide variety of different cultural and economic environments. Within these districts, we randomly selected a total of 400 villages, stratified such that the final sample consists of approximately 30 percent urban and 70 percent rural locations.[6] Within each village, we randomly selected one hamlet to be surveyed.[7] 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.

2.3. **Experimental Design.** We randomly allocated each of the 400 villages to one of two targeting methodologies: self-targeting or an automatic enrollment system, i.e. the status quo.[8]

2.3.1. *Automatic Enrollment Treatment .* In Indonesia, the automatic enrollment treatment is the status quo, and the procedure discussed in Section 2.1 was followed. Due to cost considerations, for this treatment, the automatic enrollment was only conducted in the one randomly selected

---

[5]Program PKH Bidik 1,12 Juta Rumah Tangga Miskin. Kementrian Koordinator Bidang Kesejahteraan Rakyat. October 22, 2010. Retrieved from <http://www.menkokesra.go.id/content/program-pkh-bidik-112-juta-rumah-tangga-miskin-0>, last accessed October 3, 2011.

[6]The sampling unit is a *desa* in rural areas and a *kelurahan* in urban areas. For ease of exposition, we henceforth refer to both as villages.

[7]Both desa and kelurahan are administratively divided by neighborhood into sub-villages known variously as *dusun*, *RW*, or *RT*. For ease of exposition, we henceforth call them "hamlets." In rural areas, a hamlet ranges from about 30-330 households each, while in urban areas, they range from 70-410 households each.

[8]We also randomly assigned an additional 200 villages to a "hybrid treatment" (see Alatas, Banerjee, Hanna, Olken, Purnamasari and Wai-poi (2012)).

hamlet per village that we also surveyed in the baseline.[9] For each hamlet in this treatment, the government Bureau of Statistics (BPS) enumerators were given a pre-printed list of households from the last targeting survey (PPLS 2008). When they arrived at a village, the enumerators showed the list to the village leadership and asked them to add any households to the list 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 potential interviewee, the enumerator conducted an initial five question pre-screening; those households who passed the pre-screening were given the full PMT survey.[10] Of the 6,406 households on the potential interviewee list, 16 percent were eliminated based on the initial screen, and 5,383 households (or about 37.8 percent of the hamlet) were given the full-PMT survey of 28 questions. For each household that was interviewed, a computer generated poverty score was generated using the district-specific PMT formulas.[11] A list of beneficiaries was generated by selecting all households with a predicted score below the score-cutoff for their district.

2.3.2. *Self-Targeting Treatment.* The enrollment criteria for both the demographic and consumption criteria under the self-targeting mechanism was the same as in automatic enrollment, but households were required to apply at a central registration station if they were interested in the program. The fact that households needed to self-select means that some households who might have been automatically enrolled would not receive benefits because they chose not to apply. Conversely, some households who may have been forgotten or passed over when the government compiled the list of households to be interviewed could apply and ultimately receive benefits.

The self-targeting treatment proceeded as follows: To publicize the application process, a community facilitator from a local NGO (Mitra Samya) visited each village to inform the village leaders about the program, to brainstorm about the best indicators of local poverty with the leaders, and to set a date for a series of hamlet-level meetings that were aimed at the poor.[12] In these hamlet-level meetings, the facilitator described the PKH program and explained the registration process. In particular, the facilitators stressed that the program was geared towards the very poor. For example, they listed examples of questions that would be asked during the interview (type of house, motorbike, etc), informed households that there would be a verification stage post-interview, and highlighted a set of local poverty criteria (the criteria the locals would typically use to characterize very poor households). The goal was to ensure that the households understood their chances of obtaining PKH conditional on showing up to be interviewed.

---

[9]To select beneficiaries in the other hamlets, the government used the 2008 automatic enrollment survey.

[10]The pre-screening consists of 5 questions: is the household's average income per month in the past three months more than IDR 1,000,000 (USD 110); was the average transfer received per month in the past three months more than IDR 1,000,000 (USD 110); did they own a TV or refrigerator that cost more than IDR 1,000,000 (USD 110); was the value of their livestock productive building, and large agricultural tools owned more IDR 1,500,000 (USD 167); did they own a motor vehicle; and did they own jewelry worth more than IDR 1,000,000 (USD 100). Households that answered yes on either four or five of the questions were instantly disqualified and the survey ended.

[11]The PMT formulas were determined using household survey data from SUSENAS (2010) and village survey data from PODES (2008)). On average, these regressions had an R2 of 0.52. The questions chosen for the PMT survey were those that the government were considering for the next nation-wide targeting survey (PPLS 11).

[12]The local poverty indicators generated in the meeting were not used for targeting, but were instead used by community members in the socialization process to help villagers understand how the PMT screening would operate.

Registration days for each area were scheduled in advance based on the number of applicants their relative sizes of the hamlet.[13] During the registration days, the BPS enumerators were present at the registration station from 8AM to 5PM. Households were required to come to the registration site if they wished to apply. Once they arrived, they were signed in and given a number in the queue. When their number was called, BPS interviewed the households to collect the same data that was conducted in the PMT interview.

Households who applied were subsequently categorized by eligibility based on the PMT regression formula and the district-specific very poor line, using the same PMT formula and questions as in the automatic enrollment treatment. Any household that was both classified as very poor based on the assets they disclosed in their interview and was also listed in the 2008 poverty census as very poor (about 37 percent who passed the interview at the registration site) were selected as a PKH recipient. All other households that classified as very poor based on their interview were subjected to a verification process: Government surveyors went to their homes to collect data on the same set of asset questions. The results of this home-based survey were used, with the same PMT regression formula and poverty lines, to determine final list of beneficiaries. About 68 percent of those who got to the verification stage were considered eligible.

Within self-targeting treatment villages, we varied how the self-targeting was conducted in order to vary the costs of registration. Specifically, we conducted two sub-treatments:

(1) *Distance sub-treatment*: We experimentally varied the distance to the registration site. 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 the urban areas, we randomly allocated villages to have the registration site at the sub-district office (far location) or the village office (near location). In rural areas, where distances are greater than the urban areas, villages were randomly allocated to have the registration site at the village office (far location) or in the sub-village (near location).[14]

(2) *Both spouse sub-treatment*: We experimentally varied whether one or two household members were required to come to the registration site. In half the self-targeting treatment villages, households were told that any adult in the household could do so. Given that the program was geared towards women, we expected that mostly women would apply. In the other half of the villages, we required that both the husband and the wife jointly apply at the registration site. Note that there was a fear of screening out poor households where the primary wage earner had migrated for work. Thus, if the spouse was unable to attend due to a pre-specified reasons (illness, out of village for work), the household was required to bring a letter signed by the hamlet head providing the reasons

---

[13]Specifically, we estimated the predicted number of people who would show up to be interviewed using the pilot data. We regressed the number of people who showed up on the number of households in the village and the number of poor households. BPS staff were assigned based on these predicted showup rates, assuming a capacity to interview 34 households per day and a 25 percent buffer.

[14]The distance sub-treatment was violated in four villages: in the first village, a longstanding ethnic tension caused a large subset of the village to refuse participating in interviews in a certain hamlet, so the interviewers held interview for a day in another hamlet; in the second village, a hamlet was a 4-5 hour walk away from the village office, so the interviewers set aside a day to go to that hamlet; in the third village and fourth villages, the village leaders insisted the the registration site be moved closer to the village. All analysis reports intent-to-treat effects where these four villages are categorized based on the randomization result, not actual implementation.

for the spouses' unavailability, the rationale being that obtaining the letter in advance would still be costly to households. On average, 29% of applicants in such villages provided such a letter.

2.4. **Randomization Design and Timing.** We randomly assigned each of the 400 villages to the treatments (see Table 1). In order to ensure experimental balance across geographic regions, we stratified by 58 geographic strata, where each stratum consists of all of the villages from one or more sub-districts and is entirely located in a single district. We then randomly and independently allocated each self-targeting village to the sub-treatments, with each of these two sub-treatment randomizations stratified by the previously defined strata and the main treatment.

From December 2010-March 2011, an independent survey firm (Survey Meter) collected the baseline data from one randomly selected hamlet in each village. After surveying was completed in each sub-district, the government conducted the targeting treatments. The targeting treatments thus occurred from January-April 2011.[15] SurveyMeter conducted a first follow-up survey in early August 201l, after the targeting was complete, but before the beneficiary lists were announced to the villages. Fund distribution occurred starting in late August 2011.[16] Finally, we conducted a second endline survey in January 2012 to March 2012, after two fund distributions had occurred.

2.5. **Data, Summary Statistics and Balance Test.**

2.5.1. *Data Collection.* We collected three main sources of data:

Baseline Data: The baseline survey was completed in each sub-district before any targeting occurred, and up to this point, there was no mention of the experiment in the villages. Within each village, we randomly selected one hamlet, and within that hamlet, we randomly sampled nine households from the set of those who met the demographic eligibility requirements for PKH, as well as the sub-village head, for a total of 3,997 households across the 400 villages. The survey included detailed questions on the households' consumption level, demographics, and family networks in the villages. We also collected data for all of the variables that enter the PMT formula, so that we can 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 (either at interview site or 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 registration meetings occurred, how the socialization was done in each village, etc.).

---

[15]There was no mention of the targeting process until Survey Meter 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.

[16]Note that after the experiment selected beneficiaries, the Department of Social Affairs realized it had additional funds available and decided to increase the number of people who received the program to also include households that did not pass the selection process in our experimental treatments but had been classified as very poor under the 2008 poverty census. In calculating "beneficiaries" for purposes of experimental evaluation below, we do not include these additional households, but these extra households are important to keep in mind when evaluating the program at endline.

Endline Surveys: We administered two endline surveys, both of which were conducted by SurveyMeter. The first endline survey 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 enrollment villages and 193 self-targeting villages, for a total sample of 1,045 households.[17] In this survey, we collected detailed data on the households' consumption level, as well as respondents' experience and satisfaction with the targeting process (e.g., whether they applied, how long they waited to be interviewed). In addition, for all beneficiary households, we collected additional data on demographics, family networks, relationships with local leaders, and employment. We conducted the second endline in in January 2012 to March 2012, after two rounds of PKH fund distribution. In this survey, we revisited all ten of the baseline households, collecting consumption data, as well as satisfaction with PKH.

2.5.2. *Summary Statistics and Experimental Validity.* Table 2 shows the flow of surveyed 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 who applied to be interviewed for self targeting (754 out of 2,000 or 38 percent) or were interviewed as part of the automatic enrollment treatment (706 out of 1998; or 35 percent). Column 3 shows the the number of baseline households who were ultimately chosen as beneficiaries (73 out of 2,000, or 3.65 percent, in self-targeting; 86 out of 1998, or 4.3 percent, in automatic enrollment).

Appendix Table A.1 presents summary statistics and a check on the the experimental validity using data from the baseline survey and a village census. Note that we chose all of the variables for the table prior to analyzing the data. Column 1 presents the mean and standard deviations of each variable in villages in the automatic enrollment treatment, while this information is provided for the self-targeting villages in Column 2. Column 3 shows the difference (with associated standard errors). Column 4 shows this difference after controlling for stratum fixed effects. Only 1 of the 20 differences presented is statistically significant (at the 10 percent level), confirming that the treatment villages are balanced at the baseline. At the bottom of Columns 3 and 4, we provide the p-value from a joint test of the treatment across all baseline characteristics that we consider. The p-values of 0.99 and 0.67, respectively, confirm that the groups are balanced in the baseline.

3. Model

3.1. **Model Set-up.** In this section, we briefly re-examine self-selection into a welfare program based on the expected benefits and costs of applying. We assume that households have a utility function $U(x)$, where x is current consumption. Households vary in their per period labor income, denoted by y, but for a given household this is the same number in both periods. The application cost is denoted by $c(l, y)$, where l is the distance to the registration site. Conditional on applying, households have a probability $\mu(y)$ of passing the asset-based test and actually qualifying for the program ($\mu'(y) \leq 0$). 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

---

[17]Due to safety and travel concerns that were independent of the project, the survey company asked that that we did not return to 10 villages in endline 1 and 13 villages in endline 2. These were spread among treatment and control villages.

no additional income. Finally, assume that the household starts with no assets and cannot borrow. This is consistent with the evidence that many poor, and perhaps even not so poor, households in developing countries tend to be credit constrained. This, combined with the assumption that the household discounts future utilities (the discount factor is $\delta < 1$), and the fact that in our model future consumption is always weakly higher, rules out savings, and tells us that consumption in a given period is just current income net of costs.

To complete the description of the model, assume that each person who applies receives a random utility shock, ε, that encourages him to go to register, and $F(\epsilon)$ is the distribution of ϵ.

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

$$U\left(y - c\left(l, y\right)\right) + \mu(y)\delta U\left(y + b\right) + \left(1 - \mu(y)\right)\delta U\left(y\right) + \varepsilon \tag{1}$$

If the household does not apply, expected utility is:

$$U\left(y\right) + \delta U\left(y\right) \tag{2}$$

The expected gain from applying is the difference, i.e.

$$U\left(y - c\left(l, y\right)\right) - U\left(y\right) + \mu(y)\delta\left[U\left(y + b\right) - U\left(y\right)\right] + \epsilon \tag{3}$$

It will turn out to be convenient to define:

$$g(y, l) = U\left(y - c\left(l, y\right)\right) - U\left(y\right) + \mu(y)\delta\left[U\left(y + b\right) - U\left(y\right)\right] \tag{4}$$

to denote the net gain without the shock The household will apply if the expected utility from doing so is larger than staying home, i.e. if $-g\left(y, l\right) \leq \epsilon$. The fraction of households that will apply at a particular level of income y is given by $F(-g(y, l))$. We are interested in how an increase in how a change in distance, l, affects $F(-g(y, l))$ at different levels of y.

3.2. **Analysis.** In this section, we will start with the most basic model and add elements to the model one-by-one in order to understand how each element affects the type of household that applies.

3.2.1. *The Benchmark Case.* Suppose that the utility function is linear ($U(x) = x$) and that the time cost applying is also linear in distance: τl.[18] For someone who earns a wage w, this imposes a monetary cost of $\tau l w$. If we assume that wages are proportional to income, $w = \alpha y$, then the monetary application cost can be written as $\tau l \alpha y$. Assume also that there are no shocks ($\epsilon \equiv 0$). In this case, $g\left(y\right)$ simplifies substantially, and a household applies if:

$$- \tau l \alpha y + \delta \mu(y) b \geq 0. \tag{5}$$

Since the left hand side of this expression is decreasing in y, this expression defines a cutoff value y^* such that those who have incomes less than y^* apply and those who have incomes greater than y^* do not apply. Moreover, an inspection of equation (5) shows that $\frac{\partial y^*}{\partial l} < 0$, that is, making the ordeal more onerous increases the degree of selection and means that the set of people who apply

---

[18]The linearity in time cost may be unrealistic since it includes both travel time and wait time, which are unlikely to be linear in distance (though it may be increasing in distance since the further it is the harder it is go home and come back later if the wait time is particularly long). However, nothing really turns on it and it simplifies the model.

will be poorer. This simple expression captures the basic intuition for using ordeal mechanisms for selection captured by Nichols and Zeckhauser (1982).

3.2.2. *Adding shocks.* Now, let's consider what happens if we re-introduce the utility shock term. In this case, a household applies iff:

$$\tau l\alpha y - \delta\mu(y)b \leq \varepsilon. \tag{6}$$

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

$$\frac{1 - F(\tau l\alpha y_1 - \delta\mu(y_1)b)}{1 - F(\tau l\alpha y_2 - \delta\mu(y_2)b)} \tag{7}$$

This ratio is always greater than one because the rich are less likely to sign up since their costs are higher and since their probability of getting the benefit is lower. Note that this ratio is a measure of how well targeted the application process is towards poorer individuals – the higher the ratio, the higher the fraction of the poor in the population of applicants. Making the ordeal tougher reduces the number of poor applicants and imposes dead-cost on everyone who applies, 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 costs of the program to the government.

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:

$$\frac{f(\tau l\alpha y_2 - \delta\mu(y_2)b)}{1 - F(\tau l\alpha y_2 - \delta\mu(y_2)b)}\tau\alpha y_2 - \frac{f(\tau l\alpha y_1 - \delta\mu(y_1)b)}{1 - F(\tau l\alpha y_1 - \delta\mu(y_1)b)}\tau\alpha y_1 > 0. \tag{8}$$

This formula says that 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 PDF $f(y)$ – to the number of people who are inframarginal, given by the $1 - F(y)$ term.

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

$$\frac{f(\tau l\alpha y - \delta\mu(y)b)}{1 - F(\tau l\alpha y - \delta\mu(y)b)}, \tag{9}$$

is weakly increasing with y, since if this is true then clearly $\frac{f(\tau l\alpha y - \delta\mu(y)b)}{1 - F(\tau l\alpha y - \delta\mu(y)b)}\tau\alpha y$ is increasing in y. This property holds if $F(\epsilon)$ 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 $F(\epsilon) = \frac{\epsilon^\beta}{c^\beta + \epsilon^\beta}$ where c and β are two known positive parameters and $\epsilon \geq 0$, exhibits declining hazard rates as long as $\beta \leq 1$, but not otherwise.

To gain intuition into the model, we provide a simple illustration. In Figure 1, we examine the simplest case of no shocks and linear utility. In Panel A, we draw the relationship between income and gains for registration sites that are closer versus farther away. Note that the gain is decreasing more steeply with income for higher distance; this is the standard single-crossing property common to all screening models. As Figure B shows, moving from lower to higher distance reduces the number of applicants, but only among the rich. Thus, targeting efficiency improves.

Figure 2 shows an example of how introducing shocks can overturn the benchmark intuition developed in Section 3.2.1 above. We consider a simple case where income $y \in [0, 5]$, we set $\tau\alpha = 0.2$ and $\delta\mu(y)b = 0.5$, choose the log-logistic parameters $\beta = c = 0.5$, and consider distances $l \in \{2, 3\}$. As shown in Panel A, at any given consumption level, showup rates are of course still higher at lower distances, and for any distance level, showup rates decline in income. What is important however to note however is that in this example, the initial rate of decline in showup rate with income (once the epsilons kick in) is quite high, but then slows as incomes become high. This is a consequence of the thick tails of the log-logistic distribution, which implies that $\frac{f(y)}{1-F(y)}$ is decreasing in y. This implies that increasing distance from 2 to 3 actually hurts the ratio of of poor to rich showup rates, because it has a very large impact on the takeup at low income levels (where $\frac{f(y)}{1-F(y)}$ is large) but a much smaller impact at high income levels (where $\frac{f(y)}{1-F(y)}$ is small).

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 who hence will be affected by any marginal change in ordeals.

3.2.3. *Non-linearities in the Application Cost.* Let us continue to assume linear utility, but now model a non-linearity in the cost of applying, $c(l, y)$. This non-linearity 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 $\tau l w$, 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 $\lambda l w$, where $\lambda < \tau$. Once again, λl includes waiting time. Assuming that the wage is proportional to income, $w = \alpha y$, the decision rule is:

$$D = \begin{cases} \text{bus} & \text{if } \nu + \lambda l \alpha y < \gamma l + \tau l \alpha y \\ \text{walk} & \text{if } \nu + \lambda l \alpha y \geq \gamma l + \tau l \alpha y \end{cases} \tag{10}$$

Applying is optimal if and only if:

$$- min\{\gamma l + \tau l \alpha y, \nu + \lambda l \alpha y\} + \delta\mu(y)b \geq ln\varepsilon. \tag{11}$$

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

To look at 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 $\gamma l + \tau l \alpha y_1$. For those at y_2, it is $\nu + \lambda l \alpha y_2$. The fall in utility due to an increase in distance of Δl will be greater at y_1 than y_2: $(\gamma + \tau\alpha y_1)\Delta l > (\lambda\alpha y_2)\Delta l$. Therefore, an increase in distance can increase travel costs more for the poor than for the rich.

To see this intuitively, consider the simple illustration in Figure 3. For both the rich and poor, taking the bus is initially more expensive (i.e. no bus fare), but has a lower marginal cost. Due to the higher marginal cost of their time, the rich switch to buses at lower distance than the poor (l^*). Between l^* and l^{**} (where the poor switch to the buses), one can clearly see from the figure

that the marginal travel cost when l is increased is actually *larger* for the poor than the rich. As a result, even in the case where $F(.)$ has increasing hazard rates, targeting efficiency may worsen. Note from the figure that this cannot happen if both people walk or both take the bus (i.e. travel costs are locally linear), or if the difference in incomes between them is large enough.

3.2.4. *Curvature in the Utility Function.* Finally, we introduce curvature into the utility function by letting $U(x) = lnx$. To focus on one mechanism, assume that there is no utility shock ($\epsilon \equiv 0$), that the cost of travel is linear in distance ($c(l,y) = \gamma l + \tau l\alpha y)$) and that $\mu(y)$ is a constant. In this case, the net gain from applying is:

$$
\begin{aligned}
g(y,l) &= \ln(y - c(l,y)) + \mu\delta \ln(y+b) + (1-\mu)\delta \ln y - \ln y - \delta \ln y &\qquad (12)\\
&= \ln \frac{(y - c(l,y))(y+b)^{\mu\delta} y^{(1-\mu)\delta}}{yy^\delta} &\qquad (13)
\end{aligned}
$$

The household will apply when:

$$
\frac{(y - c(l,y))(y+b)^{\mu\delta}}{yy^{\mu\delta}} \geq 1 \qquad (14)
$$

For convenience, we will work with the following function:[19]

$$
G(y,l) = \frac{(y - c(l,y))(y+b)^{\mu\delta}}{yy^{\mu\delta}}. \qquad (15)
$$

There exists a y^{min} such that $y^{min} - c\left(l, y^{min}\right) = y^{min} - \gamma l - \tau l\alpha y^{min} = 0$. Let's start the discussion at this value of y because any y below this does not make sense in our model. At just above this level of y, $\frac{y - c(l,y)}{y}$ is close to zero and as a result g must be less than one, so those with income levels in this range will not apply. As y increases, G also increases, since it starts at zero and thus can only go up). Taking the derivative of G with respect to y yields:

$$
\frac{dG}{dy} = \frac{\gamma l}{y^2}\left(1 + \frac{b}{y}\right)^{\mu\delta} - \frac{\mu\delta b}{y^2}\left(1 - \tau l\alpha - \frac{\gamma l}{y}\right)\left(1 + \frac{b}{y}\right)^{\mu\delta-1} \qquad (16)
$$

$$
= \frac{\left(1 + \frac{b}{y}\right)^{\mu\delta-1}}{y^2}\left[\gamma l\left(1 + \frac{b}{y}\right) - \mu\delta b\left(1 - \tau l\alpha - \frac{\gamma l}{y}\right)\right] \qquad (17)
$$

In the neighborhood of $y = y^{min}$, the expression in the square brackets is strictly positive. However, the expression in the square bracket goes down when y goes up and converges to $\gamma l + \tau l\alpha\mu\delta b - \mu\delta b$. If this expression is positive, then G is monotone increasing in y while if it is negative then first goes up and then goes down.

Figure 4 represents the two possible configurations of G in this case. Panel A provides the case where G first increases and then falls, while Panel B represents the case where G is monotonically increasing. In each case, the values of y for which the G curve lies above the horizontal line at $G = 1$, are those that apply. The dashed line in each figure demonstrates what happens when l goes up. In both cases the G curve shifts down – in Figure 4a this means that both the poorest

---

[19]So g, defined above, is lnG.

and richest people who were applying before the increase in l drop out, while in Figure 4b only the poorest people drop out. In the first case, the effect on targeting depends on whether more of the poor proportionally drop out than the rich, which in turn depends on how the population is distributed near the two cutoffs. In the second case, the effect is unambiguously negative, with the fraction of the rich among applicants increases when l goes up.

It is worth noting that so far in this discussion we suppressed the effect of y on $\mu(y)$ which goes in the direction of making the G function downward sloping. In particular, if there exists a y^{max} such that for $y \geq y^{max}, \mu(y) \approx 0$, as seems reasonable, then above y^{max}, $G < 1$ and no one will apply. The more realistic case is therefore probably the case in Figure 4a, and the effect of an increase in l on targeting will depend on the shape of the income distribution.

3.3. **Summary.** This exercise illustrates the complexities in designing ordeal mechanisms: once we introduce a number of realistic features into the model, such as utility shocks that may have thick-tailed distributions, alternative means of transportation, and diminishing marginal utility, the intuitive argument that ordeals induce self-selection because the poor have a lower opportunity cost of time is no longer automatically true. Increasing the costs of the ordeal can worsen self-selection under relatively standard assumptions (log utility, as we saw above, for example is enough). Note that we have not yet even introduced the more behavioral arguments for why the poor may not be able to access the programs that are intended for them, such as self-control problems (e.g., Madrian and Shea, 2001), stigma (e.g., Moffitt, 1983), as well as informational arguments, such as the fact that the poor may not learn about the programs that are available to them (e.g., Daponte, Sanders and Taylor, 1999).

Given the theoretical ambiguity, whether self-targeting improves targeting efficiency is ultimately an empirical question. Therefore, we now turn to the data.

4. Who Self-Selects?

We begin by examining whether richer or poorer households were more likely to apply for the PKH program in the 200 villages where the government implemented the self-targeting treatment. Specifically, we plot a non-parametric Fan (1992) regression of the probability of applying against baseline log per capita consumption (Figure 5). Note, again, that the consumption data was collected before any mention of targeting occurred.[20] Bootstrapped standard error bands, clustered at the village level, are shown in dashes.

Across all expenditure ranges, Figure 5 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. For example, 61 percent of households at the 5th percentile of the consumption distribution do so. The share applying falls rapidly as consumption increases: at the middle of the expenditure distribution, only

---

[20]Consumption may, of course, not be a perfect measure of welfare. First, there may be measurement error in consumption. Second, there may be alternative measures of welfare that may or may not more accurately represent a household's well-being (see Alatas, Banerjee, Hanna, Olken and Tobias 2012). We use consumption because this is often the metric that government's are trying to actually target on. Note that these measurement errors will not affect our experimental results if the variation in consumption captures relative well-being; the measurement error will simply introduce noise into our estimate.

39 percent percent of households apply, and by the 75th percentile, only 21 percent do so. At the 95th percentile of per-capita expenditure, only 13 percent of households apply.

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: 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 save the government resources since it would reduce the number of interviews that they 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 enrollment.[21] Second, there could be selection on the unobservable component of consumption: that is, conditional on a household's PMT score, households with higher unobservable consumption might also be less likely to attend. This could arise if there is self-selection based on the opportunity cost of time (as in the model), 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 enrollment.

To investigate this, we can decompose household consumption into the observable and unobservable components:

$$LNPCE_i = X_i'\beta + \epsilon_i \tag{18}$$

where $LNPCE_i$ is the household's log per capita consumption, X_i are the observable characteristics that enter the PMT formula, β are the PMT weights, and ϵ_i is the residual, or the unobserved component of consumption. We then examine the relationship between the probability of applying and both the observable component, $X_i'\beta$ and the unobservable component, ϵ_i.

We first examine these relationships graphically, presenting non-parametric Fan regressions of the probability of showing up as a function of the observable (Figure 6, Panel A) and unobservable (Panel B) components of log per-capita consumption. Bootstrapped standard 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. 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 3 provides the results from estimating the following logit equation:

$$Prob\left(SHOWUP_i = 1\right) = \frac{exp\left\{\alpha + \gamma PMT_i + \gamma \epsilon_i\right\}}{1 + exp\left\{\alpha + \gamma PMT_i + \gamma \epsilon_i\right\}} \tag{19}$$

where PMT_i is the predicted portion of a household's log per-capita consumption (equal to $X_1'\beta$ from equation (18)) and ϵ_i is the residual portion of a household's log per-capita consumption from equation (18). 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

---

[21]In reality, it is often too costly to interview everyone in the country, so most governments do some form of selection to reduce the number of people interviewed. In our experimental results, we compare self-targeting to another methodology that the government uses to cull the number of interviews (the current status quo for Indonesia). We will then compare the efficiency of self-targeting to that of a hypothetical, full census PMT, to explore this dimension further.

interpret. We show in Table A.2 that the results are qualitatively similar if we use linear probability models instead. Finally, note that all standard errors are clustered by village.

Table 3 confirms the graphical analysis and shows that there is self-selection along both margins, and that both of these forms of selection occur within both poor and richer households. Column (1) provides the coefficient estimates for the full sample. Both the observable and unobservable component of consumption significantly predict applying at the 1 percent level. The relative magnitudes suggest that the observed component of 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 based on whether the household's true consumption is above or below the "very poor" consumption line (i.e. 80 percent of the poverty line for the district). What is notable is that selection on unobservables occurs in both samples. Thus, even among the poorest 4 percent of households in our sample, those who are poorer on unobservables are more likely to apply. This strong selection on unobservables suggests that self-selection has the potential to result in a dramatically poorer distribution of beneficiaries than other methods.

5. Comparing Self-Selection and Automatic Enrollment

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.Instead, the government's choice is often to conduct self-targeting or to conduct an alternative targeting methodology.[22] Therefore, in Section 5.1, we compare self-targeting against the real government procedure, which consists of an automatic enrollment for those who pass a proxy means test among those selected by the government and local communities to be interviewed. Next, in Section 5.2, we additionally compare self-selection against a hypothetical exercise where we use the data that we have collected independently to predict selection if the automatic enrollment based on the proxy-means test was implemented universally.

5.1. **Experimental Comparison of Self-Targeting with Status Quo Targeting.** In this section, we test whether the types of individuals selected under self targeting and automatic enrollment (the current status quo procedure of the Indonesian government) differ. To do so, 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 enrollment treatment. Given the randomization, the distribution of beneficiaries and the probability of receiving benefits should be identical in the two sets of villages absent the difference in targeting, so we can ascribe the differences that we observe between the two sets of villages to the differences in targeting methodologies (see Appendix Table A.1).

---

[22]Unlike asset data, which is verifiable in an in-person interview, consumption data is completely unverifiable since it is all self-reported, so even if the government could afford to do a consumption survey for all households, it could not use such data for targeting purposes since doing so would induce people to understate their true incomes.

We begin with a graphical analysis in which we compare the distribution of beneficiaries under the self-targeting and automatic enrollments treatments (Figure 7). In Panel A, we plot the cumulative distribution function of log per-capita consumption of the final PKH beneficiaries in both sets of villages. The beneficiaries appear substantially poorer: the CDF of beneficiaries' consumption under automatic enrollment first-order stochastically dominates that under selection. A Kolmogorov-Smirnov test of equality of distributions yields a p-value of 0.103.[23]

While the results in Panel A imply that the distribution of beneficiaries are poorer under self-selection, it does not tell the full story. In particular, 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. To answer this question, we present non-parametric Fan regressions of the probability of obtaining benefits as a function of log per-capita consumption in Panel B of Figure 7. Bootstrapped 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, i.e. those with log per-capita consumption below about 12.33, the probability of receiving benefits is more than double in self targeting: 16 percent of those with log per capita consumption in the bottom 5 percent receive benefits as compared with just 7 percent in the automatic enrollment treatment. This difference is statistically significant at the 10 percent level.[24] While exclusion error is still very high – even in self-targeting, only 16 percent of households of these very poor households received benefits, meaning that 84 percent were excluded – the rate of receiving benefits is 4 times higher than the overall rate of 4 percent of households in the sample who receive benefits, and double what it is in the status quo automatic enrollment villages.

Conversely, households at higher consumption levels are substantially more likely to receive benefits in the automatic enrollment treatment. Households in the top 50 percent of the per capita expenditure distribution – none of whom should be receiving benefits – are more than twice as likely to receive benefits in automatic enrollment than in the self-targeting treatment: 2.5 percent of such households receive benefits in automatic enrollment 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 proxy means test. With self-targeting, however, most of these households do not apply, so many fewer of them slip through. In sum, Figure B suggests that self-targeting both increased the probability that very poor households received benefits and decreased the probability that richer households did so relative to the current status quo.

We now more formally quantify these effects using regression analysis, the results of which are presented in Table 4. In Column (1), we compare the difference in average log per-capita consumption of the beneficiary populations ($LNPCE_{vi}$) in the two treatments, by estimating by OLS:

$$LNPCE_{vi} = \alpha + \beta SELF_v + \vartheta_{vi} \tag{20}$$

_____

[23]This p-value is based on randomization inference methods accounting for clustering at the village level. Alternatively, abstracting from the village-level clustering yields an exact p-value of 0.06.

[24]The p-value of this difference is 0.056 in the logit model and .08 in OLS.

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). Note that this is the regression equivalent of comparing the means of the two distributions shown in Panel A of Figure 7. As suggested by the figures 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 is 21 percent lower in self-targeting as compared to automatic enrollment (significant at the 1 percent level). Including stratum fixed effects (Panel B), the difference becomes 12 percent, and the p-value increases to 0.13.[25]

In Column 2 of Table 4, we examine the probability of getting benefits ($Prob\left(BENEFIT_{vi}=1\right)$) across the treatments for different groups. Specifically, we provide estimates from the following logit model:

$$Prob\left(BENEFIT_{vi}=1\right)=\frac{exp\left\{\alpha+\beta SELF_v+\gamma LNPCE_{vi}+\eta SELF_v\times LNPCE_{vi}\right\}}{1+exp\left\{\alpha+\beta SELF_v+\gamma LNPCE_{vi}+\eta SELF_v\times LNPCE_{vi}\right\}} \quad (21)$$

The coefficient of inters is the coefficient η on $SELF_v\times LNPCE_{vi}$, which captures the degree to which there is differential targeting in the self-targeting treatment as compared with automatic enrollment (the omitted category).[26] The results confirm the overall story shown in Panel B of Figure 7: the coefficient on η is negative, large in magnitude, and statistically significant. This implies that there is much stronger targeting by consumption in the self-targeting treatment than in the automatic enrollment treatment. The magnitudes suggest that targeting is twice as strong in self-targeting: the estimates in Panel A imply that doubling consumption decreases the log-odds of receiving benefits by 0.69 in automatic enrollment, whereas it decreases the log-odds of receiving benefits by 1.37 in self-targeting.

In Columns (3) - (5), we examine alternative dependent variables to quantify the types of inclusion and exclusion error shown in Panel B of Figure 7. In Column (3) we define the overall error rate as a dummy that is equal to 1 if either exclusion error (failing to give benefits to a very poor household) or inclusion error (giving benefits to a non-very poor household) takes place. We find that the log-odds ratio of making an error is about 0.2 lower under self-targeting (the coefficients are approximately the same with and without stratum fixed effects, but it is only significant at conventional levels with the stratum FE). Column (4) examines exclusion error, defined as a dummy for a very poor households failing to receive benefits. The results in the table suggest that the log-odds of such households being excluded (i.e. failing to get benefits) are between 0.55 and 0.85 lower in self-selection, with the latter result statistically significant in the specification with stratum fixed effects. Likewise, inclusion error, defined as a non-very poor household who does receive benefits,

---

[25]In general one would expect stratum fixed effects to improve precision. However, in the regressions where we only consider beneficiaries, we have so few observations (159 observations), and hence so few observations per stratum, that including the fixed effects effectively drops many whole strata from the analysis, dramatically diminishing statistical power.

[26]We use logit models because the baseline benefit rate differs substantially by per-capita expenditure, so proportional models make more sense. Stratum fixed effects are also much more effective in proportional models given the substantially different poverty levels across strata. Appendix Table (A.3) shows that the OLS version of the same results are qualitatively similar. We cluster the standard errors in models with no fixed effects by village. For the conditional logit models where we include stratum fixed effects, for computational reasons we cluster fixed effects by stratum, which is more conservative (one stratum contains multiple villages).

is lower in self-targeting, and statistically significant in the specification with stratum fixed effects (Column 5).

On net, the non-parametric 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.

5.2. **Comparing Self-targeting to a Hypothetical Universal Automatic Enrollment Treatment.** In the automatic enrollment procedure, not all households were considered for enrollment. Instead, as discussed in Section 2.3.1, households only received the full PMT interview if they passed an initial set of screens. These pre-screening criteria were designed to save the government the cost of having to conduct a complete long-form census of every household in the country every time it wants to select beneficiaries. On net, as shown in Table 2, about 34 percent of households in the village received the full PMT interview, which is roughly comparable to the share of households who self-select to be interviewed in the self-targeting treatment.

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 pre-screening that occurs in the actual procedure. While this is less realistic (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 4 into selection on observables and selection on unobservables, we know a priori that self-targeting will perform worse than universal automatic enrollment with respect to selection on observables, because by definition universal automatic enrollment 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 still out-perform universal automatic enrollment on net if the selection on unobservables is sufficiently large.

To simulate what would have happened in universal automatic enrollment, we use our baseline data to construct PMT scores for those households who were not interviewed by the government as part of the PMT process. That is, for those households who were not interviewed as part of the real PMT treatment, we assume that they would have received benefits if their PMT score (according to the asset data we collected in our baseline survey) was below the threshold require to receive the program. We then repeat the same analysis in Figure 7 and Table 4, but instead of comparing self-targeting to the actual automatic enrollment treatment, we compare it to the constructed hypothetical universal automatic enrollment procedure.

The results are shown graphically in Figure 8 and in regression form in Table 6. Panel A of Figure 8 shows that the distribution of beneficiaries still looks poorer in self-selection than in the hypothetical universal automatic enrollment, though the difference between the two distributions is no longer statistically significant (p-value from the Kolomogorov-Smirnov test of equality of distributions, with randomization inference to cluster at village level, is 0.29). Panel B of Figure 8 reveals that automatic enrollment and self-targeting have very similar patterns in terms of the

probability of being selected at the low end of the spectrum, but that wealthier households are more likely to receive benefits under the automatic enrollment than self-targeting. This is related to selection on unobservables shown in Figure 7b – in the automatic enrollment treatment, some high consumption people make it through the PMT screen due to 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 enrollment treatment, the beneficiaries are poorer in self-targeting than in automatic enrollment (though statistical significance depends on specification.) Exclusion error looks virtually identical between the two treatments, despite the fact that many fewer people are interviewed during self-selection. 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 enrollment.

5.3. **Costs of alternative targeting approaches.** Self-targeting appears to perform better 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. And, both self-targeting and PMT 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 PMT. 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, Table 5 presents data on costs for the 200 villages in our sample in each treatment, along with the number of eligible households that do and do not receive benefits (exclusion error), the number of ineligible households that do and do not receive benefits (inclusion error), and, by way of comparison, the total annual dollar of benefits paid out to beneficiaries. We separate out costs paid by households into those paid by households who end up receiving the benefits (for whom the net cost of applying or being interviewed was therefore positive) and for those paid by households who do not end up receiving the benefits (for whom the net cost of applying or being interviewed was negative). For PMT, where we surveyed only a single neighborhood, we extrapolate to the entire village linearly; likewise, we extrapolate the costs for the hypothetical universal PMT linearly from the actual PMT costs. Finally, note that there could be economies of scale in implementing a national program. For PMT, 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.

The results show that the costs on households imposed by self-targeting for 200 villages totaled around US$70,000. The bulk of these costs (82%) were borne by non-beneficiaries, both because there were more of them and because, on average, they have a higher imputed wage rate. Administrative costs added an additional $170,000, so the total costs of targeting were around $240,000; this compares to around $1.2 million in benefits paid out in these villages per year. Since eligible

households generally receive the program for 6 years, the total targeting costs for self-targeting are about 3 percent of the total benefits given out.

The PMT treatment, which interviewed a similar number of households, imposed only US$9,300 in costs on households (just the time they spent at home taking the asset survey), and if we use the national-scale administrative costs, had a total cost of $130,000. But, as shown above, it had substantially higher rates of both inclusion and exclusion error compared to self-targeting. The hypothetical universal PMT, shown in column 3, had almost identical exclusion error as self-targeting, though it had almost double the inclusion error. The total costs imposed on households would be about $32,000 (about 45% of PMT), but the administrative costs, even using the national-scale administrative costs, are about double that of self-targeting.

This analysis suggests that, if we treat administrative costs and costs borne by households equally, self-targeting dominates the hypothetical universal PMT, in that it achieves better targeting at lower total costs. Self targeting and the status quo automatic enrollment PMT lie on very different parts of the frontier: the status quo costs as much as 40 percent less than self-targeting (though this difference could be muted if self targeting enjoyed the same nationwide economies of scale as the status quo), but has substantially higher rates of both inclusion and exclusion error. The main additional difference is that self-targeting places a higher fraction of the burden directly on households, including many who do not ultimately receive benefits. Whether the benefits of increased targeting outweigh the costs therefore depends on how one weights costs borne by households compared with administrative costs.

6. Marginal Effect of a Change in the Ordeal

Thus far, the findings suggest that self-targeting outperforms the status quo PMT procedure in identifying the poor. We next turn our discussion to what is the optimal way to design ordeal mechanisms. We showed in Section 3 that the effect of marginally increasing the intensity of ordeals on separating the rich from the poor is theoretically ambiguous. Therefore, we first experimentally test the effect of a change in the ordeal on selection. Specifically, we examine the results from experimentally varying the distance to the registration site and the number of households members required to be present at the application site, as discussed in section 2.3. Note that these experiments were carefully designed to be within a set of policy instruments that potentially could be considered by the government in their real conditional cash transfer program, under the requirements that the ordeals could not be so onerous that they would either discourage the severely credit-constrained poor from applying or that would unduly impose large application costs for the poor who might still be incorrectly screened out by the asset-test.

We then use the cross-sectional variation in our data to probe this question further: we fit a CRRA utility model of the decision to apply with logit shocks for different income groups, using a Generalized Method of Moments. The model helps understand which of the theoretical channels outlined in the model seems to be driving the results, and allowing us to predict whether one can differentially improve the selection of the poor by increasing the ordeals.

6.1. **Experimental Analysis.** We begin our discussion by exploring the effect of increasing distance. In the self-targeting villages, we experimentally chose whether the sign-up location would be

located very close or further away from the potential applicant's households. As Appendix Table A.7a shows, moving from the far to close registration sites decreased the distance from 1.88 km to 0.27 km; a reduction of 1.61 kilometers (or 1.69 kilometers controlling for strata fixed effects). [27] If the simplest version of the theory holds (See section 3.2.1 under the assumption that the utility shocks are uniformly distributed), we should expect that there should be more applicants in the close treatment and that they should be, on average, richer. Note, however, that under different model assumptions, the effect may be negative.

Table 7 explores the impact of the close treatment on targeting outcomes by estimating the following logit equation

$$Prob\left(SHOWUP_{vi} = 1\right) = \frac{exp\left\{\alpha + \beta CLOSE_v + \gamma LNPCE_{vi} + \eta CLOSE_v \times LNPCE_{vi}\right\}}{1 + exp\left\{\alpha + \beta CLOSE_v + \gamma LNPCE_{vi} + \eta CLOSE_v \times LNPCE_{vi}\right\}}$$
(22)

where $CLOSE_v$ is a dummy for the close treatment in village v, $LNPCE_{vi}$ is household i's log percapita consumption, and $CLOSE_v \times LNPCE_{vi}$ 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 (22) including only the $CLOSE_v$ variable. Including stratum fixed effects (Column (4)), the results show that the close treatment increases the log-odds of applying by 0.28 (p-value 0.101).[28] Put another way, this means that moving from far to close increases the percentage of households that apply by 15 percent (5.8 percentage points). When test for differential selection by consumption (Column (5)), we are unable to distinguish the effect of the close treatment by consumption levels from zero. Given that the theories implies that there may be non-linearities in the effect on the type of individual who applies when we alter the ordeal, we next explore potential non-linearities 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 changed the distribution of who showed up.

Similarly, as shown in Table 8, we also do not observe significant differential selection when requiring both spouses to apply in person rather than allowing either spouse to apply alone.[29] Given this, it is not surprising we find no effect either on the interaction of $BOTH$ with per-capita consumption (column 5), or when we interact the treatment with quintile bins of consumption (Column 6). One potential reason why requiring both spouses did not decrease enrollments is that

---

[27]Given differences in geography, there are treatment effect on distance varied across rural and urban locations. In rural areas, the signup station in the close treatment was located in each hamlet of the village (essentially, 0 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 Appendix Table A.7b). In urban areas, the signup 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 Appendix Table A.7c)

[28]The OLS version of this coefficient, which is clustered at the village level rather than the stratum level, is statistically significant at the 5% level (p-value 0.024); see Appendix Table A.5.

[29]In fact, the point estimates suggest that requiring both spouses to attend actually increases overall applications somewhat, perhaps because requiring both spouses means that the second spouse acts as a commitment device to show up or perhaps because it is more fun to go together.

this treatment included a provision through which spouses who were out of town and could not attend the interview could get a signed letter from a neighborhood leader to this effect, allowing the interview to proceed with only one spouse. A total of 28 percent of interviewees came with such a letter, suggesting that this provision may have been used to allow those with high opportunity costs to register anyway. This suggests that ordeals may in fact be hard to enforce in practice – loopholes such as this one, which the government put in place to be fair to those who for exogenous reasons could not possibly comply with the ordeal, can be exploited to undo the intent of the ordeal. This phenomenon seems similar to related problems observed in providing incentives to nurses in India to show up at work – a loophole that was required to exempt from incentives those who could not attend because of a legitimate outside obligation was expanded so much that it undid the entire impact of the incentive program (Banerjee, Duflo and Glennerster, 2008).

6.2. **Using the Model to Distinguish Theories and Predict Alternative Policies.** In this section, we return to the model in Section 3, estimate the unknown parameters of the model using the cross-sectional variation in the data, and use it to both understand the results thus far and to explore the effect of further increasing the ordeals on selection. The calibrated version of the model is useful for several reasons. First, it will help us understand whether the lack of differential selection we observe from experimentally increasing applications costs is consistent with what a calibrated version of the model would predict. Second, it allows us to test specifically for the different theoretical mechanisms outlined in the model (e.g. curvature in the utility function, different modes of transport for rich and poor). Finally, it allows us to consider counterfactual alternative policies to see how large the costs would have to be in order to differentially affect selection of rich and poor.

To take the model to the data, we start with equation (3), and specify a functional form for the utility function U and shock term ϵ. We assume that utility has a CRRA form ($U(x) = \frac{x^{1-\rho}}{1-\rho}$) with unknown curvature parameter ρ, and that the idiosyncratic utility shocks are drawn from a logistic distribution with mean α_ε and standard deviation β_ε. We focus on fitting these three parameters – ρ, α_ε and β_ε. [30]

To estimate the model, we exploit the cross-sectional variation in registration costs and benefits. We use data only from the far treatment group in fitting the model, so that we can explore what happens experimentally in the close treatment group as an out-of-sample validation of the model. We define registration costs as the per capita monetary cost, including foregone wages, of traveling to the registration site, waiting in line, and returning home. That is for household i, we specify:

$$c\left(y_i, l_i\right) = wage_i * \left(traveltime_i + \overline{waittime}\right) + travelmoney_i, \tag{23}$$

where $traveltime_i$ and $travelmoney_i$ are the individuals' 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 $\overline{waittime}$ by taking average wait-times by

---

[30]We opt to not estimate a fourth parameter, δ, because it turns out to have a lot of individual level heterogeneity which makes it hard to separate from the utility shocks. Choosing a reasonable value for δ is further complicated by the fact that PKH is supposed to last six years, but not everyone necessarily knows or believes that it will continue for that long. The discount factor therefore reflects that uncertainty as well as the usual impatience. We take our baseline estimate of an annual discount factor to be 0.5, which is much lower than most conventional estimates, for this reason, but show in the Appendix Table A.8 that the results are similar with other choices of δ.

treatment group and urban/rural designation calculated from the endline survey.[31] We calculate the household hourly wage rate $wage_i$ by dividing monthly household expenditure by hours worked by the household in a month.

Figure **9** 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 3.2.3.[32]

We calculate the level of benefit b_i that the household would receive if enrolled in the program based on the number and education level of their children.[33] Finally, we use a probit model to predict $\mu(y_i)$, the probability of getting the benefit conditional on applying.[34]We use the cross-sectional variation within the far treatment in $wage_i$, $traveltime_i$, $travelmoney_i$, b_i and $\mu(y_i)$ to identify the model.

We estimate the model by Generalized Method of Moments, where the moments are the mean values of the show up rates for the five quintiles of the consumption distribution in the far treatment. This gives us five moments to estimate three parameters, so we use a standard two-step GMM procedure to compute optimal weights among the five moments.

Table **9** shows the estimated parameter values. Specifically, the three estimated model parameters are $\alpha_\varepsilon = -0.0492$, $\beta_\varepsilon = 0.0643$, and $\rho = 0.0000$.[35] The result that $\alpha_\varepsilon < 0$ implies that the idiosyncratic utility shocks on average favor not showing up. The fact that $\rho = 0$, which implies that the households are expected income maximizers with linear utility, is somewhat surprising: perhaps it reflects the fact that on a monthly basis both the realized gains and the actual costs are relatively small numbers (per capita monthly benefit is on average 5.22 percent of monthly per capita expenditure for the entire sample, while total cost per capita is 0.72 percent of monthly per capita expenditure for the entire sample) Given the estimated linearity of the local utility function, it is not surprising that we get a clearly downward sloping show-up curve when we graphed show-up rates against per capita consumption in Figure 5, as the potential effect of the poor having much higher marginal utility costs of signing up discussed in section 3.2.4 do not appear to play a role empirically.

We then use these estimated parameters to predict the application rates under different assumptions for the cost function $c(y, l)$. For each possible $c(y, l)$, we simulate predicted application rates. To summarize what the model predicts, we repeat the same logit regressions we performed in Table 7

---

[31]We do not have sufficient data in to calculate separate wait times for each village.

[32]A regression of $c(y_i, l_i)$ on y_i and y_i^2 show that the coefficient on the quadratic term is statistically significant at the 5 percent level. This is not being 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. 2,000,000 per month.

[33]The benefit is calculated as follows. Each beneficiary household receives a base benefit Rp 200,000 per year. This level increases by Rp 800,000 if they have a child age less than 3 or are currently expecting, by Rp 400,000 if they have a child enrolled in primary school, and by 800,000 if they have a child in middle school. Since all beneficiaries have 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.

[34]We model the probability of receiving the benefit, conditional on applying, as a function of Log PCE. We include urban/rural interacted with district fixed effects, since the PMT cutoff for inclusion varies slightly for each urban/rural times district cell. The results are shown in Appendix Figure A.1.

[35]Note that the estimation was constrained such that $\rho \geq 0$.

on the simulated data. We also calculate the predicted showup rates for close and far subtreatments for those above and below the poverty line.[36]

The results from this exercise are shown in Table 10, and the predicted showup rates by quintile are graphed in Figure 10.. For comparison purposes, Column 1 of Table 10 and the top-left graph of Figure 10 replicates the actual empirical results (e.g. column 2 of Table 7). In addition to the empirical results from the logit model, in Panel B, we calculate the showup rates for those above and below the poverty line for both near and far treatments. In Panel C, we calculate the ratio of the poor to rich showup rates (i.e. equation (7) from the model) for both treatments, as well as the difference in this ratio between the near and far treatments (i.e. equation (8) from the model). The ratio is positive but statistically insignificant, indicating no statistically detectable differential targeting induced by moving from near to far in the experiment.[37]

In column 2 of Table 10, 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 (23), and calculate each household's predicted showup rate using the model. Since we only used the far treatment to estimate the model, comparing these simulated showup rates to actual showup rates serves as an out-of-sample check of the fit of model using the experiment. We bootstrap the standard errors using sample sizes equivalent to our actual data and with village-level clustering, so that the standard errors reported for the model-generated data are equivalent to those from the actual data. The results in column 2 thus show what we would have found had the data from our actual survey been generated by the model.[38]

Comparing the actual empirical estimates in column 1 with the estimates on the model-generated data in column 2, we find similar results on differential targeting between the treatments. In particular, even though the model seems to over-predict showup rates in the close treatment on average, the small differential effect between rich and poor showup ratios moving from near to far in the

---

[36]In order to run logits using the predicted application rates, we create 3000 copies of the data. The copies of each individual are assigned to apply or not in proportion to that individual's predicted probability of doing so. To make the standard errors comparable to the main experiment, we apply cluster bootstrap approach (clustered on villages) to this distribution, holding the total number of observations equal to the number of observations in the actual data.

[37]Note that the ratio is positive but insignificant, whereas the interaction term (the estimated coefficient on $[Close * LogPCE]$) in Panel A is negative and insignificant. The reason they are of different signs is that the logit model in Panel A is estimated using the linear $LogPCE$ variable, whereas the ratios in Panel C are based on a dummy variable for poor / non-poor. If we re-estimate the logit model using a dummy variable for rich, we obtain results with the same sign.

[38]Other recent papers that similarly use a well-identified randomized or natural experiment to provide a check of model fit include Kaboski and Townsend (2011) and Duflo, Hanna and Ryan (2012). More generally, the idea of hold out samples for validation has been used in several papers, staring with at least McFadden (1977); see Keane, Todd and Wolpin (2011). A smaller number of papers use randomized control experiments to validate a structural model. Wise (1985) estimates a model of housing demand on a control group data, and validates the model using the forecast of the effect of a housing subsidy. More recently, Todd and Wolpin (2006) used data from the PROGRESA program, a conditional cash transfer program in Mexico. Using only the control villages, they estimated a structural model of fertility, school participation and child labor. The model was validated by comparing the predicted effect of PROGRESA to the experimental estimates of program effects. Lise, Seitz and Smith (2004) use data from the Self Sufficiency Program in Canada to validate a search model of the labor market. As in Keane and Mott (1998), we estimate the model using the treatment sample because the incentive schedule provides useful variation for model identification, and use the control sample for out-of-sample model validation. Other papers which combine structural methods and experimental data (without using the control group for out of sample validation) include Attanasio, Meghir and Santiago (2012) and Ferrall (2010).

simulated data is not statistically distinguishable from what we actually observe in the experiment (Panel C; p-value 0.653). Consistent with this, the coefficients on the close dummy interacted with log per capita consumption (η in equation (22)), which is another way of capturing the degree of differential targeting between the close and far treatment, are also statistically indistinguishable between the actual experimental data in column (1) and the simulated data in column (2) (p-value 0.538). The fact that the model predictions are similar to the experimental findings provides us with greater confidence in the simulation results for alternative cost structures in the following columns.[39] A comparison of model fit can be seen by comparing the actual showup rates by quintile and treatment in the top-left of Figure 10 with the model's predicted showup rates by quintile and treatment in the top-middle of Figure 10.

6.2.1. *Distinguishing Alternate Theories.* Interestingly, even though there is strong evidence of self-selection (the poor are much more likely to show up than the rich, both on observables and unob-servables), 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 near treatment to far). We can use the model to help understand why this is not occurring, and in particular, examine the various mechanisms outlined in the model in Section 3.

Shocks. One possible explanation developed in the theory section is that, if the distribution of shocks does not have the monotone hazard rate property, it is possible that targeting could get worse as you increase distance, because the density of poor people induced to drop out by a higher marginal change is higher than the density of richer people (see Section 3.2.2). However, the version of the structural model we estimate and use in column (2) uses logit shocks, which do have the the monotone hazard rate property, yet still replicates the experimental findings. This suggests that the distribution of shocks alone are not the problem.

However, the magnitude of the shocks may explain why the response is so low. Examining equation (8), which showed the derivative of the showup 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. To assess quantitatively whether this is important, in column (3) we re-simulate the model where we cut the standard deviation of the shocks ϵ n half. Doing so doubles the impact of moving from close to far on the poor/rich showup ratio – from 0.293 in the base-case model to 0.684 – and the evidence suggests this would have been enough for such a difference to be statistically detectable. In column (4) we shut off the shocks entirely, so that everyone for whom $g(y, l) > 0$ shows up. This doubles the estimated impact on the showup ratio yet again, to 1.30. This suggests that even though the distribution of the shocks is not such that it can overturn the results and lead to the

---

[39]The one aspect of the model that does not match is that the predicted showup rates for those below the poverty line are actually higher in the far treatment than in the near treatment (69% vs 66%). 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 near group, which leads to higher showup rates. If we simulate the impact of of moving from far to close on the exact same group of beneficiaries, we indeed would obtain lower showup rates in far than in close in both rich and poor samples.

perverse case where increasing l can reduce the targeting ratio, the magnitude of the shocks is such that it severely dampens the impact of marginal increases in ordeals.

Curvature in the Utility Function. Another possible explanation given by the theory is that there may be curvature in the utility function, so that even though the marginal monetary cost of higher distance is greater for the rich, the monetary utility cost is greater for the poor (see Section 3.2.4). However, when we estimated the structural model, the found that the model was best fit with linear utility (i.e. $\rho = 0$), suggesting that this is not an important part of the explanation in our setting.[40]

Different Technologies for Overcoming Ordeals. The third explanation suggested by the model is that there are different transportation technologies used by the poor and the rich, so that the marginal monetary cost of distance is smaller for the rich (see section 3.2.3). Figure 9 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. To investigate whether this explains the lack of differential selection in response to an increase in distance, we use the model to generate simulated showup rates under the counterfactual that the poor and the rich use the same travel technology. To do so, 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 to the close and far registration places on quadratic functions of distance. We then use these predicted average travel costs – which by construction no longer allow richer households to use different transportation technologies – for all households, and re-calculate total registration costs $c(y, l)$. We then re-estimate the logit regressions and calculate the showup rates for the simulated data using these costs instead of the actual costs. Column (5) reports the results, which appear similar to the experimental findings (p-value 0.553 in Panel A; p-value of 0.670 in Panel C). The fact that the results are virtually unchanged when everyone is constrained to use similar transportation technologies suggests that the lack of differential selection between close and far is not being driven by the fact that the rich and poor use different transport technologies. The predicted showup rates using the same transport technology are show in the top right of Figure 10, and confirm that technology is not the main issue.

6.2.2. *Simulating Alternate Policies.* The results thus far suggest that, given the magnitude of the shocks, the problem is largely one of magnitudes – one might need a very large change in ordeals to impose meaningful additional self section. The remaining columns consider counterfactual experiments where, for the far group, we increase either the distance to the application site or average wait times, to see just how much of an ordeal one might need for the selection to become substantial. To simulate these counterfactual costs with increased distance, we again regress travel time and monetary costs on quadratic functions of distance from the application site, but now we do it separately for each rural/urban and income quintile bin, to allow costs to be heterogeneous by income group. We then calculate the additional costs of increased distance by adding either 3km or 6km to the actual distance, using the estimated relationships to calculated marginal time and money costs from that additional distance, and then adding that amount to the actual time and

---

[40]Appendix Figure A.2 shows the actual model fit, and alternatives where we impose higher values of ρ. As is evident from the Figure, imposing higher values of ρ leads to a more convex relationship between showup rates and income quintile than we observe in the data.

money costs reported for each individual. To simulate counterfactual costs with increased waiting time, simply the average waiting times by 3 or 6 times.

The results shown in columns (6) and (7), and graphed in the second row of Figure 10, that adding additional distance is still not enough to induce substantial differential selection – even adding 6km of distance, almost 4 times the mean mean value of 1.67 km – is not enough to induce substantial additional selection. The reason is that the marginal costs of increased distance do not appear to be that high because the costs of distance are concave – given that at such far distances almost everyone (even the poor) takes some form of motorized transportation, adding 6 km of distance raises the costs of applying by only about Rp 6,700 on average (US 70 cents) (see Appendix Table A.9).

The results in columns (8) and (9), and graphed in the third row of Figure 10, show that, by contrast, dramatically increasing wait times in the far treatment could induce differential selection. For example, when we increase wait times by a factor of 6 for the far treatment, we estimate a ratio of 3-1 for the poor-to-rich showup rates. This compares to a predicted ratio of 2.1-1 for the baseline model in column (2). What is happening is that the non-poor are dissuaded from showing up – 36 percent of non-poor show up in the baseline model, compared to only 21 percent when the costs are increased by 6. By contrast, the showup rates for the poor decrease by only 4 percentage points when the costs increase so much. Intuitively, wait times are more effective than distance in generating selection because wait times are a pure time cost, so the monetary costs are much more differential by income, while poor and rich use motorized transportation technologies after a certain distance so that the marginal cost of additional distance is relatively low for both income groups.

However, it is important to note that there are problems with long wait times in practice – the estimated wait times we needed to assume in column (7) to generate substantial differential selection averaged over 17 hours – almost two full work days of waiting in line. In a pilot for this study, when we experimented with long wait times (although still much less than 17 hours), villagers spontaneously organized themselves and assigned queuing numbers, so that people could wait at home and come back when it was their turn to be interviewed, rather than having to spend hours waiting in line. This suggests that while theoretically long wait times could be an effective screening device, actually making applicants wait for 17 hours may be very difficult in practice.

7. Conclusion

Ordeal mechanisms are often used to induce self-selection in the targeting of social programs. However, as we show in this paper, when we introduce real-world features into the model, such as credit constraints, non-linear utility functions, and non-linearities in the transport costs, the conventional wisdom that increasing ordeals improves targeting does not necessarily follow anymore. The question of how whether ordeals improve selection is therefore ultimately an empirical question.

Using data from a field experiment across 400 villages that examine targeting in Indonesia's conditional cash program (PKH), we showed that, indeed, the poor are more likely to self-select into applying than the non-poor. Interestingly, this selection occurred on two types of margins. First, we observe selection on the component of consumption is observable to governments. This implies 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 rich who may get incorrectly screened in by asset test. On net, introducing self-selection improved targeting as compared with the other targeting mechanisms that we considered, both the current status quo and a universal automatic enrollment 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 kilometer increase in distance) imposed substantial enough costs on households to lower application rates, but these costs did not differentially impact poor and rich households. Estimating our model structurally, we show that the additional time costs needed to induce differential selection of the poor are high, and out of realistic policy realm from both an implementation standpoint, and because it could induce substantial costs on the poor who may still be inaccuracy screened out by the asset test.

In short, ordeals can be a power tool to improve targeting relative to automatic enrollment systems, but that making onerous ordeals even more costly may not be the best way to improve it further. Moreover, while ordeals dominate the status quo, many of the poor still do not sign up. Further research is necessary to understand how to improve or augment design of ordeals further. For example, would increasing transparency in the rules, so that the poor know that they would indeed qualify, also allow for easier cheating of the system by the rich? Or, we know that the benefits are in the future, and we know that the poor may discount the future a lot or may have procrastination issues that would prevent them from signing up. Could small incentives to sign up increase the applications of the poor, without having perverse effects on the rich? Understanding these questions is an important direction for future research.

<center>REFERENCES</center>

Alatas, V., A. Banerjee, R. Hanna, B.A. Olken and J. Tobias. 2012. "Targeting the Poor: Evidence from a Field Experiment in Indonesia." *American Economic Review* 104 (2):1206–1240.

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 MIT.

Attanasio, O.P., C. Meghir and A. Santiago. 2012. "Education choices in Mexico: using a structural model and a randomized experiment to evaluate Progresa." *The Review of Economic Studies* 79(1):37–66.

Banerjee, A.V., E. Duflo and R. Glennerster. 2008. "Putting a Band-Aid on a corpse: Incentives for nurses in the Indian public health care system." *Journal of the European Economic Association* 6(2-3):487–500.

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

Coady, D. and S. Parker. 2009. Targeting social transfers to the poor in mexico. Working Paper 9/60 International Monetary Fund.

Currie, J. 2006. Public Policy and the Income Distribution. In *Public Policy and the Income Distribution*, ed. David E. Quigley John M. Auerbach, Alan J. Card. Russel Sage Foundation.

Daponte, B.O., S. Sanders and L. Taylor. 1999. "Why do low-income households not use food stamps? Evidence from an experiment." *Journal of Human Resources* pp. 612–628.

Duflo, E., R Hanna and Stephen P. Ryan. 2012. "Incentives Work: Getting Teachers to Come to School." *American Economic Review* 102 (4):1241–78.

Fan, Jianqing. 1992. "Design-adaptive Nonparametric Regression." *Journal of the American Statistical Association* 87(420):998–1004.

Kaboski, J.P. and R.M. Townsend. 2011. "A Structural Evaluation of a Large-Scale Quasi-Experimental Microfinance Initiative." *Econometrica* 79(5):1357–1406.

Keane, M.P., P.E. Todd and K.I. Wolpin. 2011. "The structural estimation of behavioral models: Discrete choice dynamic programming methods and applications." *Handbook of Labor Economics* 4:331–461.

Lise, J., S. Seitz and J. Smith. 2004. Equilibrium policy experiments and the evaluation of social programs. Technical report National Bureau of Economic Research.

Madrian, BC and DF Shea. 2001. "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior." *Quarterly Journal of Economics* 116(4):1149–1187.

Moffitt, R. 1983. "An economic model of welfare stigma." *The American Economic Review* 73(5):1023–1035.

Nichols, Albert L. and Richard J. Zeckhauser. 1982. "Targeting Transfers through Restrictions on Recipients." *The American Economic Review* 72(2):372–377.

Nichols, D., E. Smolensky and T.N. Tideman. 1971. "Discrimination by waiting time in merit goods." *The American Economic Review* 61(3):312–323.

Todd, P.E. and K.I. Wolpin. 2006. "Assessing the impact of a school subsidy program in Mexico: Using a social experiment to validate a dynamic behavioral model of child schooling and fertility."

The American economic review 96(5):1384–1417.

Wise, D.A. 1985. "A behavioral model versus experimentation: The effects of housing subsidies on rent." *Methods of Operations Research* 50:441–89.

TABLE 1. Experimental Design

		Both Spouse Subtreatment	Either Spouse Subtreatment	*Total*
Automatic Enrollment				200 (1,998)
Self Targeting	Close Subtreatment	50 (500)	50 (500)	100 (1,000)
	Far Subtreatment	50 (500)	50 (500)	100 (1,000)
	Total	100 (1,000)	100 (1,000)	200 (2,000)

Notes: This table provides the number of villages in each treatment cell, with the number of households in each cell shown in parentheses.

TABLE 2. Descriptive Statistics for Households Surveyed in the Baseline

	Total number of households (1)	Number of households interviewed (2)	Number of beneficiaries (3)	Percentage of households interviewed (4)	Percentage of interviewed households that received benefits (5)	Percentage of total households that received benefits (6)
Automatic Enrollment	1998	706	86	35.34%	12.18%	4.30%
Self Targeting	2000	754	73	37.70%	9.68%	3.65%

Notes: Table provides information on the flow of surveyed households through the experiment.

TABLE 3. Probability of Showing Up as a Function of the Observed and Unobserved Components of Baseline Log Per-capita Consumption

	Showed up		
	All	Very poor	Not very poor
	(1)	(2)	(3)
Observable consumption ($X_i'\beta$)	-2.217***	-0.811	-2.283***
	(0.201)	(1.981)	(0.204)
Unobservable consumption (ε_i)	-0.907***	-1.702*	-0.878***
	(0.136)	(0.877)	(0.137)
Stratum fixed effects	No	No	No
Observations	2,000	72	1,928
Mean of dependent variable	0.377	0.653	0.367

Notes: Very poor is defined as 80% of the poverty line by district. Each column shows a logit regression of showup rates on PMT score and epsilon. Robust standard errors, clustered at the village level, shown in parentheses *** p<0.01, ** p<0.05, * p<0.1

TABLE 4. Experimental Comparison of Targeting under Self Targeting and Automatic Enrollment Treatments

	Log consumption (beneficiaries) (OLS) (1)	Receives benefits (LOGIT) (2)	Error (LOGIT) (3)	Exclusion error (LOGIT) (4)	Inclusion error (LOGIT) (5)
		Panel A: No Stratum Fixed Effects			
Self targeting	-0.208***	12.142**	-0.202	-0.554	-0.308
	(0.076)	(4.894)	(0.127)	(0.412)	(0.209)
Log consumption		-1.016***			
		(0.280)			
Log consumption * Self targeting		-0.964**			
		(0.383)			
Observations	159	3,996	3,998	247	3,751
Mean of dependent variable	12.78	0.0398	0.0875	0.887	0.0349
		Panel B: With Stratum Fixed Effects			
Self targeting	-0.114	15.180***	-0.228*	-0.854*	-0.331*
	(0.077)	(5.295)	(0.136)	(0.457)	(0.193)
Log consumption		-1.042***			
		(0.283)			
Log consumption * Self targeting		-1.202***			
		(0.416)			
Observations	159	3,489	3,938	111	3,131
Mean of dependent variable	12.78	0.0456	0.0889	0.775	0.0418

Notes: Exclusion error is defined to be 1 if a household is very poor (as measured at baseline) and does not receive PKH and 0 otherwise. Inclusion error is defined to be 1 if a not-very poor household does receive PKH and 0 otherwise. Error includes either exclusion or targeting error. In Panel A, robust standard errors, clustered at the village level, are shown in parentheses. In Panel B, Columns (2) - (5), robust standard errors are clustered at the stratum level. *** p<0.01, ** p<0.05, * p<0.1

TABLE 5. Summary of targeting and costs

	Self-Targeting (1)	PMT (2)	Hypothetical Universal PMT (3)
# of eligible households that receive benefit	2046	1251	2047
# of eligible households that do not receive benefit	12399	13194	12398
# of ineligible households that receive benefit	6741	9072	11491
# of ineligible households that do not receive benefit	219569	217239	214820
Total annual benefits paid ($)	1198099	1407429	1845808
Total cost to households ($)	70149	9366	32403
Total cost to beneficiary households ($)	11978	1176	1412
Total cost to non-beneficiary households ($)	57817	8190	30995
Total administrative costs in sample ($)	170800	784083	2218978
Total administrative costs, scaled ($)	.	120378	340673

Notes: Estimates are totals for the 200 villages in our self-targeting sample. Column (1) is directly estimated using the self-targeting sample, and Columns (2) and (3) are estimated using the PMT sample. Total population in Columns (2) and (3) are scaled to match Column (1). For number of eligible/ineligible households, total annual benefits paid, and total cost to households, the percentage of eligible households in the village for Columns (2) and (3) are also scaled to Column (1). All monetary costs are reported in U.S. dollars, using an exchange rate of 9,535 IDR / 1 USD (October 2, 2012). Benefits per household are assumed to be 1.3 million IDR annually. Costs to households are calculated as the time cost of travel, waiting, and completing surveys (in PMT, just the cost of completing surveys) using the household average wage rate, as well as the cost of transportation. Total administrative costs in sample are calculated based on per-village and per-neighborhood actually incurred in by the experiment for Indonesian government surveyors in both Self-targeting and PMT, as well as an external NGO to help spread information about self-targeting; since PMT treatment was done in one neighborhood only, the actual costs are scaled up by the average number of neighborhoods in a village. Total administrative costs at scale in PMT are based on the actual Indonesian government cost of executing the PMT nationwide, when they were surveying approximately 16 million households. The costs of PMT are assumed to be linear in the number of households surveyed per village.

TABLE 6. Comparison of Targeting under Self-Selection and Hypothetical Universal Automatic Enrollment

	Log consumption (beneficiaries) (OLS) (1)	Receives benefits (LOGIT) (2)	Error (LOGIT) (3)	Exclusion error (LOGIT) (4)	Inclusion error (LOGIT) (5)
	Panel A: No Stratum Fixed Effects				
Self targeting	-0.133*	6.545	-0.275**	0.001	-0.555***
	(0.069)	(4.710)	(0.130)	(0.358)	(0.205)
Log consumption		-1.428***			
		(0.261)			
Log consumption * Self targeting		-0.552			
		(0.369)			
Observations	186	3,996	3,998	247	3,751
Mean of dependent variable	12.75	0.0465	0.0908	0.858	0.0403
	Panel B: With Stratum Fixed Effects				
Self targeting	-0.040	9.055*	-0.303**	-0.075	-0.589***
	(0.064)	(4.981)	(0.149)	(0.334)	(0.205)
Log consumption		-1.488***			
		(0.271)			
Log consumption * Self targeting		-0.749*			
		(0.393)			
Observations	186	3,489	3,938	129	3,131
Mean of dependent variable	12.75	0.0533	0.0922	0.752	0.0482

Notes: Exclusion error is defined to be 1 if a household is very poor (as measured at baseline) and does not receive PKH. Inclusion error is defined to be 1 if a not-very poor household does receive PKH. Error includes either exclusion or targeting error. Households are defined as beneficiaries of the hypothetical PMT if their PMT score defined at baseline qualifies them for PKH or they in reality received the benefit. In Panel A, robust standard errors, clustered at the village level, are shown in parentheses. In Panel B, Columns (2) - (5), robust standard errors are clustered at the stratum level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

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	0.205	1.345	0.195	0.275	0.485	0.193
	(0.146)	(2.841)	(0.238)	(0.168)	(2.920)	(0.310)
Log consumption		-1.434***			-1.446***	
		(0.143)			(0.144)	
Close subtreatment* Log consumption		-0.093			-0.023	
		(0.217)			(0.218)	
Consumption quintile 2			-0.317			-0.326
			(0.233)			(0.245)
Consumption quintile 3			-0.813***			-0.791***
			(0.231)			(0.234)
Consumption quintile 4			-1.084***			-1.072***
			(0.206)			(0.234)
Consumption quintile 5			-2.204***			-2.265***
			(0.257)			(0.279)
Close subtreatment * Consumption quintile 2			-0.271			-0.292
			(0.323)			(0.368)
Close subtreatment * Consumption quintile 3			0.255			0.321
			(0.299)			(0.325)
Close subtreatment * Consumption quintile 4			-0.385			-0.261
			(0.300)			(0.314)
Close subtreatment * Consumption quintile 5			0.174			0.277
			(0.371)			(0.387)
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	0.377	0.377	0.377	0.385	0.385	0.385

Notes: Each column present a logit regression of show-up on the close subtreatment. In Columns (1) - (3), robust standard errors are clustered at the village level. In Columns (4) - (6), robust standard errors are clustered at the stratum level. *** p<0.01, ** p<0.05, * p<0.1

TABLE 8. Experimental Results: Probability of showing up as a function of opportunity cost treatment

	No stratum fixed effects			With Stratum fixed effects		
	(1)	(2)	(3)	(4)	(5)	(6)
Both spouse subtreatment	0.196	4.303	0.461*	0.185*	3.334	0.384
	(0.146)	(2.840)	(0.237)	(0.099)	(2.857)	(0.243)
Log consumption		-1.324***			-1.343***	
		(0.145)			(0.144)	
Both spouse subtreatment * Log consumption		-0.318			-0.244	
		(0.217)			(0.217)	
Consumption quintile 2			-0.292			-0.327
			(0.212)			(0.219)
Consumption quintile 3			-0.478**			-0.470**
			(0.190)			(0.184)
Consumption quintile 4			-1.157***			-1.146***
			(0.185)			(0.205)
Consumption quintile 5			-1.871***			-1.962***
			(0.271)			(0.289)
Both spouse subtreatment * Consumption quintile 2			-0.348			-0.316
			(0.322)			(0.380)
Both spouse subtreatment * Consumption quintile 3			-0.416			-0.305
			(0.292)			(0.344)
Both spouse subtreatment * Consumption quintile 4			-0.237			-0.116
			(0.305)			(0.328)
Both spouse subtreatment * Consumption quintile 5			-0.514			-0.356
			(0.369)			(0.347)
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	0.377	0.377	0.377	0.385	0.385	0.385

Notes: Each column present a logit regression of show-up on the both spouse subtreatment. In Columns (1) - (3), robust standard errors are clustered at the village level. In Columns (4) - (6), robust standard errors are clustered at the stratum level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

TABLE 9. Estimated parameter values for the model

α_ε	β_ε	ρ
-0.0492	0.0643	0.0000
(0.0002)	(0.0012)	(0.0003)

Notes: This table reports the mean and variance of the cost shock (ε) and the coefficient of relative risk aversion (ρ). The parameters are estimated using two-step feasible GMM. The moments are defined as the average showup rates within each consumption quintile. These five moments are fit only in the far treatment villages, assuming an annual discount factor of 0.5. Bootstrapped standard errors are in parentheses.

TABLE 10. Modeled Effects of Time and Distance Costs on Showup Rates

	Show Up (Exp.)	Predicted Show Up (Model)[†]							
		Reported Total Cost	Reported Total cost, SD[eps]/2	Reported total cost, SD[eps]=0	Assuming No Differential Travel Cost	Additional Distance		Inflated Wait Time	
						Distance + 3km	Distance + 6km	Wait Time*3	Wait Time*6
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				Panel A: Logistic Regressions					
Close	1.345	-1.424	-3.617	-6.437	-1.338	-1.727	-1.720	-5.837*	-1.424
	(2.864)	(3.054)	(3.970)	(7.101)	(2.950)	(2.896)	(2.818)	(3.089)	(3.054)
Log per capita expenditure	-1.434***	-1.263***	-2.312***	-3.548***	-1.257***	-1.290***	-1.292***	-1.624***	-1.263***
	(0.144)	(0.160)	(0.232)	(0.430)	(0.160)	(0.158)	(0.153)	(0.184)	(0.160)
Close * Log per capita expenditure	-0.093	0.115	0.287	0.508	0.107	0.141	0.143	0.475*	0.115
	(0.219)	(0.234)	(0.306)	(0.559)	(0.226)	(0.221)	(0.216)	(0.237)	(0.234)
N	2000.000	5919000	5919000	6000000	5913000	5913000	5913000	5919000	5919000
P-value[‡]		0.517	0.312	0.317	0.524	0.452	0.443	0.079	0.517
				Panel B: Show-Up Rates					
Above poverty line, far	34.079	32.903	23.581	15.761	33.068	31.910	31.394	27.316	21.236
Above poverty line, close	39.001	36.386	27.891	19.341	36.383	36.386	36.386	36.386	36.386
Below poverty line, far	54.098	68.992	74.722	77.049	68.989	68.413	67.884	67.029	64.007
Below poverty line, close	57.627	65.655	69.309	69.492	65.654	65.655	65.655	65.655	65.655
				Panel C: Show-Up Rate Ratios					
Poor to rich ratio, far	1.587	2.097	3.169	4.888	2.086	2.144	2.162	2.454	3.014
	(0.222)	(0.210)	(0.342)	(0.755)	(0.204)	(0.214)	(0.216)	(0.272)	(0.372)
Poor to rich ratio, close	1.478	1.804	2.485	3.593	1.805	1.804	1.804	1.804	1.804
	(0.168)	(0.189)	(0.272)	(0.547)	(0.187)	(0.196)	(0.192)	(0.191)	(0.194)
Difference of ratios	0.110	0.293	0.684*	1.296	0.282	0.340	0.358	0.650*	1.210***
P-value	(0.285)	(0.282)	(0.436)	(0.933)	(0.280)	(0.295)	(0.292)	(0.332)	(0.422)
		0.649	0.271	0.224	0.667	0.575	0.543	0.218	0.031

Notes: In order to run logits on predicted show up rates, we create 3000 copies of the data. The copies of each individual are assigned to show up or not in proportion to his predicted probability of showing up. Bootstrapped standard errors, clustered by village, in parentheses. To compute the standard errors in columns 2-7, for each bootstrap iteration we sample 2000 households, clustered at the village level, to make the sample equivalent to that in column 1. The p-value in Panel A is the test of whether the coefficient on $[Close * LogPCE]$ is equal to the equivalent coefficient in column 1. The p-value in Panel B is the test of whether the difference in ratios is equal to the difference in ratios in column 1. *** p<0.01, ** p<0.05, * p<0.1 Significance levels not shown on first two rows of Panel C.
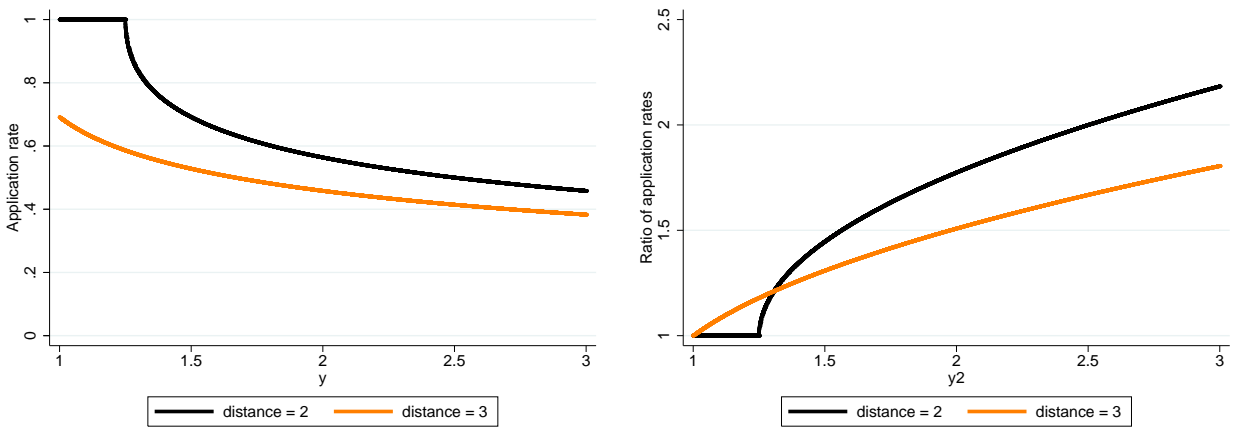
FIGURE 1. 1Illustration of utility gain with no errors
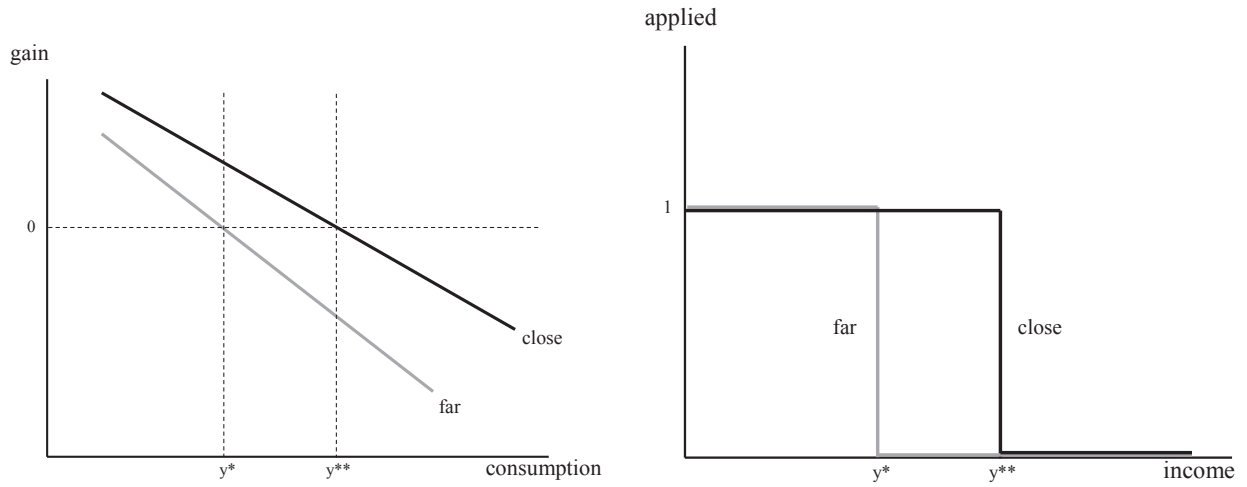
(A) Gain vs. consumption for close and far subtreatments

(B) Targeting improves as length of ordeal increases

FIGURE 2. Illustration of utility gain with log-logistic errors

(A) Showup rates with log-logistic errors

(B) Ratio of showup rates of rich (y_2) compared to poor ($y_1 = 1$)

FIGURE 3. Non-Linearities in Travel Costs

Notes: Increasing ordeal within l' to l", marginal cost for rich is lower than marginal cost for the poor.

FIGURE 4. Illustration of utility gain with concave utility
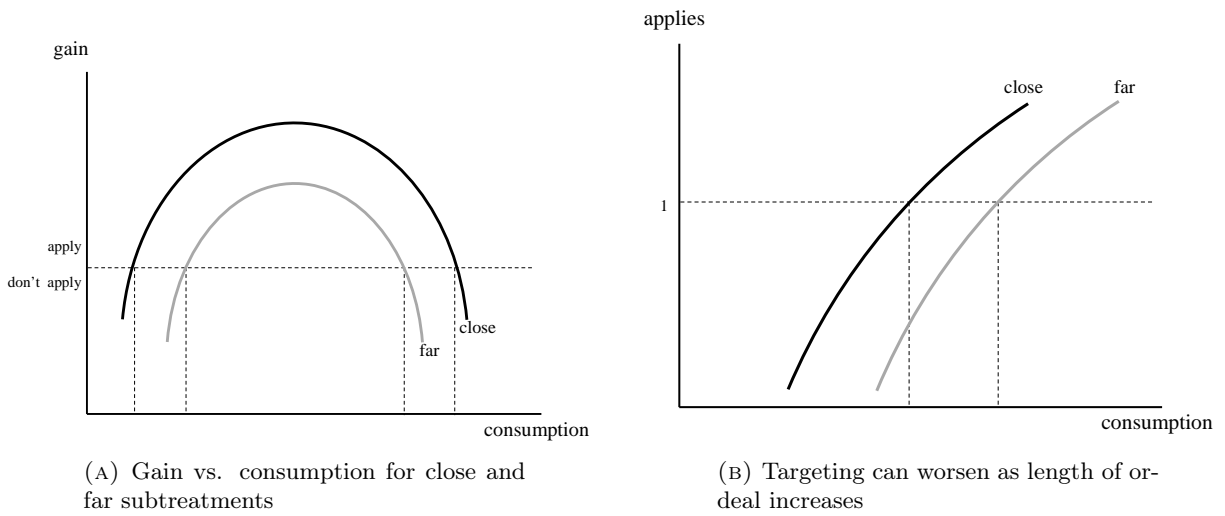
(A) Gain vs. consumption for close and far subtreatments
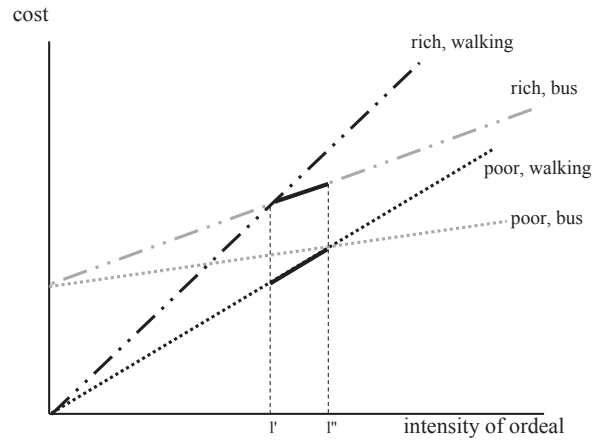
(B) Targeting can worsen as length of ordeal increases

43

FIGURE 5. Showup Rates Versus Log Per Capita Consumption
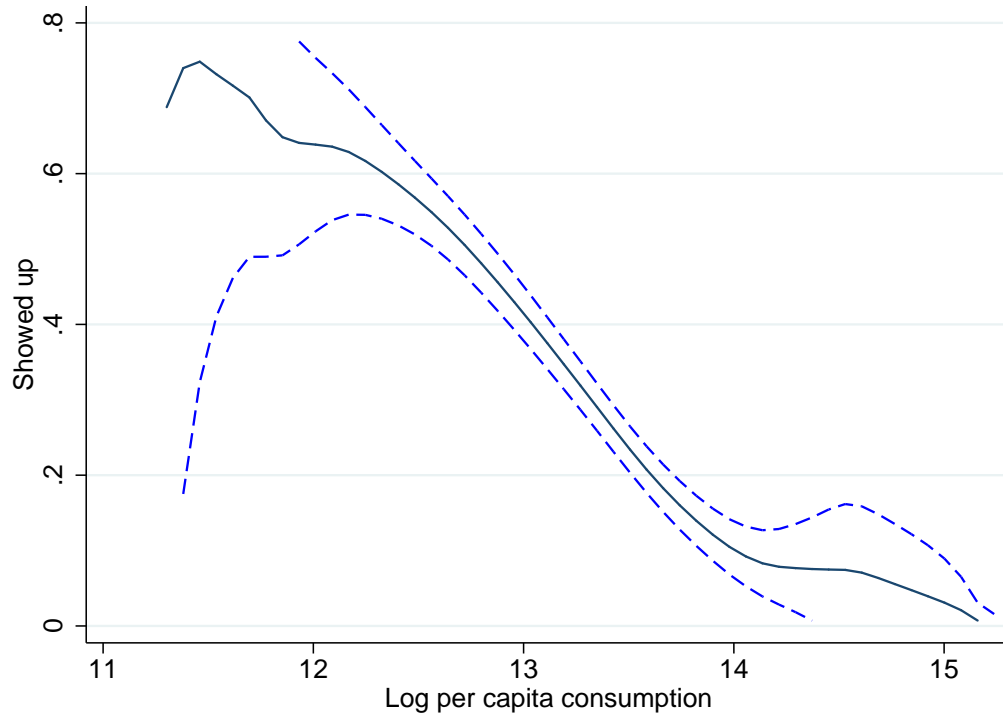
Notes: Figure provides a non-parametric fan regression of the probability of applying for PKH against baseline log per capita consumption in the 200 self-targeting villages. Bootstrapped standard error bounds, clustered at the village level, are shown in dashes.

FIGURE 6. Showup Rates Versus Observable and Unobservable Components of Log Per Capita Consumption
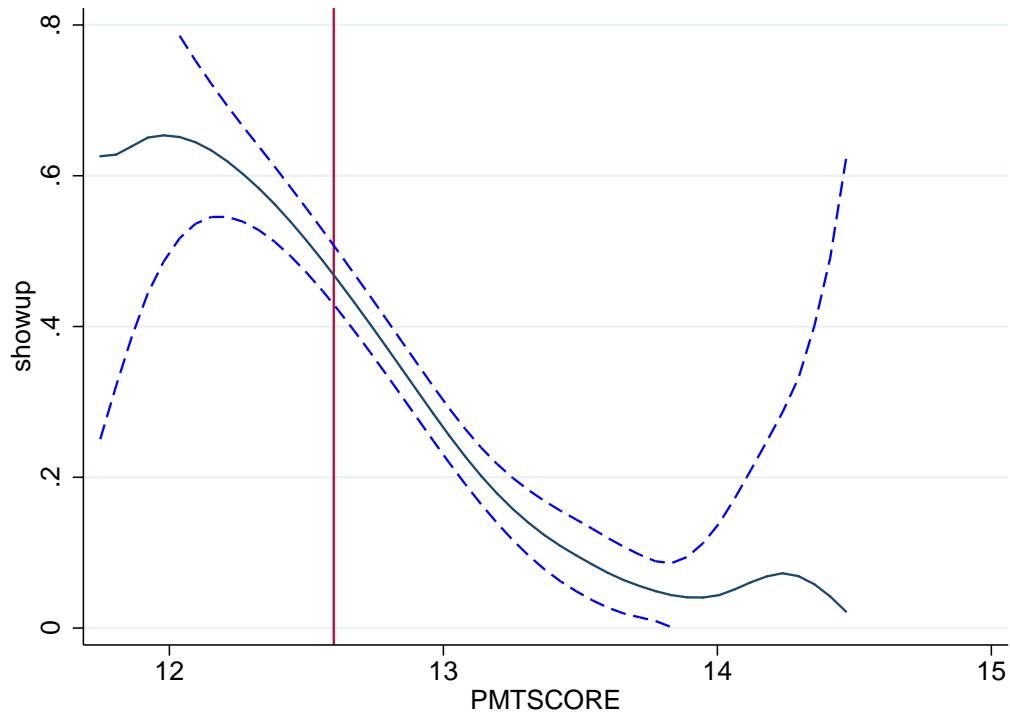
(A) Showup as a function of observable consumption $(X_i'\beta)$
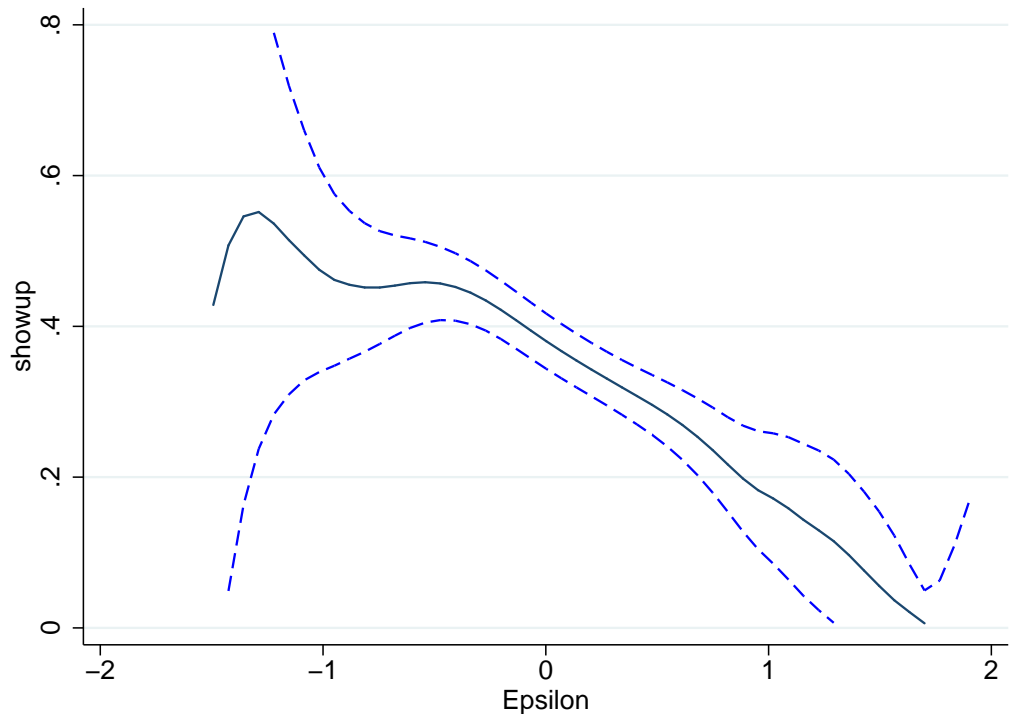
(B) Showup as a function of unobservable consumption (ε_i)

Notes: Figures provide non-parametric fan regressions of the probability of applying for PKH against components of baseline log per capita consumption in the 200 self-targeting villages. Bootstrapped standard error bounds, clustered at the village level, are shown in dashes.

FIGURE 7. Experimental Comparison of Self Targeting and Automatic Enrollment Treatments
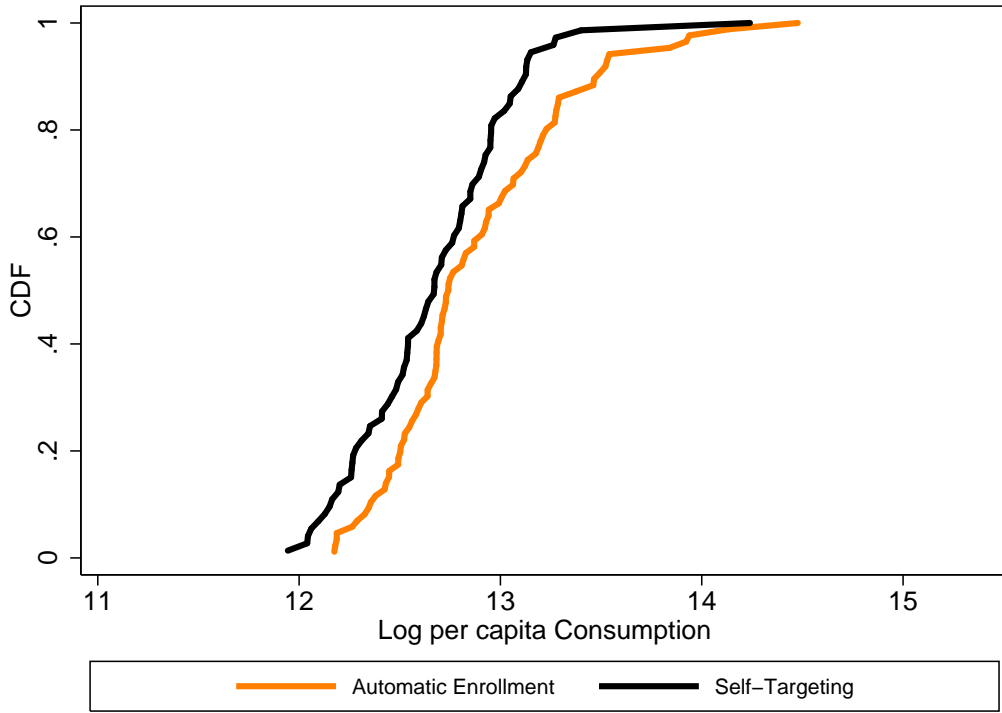
(A) CDF of log per capita consumption of beneficiaries
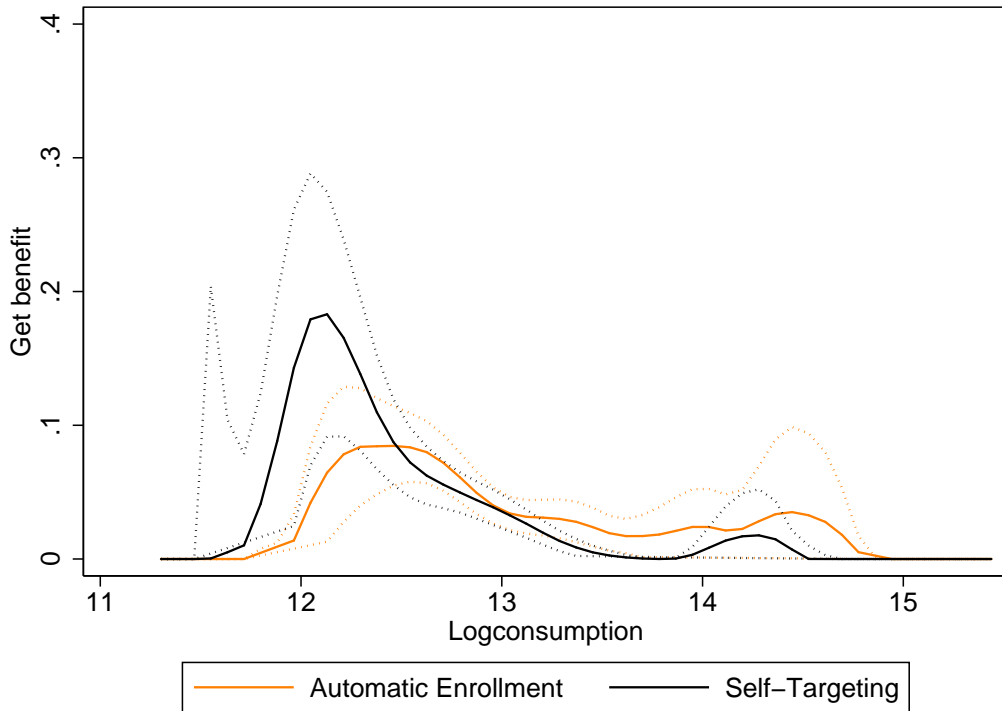
(B) Receiving benefit as a function of log per capita consumption

Notes: Panel A shows a CDF of log per capita consumption of beneficiaries. Kolmogorov-Smirnov test of equality yields a p-value of 0.10. Panel B present a non-parametric fan regression of benefit receipt on log consumption per capita. Bootstrapped standard errors, clustered at the village level, are shown in dashes.

FIGURE 8. Comparison of Self-Selection and Hypothetical Universal Automatic Enrollment
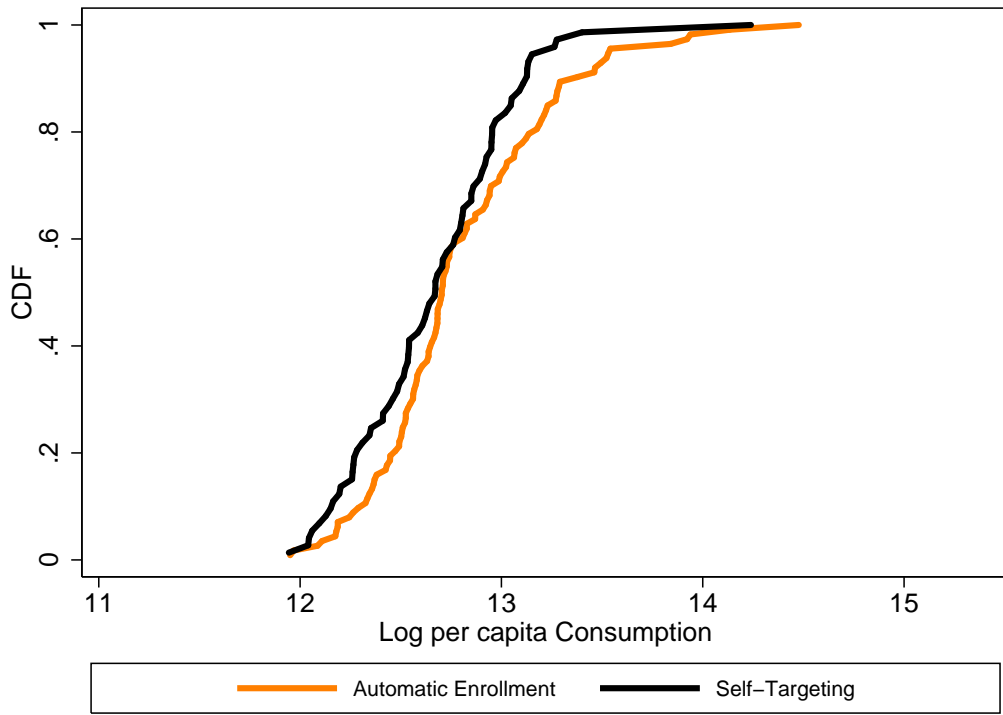
(A) CDF of consumption of beneficiaries
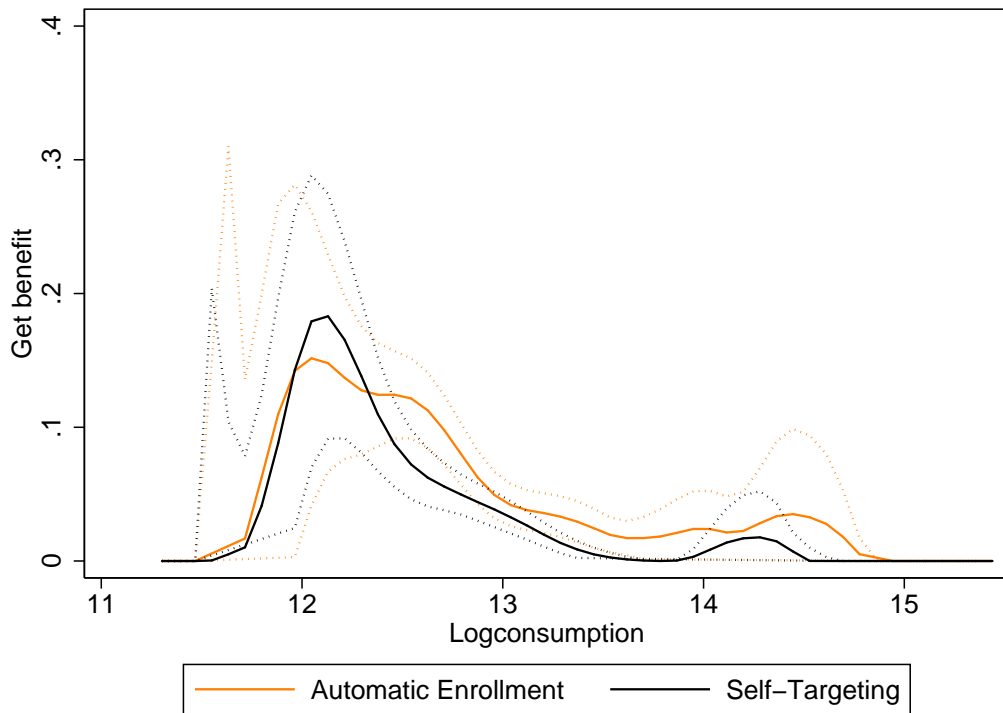
(B) Getting benefit as a function of log per capita consumption

Notes: Panel A shows a CDF of log per capita consumption of beneficiaries. Kolmogorov-Smirnov test of equality yields a
p-value of 0.29. Panel B present a non-parametric fan regression of benefit receipt on log consumption per capita.
Bootstrapped standard errors, clustered at the village level, are shown in dashes.

FIGURE 9. Cost of Applying by Consumption
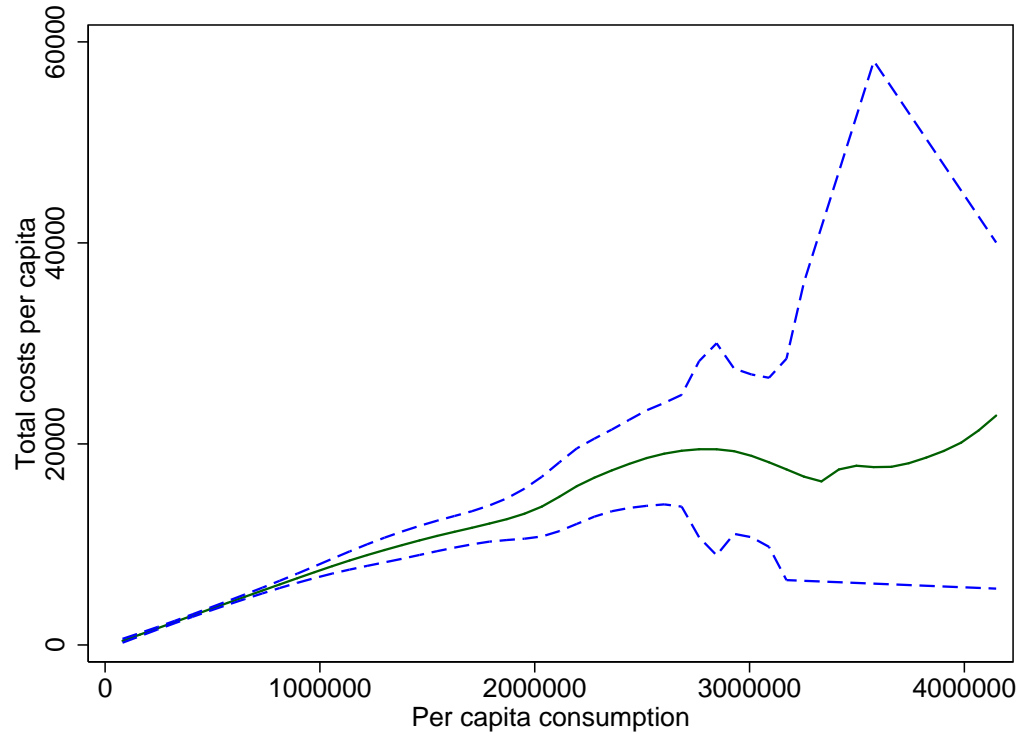
Notes: Figure shows a non-parametric fan regression of total costs incurred in applying for PKH against per capita consumption. Bootstrapped standard errors, clustered at the village level, are shown in dashes.

FIGURE 10. Model Fit and Counterfactuals
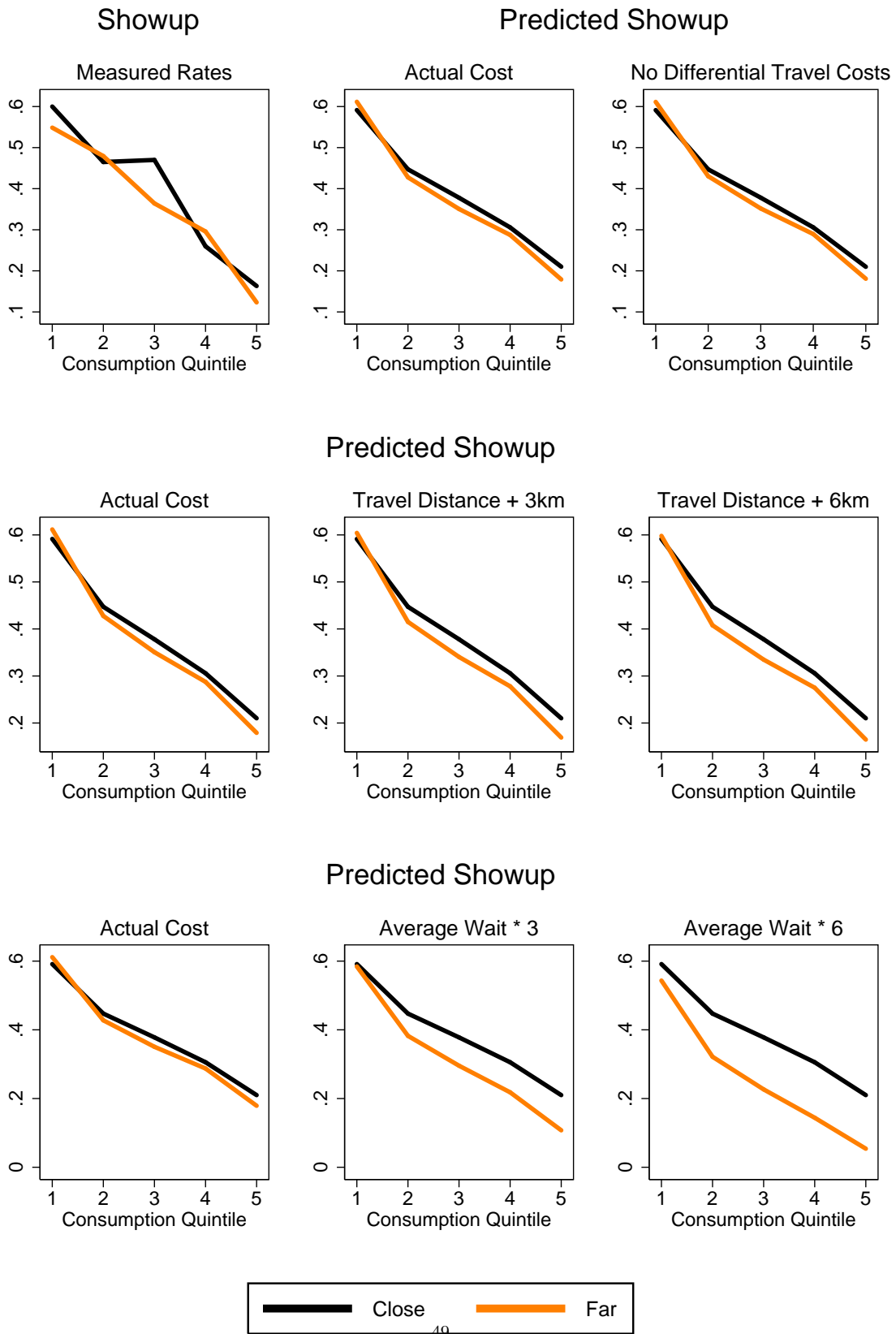
APPENDIX A. ADDITIONAL TABLES AND FIGURES

TABLE A.1. Experiment Validation

	Mean in automatic enrollment (1)	Mean in self targeting (2)	No stratum fixed effect (3)	With stratum fixed effect (4)
Log Per Capita Consumption	13.112	13.105	-0.007	-0.001
	(0.228)	(0.251)	(0.024)	(0.021)
Years of education: household head	7.440	7.329	-0.112	-0.083
	(2.203)	(2.055)	(0.213)	(0.167)
PMT score	12.798	12.796	-0.002	0.003
	(0.228)	(0.251)	(0.024)	(0.019)
Percentage of households in agriculture	0.073	0.071	-0.002	-0.004
	(0.068)	(0.063)	(0.007)	(0.005)
Years of education: hamlet head	8.307	8.181	-0.074	-0.105
	(3.697)	(3.334)	(0.182)	(0.311)
Log of number of households in hamlet	4.227	4.241	-0.126	0.031
	(0.520)	(0.468)	(0.353)	(0.045)
Distance to kec	7.434	6.404	-1.031	-1.038
	(21.919)	(8.184)	(1.654)	(1.615)
Log of village size	4.038	3.925	-0.113	-0.129*
	(1.574)	(1.476)	(0.153)	(0.067)
Religious building per household	0.005	0.005	0.000	-0.000
	(0.003)	(0.003)	(0.000)	(0.000)
Primary school per household	0.003	0.003	-0.000	-0.000
	(0.001)	(0.002)	(0.000)	(0.000)
Observations	200	200	400	400
Joint significance test (chi squared):			2.49	7.58
p-value			0.99	0.67

TABLE A.2. Probability of Showing Up as a Function of the Observed and Unobserved Components of Baseline Log Per-capita Consumption (OLS)

	Showed up		
	All	Very poor	Not very poor
	(1)	(2)	(3)
Observable consumption ($X_i'\beta$)	-0.457***	-0.410	-0.459***
	(0.029)	(0.895)	(0.032)
Unobservable consumption (ε_i)	-0.126***	-0.137	-0.124***
	(0.022)	(0.246)	(0.022)
Observations	2,000	72	1,928
Mean of dependent variable	0.377	0.653	0.367

Notes: OLS version of 3. Very poor is defined as 80% of the poverty line by district. Each column shows an OLS regression of showup rates on PMT score and epsilon.

Robust standard errors, clustered at the village level, shown in parentheses *** p<0.01, ** p<0.05, * p<0.1

TABLE A.3. Experimental Comparison of Targeting under Self-Targeting and Automatic Enrollment Treatments (OLS)

	Log consumption (beneficiaries) (OLS) (1)	Receives benefits (OLS) (2)	Error (OLS) (3)	Exclusion error (OLS) (4)	Inclusion error (OLS) (5)
		Panel A: No Stratum Fixed Effects			
Self targeting	-0.208***	0.269	-0.016	-0.055	-0.010
	(0.076)	(0.178)	(0.010)	(0.041)	(0.007)
Log consumption		-0.037***			
		(0.009)			
Log consumption * Self targeting		-0.021			
		(0.013)			
Observations	159	3,996	3,998	247	3,751
Mean of dependent variable	12.78	0.0398	0.0875	0.887	0.0349
		Panel B: With Stratum Fixed Effects			
Self targeting	-0.114	0.297*	-0.018*	-0.078*	-0.011*
	(0.077)	(0.171)	(0.009)	(0.045)	(0.006)
Log consumption		-0.036***			
		(0.009)			
Log consumption * Self targeting		-0.023*			
		(0.013)			
Observations	159	3,996	3,998	247	3,751
Mean of dependent variable	12.78	0.0398	0.0875	0.887	0.0349

Notes: OLS version of 4. Exclusion error is defined to be 1 if a household is very poor (as measured at baseline) and does not receive PKH. Inclusion error is defined to be 1 if a not-very poor household does receive PKH. Error includes either exclusion or targeting error. Robust standard errors, clustered at the village level, in parentheses. *** p<0.01, ** p<0.05, * p<0.1

TABLE A.4. Comparison of Targeting under Self-Targeting and Hypothetical Universal Automatic Enrollment (OLS)

	Log consumption (beneficiaries) (OLS) (1)	Receives benefits (OLS) (2)	Error (OLS) (3)	Exclusion error (OLS) (4)	Inclusion error (OLS) (5)
Panel A: No Stratum Fixed Effects					
Self targeting	-0.133*	-0.103	-0.023**	0.000	-0.021***
	(0.069)	(0.194)	(0.011)	(0.044)	(0.008)
Log consumption		-0.064***			
		(0.011)			
Log consumption * Self targeting		0.006			
		(0.014)			
Observations	159	3,996	3,998	247	3,751
Mean of dependent variable	12.75	0.0465	0.0908	0.858	0.0403
Panel B: With Stratum Fixed Effects					
Self targeting	-0.040	-0.085	-0.025**	-0.009	-0.022***
	(0.064)	(0.187)	(0.010)	(0.049)	(0.007)
Log consumption		-0.063***			
		(0.010)			
Log consumption * Self targeting		0.005			
		(0.014)			
Observations	186	3,996	3,998	247	3,751
Mean of dependent variable	12.75	0.0465	0.0908	0.858	0.0403

Notes: OLS version of 6. Exclusion error is defined to be 1 if a household is very poor (as measured at baseline) and does not receive PKH. Inclusion error is defined to be 1 if a not-very poor household does receive PKH. Error includes either exclusion or targeting error. Robust standard errors, clustered at the village level, in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

TABLE A.5. Experimental Results: Probability of showing up as a function of distance (OLS)

	No stratum fixed effects			With stratum fixed effects		
	(1)	(2)	(3)	(4)	(5)	(6)
Close subtreatment	0.048	0.451	0.047	0.058**	0.280	0.043
	(0.034)	(0.471)	(0.058)	(0.026)	(0.457)	(0.053)
Log consumption		-0.277***			-0.254***	
		(0.023)			(0.022)	
Close subtreatment* Log consumption		-0.032			-0.019	
		(0.035)			(0.034)	
Consumption quintile 2			-0.079			-0.079
			(0.058)			(0.056)
Consumption quintile 3			-0.199***			-0.183***
			(0.055)			(0.051)
Consumption quintile 4			-0.259***			-0.236***
			(0.047)			(0.048)
Consumption quintile 5			-0.435***			-0.404***
			(0.044)			(0.045)
Close subtreatment * Consumption quintile 2			-0.066			-0.064
			(0.080)			(0.073)
Close subtreatment * Consumption quintile 3			0.061			0.068
			(0.072)			(0.066)
Close subtreatment * Consumption quintile 4			-0.086			-0.054
			(0.067)			(0.063)
Close subtreatment * Consumption quintile 5			-0.002			0.014
			(0.065)			(0.064)
Stratum fixed effects	No	No	No	Yes	Yes	Yes
Observations	2,000	2,000	2,000	2,000	2,000	2,000
Mean of dependent variable	0.377	0.377	0.377	0.377	0.377	0.377

Notes: OLS version of 7. Robust standard errors, clustered at the village level, in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

TABLE A.6. Experimental Results: Probability of showing up as a function of opportunity cost (OLS)

	No stratum fixed effects			With Stratum fixed effects		
	(1)	(2)	(3)	(4)	(5)	(6)
Both spouse subtreatment	0.046	0.953**	0.112*	0.039	0.794*	0.091*
	(0.034)	(0.464)	(0.057)	(0.025)	(0.434)	(0.050)
Log consumption		-0.259***			-0.235***	
		(0.024)			(0.024)	
Both spouse subtreatment * Log consumption		-0.070**			-0.058*	
		(0.034)			(0.033)	
Consumption quintile 2			-0.073			-0.075
			(0.053)			(0.049)
Consumption quintile 3			-0.118**			-0.109**
			(0.047)			(0.045)
Consumption quintile 4			-0.267***			-0.240***
			(0.042)			(0.040)
Consumption quintile 5			-0.380***			-0.358***
			(0.047)			(0.046)
Both spouse subtreatment * Consumption quintile 2			-0.083			-0.076
			(0.079)			(0.073)
Both spouse subtreatment * Consumption quintile 3			-0.101			-0.078
			(0.070)			(0.064)
Both spouse subtreatment * Consumption quintile 4			-0.066			-0.045
			(0.068)			(0.063)
Both spouse subtreatment * Consumption quintile 5			-0.118*			-0.086
			(0.064)			(0.061)
Stratum fixed effects	No	No	No	Yes	Yes	Yes
Observations	2,000	2,000	2,000	2,000	2,000	2,000
Mean of dependent variable	0.377	0.377	0.377	0.377	0.377	0.377

Notes: OLS version of 8. Robust standard errors, clustered at the village level, in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

TABLE A.7. Effect of Close Subtreatment on Distance

(A) All Villages

VARIABLES	Reported distance		GPS distance	
	(1)	(2)	(3)	(4)
Close subtreatment	-1.686***	-1.077	-0.963***	-1.317
	(0.167)	(1.688)	(0.209)	(3.496)
Log per capita consumption		-0.125		-0.173
		(0.090)		(0.185)
Close subtreatment * Log per capita consumption		-0.048		0.026
		(0.132)		(0.257)
Observations	1,999	1,999	1,847	1,847
Mean of dependent variable	1.079	1.079	0.652	0.652

(B) Rural Villages

VARIABLES	Reported distance		GPS distance	
	(1)	(2)	(3)	(4)
Close subtreatment	-1.248***	-2.032	-0.897***	-4.574
	(0.154)	(1.603)	(0.199)	(3.952)
Log per capita consumption		-0.159*		-0.191
		(0.096)		(0.244)
Close subtreatment * Log per capita consumption		0.059		0.281
		(0.123)		(0.289)
Observations	1,320	1,320	1,319	1,319
Mean of dependent variable	0.606	0.606	0.463	0.463

(C) Urban Villages

VARIABLES	Reported distance		GPS distance	
	(1)	(2)	(3)	(4)
Close subtreatment	-2.639***	-3.055	-1.418***	-0.974
	(0.412)	(3.327)	(0.427)	(3.238)
Log per capita consumption		-0.222		-0.249**
		(0.174)		(0.101)
Close subtreatment * Log per capita consumption		0.030		-0.036
		(0.263)		(0.246)
Observations	679	679	528	528
Mean of dependent variable	1.997	1.997	1.124	1.124

TABLE A.8. Modeled Effects of Time and Distance Costs on Showup Rates (Discount Rate Robustness Check)

	Show Up (Exp.)		`Predicted Show Up					
		δ =	0.4	0.5	0.6	0.7	0.8	0.9
	(1)		(2)	(3)	(4)	(5)	(6)	(7)
		Panel A: Coefficient on Interaction Term from Logit Regressions						
β(Close * Log PCE):	-0.093	Reported Total Cost	0.129	0.105	0.088	0.080	0.073	0.053
	(0.219)		(0.222)	(0.229)	(0.230)	(0.235)	(0.227)	(0.227)
		No Differential Travel Costs	0.119	0.098	0.082	0.075	0.068	0.050
			(0.233)	(0.222)	(0.237)	(0.226)	(0.235)	(0.232)
		Distance + 6km	0.164	0.135	0.112	0.099	0.088	0.063
			(0.237)	(0.220)	(0.237)	(0.232)	(0.230)	(0.229)
		Wait Time*6	1.081***	0.895***	0.731***	0.610**	0.496*	0.340
			(0.293)	(0.266)	(0.252)	(0.251)	(0.248)	(0.243)
		Panel B: Show-Up Rates (Using Reported Total Cost)						
Above poverty line, far			34%	34%	34%	34%	33%	32%
Above poverty line, close			38%	38%	37%	37%	36%	35%
Below poverty line, far			71%	72%	72%	74%	74%	72%
Below poverty line, close			69%	69%	69%	70%	71%	68%
		Panel C: Difference in Show-Up Rate Ratios						
(Poor/rich, far) –	0.110	Reported Total Cost	0.297	0.285	0.278	0.284	0.290	0.268
(Poor/rich, close)	(0.278)		(0.271)	(0.263)	(0.274)	(0.268)	(0.288)	(0.292)
		No Differential Travel Costs	0.283	0.274	0.269	0.277	0.284	0.264
			(0.266)	(0.270)	(0.268)	(0.273)	(0.273)	(0.280)
		Distance + 6km	0.389	0.353	0.328	0.324	0.320	0.287
			(0.271)	(0.278)	(0.274)	(0.282)	(0.271)	(0.286)
		Wait Time*6	1.568***	1.215***	0.967***	0.818**	0.694*	0.529*
			(0.462)	(0.397)	(0.349)	(0.341)	(0.338)	(0.302)

Notes: Each cell represents the coefficient on close*logPCE from a separate logit regression, showing varying values of delta. Column (3) matches the results shown in 10. Bootstrapped standard errors, clustered at the village level., in parentheses. *** p<0.01, ** p<0.05, * p<0.1

TABLE A.9. Summary Statistics of Modeled Registration Costs

	In-Sample Statistics		Out-Of-Sample Statistics				
	Reported Total Cost		Assuming No Differential Travel Costs	Additional Distance		Inflated Wait Time	
				Distance + 3km	Distance + 6km	Wait Time*3	Wait Time*6
	(Close)	(Far)	(Far)	(Far)	(Far)	(Far)	(Far)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Mean total monetary costs to registration (Rp. thousand)	3.24	5.01	4.87	6.15	6.74	12.65	24.10
Mean distance to registration site (km)	0.28	1.88	1.88	4.88	10.88	1.88	1.88
Mean wait time (mins)	156.55	175.73	175.73	175.73	175.73	527.19	1054.38

FIGURE A.1. Probability of obtaining benefits (μ) vs. log per-capita consumption
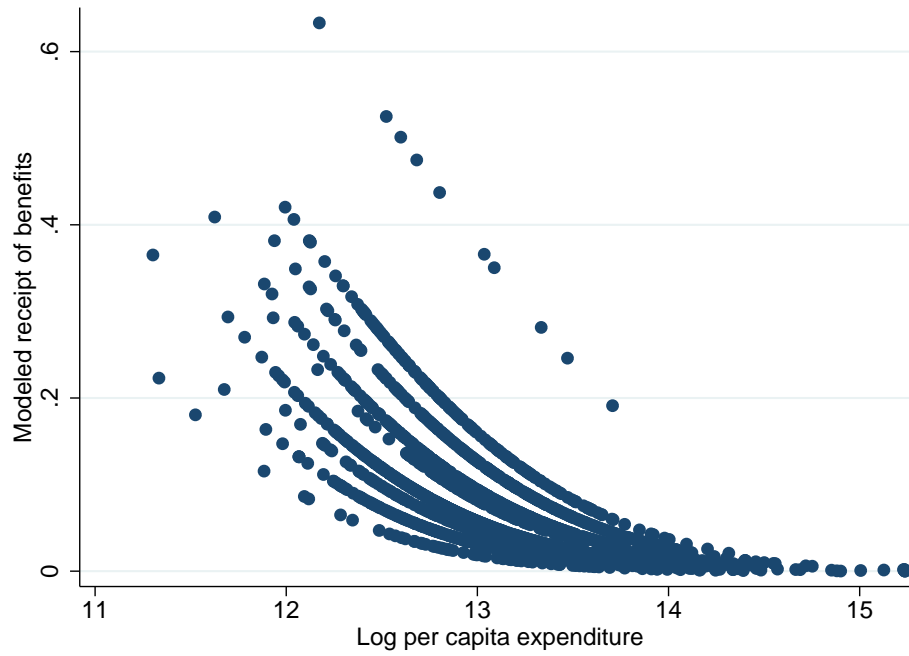
Notes: This figure shows the predicted of receiving the benefit, conditional on applying, from a probit model of receiving a benefit as a function of Log PCE. We include urban/rural interacted with district fixed effects in the probit, since the PMT cutoff for inclusion varies slightly for each urban/rural times district cell. These predicted values are the $\mu(y_i)$ that we use in the model.

FIGURE A.2. Model fit imposing different values of ρ
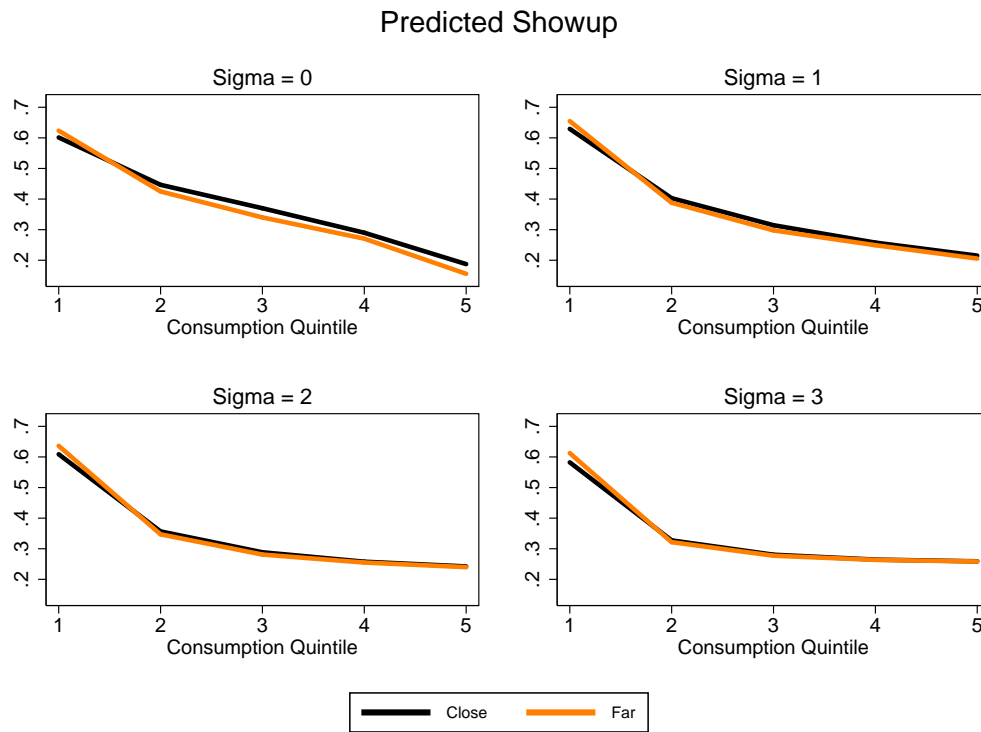
Predicted Showup

Notes: This figure shows predicted showup rates from the model for different values of ρ. As is evident from the figure, higher values of ρ lead to a convex showup pattern by income quintile, which is not consistent with the actual showup pattern shown in Figure 10.