PMT vulnerability to measurement errors: Evidence from a survey experiment in Tanzania

Gazeaud Jules*

PRELIMINARY AND INCOMPLETE - PLEASE DO NOT CITE

This Version: January 30, 2018

Abstract

Proxy Means Testing (PMT) is a popular method to target the poor in developing countries. PMT usually relies on survey-based consumption data and assumes that they are measured with random errors—even though this assumption has been challenged by recent literature. This paper brings causal evidence on the impact of non-random errors in consumption on PMT performances. I rely on a survey experiment conducted in Tanzania in which eight alternative consumption questionnaires were randomly distributed across households, and compare the performances of PMT relying on error-prone consumption data with those of a PMT using gold standard consumption data. Results consistently show that non-random errors in consumption data reduce both the predictive and targeting performances of PMT. Specifically, I find that the rate of targeting error increases by a magnitude ranging from 13 to 33%, depending on how consumption data are collected and how the poverty line is defined. I argue that this reduction in PMT performances is non trivial and has the potential to influence assessments of PMT performances and their comparisons with other targeting methods.

JEL Codes: I32, I38, O12, O20

Keywords: Social Protection, Targeting, Proxy Means Testing (PMT), Measurement Errors, Tanzania.

*Ph.D Student at CERDI, University of Auvergne. Bd. F. Mitterrand, 65, F-63000, Clermont-Fd; email: jules.gazeaud@etu.udamail.fr. I am grateful to Joachim De Weerdt for sharing the data underlying this research. Viewpoints and errors are my own.

1 Introduction

Social safety nets programs (SSNP) such as cash and in-kind transfers have become an important tool for achieving poverty alleviation in developing countries. Based on the World Bank Aspire database, in the last two decades, the number of developing countries with SSNP doubled from 72 to 149 so that today almost every country has at least one SSNP.[1] However, with an average spending of 1.6% of GDP, coverage is far from universal, and governments and development practitioners often use targeting tools in an effort to concentrate the benefits of SSNP on the poorest. Poor households targeting is an inherently inexact and challenging practice, especially in developing countries which face a lack of verifiable records on earnings. This lack of records makes means-testing impractical in the context of developing countries. Consequently, Proxy Means Testing (PMT), which does not rely on directly observed income, has become an increasingly popular targeting method. PMT is currently implemented in SSNP in Sub-Saharan African countries such as Burkina Faso, Ethiopia, Ghana, Malawi, Niger, Nigeria and Tanzania.[2]

In PMT targeting, a survey-based measure of well-being (usually consumption) is regressed on household covariates to estimate a proxy for well-being, and this proxy is in turn used for targeting out of the sample. In particular, the implementation of PMT has two distinct phases. First, an in-depth survey is administered to a *sample* of households to collect data on consumption as well as some easily observable and verifiable correlates of consumption (such as demographic characteristics and home attributes). These data are used to estimate a regression of log consumption per capita on correlates of consumption. Second, a short survey is administered to *all* potential beneficiary households to collect information on the same correlates of consumption, compute PMT scores based on coefficients estimates, and determine the list of beneficiaries based on resulting PMT scores.

Recently, PMT has been the subject of a lively debate among policy makers, civilian stakeholders and academics. The most debated issue is probably the claim that PMT is one of the best mechanisms, if not the best mechanism available for identifying households living in poverty. PMT advocates argue that it "can accurately and cost-effectively target the chronic poor" (Del Ninno and Mills, 2015, p.20), and that its performances can be improved by combining PMT with other targeting methods such as demographic targeting, geographical targeting or CBT (Coady et al., 2004; Grosh et al., 2008; Leite, 2014; Stoeffler et al., 2016). A recent

[1] ASPIRE database–`www.worldbank.org/aspire`.

[2] In addition to means tests and PMT, other methods include demographic targeting (targeting of specific categories such as elderly, widowed and children), community-based targeting or CBT (groups of community leaders and members determine eligibility), geographic targeting (location determines eligibility) and self-targeting (benefits and transaction costs are set so that only needy households enrol). For a detailed overview on PMT and other targeting methods used in developing countries see Grosh (1994); Grosh et al. (2008); Del Ninno and Mills (2015); Devereux et al. (2017).

World Bank report in Namibia recommends the use of PMT to target beneficiaries of social benefits because it "could provide better coverage at existing spending levels, providing a greater poverty and inequality impact" (Sulla et al., 2017). In contrast, critics often point to PMT high built-in errors, implementation issues and lack of transparency (Kidd and Wylde, 2011).[3] Kidd et al. (2017) argue PMT "is best understood as a rationing mechanism, attempting to select households in a context of limited resources in a moderately 'pro-poor' manner, while excluding the majority of those in need".

This debate has been feeded by a surge of careful studies assessing the performances of targeting methods. In these studies, performances are typically displayed in terms of "errors of inclusion" (providing benefits to households which should not be eligible) and "errors of exclusion" (not providing benefits to households that should be eligible). Brown et al. (2016) provide a systematic assessment of PMT performances for nine countries in Sub-Saharan Africa (SSA). Overall, they find that PMT reduce inclusion errors but this comes at the cost of substantial exclusion errors. Comparisons of PMT with CBT suggest some gains in terms of accuracy but some loss in terms of community satisfaction with the beneficiary list (Alatas et al., 2012; Basurto et al., 2017; Karlan and Thuysbaert, 2016; Stoeffler et al., 2016). For example, Alatas et al. (2012) in Indonesia report that PMT allowed a 10 percent reduction in the error rate relative to CBT, while CBT resulted in 60 percent fewer complaints than PMT.[4]

Assessments of PMT performances rely (often implicitly) on the assumption that consumption data underlying PMT regressions are error-free or measured with random error. However, this assumption has been challenged by Gibson et al. (2015) which showed that measurement errors have a mean-reverting negative correlation with true consumption. According to the typical textbook on the impact of measurement errors, a violation of this assumption would lead to biased PMT estimates.[5]

[3]Technical reports commisionned by development agencies outlined many narratives about communities' perceptions of PMT's stated errors and lack of transparancy. Adato and Roopnaraine (2004), in Nicaragua, argue that "the targeting process as a whole is poorly understood at the community level in both geographical- and household-targeted communities. When asked why some households were beneficiaries and others not, informants offered a range of explanations, from divine intervention to a random lottery. For example, one informant from a geographically-targeted community noted: 'Well, some people wonder why they weren't targeted even though they live in this same area. So we tell them that the Bible says that many are called but few are chosen."' Fitzgibbon (2014), in Northern Kenya, reports "bafflement as to what basis the computer had selected people" and cites informants: "See this old lady she is totally blind and lives by herself with no family, when we ranked households in this village she was number 1 (the poorest) yet she is not on the list"; "We don't want a computer to pick our poorest we know better than anyone else who is needy here and we should be able to identify them."

[4]Some studies assess PMT targeting outcomes beyond accuracy and satisfaction. Cameron and Shah (2013) show that PMT had significant negative social consequences such as an increase in the prevalence of crime within communities and a decline of the participation in community groups. In the context of a subsidy program in Malawi, Basurto et al. (2017) report that local leaders allocate input subsidies to farmers with larger returns.

[5]See Bound et al. (2001) for a good discussion on the impact of measurement errors on regression estimates.

The goal of this paper is to explore the impact of non-random measurement errors in consumption on PMT performances. As with many impact evaluation, the key challenge here is to construct the most credible counterfactual of what would happen with random measurement errors in consumption. I exploit a unique survey experiment which randomly assigned eight different designs of consumption module to more than 4000 households in Tanzania. This experiment has been used to explore the relative performances of different survey designs in terms of mean consumption, inequality, poverty, the prevalence of hunger and measurement errors (Beegle et al., 2012; De Weerdt et al., 2016; Friedman et al., 2016; Gibson et al., 2015), but never with an explicit focus on the implications for targeting accuracy. One design of the consumption module involved the distribution of individual diaries to each adult member of households to track all commodity in-flows (harvests, purchases, gifts, destocking) and outflows (sales, gifts, restocking, food fed to animals). In addition, each adult member was provided with tight supervision by interviewers specifically trained to cross-check and query reported information. This resource intensive design is believed to approximate a "gold standard" for consumption estimates in that it minimizes the prevalence of various sources of measurement errors. My empirical strategy compares the performances of PMT relying on the gold standard consumption data with those of PMT using the more error-prone consumption data.

This paper contributes to the lively debate on the methods to target poor households. It provides empirical evidence on one largely ignored aspect of PMT targeting, namely its vulnerability to non-random measurement error in survey-based consumption data. I estimate that both PMT scores and PMT targeting accuracy are reduced in the presence of non-random errors in consumption. Specifically, the targeting error rate increases by a magnitude ranging from 12 to 33%, depending on how consumption data is collected and how the poverty line is defined. Overall, the results show that the targeting accuracy of PMT has been significantly overestimated in the literature. This reduction in PMT accuracy is non trivial and has the potential to influence assessments of PMT performances and their comparisons with other targeting methods. For example, the 10 percent lower targeting error rate of PMT relative to CBT that Alatas et al. (2012) found in Indonesia stands on the lower-hand of my estimates.

It is always difficult to extrapolate the results derived from one context and the findings presented in this article may not hold in other contexts. However, I should mention that my focus is on measurement errors due to survey design, which are more likely to have external validity. This provides some reassurance that the results are not too specific. In addition, measurement errors due to data fabrication by interviewers are not considered here, meaning that these findings should be interpreted as lower bounds of the impact of measurement errors on PMT performances.

The remainder of the paper is organised as follows. Section 2 walks through a number of

In section 2.2 below, I present in more details how Bound et al. (2001) speak to the present study.

error sources that can be expected to arise when measuring consumption and how some of these sources likely differ by survey design. That section also introduces the expected impact of measurement error on PMT performances. Section 3 describes the experimental set-up. Section 4 presents the empirical strategy. Section 5 reports the findings. Section 6 draws the main conclusions.

2 Measurement Errors and PMT Performances

Before moving on to a description of the survey experiment, I briefly introduce some of the evidence found in the literature on measurement errors in household consumption. Consumption is generally used in PMT as a measure of household well-being, i.e. the dependent variable in PMT regressions, so I will focus on how measurement error in the dependent variable is expected to affect model estimates, and thus how PMT performances may be impacted by errors in consumption.

2.1 Consumption measurement errors

Consider the following typical survey questions about some consumed item X:

> "How much X did your household consume in the past 14 days? How much came from purchases? How much did you spend? How much came from own-production? How much came from gifts and other sources?"

Often, individuals trying to answer these questions will struggle to give accurate figures, leading to imprecise data.

Why should one expect consumption estimates to deviate from actual consumption?[6] First, it is well documented in the literature that retrospective reports on expenditures can cause both recall and telescoping errors. In the former, a household under-reports true consumption over the period of recall because it may forget certain events. The longer the period of recall the greater the likelihood events are forgotten or not precisely remembered. In the latter, a respondent recalls consumption that occurred before the reference period, resulting in reported

[6] I only consider deviations that are caused by the insufficient ability of respondents to acquire, process and recall information. However, it should be noted that deviations can also arise from other sources, such as social desirability bias (e.g. under-reporting of "bad" consumption such as spending on alcohol or cigarettes), strategic responses (e.g. understatements of consumption because of the belief that responses may be used to determine eligibility for some future social program; negative answers bias in order to avoid follow-up questions) and untrained, inadvertent or strategic enumerators (e.g. enumerators guiding respondents to give answers that minimize interview length).

consumption greater than the actual one. A second source of error is the inability of respondents to accurately report individual consumption by other household members, which may be particularly salient in the context of SSA where households are larger and the unitary model has been challenged empirically.[7] This source of error is likely to be more compelling for certain types of consumption such as alcohol, tobacco, meals eaten outside the home, telecommunication or personal toiletries. Lastly, for longer recall periods or items involving frequent transactions, respondents may resort to inference rather than memory to estimate consumption, resulting in what can be termed rule of thumb errors. This source of error has no obvious direction of bias but it is probably more important in hypothetical scenarios requiring high cognitive readiness such as consumption during a usual month.

These various sources of errors may be more or less prevalent depending on the features of data collection instruments used. In recent years, a number of empirical studies confirmed that measurement of consumption is fairly sensitive to survey design. I focus here on evidence on four key dimensions in which survey design vary: the method of data capture (diary versus recall questionnaires), the length of the recall period, the number of items on which data are collected and the level of respondent (individual versus household). This focus is motivated by the specific experiment exploited in this paper and described in the next section, which randomly assigned households to eight survey designs differing along the four dimensions above-mentionned.[8]

While diaries are generally believed to overcome some sources of error such as recall errors or rule of thumb errors, some concerns related to their implementation in the field exist. Specifically, in the case of illiterate, unmotivated or non-cooperative respondents, a diary survey with a lack of supervision may be equivalent to a recall survey if the information is gathered by the enumerator at the end of the period. In a survey in Canada, where households reported their food expenditures during the past month and then filled in a diary during the following two weeks, Ahmed et al. (2006) identify substantial measurement errors in recall food consumption with properties inconsistent with classical measurement error. However, it also found some discrepancies in the diary survey and concludes that the "superiority of the diary may not be as obvious as the literature suggests." Implementation of diary in developping countries may be even more challenging. Beegle et al. (2012) mention stylized facts from two diary household budget surveys in Tanzania (2002) and Malawi (1998) consistent with poor supervision, respondent fatigue and incomplete or unreliable data. The same authors wrap up that "the implications of variation in literacy, motivation, and other factors, although not well-documented, suggest it can be quite difficult to conduct high-quality diary survey."

[7] Anderson and Baland (2002) and Duflo and Udry (2004) are two examples of empirical evidence inconsistent with the unitary model of household decision making, which assumes all household members have the same utility function.

[8] For more detailed discussions on the sensitivity of consumption expenditures to survey design, see for instance Deaton (1997), Deaton and Grosh (2000), Gibson and Kim (2007) and Beegle et al. (2012).

There is a wide understanding that an inverse relationship exists between the length of time over which respondents are asked to recall events and the accuracy of the reported estimates. Events are less likely to be precisely remembered with time due to recall errors and telescoping (see above). While these errors work in opposite directions, experimental studies of self-reported consumption show that under-reporting is more widespread than over-reporting. In an experiment in Ghana, Scott and Amenuvegbe (1991) varied recall periods and find that the reported spending on a basket of the 13 most frequently purchased items decreased by 2.9% for every additional day of recall. Similarly, Beegle et al. (2012) in Tanzania report that a 7-day recall design measured a 11% higher mean food consumption than a 14-day recall design. Hence, both studies suggest that telescoping is less prevalent than recall error.

Shorter versus longer lists of items included in questionnaires has also been shown to influence consumption estimates. Observational work by Lanjouw and Ravallion (1996) in Ecuador estimated a decline in poverty of seven percentage points between 1994 and 1995 while the country did not experience ant policy to reduce poverty nor significant growth, suggesting that the observed decline in poverty was more related to the change of design in the questionnaire (more than 25 percent additional items was added between the two survey rounds). This positive relationship between the number of items and the level of recorded consumption has been confirmed by experimental work in El Salvador by Jolliffe (2001), where longer, more detailed questions on consumption resulted in an estimate of mean household consumption that was 31% higher than estimates derived from a condensed version of the questionnaire.

Finally, who is responding to survey questions may influence consumption records due to the difficulty for a sole respondent to perfectly capture the consumption by other household members for items such as alcohol, tobacco, meals eaten outside the home, telecommunication or personal toiletries. As reported by Beegle et al. (2012), personal diaries have been used in Russia for a random sample of households during the 2003 Household Budget Survey, and this yielded 6–11% higher expenditure levels, even if the survey was plagued with non-respondent problems.

These examples of diverging consumption estimates when different survey designs are used in the same setting are indicative of measurement error. However, because of a lack of data on actual consumption, there is only scant evidence on the nature of measurement error in estimates of household consumption. Gibson et al. (2015) made huge efforts to field a survey experiment and collect benchmark consumption data allowing them to make such investigations.[9] They find that errors in measured consumption are non-random and negatively correlated with true values–a pattern that Bound and Krueger (1991) also found for earnings data and labelled *mean-reverting measurement error*. In what follows, I present how non-random

[9] As noted above, this paper rests on the same data as Gibson et al. (2015). More details on the design of the survey are presented in the next section.

error in consumption may affect PMT performances.

2.2 The impact of non-random error in the dependent variable on parameter estimates

A significant amount of attention has been devoted to measurement error and its effects on model estimates. Because this paper is primarily interested in measurement error in consumption, which is used as a left-hand-side variable in PMT regressions, I confine attention to the impact of errors in the dependent variable.[10] Assume the true model is:

$$y = \alpha + \beta X + \varepsilon \tag{1}$$

where y is the dependent variable, X a vector of independent variables, β the associated coefficients and ε a pure random error which is assumed to be uncorrelated with X.[11] Instead of y, the observed value of the outcome variable is y^*, which is related to the true value y by:

$$y^* = \theta + \lambda y + v \tag{2}$$

The estimator of the response coefficient with the error-ridden dependent variable is:

$$\beta_{y^*X} = \frac{cov(y^*, X)}{var(X)} = \frac{cov(\lambda\alpha + \lambda\beta X + \lambda\varepsilon - v, X)}{var(X)} \tag{3}$$

One has to assume random error in order to get consistent estimates of β from equation 3. Random error is a special case which adds variability to the data but does not affect average performance for the sample. The following assumptions are made under random error: $\theta = 0$, $\lambda = 1$ and $E(v) = cov(y, v) = cov(X, v) = cov(\varepsilon, v) = 0$. In contrast, mean-reverting measurement error in y^* assumes $0 < \lambda < 1$ which makes estimates of β inconsistent: from equation 3 it is now equal to $\lambda\beta$.

Thus, with $0 < \lambda < 1$, estimates of equation 1 will be attenuated. In other words, mean-reverting measurement error in consumption data is expected to bias downward the coefficients of consumption correlates derived from PMT estimates. As noted in the introduction, some general assessments of PMT targeting are already available in the literature. However, I am not aware of any previous work looking at the severity of this bias and to what extent it affects PMT performances. One exception is Brown et al. (2016), which exploit panel data in Ethiopia, Malawi, Nigeria, Tanzania and Uganda to reduce any bias due to measurement errors. They use time-mean consumption instead of current consumption and find that PMT performances

[10] The framework presented in this section is adapted from Bound et al. (2001), Hausman (2001) and Gibson et al. (2015).

[11] This assumption is motivated by a focus on the impact of measurement error on PMT estimates, meaning that the question of whether ε is correlated with X is not crucial here.

slightly improve. However, as noted by the authors themselves, this exercise is unlikely to completely eliminate measurement errors.[12]

3 The Survey Experiment

As mentionned in the introduction, I exploit the same survey experiment as Beegle et al. (2012); De Weerdt et al. (2016); Friedman et al. (2016); Gibson et al. (2015). It is a unique experiment developed by the Living Standards Measurement Study (LSMS) Team in the World Bank in collaboration with the University of Dar es Salaam and the Economic Development Initiatives (EDI), a leading research company established in 2002 in Tanzania. This section summarizes the experiment and its implementation. More details can be found in Beegle et al. (2012).

3.1 Sample

The sample for the experiment consists of 4,032 households spread across seven Tanzanian districts: one district in the regions of Domona, Pwani, Dar es Salaam, Manyara, and Shinyanga and two districts in the Kagera Region. While the districts in the regions of Dodoma and Dar es Salaam are urban areas, other districts are rural.[13] Within these seven districts, a probability-proportional-to-size sample of 24 villages was selected using data from the 2002 Census. In each selected village, Enumeration Areas (EA) were listed in cooperation with local informants, and one of these EA was randomly chosen for the experiment. These EA are best thought of as sub-villages or neighbourhoods. Finally, in each selected EA, all households were listed, and 24 households were randomly sampled for the survey experiment. According to Beegle et al. (2012), "the sample was constructed to be representative at the district level, but not at the national level," however "the basic characteristics of the sampled households generally match the nationally representative estimates from the 2006/2007 Household Budget Survey."

[12] Griliches and Hausman (1986) argue that what matters in such cases is the correlation over time in the true values of the dependent variable (y in equation 1) and in the measurement errors (v in equation 2). Specifically, if true values of y are highly correlated over time, while the measurement errors v are more or less uncorrelated, moving from cross-sectional estimates to panel estimates would actually intensify the bias due to measurement errors in y.

[13] According to Beegle et al. (2012), "districts were purposively selected to capture variations between urban and rural areas as well as across socio-economic dimensions to inform survey design related to labor statistics and consumption expenditure for low-income settings." Table 1 shows basic descriptive statistics.

3.2 Experimental design

In each sub-village, three households were randomly assigned to each of the eight consumption modules summarized in Table 3. Each household was assigned to a single module to prevent potential cross-module spillovers. The designs of these eight modules vary along five key dimensions: the method of data capture (diary versus recall questionnaires), the length of the recall period, the number of items in the recall list, the level of respondent (individual versus household) and the degree of supervision received. These eight survey designs were strategically selected to reflect the most common methods used in low-income-countries and the scope of variation one is likely to find in practice (Beegle et al., 2012).

Modules 1–5 relied on a recall design and modules 6–8 on diaries. Modules 1 and 2 used a long list of 58 commodities with a recall period of 14 and 7 days respectively. Module 3 used a subset list consisting of the 17 most important commodities and representing 77 percent of the food consumption expenditure in Tanzania (based on the national Household Budget Survey 2000-2001).[14] Module 4 included a list of 11 comprehensive categories, which was an aggregated version of the list of 58 commodities. Module 5 asked for "usual" consumption over the list of 58 commodities. In particular, households reported the number of months in which the commodity is typically consumed, the quantity usually consumed, and the average value of what is consumed in those months. Module 6 and 7 were both household diaries (i.e. a single diary was used to record all household consumption) with different frequency of the supervision received. Households allocated to module 6 were visited by a trained survey staff every other day, while those allocated to module 7 were only visited weekly. Module 8 was a personal diary in which each adult member was provided with his or her own diary while children were placed on the diaries of the adults who knew most about their daily activities. Each adult was visited every other day.

Non-food items were divided into two categories based on frequency of purchase. Frequently purchased items such as charcoal, soap, cigarettes and communications were collected using a 14-day recall period for modules 1–5 and the 14-day diary for modules 6–8. Non-frequent expenditures such as durables, education and health were collected using the same design across modules (i.e. a one or 12-month recall period depending on the item in question).

3.3 Data

The data were collected between September 2007 and August 2008 by EDI. Each interviewer implemented all eight modules in equal proportion in order to avoid to confound module effects with interviewer effects. In each EA, households assigned to the recall modules were surveyed in

[14] To make data comparable, reported expenditures for module 3 were scaled up by a factor equal to 1/0.77, as is commonly done in practice.

the span of the 14 days the survey team was in the EA to collect the data based on the diaries. Interviewers were provided with an extensive training starting in June 2007 and including intensive sessions on how to check and correct individual diaries for the issue of double-counting. The survey was administered on paper but maximum control was made possible by the relatively small number of a dozen interviewers and the long 12-month period of data collection. Specifically, back-checks as well as direct observations were carried out on regular basis by supervisors. The same double blind data entry protocol was used for all modules in order to avoid any systematic error to arise and bias the results. Refusal and attrition were negligible: there were only 13 replacements due to refusals and only three households that started a diary were dropped because they did not complete their final interview. Another five households were dropped because of missing data, yielding a final sample size of 4,025 households.

4 The Empirical Strategy

I seek to quantify the impact of measurement error in consumption on PMT estimates and to assess how PMT performances are impacted. As with many impact evaluation, the key challenge here consists in constructing the most credible counterfactual of what would happen without measurement errors in consumption. Ideally, we would like to have error-free *and* error-prone consumption data for each household. Most studies on measurement error rely on validation data such as administrative records for income (Bound and Krueger, 1991). However, the lack of data on actual consumption makes validation studies impractical for consumption. The survey experiment described in section 3 offers a rare opportunity to study measurement errors in consumption.

4.1 Identification strategy

While in the survey experiment there are no validation data available, the personal diary (module 8) is believed the most accurate and to approximate a "gold standard" or a "benchmark". In the personal diary, there is a smaller scope for recall errors, telescoping and missed individual consumption. In addition, three measures have been undertaken to avoid double counting—the main stated weakness of personal diaries. First, the personal diary has been designed as a record of food brought into the household instead of food consumed, which is likely to reduce the scope for double-counting purchased or self-produced items. Second, as discussed, interviewers were trained to cross-check individual diaries for similar items and apply the appropriate corrections when the same item was accidentally recorded by two individuals. Third, each adult member was visited every other day in order to provide him or her with adequate supervision. Reassuringly, some statistics, such as the daily consumption, show no diary fatigue.

My identification strategy exploits this benchmark consumption and the random assignment of the different survey designs across households. Table 2 shows the results of randomization balance checks across a set of core household characteristics. Overall, randomized assignment of households to the eight different designs was successful (Beegle et al., 2012). Any systematic difference in measured consumption across modules can be attributed to measurement error due to alternative survey designs, thus comparisons of error prone survey designs with the benchmark identify the effect of measurement error on PMT targeting.

4.2 Estimation procedure and construction of the outcomes of interest

While I recognize that poverty is multidimensional in nature, I rely on per capita consumption as the main welfare indicator for the analysis because it is generally considered as a good predictor of neediness (Deaton, 1997) and because it is used in most PMT targeting exercises.[15] Per capita consumption is aggregated on an annual basis using data collected on food consumption and frequent non-food consumption.[16] Total food consumption from module 3 is scaled up by a factor equal to 1/0.77 (i.e. 29.87%) to make data comparable across modules.

First, I create a set of 25 variables that are long-term determinants or correlates of poverty, encompassing household's demographic characteristics (household size, number of children, etc.), home attributes (floor type, wall type, etc.) and household head's features (education, occupation, etc.). Then, using OLS, I estimate the relationship between this set of variables and log consumption per capita (the so-called PMT formula) by module type. The following regression is estimated eight times (one for each module type k):[17]

$$y_i = \alpha_k + \beta_k x_i + \varepsilon_i \tag{4}$$

where y_i is the log consumption per capita of household i (with $i = 1, ..., N_k$), N_k the sample size of households assigned to module k (with $k = 1, ..., 8$) and x_i the set of 25 correlates of

[15] I have also done the analysis using consumption per adult equivalent and the results are similar (results available upon request).

[16] I also considered the option of using only food consumption and results are similar. In both cases, the consumption of non-frequently purchased items such as durables, education and health was excluded because it was collected using the same design across modules (i.e. there is no benchmark) and because it is usually not included in PMT. That said, measurement error for these items is likely more prevalent because of the longer recall period (one month or 12 months depending on the items considered), and hence, the impact of measurement error on PMT performances would be more pronounced if non-frequent consumption was included. Unfortunately, I am not able to check this assumption because there are no benchmark data on actual non-frequent consumption.

[17] In an extended version, I also include variables on assets and livestock which are good correlates of consumption but are more difficult to verify and may be vulnerable to strategic responses. The results are largely similar.

consumption I created. Estimates from equation 4 are then used to predict PMT scores of household i for each PMT formula k:

$$\hat{y}_{ik} = \hat{\alpha}_k + \hat{\beta}_k x_i \tag{5}$$

Note that each PMT formula is used to compute PMT scores "in and out of sample" (e.g. formula 1 is used to predict PMT scores of households assigned to module 1 but also for the sample of households assigned to the other modules). As a result, I obtain eight PMT scores per household (one from each of the eight PMT formulas) which form the basis to assess the impact of measurement errors on PMT performances. Then, I restrict the analysis to the sample of households assigned to the personal diary (module 8), and compare how each formula perform to predict measured consumption and which households are poor. Under the identifying assumption that the personal diary approximates a benchmark for true consumption, formula 8 can be interpreted as the closest to the counterfactual scenario, i.e. the PMT formula one would obtain if consumption was measured without errors.

In a first part, I compare the predictive performances of the alternative PMT formulas. I estimate an equation of the following form using a linear probability model:

$$\hat{y}_{ik} = \gamma_k M_{ik} + v_{ik} \tag{6}$$

where $\hat{y}_{ik}$ is the PMT score of household i derived from formula k (see equation 5) and M_{ik} a vector of dummy variables indicating if $\hat{y}_{ik}$ is derived from formula k. I also compute the mean squared prediction error $\hat{\mu}_{ik} = (y_i - \hat{y}_{ik})^2$, where y_i is the actual consumption, and regress it on the same variables:

$$\hat{\mu}_{ik} = \gamma_k M_{ik} + v_{ik} \tag{7}$$

In both estimates, standard errors are clustered at the village level to account for the correlation between the error terms of observations from the same village. The comparison of γ_k ($k = 1, ..., 7$) with γ_8 will give a sense of the impact of measurement errors on PMT predictive performances by survey design.

In a second part, I compare the performances of the alternative PMT formulas against different measures of targeting accuracy. As discussed in the introduction, there are two types of targeting errors: Inclusion Errors (IE), i.e. identifying a non-poor household as poor, and Exclusion Errors (EE), i.e. identifying a poor household as non-poor. The Inclusion Error Rate (IER), defined as the proportion of the non-poor households identified as poor, for module k, can be written as:[18]

$$IER_k = \frac{\sum_{i=1}^{N_k} \mathbb{1}(\hat{y}_{ik} \leq z \mid y_i > z)}{\sum_{i=1}^{N_k} \mathbb{1}(y_i > z)} \tag{8}$$

where N_k is the sample size, z the poverty line, y_i the measured per capita consumption of household i, $\hat{y}_{ik}$ its PMT score using PMT formula k and $1(.)$ an indicator function which takes the value one when the condition in parentheses is true and zero otherwise. Similarly, the Exclusion Error Rate (EER), defined as the proportion of the poor households not identified as poor, can be written as:

$$EER_k = \frac{\sum_{i=1}^{N_k} \mathbb{1}(\hat{y}_{ik} > z \mid y_i \leq z)}{\sum_{i=1}^{N_k} \mathbb{1}(y_i \leq z)} \tag{9}$$

The IER and the EER do not consider how far from the poverty line beneficiary and non-beneficiary households lie. For instance, the EER would be the same if a given household i, excluded by mistake, is just below or very far below the poverty line. Hence, mean squared error, which allocate a higher weight for errors farther from the poverty line, is perhaps richer for measuring targeting errors. The Mean Squared IE (MSIE) and Mean Squared EE (MSEE) for module k are given by:

$$MSIE_k = \frac{\sum_{i=1}^{N_k} \mathbb{1}(\hat{y}_{ik} \leq z \mid y_i > z) * (z - y_i)^2}{\sum_{i=1}^{N_k} \mathbb{1}(y_i > z)} \tag{10}$$

$$MSEE_k = \frac{\sum_{i=1}^{N_k} \mathbb{1}(\hat{y}_{ik} > z \mid y_i \leq z) * (z - y_i)^2}{\sum_{i=1}^{N_k} \mathbb{1}(y_i \leq z)} \tag{11}$$

The IER, the EER, the Targeting Error Rate (TER), defined as the weighted sum of the IER and the EER (weights are the share of poor/non-poor households), the MSIE, the MSEE and the MSTE, defined as the weighted sum of the MSIE and the MSEE, form the basis to assess the targeting performances of the alternative PMT formulas. From the rates and means defined hereinabove, I construct variables which can fit in typical regression frameworks. Specifically, for the IER, the EER and the TER, I create dummies equal to one if household i with consumption derived from formula k is mistargeted, and zero otherwise. For instance, IE_{ik} is equal to one for all households i which are considered as poor by mistake using PMT formula k. Similarly, for the MSIE, the MSEE and the MSTE, I create variables equal to the squared targeting error if household i with consumption derived from PMT formula k is mistargeted, and zero otherwise. For instance, IE_{ik}^2 is equal to the squared inclusion error (i.e.

[18] I use the same definition as Alatas et al. (2012). IER could also be defined as Brown et al. (2016), i.e. the proportion of those identified as poor who are not poor. I argue that the latter definition is less practical in the present study. Indeed, the sample of households identified as poor is likely to vary across PMT formula. However, as a robustness check, I adapt the definition of Brown et al. (2016) by including each household considered as poor from at least one PMT formula in the sample of poor households. This does not change the sign and magnitude of the results.

$(y_i - \hat{y}_{ik})^2)$ for all households i which are considered as poor by mistake using PMT formula k. Each of these outcomes of interest is estimated with the same specification as equations 6 and 7. Importantly, the poverty line z in equations 8–11 can be defined in absolute or in relative terms. With a poverty line defined in absolute terms, e.g. PPP$1.25, beneficiaries are those with a PMT score below PPP$1.25. With a poverty line defined in relative terms, e.g. the poorest 30%, beneficiaries are those with a PMT score equal or below the PMT score of the 30th percentile. I start by assessing PMT targeting performances with respect to the typical PPP$1.25 poverty line. In practice, because of fixed resources and operational constraints, the poverty line is often defined in relative terms in SSNP, so I also apply a typical 30% threshold.[19] Specifically, for each PMT formula, I rank households from lowest to highest PMT scores and consider as eligible those with PMT scores equal or below the PMT score of the 30th percentile.

5 Results

Table 4 presents the results of PMT regressions by module type. Adjusted-R^2 ranges from 0.48 for the sample assigned to household diary with infrequent supervision (module 7), to 0.67 for the sample assigned to the usual month recall (module 5). Column 9 displays the PMT on the full sample of households and Adjusted-R^2 is 0.56, which is similar to the 0.59 obtained by Brown et al. (2016) in Tanzania using LSMS-ISA data, and somewhat higher than the 0.40 obtained in Indonesia by Alatas et al. (2012).

5.1 PMT predictive performances

The impact of measurement error in consumption on the predictive performances of PMT is highlighted in Table 5. Overall, formulas 1–7 yield significantly lower PMT scores and higher squared prediction errors than formula 8 (which is derived from the benchmark personal diary). The results in column 1 show that formulas 1–7 predict between 6 and 28% lower PMT scores compared with the benchmark formula 8. Similarly, formulas 1–7 produce mean prediction errors between 16 and 51% higher compared with the benchmark. PMT formulas derived from the long list 7-day recall (module 2) and the subset list (module 3) appear to yield slightly better predictions compared with formulas derived from the collapsed list (module 4) and usual month (module 5) appear to perform worse.

[19] As a robustness check, I also use a 50% threshold and results are very similar.

5.2 PMT targeting performances

Regressions in Table 6 compare the targeting performances of each of the seven formulas against the benchmark formula derived from the sample of households assigned to the personal diary (module 8), using the PPP$1.25 poverty line. The results in column 1 show that measurement error in formulas 1–7 increases the TER by a magnitude ranging from 2.98 and 8.35 percentage point. Given that the TER derived from formula 8 is 23.5%, these effects are equivalent to an increase in TER ranging from 13 to 33%. In columns 2 and 3, I examine the error rates separately for the non-poor and the poor (defined as the households below vs. above the PPP$1.25 poverty line). The results show that the IER increase and the EER decrease for all formulas compared with formula 8, which is not surprising given that I found in table 5 that formulas 1–7 predict lower PMT scores. This means that the number of poor households is overestimated when formulas 1–7 are used.

I further investigate whether this pattern (higher TER and EER and lower IER) hold when the poverty line is defined in relative terms, using a typical 30% threshold. The results can be seen in Table 7 and show an increase in TER ranging from 2.39 to 4.37 percentage point using formulas 1–7, which is equivalent to a 12 to 23% increase in TER. This time both the IER and the EER are higher but the coefficients are mostly statistically insignificant (columns 2 and 3). I find IER between 1.70 and 3.12 percentage point higher (equivalent to a 12 to 25% increase in IER). EER are between 3.97 and 7.28 percentage point higher (equivalent to a 10 to 18% increase in EER). These results provide suggestive evidence that measurement errors in consumption not only affect the probability that a household is deemed poor but also the overall distribution of poor households. In other words, measurement errors in consumption seem to have implications on the distribution of PMT scores.

In Table 8, I present the results by quartile of consumption. While targeting errors seem to increase for almost all formulas and all quartiles of consumption, the magnitude and statistical significance of the results are higher for the poorest households, with an increase in EER ranging from 5.56 to 8.73 percentage point (equivalent to 15 to 23%).

Finally, the results presented columns 4–6 in tables 6 and 7 suggest the effects of measurement error in consumption on the MSTE, the MSIE and the MSEE are similar to those found for the TER, the IER and the EER, with overall increases—even though the estimated coefficients are lack significance.

How should one interpret these results in terms of "economic" significance? Alatas et al. (2012) found that PMT targeting reduced the TER by three percentage point (10%) compared with CBT, which is in the order of magnitude of my estimate of the effect of measurement error on TER. Hence, it seems the impact I found is non trivial in that the conclusions of targeting methods assessments could be influenced.

6 Conclusion

In this paper, I have determined the impact of non-random measurement error on PMT performances. Assessments of PMT performances rely (often implicitly) on the assumption that consumption data underlying PMT regressions are error-free or measured with random error—an assumption that has been challenged by recent literature. Using a unique survey experiment in Tanzania, I show that the presence of non-random measurement error in consumption reduces PMT accuracy. Specifically, I estimate that measurement error increases the TER by a magnitude ranging from 12 to 33%, depending on how consumption data is collected and how the poverty line is defined. I argue that this reduction in PMT accuracy is non trivial and has the potential to influence assessments of PMT performances and their comparisons with those of other targeting methods. For instance, Alatas et al. (2012) found that PMT targeting caused a three percentage point (10%) lower TER compared with CBT, which is in the lower-hand of my estimates of the effect of measurement error on TER.

Although these results provide us with evidence on the vulnerability of PMT to measurement errors, some unresolved questions remain. First, I only considered measurement errors in the dependent variable, while measurement errors in the independent variables could also have impact on PMT performances. Another issue I have not addressed in this paper is that of assessing the vulnerability of PMT to data fraud or data fabrication by interviewers. I chose instead to focus on measurement errors due to survey design, which are likely to have more external validity. However, the problem of data fabrication in surveys has been shown to be prevalent (Finn and Ranchhod, 2017) and documenting the vulnerability of PMT to this other form of measurement errors should be a priority for future work.

Nevertheless, the results presented here provide empirical evidence on one largely ignored aspect of PMT, namely its vulnerability to non-random errors due to survey design. The results may be of relevant interest to researchers in their assessments of PMT performances or in their comparisons of the different targeting mechanisms available. It also has implications for development practitioners and governments designing the targeting devices of the many SSNP implemented in developing countries.

References

Adato, M. and Roopnaraine, T. (2004). A social analysis of the Red de Protección Social (RPS) in Nicaragua. IFPRI report, Washington, DC. p.15.

Ahmed, N., Brzozowski, M., and Crossley, T. F. (2006). Measurement errors in recall food consumption data. Working Paper 06/21, Institute for Fiscal Studies (IFS).

Alatas, V., Banerjee, A., Hanna, R., Olken, B. A., and Tobias, J. (2012). Targeting the poor: Evidence from a field experiment in Indonesia. *The American Economic Review*, 102(4):1206–1240.

Anderson, S. and Baland, J.-M. (2002). The economics of roscas and intrahousehold resource allocation. *The Quarterly Journal of Economics*, 117(3):963–995.

Basurto, P. M., Dupas, P., and Robinson, J. (2017). Decentralization and efficiency of subsidy targeting: Evidence from chiefs in rural Malawi. Working Paper 23383, National Bureau of Economic Research.

Beegle, K., De Weerdt, J., Friedman, J., and Gibson, J. (2012). Methods of household consumption measurement through surveys: Experimental results from Tanzania. *Journal of development Economics*, 98(1):3–18.

Bound, J., Brown, C., and Mathiowetz, N. (2001). Measurement error in survey data. *Handbook of Econometrics*, 5:3705–3843.

Bound, J. and Krueger, A. B. (1991). The extent of measurement error in longitudinal earnings data: Do two wrongs make a right? *Journal of Labor Economics*, 9(1):1–24.

Brown, C., Ravallion, M., and Van De Walle, D. (2016). A poor means test? Econometric targeting in africa. Working paper 22919, National Bureau of Economic Research.

Cameron, L. and Shah, M. (2013). Can mistargeting destroy social capital and stimulate crime? Evidence from a cash transfer program in Indonesia. *Economic Development and Cultural Change*, 62(2):381–415.

Coady, D., Grosh, M. E., and Hoddinott, J. (2004). *Targeting of transfers in developing countries: Review of lessons and experience*, volume 1. World Bank Publications.

De Weerdt, J., Beegle, K., Friedman, J., and Gibson, J. (2016). The challenge of measuring hunger through survey. *Economic Development and Cultural Change*, 64(4):727–758.

Deaton, A. (1997). *The analysis of household surveys: a microeconometric approach to development policy.* World Bank Publications.

Deaton, A. and Grosh, M. (2000). Consumption. In *Designing Household Questionnaires for Developing Countries: Lessons from Fifteen Years of the Living Standard Measurement Study*, volume 15, pages 91–133.

Del Ninno, C. and Mills, B. (2015). *Safety Nets in Africa: Effective Mechanisms to Reach the Poor and Most Vulnerable.* World Bank and Agence Française de Développement Publications.

Devereux, S., Masset, E., Sabates-Wheeler, R., Samson, M., Rivas, A.-M., and te Lintelo, D. (2017). The targeting effectiveness of social transfers. *Journal of Development Effectiveness*, 9(2):162–211.

Duflo, E. and Udry, C. (2004). Intrahousehold resource allocation in Cote d'Ivoire: Social norms, separate accounts and consumption choices. Working paper 10498, National Bureau of Economic Research.

Finn, A. and Ranchhod, V. (2017). Genuine fakes: The prevalence and implications of data fabrication in a large South African survey. *The World Bank Economic Review*, 31(1):129–157.

Fitzgibbon, C. (2014). Registration and targeting: Lessons learned and recommendations. Final report, UK Department of International Development (DFID). p.37.

Friedman, J., Beegle, K., De Weerdt, J., and Gibson, J. (2016). Decomposing response errors in food consumption measurement. *Food Policy.* Forthcoming.

Gibson, J., Beegle, K., De Weerdt, J., and Friedman, J. (2015). What does variation in survey design reveal about the nature of measurement errors in household consumption? *Oxford Bulletin of Economics and Statistics*, 77(3):466–474.

Gibson, J. and Kim, B. (2007). Measurement error in recall surveys and the relationship between household size and food demand. *American Journal of Agricultural Economics*, 89(2):473–489.

Griliches, Z. and Hausman, J. A. (1986). Errors in variables in panel data. *Journal of Econometrics*, 31(1):93–118.

Grosh, M., Del Ninno, C., Tesliuc, E., and Ouerghi, A. (2008). *For protection and promotion: The design and implementation of effective safety nets.* World Bank Publications.

Grosh, M. E. (1994). *Administering targeted social programs in Latin America: From platitudes to practice*, volume 94. World Bank Publications.

Hausman, J. (2001). Mismeasured variables in econometric analysis: problems from the right and problems from the left. *The Journal of Economic Perspectives*, 15(4):57–67.

Jolliffe, D. (2001). Measuring absolute and relative poverty: the sensitivity of estimated household consumption to survey design. *Journal of Economic and Social Measurement*, 27(1, 2):1–23.

Karlan, D. and Thuysbaert, B. (2016). Targeting ultra-poor households in Honduras and Peru. *The World Bank Economic Review).*

Kidd, S., Gelders, B., and Bailey-Athias, D. (2017). Exclusion by design. Social protection department, International Labour Office, Geneva.

Kidd, S. and Wylde, E. (2011). Targeting the poorest: An assessment of the proxy means test methodology. Ausaid research paper, Australian Agency for International Development, Canberra, Australia.

Lanjouw, P. and Ravallion, M. (1996). How should we assess poverty using data from different surveys? *Poverty Lines*, (3).

Leite, P. (2014). Effective targeting for the poor and vulnerable. Social protection and labor technical note, The World Bank.

Scott, C. and Amenuvegbe, B. (1991). Recall loss and recall duration: an experimental study in Ghana. *Inter-Stat*, 4(1):31–55.

Stoeffler, Q., Mills, B., and Del Ninno, C. (2016). Reaching the poor: Cash transfer program targeting in Cameroon. *World Development*, 83:244–263.

Sulla, V., Zikhali, P., Schuler, P., and Jellema, J. (2017). Does fiscal policy benefit the poor and reduce inequality in Namibia? Working Paper 116029, World Bank.

Table 1: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Ln conso	12.48	0.79	10.21	15.65	4025
Hhsize	5.28	2.88	1	23	4025
Young Children	1.08	1.1	0	8	4025
Children	1.42	1.41	0	12	4025
Elderly	0.33	0.59	0	3	4025
Married	0.74	0.44	0	1	4025
Widowed	0.13	0.34	0	1	4025
Age	46.65	16.3	17	96	4025
Male	0.8	0.4	0	1	4025
Litteracy	0.65	0.48	0	1	4025
Primary	0.72	0.45	0	1	4025
Secondary	0.09	0.29	0	1	4025
Litteracy Max	0.86	0.35	0	1	4025
Primary Max	0.92	0.26	0	1	4025
Secondary Max	0.2	0.4	0	1	4025
Mud/Dirt Floor	0.57	0.49	0	1	4025
Thatch Roof	0.34	0.47	0	1	4025
Mud Walls	0.72	0.45	0	1	4025
N Rooms	3.57	1.8	1	18	4025
Electricity	0.14	0.34	0	1	4025
Urban	0.2	0.4	0	1	4025
Water	0.27	0.44	0	1	4025
Flushed Toilet	0.1	0.31	0	1	4025
Cooking	0.22	0.42	0	1	4025

Table 2: Balance Table

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)				T-test			
	Module 1	Module 2	Module 3	Module 4	Module 5	Module 6	Module 7	Module 8				Difference			
Variable	Mean/SE	Mean/SE	Mean/SE	Mean/SE	Mean/SE	Mean/SE	Mean/SE	Mean/SE	(8)-(1)	(8)-(2)	(8)-(3)	(8)-(4)	(8)-(5)	(8)-(6)	(8)-(7)
Hhsize	5.227	5.153	5.155	5.460	5.282	5.337	5.317	5.280	0.054	0.128	0.126	-0.180	-0.001	-0.056	-0.037
	(0.157)	(0.140)	(0.140)	(0.142)	(0.139)	(0.153)	(0.157)	(0.151)							
Young Children	1.083	1.016	1.065	1.141	1.069	1.068	1.070	1.093	0.010	0.078	0.028	-0.047	0.024	0.026	0.024
	(0.060)	(0.053)	(0.055)	(0.053)	(0.053)	(0.059)	(0.058)	(0.058)							
Children	1.429	1.333	1.440	1.498	1.444	1.444	1.355	1.433	0.004	0.100	-0.007	-0.065	-0.011	-0.011	0.078
	(0.071)	(0.062)	(0.071)	(0.070)	(0.069)	(0.067)	(0.075)	(0.071)							
Elderly	0.370	0.339	0.278	0.312	0.347	0.331	0.339	0.336	-0.034	-0.003	0.058	0.024	-0.011	0.005	-0.003
	(0.030)	(0.029)	(0.025)	(0.026)	(0.025)	(0.030)	(0.030)	(0.029)							
Married	0.742	0.732	0.720	0.730	0.718	0.747	0.737	0.763	0.022	0.031	0.043	0.033	0.045	0.016	0.027
	(0.022)	(0.021)	(0.022)	(0.021)	(0.021)	(0.020)	(0.021)	(0.021)							
Widowed	0.125	0.113	0.133	0.135	0.143	0.124	0.160	0.135	0.010	0.022	0.002	0.000	-0.008	0.012	-0.024
	(0.015)	(0.015)	(0.016)	(0.015)	(0.016)	(0.016)	(0.017)	(0.016)							
Age	47.628	46.192	46.048	46.419	46.532	46.629	46.988	46.803	-0.825	0.611	0.756	0.385	0.271	0.174	-0.185
	(0.838)	(0.756)	(0.762)	(0.751)	(0.717)	(0.765)	(0.809)	(0.809)							
Male	0.811	0.802	0.794	0.792	0.784	0.819	0.788	0.791	-0.020	-0.010	-0.002	-0.000	0.008	-0.027	0.003
	(0.019)	(0.019)	(0.019)	(0.018)	(0.019)	(0.018)	(0.019)	(0.019)							
Litteracy	0.674	0.681	0.615	0.649	0.655	0.669	0.647	0.618	-0.056**	-0.062**	0.003	-0.031	-0.036	-0.051*	-0.028
	(0.022)	(0.023)	(0.026)	(0.025)	(0.023)	(0.023)	(0.025)	(0.024)							
Primary	0.710	0.726	0.738	0.710	0.708	0.731	0.715	0.706	-0.004	-0.020	-0.032	-0.005	-0.003	-0.025	-0.009
	(0.022)	(0.022)	(0.023)	(0.023)	(0.022)	(0.021)	(0.024)	(0.022)							
Secondary	0.082	0.101	0.087	0.097	0.091	0.088	0.102	0.099	0.018	-0.002	0.012	0.002	0.008	0.012	-0.002
	(0.014)	(0.014)	(0.015)	(0.016)	(0.015)	(0.014)	(0.016)	(0.017)							
Litteracy Max	0.863	0.869	0.865	0.877	0.877	0.865	0.838	0.829	-0.034	-0.040*	-0.036	-0.048***	-0.048**	-0.036	-0.009
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.017)	(0.019)	(0.017)							
Primary Max	0.913	0.919	0.927	0.933	0.940	0.914	0.942	0.913	0.000	-0.006	-0.014	-0.020	-0.028*	-0.002	-0.030*
	(0.013)	(0.013)	(0.011)	(0.012)	(0.010)	(0.014)	(0.011)	(0.012)							
Secondary Max	0.195	0.208	0.208	0.196	0.204	0.211	0.210	0.195	0.000	-0.014	-0.014	-0.002	-0.010	-0.016	-0.015
	(0.020)	(0.022)	(0.021)	(0.021)	(0.021)	(0.021)	(0.022)	(0.021)							
Mud/Dirt Floor	0.571	0.577	0.581	0.565	0.571	0.572	0.561	0.577	0.006	-0.001	-0.005	0.011	0.005	0.005	0.016
	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)	(0.034)	(0.034)	(0.033)							
Thatch Roof	0.328	0.321	0.331	0.345	0.323	0.339	0.357	0.340	0.012	0.019	0.009	-0.005	0.017	0.001	-0.017
	(0.028)	(0.027)	(0.028)	(0.029)	(0.029)	(0.030)	(0.028)	(0.028)							
Mud Walls	0.698	0.714	0.712	0.732	0.710	0.735	0.739	0.702	0.004	-0.012	-0.011	-0.030	-0.009	-0.033*	-0.037**
	(0.029)	(0.029)	(0.029)	(0.028)	(0.029)	(0.029)	(0.029)	(0.030)							
N Rooms	3.529	3.581	3.492	3.615	3.558	3.614	3.607	3.598	0.070	0.017	0.106	-0.017	0.041	-0.015	-0.008
	(0.103)	(0.100)	(0.087)	(0.095)	(0.089)	(0.103)	(0.095)	(0.095)							
Electricity	0.161	0.125	0.133	0.129	0.151	0.135	0.128	0.133	-0.028	0.008	0.000	0.004	-0.018	-0.002	0.005
	(0.022)	(0.020)	(0.021)	(0.021)	(0.022)	(0.021)	(0.020)	(0.021)							
Urban	0.203	0.202	0.202	0.202	0.202	0.203	0.200	0.201	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	0.001
	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)							
Water	0.264	0.264	0.266	0.268	0.274	0.267	0.267	0.270	0.006	0.006	0.005	0.003	-0.003	0.003	0.003
	(0.031)	(0.031)	(0.032)	(0.031)	(0.032)	(0.032)	(0.031)	(0.031)							
Flushed Toilet	0.119	0.097	0.111	0.103	0.111	0.092	0.108	0.093	-0.026*	-0.004	-0.018	-0.010	-0.018	0.002	-0.014
	(0.021)	(0.017)	(0.020)	(0.019)	(0.020)	(0.018)	(0.020)	(0.018)							
Cooking	0.229	0.240	0.234	0.222	0.226	0.221	0.204	0.215	-0.014	-0.025**	-0.019	-0.008	-0.011	-0.006	0.011
	(0.031)	(0.031)	(0.030)	(0.029)	(0.030)	(0.030)	(0.029)	(0.030)							
N	503	504	504	504	504	502	501	503							
Clusters	168	168	168	168	168	168	168	168							
F-test of joint significance (F-stat)									1.496*	1.798**	1.444*	1.277	1.138	1.291	1.247
F-test, number of observations									1006	1007	1007	1007	1007	1005	1004

Note: The value displayed for t-tests are the difference in means between households assigned to module 8 and households assigned to each of the other modules. The value displayed for F-tests are the F-statistics. Standard errors in parentheses are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Survey experiment consumption modules

Module	Consumption measurement	Food content	N households
1	Long list (58 food items) 14 day	Quantity from purchases, own-production, and gifts/other sources; Tshilling value of consumption from purchases	504
2	Long list (58 food items) 7 day	Quantity from purchases, own-production, and gifts/other sources; Tshilling value of consumption from purchases	504
3	Subset list (17 food items; subset of 58 foods) 7 day	Quantity from purchases, own-production, and gifts/other sources; Tshilling value of consumption from purchases	504
4	Collapsed list (11 food items covering universe of food categories) 7 day	Tshilling value of consumption	504
5	Long list (58 food items) Usual 12 month	Consumption from purchases: number of months consumed, quantity per month, Tshilling value per month Consumption from own-production: number of months consumed, quantity per month, Tshilling value per month Consumption from gifts/other sources: total estimated value for last 12 months	504
6	Household diary, frequent visits 14 day diary		503
7	Household diary, infrequent visits 14 day diary		503
8	Personal diary, frequent visits 14 day diary		503
			4029

Notes: Frequent visits entailed daily visits by the local assistant and visits every other day by the survey enumerator for the duration of the 2-week diary. Infrequent visits entail 3 visits: to deliver the diary (day 1), to pick up week 1 diary and drop off week 2 diary (day 8), and to pick up week 2 diary (day 15). Households assigned to the infrequent diary but who had no literate members (about 18% of the 503 households) were visited every other day by the local assistant and the enumerator.
Non-food items are divided into two groups based on frequency of purchase. Frequently purchased items (charcoal, firewood, kerosene/paraffin, matches, candles, lighters, laundry soap, toilet soap, cigarettes, tobacco, cell phone and internet, transport) were collected by 14-day recall for modules 1–5 and in the 14-day diary for modules 6–8. Non-frequent non-food items (utilities, durables, clothing, health, education, contributions, and other; housing is excluded) are collected by recall identically across all modules at the end of the interview (and at the end of the 2-week period for the diaries) and over the identical one or 12-month reference period, depending on the item in question.

Source: Gibson et al. (2015)

Table 4: PMT Regressions

	(1) Module 1	(2) Module 2	(3) Module 3	(4) Module 4	(5) Module 5	(6) Module 6	(7) Module 7	(8) Module 8	(9) All
Hhsize	-0.245***	-0.268***	-0.200***	-0.237***	-0.243***	-0.143***	-0.181***	-0.200***	-0.209***
	(0.040)	(0.040)	(0.045)	(0.037)	(0.044)	(0.028)	(0.031)	(0.043)	(0.014)
$Hhsize^2$	0.0112***	0.0128***	0.0107***	0.00975***	0.0112***	0.00624***	0.00750***	0.00902***	0.00928***
	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.001)	(0.001)	(0.002)	(0.001)
Young Children	-0.0739*	-0.0662*	-0.124***	-0.0539	-0.108***	-0.101***	-0.0646**	-0.0859**	-0.0864***
	(0.038)	(0.034)	(0.035)	(0.036)	(0.032)	(0.034)	(0.033)	(0.037)	(0.013)
Children	-0.00362	0.00426	-0.0424	-0.00369	0.000972	-0.0183	-0.0144	-0.0491	-0.0166
	(0.029)	(0.028)	(0.033)	(0.030)	(0.034)	(0.032)	(0.027)	(0.033)	(0.012)
Elderly	0.0510	0.116**	0.00636	-0.0393	-0.0198	-0.0115	-0.0356	-0.115	0.000132
	(0.061)	(0.058)	(0.072)	(0.062)	(0.060)	(0.060)	(0.061)	(0.070)	(0.024)
Mud/Dirt Floor	-0.0911	-0.0423	-0.0722	-0.0626	-0.0414	-0.114**	-0.0604	0.0640	-0.0531*
	(0.077)	(0.062)	(0.064)	(0.066)	(0.065)	(0.056)	(0.066)	(0.067)	(0.030)
Thatch Roof	0.0917	0.00200	-0.0437	-0.155***	-0.0176	-0.0890*	-0.127**	-0.0905	-0.0580**
	(0.064)	(0.065)	(0.057)	(0.059)	(0.062)	(0.049)	(0.056)	(0.059)	(0.025)
Mud Walls	-0.253***	-0.0820	-0.255***	-0.185***	-0.214***	-0.132	-0.0421	-0.110	-0.165***
	(0.086)	(0.074)	(0.068)	(0.070)	(0.066)	(0.084)	(0.087)	(0.088)	(0.033)
N Rooms	0.0229	0.0463**	0.0158	0.0452**	0.0264	0.0258	0.0675***	0.0768***	0.0412***
	(0.021)	(0.020)	(0.017)	(0.018)	(0.026)	(0.017)	(0.017)	(0.018)	(0.007)
Electricity	-0.121	0.0837	-0.145**	0.154*	-0.135	0.234***	0.306***	0.346***	0.0598
	(0.096)	(0.091)	(0.070)	(0.091)	(0.094)	(0.080)	(0.099)	(0.087)	(0.042)
Urban	0.335***	0.446***	0.231***	0.266**	0.320***	0.210*	0.0192	0.329***	0.274***
	(0.099)	(0.097)	(0.088)	(0.120)	(0.103)	(0.116)	(0.126)	(0.115)	(0.061)
Water	-0.0645	0.0583	0.0542	-0.122	-0.100	-0.00693	0.00969	-0.00792	-0.0231
	(0.084)	(0.072)	(0.060)	(0.076)	(0.085)	(0.087)	(0.076)	(0.072)	(0.036)
Flushed Toilet	0.0855	0.0821	0.0615	-0.00619	0.157	0.109	0.279***	0.122	0.129***
	(0.092)	(0.107)	(0.096)	(0.089)	(0.099)	(0.109)	(0.101)	(0.106)	(0.048)
Cooking	0.534***	0.359***	0.454***	0.555***	0.697***	0.0637	0.0255	0.121	0.354***
	(0.107)	(0.080)	(0.088)	(0.111)	(0.108)	(0.127)	(0.111)	(0.105)	(0.044)
Married	-0.255**	-0.124	-0.210**	-0.0789	0.0385	-0.262***	-0.359***	0.132	-0.146***
	(0.116)	(0.088)	(0.091)	(0.106)	(0.104)	(0.095)	(0.121)	(0.128)	(0.041)
Widowed	-0.158	-0.182*	0.0489	-0.128	0.0222	-0.0787	-0.111	-0.0159	-0.0969**
	(0.103)	(0.095)	(0.075)	(0.098)	(0.112)	(0.093)	(0.118)	(0.108)	(0.038)
Age	0.00369	0.0191**	0.00793	0.0196*	-0.000395	-0.00938	0.00246	-0.00125	0.00404
	(0.009)	(0.009)	(0.009)	(0.010)	(0.008)	(0.009)	(0.009)	(0.011)	(0.003)
Age^2	-0.0000	-0.0002***	-0.0000	-0.0002**	-0.0000	0.0000	-0.0000	-0.0000	-0.0001*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Male	0.330***	0.0588	0.337***	0.180	0.0950	0.290***	0.263**	-0.0241	0.177***
	(0.120)	(0.086)	(0.085)	(0.114)	(0.106)	(0.110)	(0.126)	(0.121)	(0.039)
Litteracy	-0.0215	-0.0178	0.0767	0.0802	0.0987	0.0429	-0.105	0.121	0.0375
	(0.111)	(0.100)	(0.054)	(0.091)	(0.085)	(0.081)	(0.080)	(0.084)	(0.032)
Primary	0.0184	0.0903	-0.0655	0.00817	-0.0718	0.0935	0.180**	0.00251	0.0396
	(0.111)	(0.094)	(0.067)	(0.094)	(0.089)	(0.081)	(0.073)	(0.077)	(0.035)
Secondary	0.165	0.158	0.221**	0.0548	0.336***	0.0902	0.297**	0.0321	0.155***
	(0.119)	(0.099)	(0.103)	(0.103)	(0.127)	(0.100)	(0.115)	(0.118)	(0.042)
Litteracy Max	0.0317	0.199	0.0511	-0.0120	0.103	0.182*	0.370***	0.155	0.133***
	(0.142)	(0.123)	(0.131)	(0.099)	(0.092)	(0.096)	(0.084)	(0.099)	(0.038)
Primary Max	0.134	-0.154	0.0529	0.0206	0.155	-0.217	-0.250**	-0.190	-0.0832*
	(0.162)	(0.151)	(0.147)	(0.130)	(0.115)	(0.131)	(0.122)	(0.127)	(0.047)
Secondary Max	0.129	0.176**	0.125*	0.109	0.0884	0.0416	-0.0535	0.144	0.0974***
	(0.083)	(0.077)	(0.071)	(0.089)	(0.087)	(0.077)	(0.078)	(0.092)	(0.028)
Adjusted-R^2	0.622	0.638	0.638	0.627	0.665	0.510	0.483	0.508	0.557
Observations	503	504	504	504	504	502	501	503	4025

Note: This table reports regressions of per capita consumption (in log) as reported in different survey designs. The sample in columns 1–8 is restricted to households assigned to a certain consumption module. Results for the full sample are reported in column 9. OLS estimator is used for all regressions. Standard errors in parentheses are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Predictive Performances

	(1)	(2)
	$\hat{y}_{ik}$	$\hat{\mu}_{ik}$
Formula 1	-0.188***	0.108***
	(0.012)	(0.017)
Formula 2	-0.0585***	0.0441***
	(0.009)	(0.010)
Formula 3	-0.0868***	0.0671***
	(0.011)	(0.013)
Formula 4	-0.277***	0.115***
	(0.009)	(0.018)
Formula 5	-0.239***	0.144***
	(0.013)	(0.020)
Formula 6	-0.222***	0.0759***
	(0.007)	(0.013)
Formula 7	-0.149***	0.0717***
	(0.010)	(0.013)
F-statistics	353.84***	9.38***
Observations	4024	4024
Number of Households	503	503
Mean in *Formula* 8	12.621	0.282

Note: This table reports regressions of predictive performances of PMT by survey design. $\hat{y}_{ik}$ is the predicted value of the log consumption per capita (PMT score) of household i for formula k. $\hat{\mu}_{ik}$ is the squared prediction error for household i and formula k. Formula k (with $k = \{1, 8\}$) is a dummy variable taking the value of 1 if PMT Formula k is used to derive $\hat{y}_{ik}$. All coefficients are interpretable relative to formula 8, which is the omitted category and the benchmark to assess the impact of measurement error on the predictive performances by survey design. OLS estimator is used for both regressions. Robust standard errors clustered at the village level in parentheses. F-test is performed on the null hypothesis that the coefficients of all controls are jointly zero. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Targeting Performances, \$1.25 Poverty Line

	(1)	(2)	(3)	(4)	(5)	(6)
	TE_{ik}	IE_{ik}	EE_{ik}	TE^2_{ik}	IE^2_{ik}	EE^2_{ik}
Formula 1	0.0795***	0.212***	-0.209***	0.0430***	0.0758***	-0.0284*
	(0.022)	(0.026)	(0.035)	(0.013)	(0.017)	(0.014)
Formula 2	0.0298*	0.110***	-0.146***	0.0136*	0.0325***	-0.0275**
	(0.018)	(0.020)	(0.036)	(0.008)	(0.010)	(0.012)
Formula 3	0.0517***	0.128***	-0.114***	0.0271**	0.0362***	0.00701
	(0.018)	(0.022)	(0.031)	(0.011)	(0.012)	(0.023)
Formula 4	0.0835***	0.258***	-0.297***	0.0435***	0.0873***	-0.0522***
	(0.025)	(0.027)	(0.036)	(0.014)	(0.018)	(0.014)
Formula 5	0.0835***	0.246***	-0.272***	0.0462***	0.0877***	-0.0443***
	(0.024)	(0.027)	(0.037)	(0.013)	(0.018)	(0.013)
Formula 6	0.0537***	0.162***	-0.184***	0.0183*	0.0456***	-0.0415**
	(0.019)	(0.023)	(0.034)	(0.010)	(0.011)	(0.016)
Formula 7	0.0477***	0.0986***	-0.0633**	0.0148*	0.0317***	-0.0222*
	(0.016)	(0.019)	(0.029)	(0.008)	(0.010)	(0.012)
F-statistics	2.64**	14.74***	12.24***	2.20**	5.09***	3.83***
Observations	4024	2760	1264	4024	2760	1264
Number of Households	503	345	158	503	345	158
Mean in *Formula* 8	0.235	0.136	0.449	0.0489	0.0298	0.0905

Note: This table reports regressions of targeting performances of PMT by survey design. The dependent variable in column 1 is a dummy equal to 1 if household i with consumption derived from PMT Formula k is mistargeted, and 0 otherwise. Dependent variable in column 4 is equal to mean squared error if household i with consumption derived from PMT Formula k is mistargeted, and 0 otherwise. Columns 2–3 and 5–6 disaggregate the results by error type. Formula k (with $k = \{1, 8\}$) is a dummy variable taking the value of 1 if PMT Formula k is used to predict Y_{ik}. All coefficients are interpretable relative to formula 8, which is the omitted category and the benchmark to assess the impact of measurement error on the predictive performances by survey design. The mean of the dependent variable in formula 8 is shown in the bottom row. LPM is used for regressions 1–3. OLS is used for regressions 4–6. Standard errors in parentheses are clustered at the village level. F-test is performed on the null hypothesis that the coefficients of all controls are jointly zero. *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Targeting Performances, 30% Poverty Threshold

	(1) TE_{ik}	(2) IE_{ik}	(3) EE_{ik}	(4) TE^2_{ik}	(5) IE^2_{ik}	(6) EE^2_{ik}
Formula 1	0.0398***	0.0284	0.0662*	0.0131	0.00634	0.0289
	(0.014)	(0.017)	(0.034)	(0.009)	(0.009)	(0.023)
Formula 2	0.0437***	0.0313*	0.0728***	0.0147**	0.0142	0.0156
	(0.012)	(0.016)	(0.028)	(0.007)	(0.009)	(0.011)
Formula 3	0.0239*	0.0170	0.0397	0.0137	0.00628	0.0309
	(0.013)	(0.016)	(0.034)	(0.009)	(0.009)	(0.023)
Formula 4	0.0239**	0.0170	0.0397	0.00954	0.00297	0.0248
	(0.011)	(0.013)	(0.026)	(0.006)	(0.005)	(0.019)
Formula 5	0.0239*	0.0170	0.0397	0.0100	0.000461	0.0323
	(0.013)	(0.016)	(0.031)	(0.007)	(0.005)	(0.020)
Formula 6	0.0278**	0.0199	0.0464*	0.00472	-0.000209	0.0162
	(0.012)	(0.015)	(0.025)	(0.004)	(0.004)	(0.010)
Formula 7	0.0278*	0.0199	0.0464	0.00761	0.00883	0.00476
	(0.014)	(0.016)	(0.028)	(0.007)	(0.007)	(0.014)
F-statistics	2.34**	0.84	1.54	0.86	0.68	0.77
Observations	4024	2816	1208	4024	2816	1208
Number of Households	503	352	151	503	352	151
Mean in *Formula* 8	0.239	0.170	0.397	0.052	0.036	0.089

Note: Inclusion threshold is adjusted to obtain 30% of the household targeted for each module. LPM is used for regressions 1–3. OLS is used for regressions 4–6. Standard errors in parentheses are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1. See notes to Table 6 for other details.

Table 8: Targeting Errors by Quartile of Consumption

	(1) Q1	(2) Q2	(3) Q3	(4) Q4
Formula 1	0.0873**	0.0476	0.0238	3.69e-17
	(0.035)	(0.033)	(0.029)	(0.016)
Formula 2	0.0794***	0.0317	0.0317	0.0320
	(0.028)	(0.030)	(0.027)	(0.020)
Formula 3	0.0635*	-1.28e-15	0.0238	0.00800
	(0.037)	(0.034)	(0.029)	(0.018)
Formula 4	0.0556*	0.0238	0.00794	0.00800
	(0.028)	(0.029)	(0.024)	(0.008)
Formula 5	0.0794**	0.0159	7.44e-16	1.90e-17
	(0.030)	(0.038)	(0.025)	(0.011)
Formula 6	0.0635**	0.0317	0.0238	-0.00800
	(0.029)	(0.030)	(0.029)	(0.008)
Formula 7	0.0556*	-1.27e-15	0.0476	0.00800
	(0.031)	(0.032)	(0.034)	(0.014)
F-statistics	2.02*	0.70	0.70	.
Observations	1008	1008	1008	1000
Number of Households	126	126	126	125
Mean in *Formula* 8	0.389	0.357	0.167	0.040

Note: This table disaggregates the results presented in Table 7 - Column 1 by splitting households into quartiles of measured consumption. Dependent variables in columns 1–4 are dummies equal to 1 if household i is mistargeted. LPM is used for all regressions. Standard errors in parentheses are clustered at the village level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. See notes to Tables 6 and 7 for other details.