The Journal of Development Studies

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/fjds20

Proxy Means Testing Vulnerability to Measurement Errors?

Jules Gazeaud

To cite this article: Jules Gazeaud (2020) Proxy Means Testing Vulnerability to Measurement Errors?, The Journal of Development Studies, 56:11, 2113-2133, DOI: 10.1080/00220388.2020.1715942

To link to this article: https://doi.org/10.1080/00220388.2020.1715942

The Journal of Development Studies, 2020
Vol. 56, No. 11, 2113–2133, https://doi.org/10.1080/00220388.2020.1715942

Proxy Means Testing Vulnerability to Measurement Errors?

JULES GAZEAUD
Université Clermont Auvergne, CNRS, IRD, CERDI, Clermont-Ferrand, France

(Original version submitted November 2018; final version accepted January 2020)

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 random measurement errors – an assumption that has been challenged by recent literature. Using a survey experiment conducted in Tanzania, this paper brings causal evidence on the impact of non-random errors on PMT performances. Results show that non-random errors bias the coefficients from PMT models, resulting in a 5 to 27 per cent reduction in PMT predictive performances. Moreover, non-random errors induce a 10 to 34 per cent increase in the incidence of targeting errors when poverty is defined in absolute terms. More reassuringly, impacts on the ranking of households are smaller and essentially non-significant. Taken together, these results indicate that PMT performances are quite vulnerable to non-random errors when the objective is to target absolutely poor households, but remain largely unaffected when the objective is to target a fixed share of the population.*

1. Introduction

Social safety nets programmes (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, the number of developing countries with SSNP doubled from 72 to 149 in the last two decades.[1] However, with an average spending of 1.6 per cent of GDP, coverage is far from universal. Governments and development practitioners often use targeting tools in an effort to concentrate the benefits of SSNP on the poorest, but poor households targeting is an inherently inexact and challenging practice, especially in low-income countries because of the lack of verifiable records on earnings. This lack of records often makes means-testing impractical.

Against this backdrop, Proxy Means Testing (PMT hereafter) has become an increasingly popular targeting method. PMT has been implemented in large countries such as Indonesia, Pakistan, Mexico, and the Philippines, as well as in a number of smaller countries, ranging from Ecuador to Jamaica, and more recently to at least 20 African countries (Cirillo & Tebaldi, 2016; Fiszbein & Schady, 2009). In PMT, 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. Typically, 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

Correspondence Address: Jules Gazeaud, Université Clermont Auvergne, CNRS, IRD, CERDI, 26 Avenue Léon Blum, Clermont-Ferrand, 63000, France. Email: jules.gazeaud@etu.uca.fr
Supplementary Materials are available for this article which can be accessed via the online version of this journal available at https://doi.org/10.1080/00220388.2020.1715942

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.

PMT is subject to 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. Del Ninno and Mills (2015, p. 20) argue that it '*can accurately and cost-effectively target the chronic poor*'. A recent World Bank report recommends the use of PMT to target beneficiaries of social benefits in Namibia because it '*could provide better coverage at existing spending levels, providing a greater poverty and inequality impact*' (Sulla, Zikhali, Schuler, & Jellema, 2017, p. 63). In contrast, critics often point to PMT's high built-in errors, implementation issues and lack of transparency. For instance, Kidd and Wylde (2011, p. 2) argue that '*PMT is inherently inaccurate, especially at low levels of coverage, and it relatively arbitrarily selects beneficiaries*', while '*other methods (...) may be better at including intended beneficiaries*'. Other targeting 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).[2]

This debate has been fed by a surge of recent studies assessing the performances of PMT. In these studies, performances are typically displayed in terms of 'errors of inclusion' (providing benefits to households that should not be eligible) and 'errors of exclusion' (not providing benefits to households that should be eligible). Brown, Ravallion, and van de Walle (2018) provide a systematic assessment of PMT performances for nine countries in Sub-Saharan Africa (SSA). The authors find that PMT yields relatively low inclusion errors but high exclusion errors. In the context of Ghana's fertiliser subsidy programme, Houssou, Asante-Addo, Andam, and Ragasa (2018) show that PMT would be more efficient and more cost-effective than a universal allocation. In Sri Lanka, Sebastian et al. (2018) indicate that switching from self-reported income to PMT could considerably improve the targeting performance of Samurdhi, Sri Lanka's flagship SSNP, and would significantly improve the poverty impact of the programme. Comparisons of PMT with Community-Based Targeting (CBT) suggest some gains in terms of accuracy but some loss in terms of community satisfaction with the beneficiary list (Alatas, Banerjee, Hanna, Olken, & Tobias, 2012; Basurto, Dupas, & Robinson, 2017; Karlan & Thuysbaert, 2016; Stoeffler, Mills, & Del Ninno, 2016).[3] For instance, Alatas et al. (2012) in Indonesia report that PMT allowed a 10 per cent reduction in the error rate relative to CBT, while CBT resulted in 60 per cent fewer complaints than PMT.[4]

An implicit assumption made by these studies is that consumption data underlying PMT regressions are error-free or measured with random errors. However, this assumption has been challenged by recent literature. In particular, Gibson, Beegle, De Weerdt, and Friedman (2015) show that measurement errors in consumption have a mean-reverting negative correlation with true values. According to the typical textbook on the impact of measurement errors, this would lead to biased PMT estimates.[5] However, the magnitude of the bias and its implications on targeting accuracy are not clear.

The goal of this paper is to assess the effects of a violation of this assumption on PMT performances. As with many impact evaluations, the key challenge here is to construct the most credible counterfactual of what would happen with error-free or random measurement errors in consumption. I rely on a unique survey experiment that randomly assigned eight different designs of consumption module to more than 4,000 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, De Weerdt, Friedman, & Gibson, 2012; De Weerdt, Beegle, Friedman, & Gibson, 2016; Friedman, Beegle, De Weerdt, & Gibson, 2017; 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 minimises 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 ongoing 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 errors in survey-based consumption data. I estimate that coefficients from PMT regressions are biased in the presence of non-random errors. This results in a reduction in both the predictive and targeting performances of PMT. The predictive performances of PTM decrease by 5 to 27 per cent depending on how consumption data are collected. Moreover, using the typical $1.25 poverty line, the incidence of targeting errors increase by a magnitude ranging from 10 to 34 per cent. This latter result is largely driven by an increase in inclusion errors, which suggests that PMT typically overestimates poverty rates. More reassuringly, I find rather small and non-significant effects on targeting performances when poverty is defined in relative terms (such as with the typical 30 per cent threshold used in many development projects). This means that non-random errors in consumption have if anything, a limited impact on the ranking of households.

It is always difficult to extrapolate the results derived from one context and one may be concerned that the findings presented in this paper may not hold in other contexts. However, the focus on measurement errors due to survey design (as opposed to other type of errors such as fraud or fabrication) provides some reassurance that the results are not too specific. Indeed, it is quite reasonable to assume that survey design has a core mechanism that affects respondents answers regardless of the context.

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

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 (Beegle et al., 2012; Deaton & Grosh, 2000; Scott & Amenuvegbe, 1991). Recall errors refer to situations where respondents forget consumption resulting in under-reports of consumption. Telescoping errors refer to situations where respondents over-report consumption because of the perception that some items have been consumed more recently than they were. The longer the period of recall the greater the likelihood events are forgotten or not precisely remembered. 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 sub-Saharan Africa 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.

These various sources of errors may be more or less prevalent depending on the design of data collection instruments. In recent years, a number of empirical studies confirmed that measurement of consumption is sensitive to survey design. I focus here on evidence on four key dimensions in which survey design varies: 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. This experiment randomly assigned households to eight survey designs differing along the four dimensions above-mentioned.[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 have been raised. 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 Canada, where households reported their food expenditures during the past month and then filled in a diary during the following two weeks, Ahmed, Brzozowski, and Crossley (2006) identify substantial measurement errors in recall food consumption with properties inconsistent with random 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 developing countries may be even more challenging. Beegle et al. (2012) mention two diary household surveys conducted in Tanzania and Malawi where stylised facts are consistent with poor supervision, respondent fatigue, and incomplete or unreliable data. The authors conclude 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*'.

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. 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 per cent for every additional day of recall. Similarly, Beegle et al. (2012) in Tanzania report that a 7-day recall design measured a 11 per cent higher mean food consumption than a 14-day recall design.

Shorter versus longer lists of items included in questionnaires have 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 any 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% additional items was added between the two survey rounds). Jolliffe (2001) confirmed this positive relationship between the number of items and the level of recorded consumption in El Salvador. The author found that more detailed questions on consumption result in an estimate of mean household consumption 31 per cent higher than estimates derived from a condensed version of the questionnaire.

Finally, the identity of the respondent 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 per cent 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. One of the main contribution of the survey experiment conducted by Beegle et al. (2012) and Gibson et al. (2015) is that they collect benchmark consumption data allowing them to make such investigations.[9] Gibson et al. (2015) 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*. Beegle et al. (2012) look at whether measurement errors in consumption are correlated to specific household characteristics. They find evidence that the mis-measurement in diaries with infrequent visits is more prevalent for illiterate households. In addition, they examine how the performances of the different modules relate to total household size, the number of adult household members, urban location, education of the household head, and an asset index. They find that misreports of consumption data are positively associated with household size and the number of adults, and negatively associated with the asset index. However, looking at the education of the household head and the urban location, they find no systematic pattern of errors. In what follows, I present how this pattern 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. 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 that adds variability to the data but does not affect average performance for the sample. The following assumptions are made under random error: $E(\theta) = 0$, $E(\lambda) = 1$ and $E(v) = cov(y, v) = cov(X, v) = cov(\varepsilon, v) = 0$. In contrast, mean-reverting measurement error in y^* assumes $0 < E(\lambda) < 1$, which makes estimates of β inconsistent – from Equation 3 it is now equal to $E(\lambda)\beta$.

Thus, with $0 < E(\lambda) < 1$, 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 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.[11]

3. Survey experiment

I exploit the same survey experiment as Beegle et al. (2012), De Weerdt et al. (2016), Friedman et al. (2017), and 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 hereafter), a leading research company established in 2002 in Tanzania. This section summarises 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 Dodoma, 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.[12] 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*'.

3.2. Experimental design

In each sub-village, three households were randomly assigned to each of the eight consumption modules summarised in Table S4 (Supplementary Materials). Households were 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 rely on a recall design and modules 6–8 on diaries. Modules 1 and 2 use a long list of 58 commodities with a recall period of 14 and 7 days, respectively. Module 3 uses a subset list consisting of the 17 most important commodities and representing 77 per cent of the food consumption expenditure in Tanzania (based on the national Household Budget Survey 2000–2001).[13] Module 4 includes a list of 11 comprehensive categories that correspond to an aggregated version of the list of 58 commodities. Module 5 inquires about '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. Modules 6 and 7 are household diaries (that is, a single diary was used to record all household consumption) with different intensities of supervision. Households assigned to module 6 were visited by a trained survey staff every other day, while those assigned to module 7 were only visited weekly. Module 8 is 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 (that is, 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 confounding module effects with interviewer effects. In each EA, households assigned to the recall modules were surveyed in 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. A summary of key statistics for the sample is reported in Table S1 (Supplementary Materials).

4. Empirical strategy

This paper seeks to quantify the impact of non-random measurement errors 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 & 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

A key assumption of the identification strategy is that the personal diary (module 8) provides 'gold standard' (or 'benchmark') data on consumption. 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.

The identification strategy exploits this benchmark consumption and the random assignment of the different survey designs across households. Table S3 (Supplementary Materials) shows the results of randomisation balance checks across a set of core household characteristics. Overall, randomised assignment of households to the eight different designs seems successful. Six of the differences are statistically significant at conventional levels. These differences, while significant, are not too worrying because they are small in size. Consequently, any systematic difference in measured consumption across modules can be attributed to measurement error due to alternative survey designs. Comparisons of error-prone survey designs with the benchmark give estimates of the effect of measurement error on PMT targeting.

4.2. Estimation procedure and construction of the outcomes of interest

While I recognise 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. Per capita consumption is aggregated on an annual basis using data collected on food consumption and frequent non-food consumption.[14] Total food consumption from module 3 is scaled up by a factor equal to 1/ 0.77 (that is, 29.87%) to make data comparable across modules.

First, I create a set of variables that are long-term determinants or correlates of poverty, encompassing household's demographic characteristics (household size, number of children, and so on), home attributes (floor type, wall type, and so on), and household head's features (education, occupation, and so on). These variables have been selected to be representative of the variables typically included in PMT targeting.[15] Then, using OLS with a backward stepwise selection of the variables, I estimate the relationship between this set of variables and log consumption per capita (the so-called PMT formula) by module type.[16] The following regression is estimated eight times (one for each sample of households assigned to module type k):

$$y_{ik} = \alpha_k + \beta_k X_i + \varepsilon_{ik} \tag{4}$$

where y_{ik} is the log consumption per capita of household i (with $i = 1, \ldots, N_k$; N_k the sample size of households assigned to module k; $k = 1, \ldots, 8$), X_i the set of correlates of consumption. 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' (for example, 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) that form the basis to assess the impact of measurement errors on PMT performances. Then, I restrict the analysis to the sample of households for which benchmark consumption data are available, that is, those assigned to the personal diary (module 8), and compare how each formula perform to predict their 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, that is, the PMT formula one would obtain if consumption was measured without errors.

In the first part, I compare the predictive performances of the alternative PMT formulas. I estimate an equation of the following form using ordinary least squares:

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

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 individual diary 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 $(k = 1, \ldots, 7)$ with γ_8 gives the impact of measurement errors on PMT predictive performances by survey design.

In the 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), that is, identifying a non-poor household as poor, and Exclusion Errors (EE), that is, 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:

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

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.[17]

Similarly, the Exclusion Error Rate (EER), defined as the proportion of the poor households not identified as poor, can be written as:

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

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 errors, which allocate a higher weight for errors farther from the poverty line, are perhaps richer for measuring targeting errors. The Mean Squared IE (MSIE) and Mean Squared EE (MSEE) for module k are given by:

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

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

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 above, I construct variables that 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 that 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 (that is, $(y_i - \hat{y}_{ik})^2$) for all households i that 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 using a linear probability model. 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, for example, PPP\$1.25, beneficiaries are those with a PMT score below PPP\$1.25. With a poverty line defined in relative terms, for example, the poorest 30 per cent, 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. Then I use a poverty line defined in relative terms using the 30 per cent threshold used in many SSNP. 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

5.1. PMT estimates

Table 1 presents the results of PMT regressions by module type. Column 9 displays the PMT on the full sample of households and R-squared is 0.54, which is slightly lower than the 0.59 R-squared obtained by Brown et al. (2018) in Tanzania using LSMS-ISA data, and somewhat higher than the 0.40 obtained in Indonesia by Alatas et al. (2012). R-squared values range from 0.45 for the sample assigned to household diary with infrequent supervision (module 7), to 0.64 for the sample assigned to the usual month recall (module 5).

Somewhat surprisingly, module 8 yields a relatively lower R-squared. This could be due to a smaller number of covariates in the model, but even when PMT estimates rely on a fixed set of variables the R-squared of module 8 remains relatively lower (Table S7 in Supplementary Materials). Alternatively, it could be related to some specificities of diaries. PMT models derived from diaries (modules 6–8) yield consistently lower R-squared than PMT models derived from recall questionnaires (modules 1–5). There are two potential candidates that could explain this pattern of lower R-squared in PMT using diaries. First, consumption data in diaries could include variations that are harder to capture using traditional sets poverty correlates. Second, measurement errors of consumption from the recall questionnaires may be correlated with observable characteristics included in PMT estimates, which would therefore inflate the R-squared artificially. These two explanations are not mutually exclusive. Unfortunately, the data does not allow me to test them empirically. However, it is worth noting that within the family of diary modules, model 8 yields a slightly higher R-squared (0.47 against 0.45 for model 7 and 0.46 for model 6).

There are at least two places where non-random errors in consumption could affect PMT estimates: (i) coefficient estimates; (ii) variable selection. Mis-measured consumption data can lead to bias coefficients, and to exclude relevant correlates of consumption from PMT or to include irrelevant correlates of consumption. Interestingly, while coefficients and variables selected through the backward stepwise procedure vary across specifications, signs do not change (with a few exceptions). Overall, coefficients are larger for households assigned to the benchmark (module 8), which is consistent with the assumption that non-random measurement errors in consumption data bias downward PMT estimates (see Section 2.2 above). Module 8 appears to yield one of the most sparse formula, and this could be due to the fact that the distribution of consumption is tighter (Figure S1 in Supplementary Materials).

5.2. PMT predictive performances

Simple comparisons of distributions of PMT scores using different formulas in Figure 1 show that the benchmark PMT formula yields relatively higher scores. The distribution of scores is shifted to the right compared to other formulas.[18] This is confirmed by the results of regressions presented in Table 2. Overall, formulas 1–7 yield significantly lower PMT scores and higher squared prediction errors than formula 8 (derived from the benchmark personal diary). The results in column 1 show that formulas 1–7 predict between 5 and 27 per cent lower PMT scores compared with the benchmark formula 8. Similarly, formulas 1–7 produce mean prediction errors between 12 and 49 per cent 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

Table 1. PMT regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Module 1	Module 2	Module 3	Module 4	Module 5	Module 6	Module 7	Module 8	All
Hhsize	−0.249***	−0.289***	−0.176***	−0.231***	−0.212***	−0.158***	−0.184***	−0.249***	−0.197***
	(0.033)	(0.031)	(0.041)	(0.030)	(0.036)	(0.019)	(0.022)	(0.027)	(0.014)
Hhsize^2	0.011***	0.014***	0.010***	0.009***	0.011***	0.007***	0.007***	0.009***	0.009***
	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.001)	(0.001)	(0.002)	(0.001)
Elderly		0.110*						−0.156***	
		(0.056)						(0.048)	
Young Children	−0.076***	−0.072**	−0.150***	−0.063**	−0.132***	−0.089***	−0.059**		−0.100***
	(0.029)	(0.030)	(0.030)	(0.030)	(0.029)	(0.028)	(0.027)		(0.012)
Children			−0.049*						−0.023**
			(0.029)						(0.012)
Mud/Dirt Floor		−0.118*				−0.146***			−0.074**
		(0.060)				(0.056)			(0.031)
Thatch Roof				−0.159***		−0.134**	−0.187***		−0.066**
				(0.060)		(0.052)	(0.057)		(0.026)
Mud Walls	−0.308***		−0.326***	−0.253***	−0.266***	−0.249***		−0.164**	−0.193***
	(0.082)		(0.058)	(0.071)	(0.065)	(0.076)		(0.081)	(0.031)
N Rooms		0.042**		0.041**			0.072***	0.098***	0.039***
		(0.020)		(0.018)			(0.016)	(0.017)	(0.008)
Electricity		0.234**	−0.138**	0.186**		0.307***	0.343***	0.393***	0.072*
		(0.096)	(0.069)	(0.091)		(0.080)	(0.096)	(0.092)	(0.044)
Urban	0.126*	0.289***						0.283***	0.101**
	(0.076)	(0.072)						(0.085)	(0.040)
Water			0.122**						
			(0.061)						
Flushed Toilet					0.185*	0.206**	0.356***		0.161***
					(0.098)	(0.097)	(0.090)		(0.052)
Cooking	0.575***	0.505***	0.590***	0.665***	0.787***			0.216**	0.418***
	(0.112)	(0.086)	(0.070)	(0.090)	(0.073)			(0.105)	(0.050)
Married	−0.238**		−0.223***			−0.236**	−0.353***	0.189***	−0.149***
	(0.117)		(0.082)			(0.094)	(0.117)	(0.060)	(0.042)
Widowed	−0.171*	−0.129*							−0.105***
	(0.102)	(0.074)							(0.039)

(*continued*)

Table 1. (*Continued*)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Module 1	Module 2	Module 3	Module 4	Module 5	Module 6	Module 7	Module 8	All
Age		0.023**		0.019*					
		(0.009)		(0.010)					
Age^2		−0.000***		−0.000**	−0.000***				−0.000***
		(0.000)		(0.000)	(0.000)				(0.000)
Male	0.347***		0.315***	0.195***		0.370***	0.332***		0.179***
	(0.119)		(0.081)	(0.063)		(0.097)	(0.117)		(0.039)
Primary		0.120**				0.142**	0.197***		0.077***
		(0.058)				(0.056)	(0.054)		(0.025)
Secondary			0.287***		0.449***		0.291***		0.154***
			(0.109)		(0.097)		(0.092)		(0.041)
Primary Max	0.222**				0.260**				
	(0.096)				(0.105)				
Secondary Max	0.249***	0.260***	0.151**	0.165**				0.235***	0.106***
	(0.068)	(0.069)	(0.070)	(0.066)				(0.071)	(0.028)
Ajusted-R^2	0.58	0.60	0.62	0.59	0.64	0.46	0.45	0.47	0.54
Observations	503	504	504	504	504	502	501	503	4025

Notes: This table reports regressions of per capita consumption (in log) as reported in different survey designs. Sequential selection of variables has been done using backward stepwise regression. Definition of the variables is provided in Table S2 (Supplementary Materials). 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. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

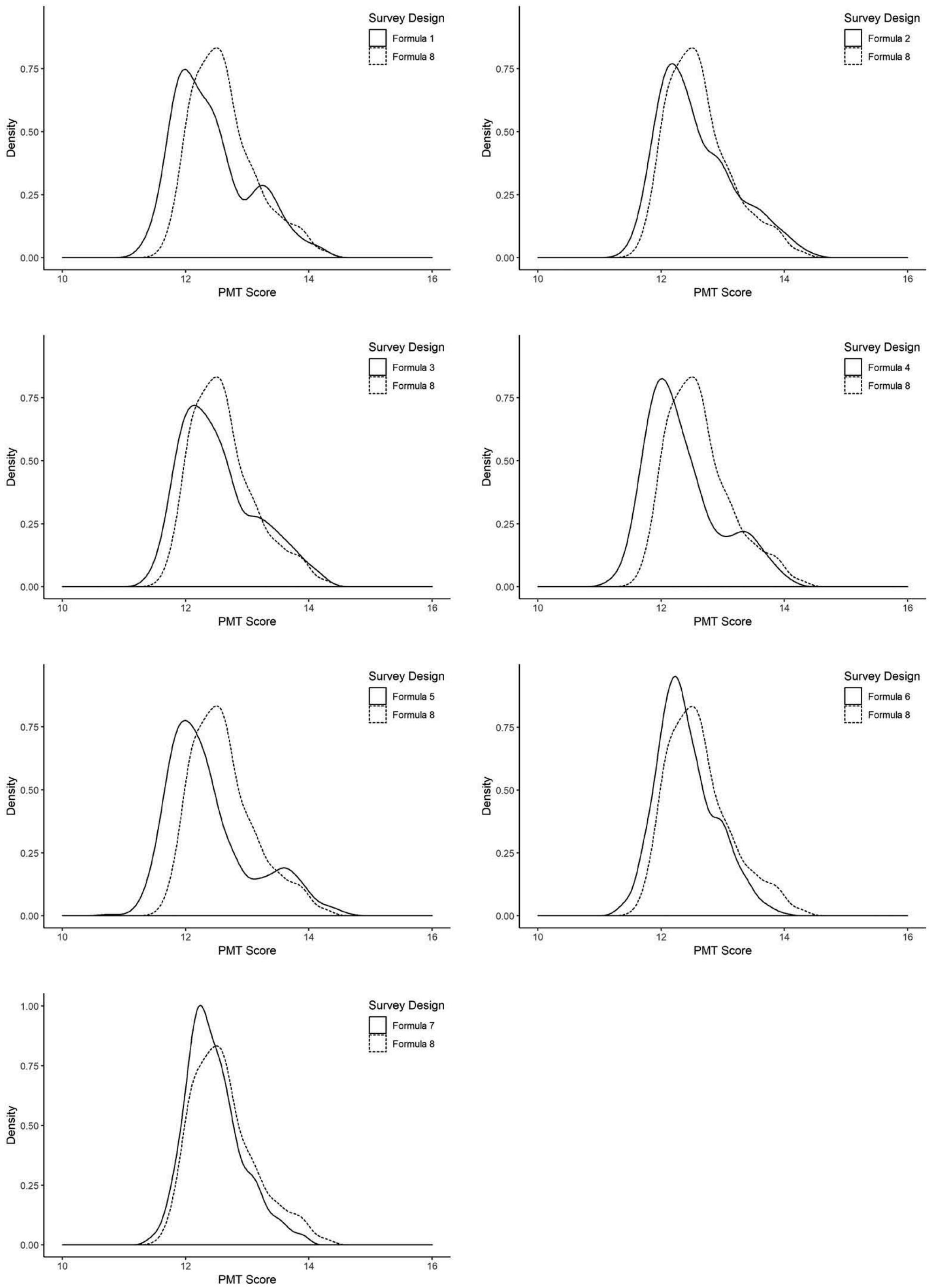

Figure 1. Comparisons of the distributions of PMT scores by survey design.
Notes: Each sub-figure compares the distribution of PMT scores derived from Formula 8 (the benchmark) with distributions of PMT scores of households derived from Formula k (with $k = \{1, 7\}$).

collapsed list (module 4) and usual month (module 5). Non-random measurement errors in consumption have thus a significant and rather large impact on the predictive performances of PMT. In the next section, I will investigate how these relate to targeting performances.

Table 2. Predictive performances

	(1)	(2)
	$\hat{y}_{ik}$	$\hat{\mu}_{ik}$
Formula 1	−0.195***	0.100***
	(0.014)	(0.017)
Formula 2	−0.050***	0.034***
	(0.011)	(0.012)
Formula 3	−0.089***	0.073***
	(0.015)	(0.015)
Formula 4	−0.269***	0.105***
	(0.013)	(0.019)
Formula 5	−0.239***	0.145***
	(0.020)	(0.024)
Formula 6	−0.210***	0.089***
	(0.014)	(0.018)
Formula 7	−0.159***	0.080***
	(0.017)	(0.018)
F-statistics	193.98***	7.20***
Observations	4024	4024
Number of Households	503	503
Mean in *Formula* 8	12.621	0.293

Notes: 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.

5.3. *PMT targeting performances*

Regressions in Table 3 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 errors in formulas 1–7 increase the TER by a magnitude ranging from 2.4 and 8.3 percentage point. Given that the TER derived from formula 8 is 24.7 per cent, these effects are equivalent to an increase in TER of 10 to 34 per cent. In columns 2 and 3, I examine the error rates separately for the non-poor and the poor (defined as the households above/below 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 Table 2 found that formulas 1–7 predict lower PMT scores. This means that the number of poor households is overestimated when formulas 1–7 are used.[19]

I further investigate whether this pattern (higher TER and EER and lower IER) holds when the poverty line is defined in relative terms. Figure 2 looks at whether a household position in the distribution of PMT scores is influenced to some extent by the formula being used to predict her score. Each point in the graphs represent the percentile of a household in the consumption distribution when PMT formula 8 is used against the percentile in the consumption distribution for the same household when PMT formula k (with $k = \{1, 7\}$) is used. If measurement errors had no distributive impacts, each point should be on the diagonal. Households are relatively well distributed around the diagonal, even though large deviations exist for some households. Spearman correlations range from 0.83 for PMT scores derived from formula 7 (household diary with infrequent supervision) to 0.93 for PMT scores derived from formula 2 (7-day recall with the long list of items) and formula 4 (7-day recall with the collapsed list of items). Table 4 refines the insights from Figure 2 by investigating the

Table 3. 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	0.054**	0.200***	−0.266***	0.034***	0.072***	−0.049***
	(0.023)	(0.025)	(0.036)	(0.012)	(0.016)	(0.014)
Formula 2	0.040**	0.116***	−0.127***	0.021**	0.046***	−0.034**
	(0.017)	(0.020)	(0.032)	(0.010)	(0.013)	(0.014)
Formula 3	0.050**	0.136***	−0.139***	0.021**	0.044***	−0.029**
	(0.020)	(0.024)	(0.037)	(0.010)	(0.013)	(0.012)
Formula 4	0.058**	0.238***	−0.335***	0.037***	0.085***	−0.070***
	(0.024)	(0.026)	(0.038)	(0.014)	(0.017)	(0.018)
Formula 5	0.083***	0.261***	−0.304***	0.060***	0.120***	−0.070***
	(0.024)	(0.027)	(0.039)	(0.016)	(0.020)	(0.019)
Formula 6	0.024	0.130***	−0.209***	0.007	0.040***	−0.065***
	(0.021)	(0.024)	(0.038)	(0.011)	(0.012)	(0.023)
Formula 7	0.044**	0.104***	−0.089**	0.014	0.041***	−0.043*
	(0.019)	(0.022)	(0.043)	(0.011)	(0.010)	(0.026)
F-statistics	2.25**	17.75***	13.74***	2.68**	6.99***	3.77***
Observations	4024	2760	1264	4024	2760	1264
Number of Households	503	345	158	503	345	158
Mean in *Formula* 8	0.247	0.139	0.481	0.059	0.029	0.124

Notes: 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.

results of regressions. Poverty is defined in relative terms, using a typical 30 per cent threshold. Interestingly, the coefficients are now much smaller in magnitudes and not statistically significant (except formula 5, derived from the usual month recall module). Point estimates correspond to a 0.4 to 3.2 percentage point increase in TER using formulas 1–7. Similarly, both the IER and EER estimates find small and statistically insignificant coefficients (columns 2 and 3). These results provide evidence that if anything measurement errors in consumption does not affect to a great extent the distribution of poor households. In other words, measurement errors in consumption seem to have relatively weak implications on the distribution of PMT scores.

The results presented in columns 4–6 in Tables 3 and 4 suggest the effects of measurement errors in consumption on the MSTE, the MSIE, and the MSEE are similar to those found for the TER, the IER, and the EER.

One limitation of the results reported in Tables 3 and 4 relates to the fact that the PMT scores used to assess targeting performances have been estimated only on the sample of households assigned to module 8 (for which the benchmark consumption is available). Formula 8 therefore estimates completely within the sample, while formulas 1–7 estimate completely out-of-sample. This could give an advantage to formula 8 targeting performances.[20] To test whether it drives the results, I replicate the estimates using a split sample approach. More specifically, for each module, I randomly split the sample into two parts: one part to estimate PMT formulas (Equation 4); the other part to test predictive and targeting performances. I repeat these operations one hundred times. Results are presented in Tables S13, S18, and S27 (Supplementary Materials). Predictive and targeting performances of the different formulas are now estimated exclusively out-of-sample. The

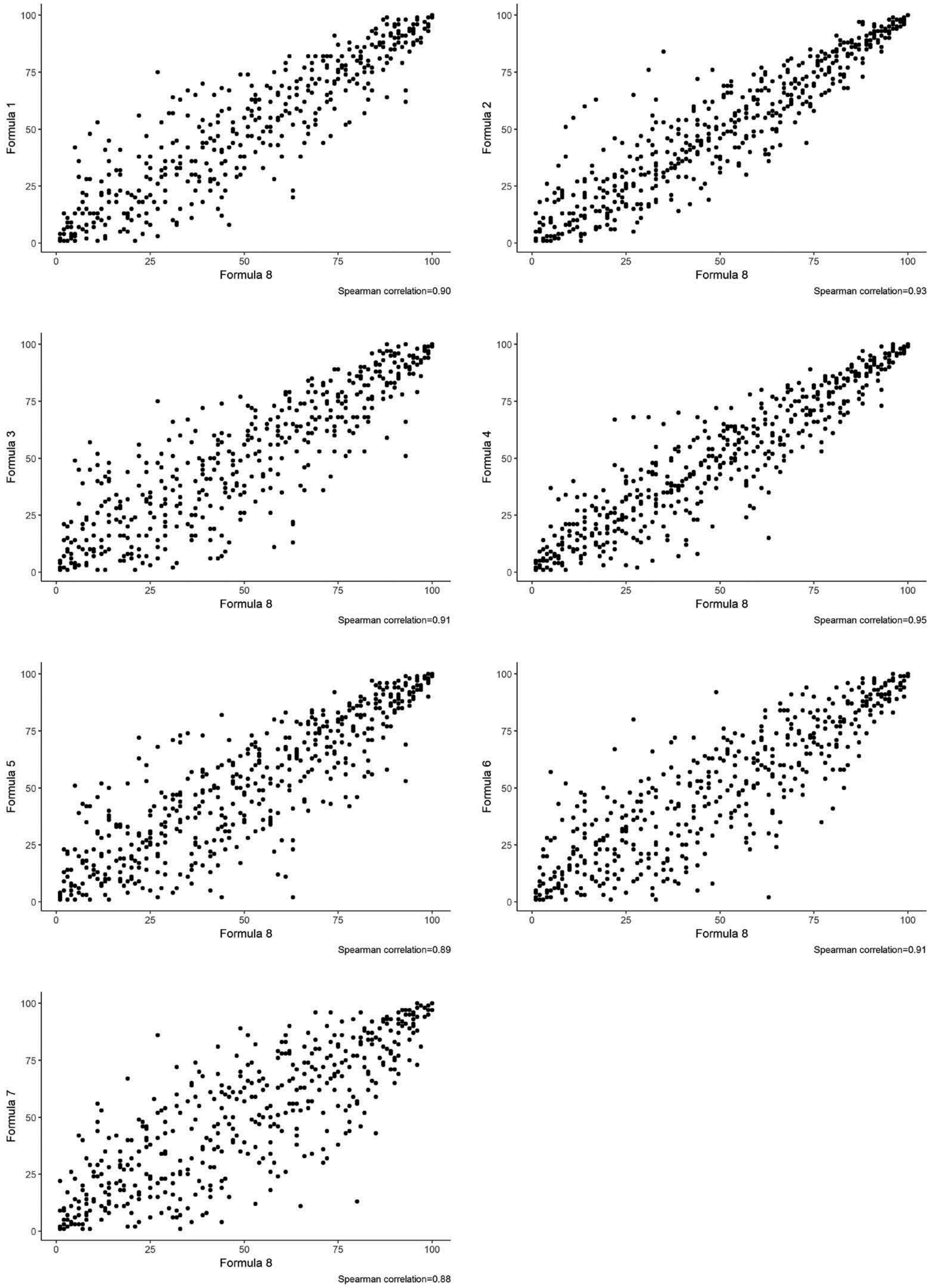

Figure 2. Correlation between PMT score's percentile predicted by the benchmark PMT formula (formula 8) and the seven other formulas.

Notes: Each point in the graphs represent the percentile of the household in the consumption distribution when PMT formula 8 is used (x-axis) against the percentile in the consumption distribution for the same household when PMT formula k (with $k = \{1, 7\}$) is used (y-axis).

magnitude of the coefficients is similar to those reported in Tables 3 and 4. Coefficients are less precisely estimated and thus less significant, but this can be attributed to the reduction in sample

Table 4. Targeting performances, 30 per cent poverty threshold

	(1)	(2)	(3)	(4)	(5)	(6)
	TE_{ik}	IE_{ik}	EE_{ik}	TE^2_{ik}	IE^2_{ik}	EE^2_{ik}
Formula 1	0.024	0.026	0.020	0.006	0.002	0.015
	(0.016)	(0.018)	(0.036)	(0.008)	(0.010)	(0.014)
Formula 2	0.004	0.003	0.007	0.004	0.007	−0.002
	(0.012)	(0.013)	(0.031)	(0.006)	(0.007)	(0.014)
Formula 3	0.024	0.017	0.040	0.002	0.000	0.008
	(0.016)	(0.020)	(0.040)	(0.008)	(0.009)	(0.015)
Formula 4	0.012	0.009	0.020	0.002	−0.003	0.013
	(0.015)	(0.018)	(0.035)	(0.005)	(0.005)	(0.014)
Formula 5	0.032*	0.023	0.053	0.008	0.003	0.018
	(0.016)	(0.020)	(0.038)	(0.007)	(0.008)	(0.014)
Formula 6	0.022	0.017	0.033	0.000	0.001	−0.001
	(0.019)	(0.022)	(0.047)	(0.010)	(0.010)	(0.026)
Formula 7	0.010	0.009	0.013	0.000	0.009	−0.019
	(0.018)	(0.021)	(0.048)	(0.010)	(0.010)	(0.026)
F-statistics	0.65	0.40	0.54	0.62	0.59	0.82
Observations	4024	2816	1208	4024	2816	1208
Number of Households	503	352	151	503	352	151
Mean in *Formula* 8	0.258	0.185	0.430	0.065	0.044	0.114

Notes: Inclusion threshold is adjusted to obtain 30 per cent 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 3 for other details.

size. Overall, these results suggest that estimates of the impact of measurement errors on predictive and targeting performances reported in the paper are not driven by an in-sample advantage of formula 8.

Measurement errors in consumption may be correlated with household characteristics. For instance, the number of adults in the household may affect the relative prevalence of measurement errors across modules (individual consumption from other adult household members may be missed in designs with a sole respondent). Table 5 explores the potential effects that interactions between key household characteristics (household size, number of adults, literacy, urban/rural location) and the formulas dummies could have on the TER. The poverty line is defined in absolute terms in Panel A and in relative terms in Panel B. Household size and the number of adult members do not seem to mediate the impact of measurement errors on targeting accuracy. No interaction term is significant for any formula except formula 5 (the usual month recall) in Panel A and formulas 4–6 in Panel B. Column 3 shows that literate households seem more vulnerable to targeting errors (due to measurement errors) vis-à-vis the benchmark (module 8). For six out of seven formulas in Panel A interaction terms are significantly different from zero. Finally, household vulnerability to targeting errors (due to measurement errors) does not seem to depend on urban/rural location.

6. Conclusions

In this paper, I have investigated the impact of non-random measurement error on PMT performances. Assessments of PMT performances rely on the assumption that consumption data underlying PMT regressions are error-free or measured with random error, even though this assumption 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 the predictive performances of PMT by a magnitude ranging from 5 to 27 per cent, which in turn induces a 10 to 34 per cent increase in the incidence of targeting errors (using the typical PPP $1.25 poverty line). More reassuringly,

Table 5. Interaction of PMT formula and select household characteristics

	(1)	(2)	(3)	(4)
	Household size	Number of adults	Literacy	Urban
Panel A:				
Targeting Error ($1.25 poverty line)				
Interaction 1	0.009	0.005	0.078*	−0.038
	(0.006)	(0.011)	(0.041)	(0.043)
Interaction 2	0.002	−0.002	0.014	−0.017
	(0.004)	(0.010)	(0.035)	(0.030)
Interaction 3	0.010	0.010	0.080**	−0.014
	(0.006)	(0.013)	(0.040)	(0.039)
Interaction 4	0.004	−0.001	0.110**	0.009
	(0.008)	(0.014)	(0.047)	(0.046)
Interaction 5	0.014**	0.031**	0.101**	0.014
	(0.007)	(0.013)	(0.048)	(0.047)
Interaction 6	0.009	0.010	0.072*	0.017
	(0.007)	(0.014)	(0.043)	(0.043)
Interaction 7	0.004	−0.008	0.088**	0.013
	(0.006)	(0.012)	(0.043)	(0.036)
Panel B:				
Targeting Error (30% poverty threshold)				
Interaction 1	0.008	0.011	0.022	0.017
	(0.005)	(0.011)	(0.036)	(0.031)
Interaction 2	0.003	0.002	0.015	0.020
	(0.004)	(0.011)	(0.030)	(0.021)
Interaction 3	0.006	0.002	0.047	0.043
	(0.005)	(0.012)	(0.038)	(0.033)
Interaction 4	0.012**	0.017*	0.036	0.043
	(0.005)	(0.009)	(0.031)	(0.031)
Interaction 5	0.017***	0.032***	0.026	0.013
	(0.006)	(0.012)	(0.035)	(0.034)
Interaction 6	0.013*	0.011	0.052	0.011
	(0.007)	(0.014)	(0.039)	(0.040)
Interaction 7	0.002	−0.010	0.058	0.029
	(0.006)	(0.012)	(0.042)	(0.036)
Observations	4024	4024	4024	4024
Number of Households	503	503	503	503

Notes: This table represents the results of (separate) LPM estimates of a selected measure of targeting performances (mentioned in panels' title) on PMT formula dummies, a single selected household characteristic (mentioned in the column headings) and their interactions. Only the interaction terms are reported due to space limitations. Interaction k (with $k = \{1, 8\}$) is an interactive variable between the characteristic mentioned in the column heading and formula k dummy. Standard errors in parentheses are clustered at the village level. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. See notes to Tables 3 and 4 for other details.

when poverty is defined in relative terms, impacts on the relative distribution of households are small and non-significant, meaning that measurement errors in consumption have weak implications on the distribution of PMT scores.

Some unresolved questions remain. First, I only discussed one dimension of PMT, that is, its predictive and targeting performances, and more attention on cost-efficiency, transparency, fairness, and acceptance would be welcome. Second, I focused on measurement errors in the dependent variable, while measurement errors in the independent variables could also impact PMT performances. Third, I do not take into account PMT vulnerability 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 & Ranchhod, 2017). Finally, recent studies such as

McBride and Nichols (2016) have shown that new tools from machine learning applied to poverty prediction outperform PMT. These new tools are typically trained on survey-based data and documenting whether they are also vulnerable to measurement errors is a potential avenue for future research.

Nevertheless, the results presented in this paper 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. If the objective is to target people below the poverty line, then PMT performances are quite vulnerable to measurement errors. However, if the objective is to target a fixed share of the population (regardless of whether they are above or below the poverty line), PMT performances are quite robust to the presence of measurement errors.

Acknowledgements

I thank Catherine Araujo Bonjean and Vianney Dequiedt for their guidance and support. This paper has benefited from the thoughts of Caitlin Brown, Johanna Choumert, John Gibson, Douglas Gollin, Flore Gubert, Zaneta Kubik, Linden McBride, Fouzi Mourji, Elisabeth Sadoulet, William Seitz, as well as audiences at the Journées Doctorales du Développement (Clermont-Ferrand, December 2017) and CSAE Conference (Oxford, March 2018). I am especially grateful to Joachim De Weerdt for sharing the data underlying this research. Finally, I thank the Editor of this journal, Richard Palmer-Jones, and two extremely helpful anonymous reviewers for their suggestions. The usual disclaimers apply. Data and replication files are available upon request.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Agence Nationale de la Recherche of the French government through the programme 'Investissements d'avenir' [grant number ANR-10-LABX-14-01].

Notes

1. ASPIRE database – Consulted on: www.worldbank.org/aspire. See also Beegle, Coudouel, and Monsalve (2018) for a focus on Africa.
2. For a detailed overview on PMT and other targeting methods used in developing countries see Del Ninno and Mills (2015), Devereux et al. (2017), Grosh, Del Ninno, Tesliuc, and Ouerghi (2008), Grosh (1994), and Hanna and Olken (2018).
3. A notable exception is Premand and Schnitzer (2018) who find that local populations in Niger had a slight preference for PMT compared to CBT.
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 programme in Malawi, Basurto et al. (2017) report that local leaders allocate input subsidies to farmers with larger returns compared to PMT.
5. See Bound, Brown, and Mathiowetz (2001) for a discussion on the impact of measurement errors on regression estimates. In Section 2.2, I present in more detail how Bound et al. (2001) speak to the present study.
6. I only consider deviations 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 (for example, under-reporting of 'bad' consumption such as spending on alcohol or cigarettes), strategic responses (for example,

understatements of consumption because of the belief that responses may be used to determine eligibility for some future social programme; negative answers bias in order to avoid follow-up questions) and untrained, inadvertent, or strategic enumerators (for example, enumerators guiding respondents to give answers that minimise interview length).

7. The unitary household model assumes that all household members have the same utility function. Anderson and Baland (2002) and Duflo and Udry (2004) are two examples of empirical evidence inconsistent with the unitary model of household decision-making.
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).
9. As noted above, this paper rests on the same data as Beegle et al. (2012) and Gibson et al. (2015). More details on the design of the survey are presented in the next section.
10. The framework presented in this section is adapted from Bound et al. (2001), Hausman (2001), and Gibson et al. (2015).
11. One exception is Brown et al. (2018), which exploits panel data in Ethiopia, Malawi, Nigeria, Tanzania, and Uganda to reduce any bias due to measurement errors. The authors use time-mean consumption instead of current consumption and find that PMT performances slightly improve. However, Griliches and Hausman (1986) argue that a crucial parameter 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.
12. 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 S1 in Supplementary Materials shows basic descriptive statistics.
13. To make data comparable across modules, and because surveys are typically interested in total food consumption, estimates from module 3 are scaled up by a factor equal to 1/0.77.
14. Results are robust using food consumption only (see Tables S5, S9, S14, and S23 in Supplementary Materials) or consumption per adult equivalent (see Tables S6, S10, S15, and S24 in Supplementary Materials). 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 (that is, there is no benchmark) and because it is usually not included in PMT. That said, it would be quite reasonable to assume that measurement errors for these items are more prevalent because of the longer recall period (1 month or 12 months depending on the items considered). Unfortunately, I am not able to check this assumption because there are no benchmark data on actual non-frequent consumption.
15. 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 similar (see Tables S7, S11, S16, and S25 in Supplementary Materials).
16. I set the significance level at which variables can enter the model at 10 per cent. The backward selection model starts with all the variables that may be correlated with consumption. At each step, the variable that is the least significant is removed. This process continues until only significant variables remain in the model. Tables S8, S12, S17, and S26 in Supplementary Materials show that results are largely similar without the stepwise option.
17. I use the same definition as Alatas et al. (2012). IER could also be defined as Brown et al. (2018), that is, the proportion of those identified as poor who are not poor. The latter definition is less practical in the present study. Indeed, the sample of households identified as poor is likely to vary across PMT formulas.
18. Figure S1 in Supplementary Materials similarly compares raw distributions of consumption before PMT regressions and shows that households assigned to module 8 have higher scores.
19. Results presented in Table 3 may depend to some extent on the poverty line that is being used. Table 3 reports the results for a poverty line of 1.25$ per capita. As a robustness check, I report the results using different levels of poverty lines in Tables S19–S22 (Supplementary Materials). Results are largely similar.
20. I thank an anonymous referee for spotting this limitation.

References

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

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

Anderson, S., & 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., & Robinson, J. (2017). *Decentralization and efficiency of subsidy targeting: Evidence from chiefs in rural Malawi* (Working Paper 23383). Cambridge, MA: National Bureau of Economic Research.

Beegle, K., Coudouel, A., & Monsalve, E. (2018). *Realizing the full potential of social safety nets in Africa*. Washington, DC: World Bank.

Beegle, K., De Weerdt, J., Friedman, J., & 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., & Mathiowetz, N. (2001). Measurement error in survey data. *Handbook of Econometrics*, *5*, 3705–3843.
Bound, J., & 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., & van de Walle, D. (2018). A poor means test? Econometric targeting in Africa. *Journal of Development Economics*, *134*, 109–124.
Cameron, L., & 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.
Cirillo, C., & Tebaldi, R. (2016). *Social protection in Africa: Inventory of non-contributory programmes*. Brasilia: IPC-IG.
De Weerdt, J., Beegle, K., Friedman, J., & 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*. Washington, DC: World Bank Publications.
Deaton, A., & Grosh, M. (2000). Consumption. In *Designing household questionnaires for developing countries: Lessons from fifteen years of the living standard measurement study* (Vol. 15, pp. 91–133). Washington, DC: World Bank.
Del Ninno, C., & Mills, B. (2015). *Safety nets in Africa: Effective mechanisms to reach the poor and most vulnerable*. Washington, DC: World Bank and Agence Française de Développement Publication.
Devereux, S., Masset, E., Sabates-Wheeler, R., Samson, M., Rivas, A.-M., & Te Lintelo, D. (2017). The targeting effectiveness of social transfers. *Journal of Development Effectiveness*, *9*(2), 162–211.
Duflo, E., & Udry, C. (2004). *Intrahousehold resource allocation in Cote d'Ivoire: Social norms, separate accounts and consumption choices* (Working paper 10498). Cambridge, MA: National Bureau of Economic Research.
Finn, A., & 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.
Fiszbein, A., & Schady, N. R. (2009). *Conditional cash transfers: Reducing present and future poverty*. Washington, DC: The World Bank.
Friedman, J., Beegle, K., De Weerdt, J., & Gibson, J. (2017). Decomposing response errors in food consumption measurement. *Food Policy*, *72*, 94–111.
Gibson, J., Beegle, K., De Weerdt, J., & 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., & 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., & Hausman, J. A. (1986). Errors in variables in panel data. *Journal of Econometrics*, *31*(1), 93–118.
Grosh, M., Del Ninno, C., Tesliuc, E., & Ouerghi, A. (2008). *For protection and promotion: The design and implementation of effective safety nets*. Washington, DC: World Bank Publications.
Grosh, M. E. (1994). *Administering targeted social programs in Latin America: From platitudes to practice* (Vol. 94). Washington, DC: World Bank Publications.
Hanna, R., & Olken, B. A. (2018). Universal basic incomes vs. targeted transfers: Anti-poverty programs in developing countries. *Journal of Economic Perspectives*, *32*(4), 201–226.
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.
Houssou, N., Asante-Addo, C., Andam, K. S., & Ragasa, C. (2018). How can African governments reach poor farmers with fertiliser subsidies? Exploring a targeting approach in Ghana. *The Journal of Development Studies*, *55*, 1983–2007.
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., & Thuysbaert, B. (2016). Targeting ultra-poor households in Honduras and Peru. *The World Bank Economic Review*, *33*, 63–94.
Kidd, S., & Wylde, E. (2011). *Targeting the poorest: An assessment of the Proxy Means Test methodology* (Ausaid Research Paper). Canberra, Australia: Australian Agency for International Development.
Lanjouw, P., & Ravallion, M. (1996). How should we assess poverty using data from different surveys? *Poverty Lines*, (3).
McBride, L., & Nichols, A. (2016). Retooling poverty targeting using out-of-sample validation and machine learning. *The World Bank Economic Review*, *32*, 531–550.
Premand, P., & Schnitzer, P. (2018). *Efficiency, legitimacy and impacts of targeting methods: Evidence from an experiment in Niger* (Policy Research Working Paper No. 8412). Washington, DC: World Bank.
Scott, C., & Amenuvegbe, B. (1991). Recall loss and recall duration: An experimental study in Ghana. *Inter-Stat*, *4*(1), 31–55.
Sebastian, A., Shivakumaran, S., Silwal, A. R., Newhouse, D., Walker, T., & Yoshida, N. (2018). *A Proxy Means Test for Sri Lanka* (Policy Research Working Paper No. 8605). Washington, DC: World Bank.
Stoeffler, Q., Mills, B., & Del Ninno, C. (2016). Reaching the poor: Cash transfer program targeting in Cameroon. *World Development*, *83*, 244–263.
Sulla, V., Zikhali, P., Schuler, P., & Jellema, J. (2017). *Does fiscal policy benefit the poor and reduce inequality in Namibia?* Washington, DC: World Bank Document.