

DISCUSSION PAPER SERIES

IZA DP No. 14606

**Estimating Poverty among Refugee
Populations:
A Cross-Survey Imputation Exercise for
Chad**

Theresa Beltramo
Hai-Anh H. Dang
Ibrahima Sarr
Paolo Verme

JULY 2021

DISCUSSION PAPER SERIES

IZA DP No. 14606

Estimating Poverty among Refugee Populations: A Cross-Survey Imputation Exercise for Chad

Theresa Beltramo

United Nations High Commissioner for Refugees

Hai-Anh H. Dang

The World Bank and IZA

Ibrahima Sarr

United Nations High Commissioner for Refugees

Paolo Verme

The World Bank

JULY 2021

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Estimating Poverty among Refugee Populations: A Cross-Survey Imputation Exercise for Chad¹

Household consumption surveys do not typically cover refugee populations, and poverty estimates for refugees are rare. This paper tests the performance of a recently developed cross-survey imputation method to estimate poverty for a sample of refugees in Chad, combining survey and administrative data collected by the United Nations High Commissioner for Refugees. The proposed method offers poverty estimates based on administrative data that are mostly statistically insignificantly different from those based on survey consumption data. This result is robust to different poverty lines, sets of regressors, and modeling assumptions of the error terms. We find the method to outperform common targeting methods, such as proxy means tests and the targeting method that is currently used by humanitarian organizations in Chad.

JEL Classification: C15, F22, I32, O15, O20

Keywords: refugees, forced displacement, poverty, imputation, targeting, Chad

Corresponding author:

Hai-Anh H. Dang
Data Production & Methods Unit Development Data Group
World Bank
1818 H St. N.W.
Washington, D.C. 20433
USA
E-mail: hdang@worldbank.org

¹ The authors are grateful to Andrew Dabalen, Giacomo De Giorgi, Kristen Himelein, Aly Sanoh, colleagues at Global Poverty Practice at the World Bank, and participants at the American Economic Association's session on "Comparing Poverty of Refugees and Their Hosts" held in San Diego on January 5th, 2020 for helpful comments on previous versions. This work is part of the program "Building the Evidence on Protracted Forced Displacement: A Multi-Stakeholder Partnership". The program is funded by UK aid from the United Kingdom's Department for International Development (DFID), it is managed by the World Bank Group (WBG) and was established in partnership with the United Nations High Commissioner for Refugees (UNHCR). The scope of the program is to expand the global knowledge on forced displacement by funding quality research and disseminating results for the use of practitioners and policy makers. We further thank DFID for additional funding through its Knowledge for Change (KCP) program. This work does not necessarily reflect the views of DFID, the WBG or UNHCR.

I. Introduction

The UN General Assembly’s Sustainable Development Goal (SDG) 1—*End poverty in all its forms* by 2030—explicitly pledges that “no one will be left behind.” To achieve this goal, the availability of high-quality household consumption surveys is essential, and it is equally important for these surveys to be inclusive and cover marginal populations, such as refugees and Internally Displaced Persons (IDPs). Unfortunately, household consumption surveys rarely include forcibly displaced populations, despite the fact that these populations are among the most vulnerable and deprived. They typically lack fundamental rights such as freedom of movement and the right to work, have eroded human and physical capital, and face more frequent shocks than surrounding host communities.

This is a significant and growing challenge, particularly in Sub-Saharan Africa. The global number of forcibly displaced persons almost doubled from 43.3 million in 2009 to 82.4 million in 2020. Among them, there are 26.4 million refugees, 4.1 million asylum seekers, 48 million IDPs (UNHCR 2021), and other displaced populations under the United Nations High Commissioner for Refugees (UNHCR) protection. Almost four out of five refugees live in countries neighboring the place of origin, and some 84% of them live in developing countries. Sub-Saharan Africa hosts around one-third of the world’s refugee population, half of the 10 countries with the largest refugee population relative to the national population, and six of the 10 countries with the largest numbers of IDPs.

Yet, household consumption data for the region are not collected frequently, they are often of low quality and rarely include displaced populations. There can be different reasons for this challenge. One reason is that the region has the highest poverty rates in the world, which are strongly associated with fewer household surveys (Beegle et al. 2016). Another reason is that displaced populations are often scattered in different places; this hard-to-reach feature poses technical and logistic issues for survey implementation. A third reason may be the lack of political will or financial resources to cover non-citizens in national surveys. Consequently, measuring

poverty among displaced populations in Sub-Saharan Africa is an important undertaking that is severely hampered by missing household consumption data.²

To address this challenge, imputation methods have been widely employed in economics to fill data gaps where a variable is missing in one survey but available in another survey, which are both representative of the same population. These methods are known as “cross-survey” imputation methods and have been used to estimate poverty trends across time periods in the context of repeated cross-section or panel data (e.g., Dang et al 2019), across geographical areas in the context of poverty mapping exercises (e.g., Elbers et al. 2003), or across types of surveys such as consumption and labor force surveys (e.g., Doudich et al. 2016). Yet, there is barely any study that applies cross-survey imputation to address the data challenge in the context of refugees. To our knowledge, Dang and Verme (2021) is the first study to propose cross-survey imputation methods to estimate poverty for the Syrian refugees living in Jordan, using survey and administrative data provided by the UNHCR. Their findings suggest that, in the absence of actual household consumption survey data, cross-survey imputation methods can provide encouraging results. Altindag et al. (2021) is the only other article we could find that estimates refugee household welfare using administrative data from humanitarian organizations. Their paper focuses on Syrian refugees in Lebanon and, similarly to Dang and Verme’s (2021) article on Syrian refugees in Jordan, they find that administrative data held by humanitarian organizations can be used to estimate refugee household welfare accurately.

This paper adds to the nascent literature on measuring poverty for refugees in the context of missing household consumption data. While we use the same cross-survey imputation method employed in Dang and Verme’s (2021) study on Jordan, we extend the existing literature in different respects. First, we analyze data from Chad, which provides a completely different geographic, economic, and refugee context from Jordan. Chad is a large, landlocked, Sub-Saharan country and is one of the poorest countries in the world. It also hosts a multitude of refugees that come from different countries and are located in remote refugee camps. In these respects, Chad

² Missing data issues are not a problem limited to displaced populations, but can emerge because of lack of survey data on a particular topic of interest, population group, or time period. These issues can also be caused by sampling errors, incomplete data due to unit or item nonresponse, data input errors, or post-survey data manipulations such as top-coding or censoring.

provides a counterfactual to Jordan, which is a small, not landlocked, Middle-East, middle-income country with refugees who are mostly Syrians living outside camps.

Second, we exploit a richer and more diverse set of data than the Jordan study, which includes registration data, census-type targeting data, and a household consumption survey which were collected at about the same time. This triangulation between several different data sets offers an opportunity to obtain a more nuanced validation of the proposed imputation method. We also rigorously examine the imputation method against different poverty lines, including the food poverty line, the national poverty line, the international poverty line, and various other simulated poverty thresholds. Finally, in addition to the poverty imputation exercise, we test the performance of the proposed method for targeting purposes. In particular, we compare the targeting performance of our method with the targeting method currently used in Chad to administer cash assistance to refugees. This exercise helps provide clear policy indications on how to improve targeting for refugees.

The estimation results indicate that the limited set of variables that are provided in the UNHCR registration system predict household consumption (welfare) reasonably well. Estimates from the three sets of data available for the analysis produce similar welfare figures. The current targeting strategy in Chad, which is used jointly by the National Commission on the Welcoming and Resettlement of Refugees (CNARR), UNHCR, and World Food Programme (WFP), is fairly accurate in predicting household welfare. However, our results suggest that this targeting strategy could be further improved by reducing the inclusion and exclusion errors. If these encouraging results are replicated in other contexts, poverty predictions for refugees can be expanded at scale, with good prospects for the improvement of targeted programs.³

³ The poverty estimates used in this paper do not reflect the official poverty estimates monitored by the government and the international community. The interest of this paper is to test a cross-survey imputation methodology, and, for this purpose, we use a subsample of UNHCR refugee data that are not nationally representative of the refugee population in Chad. By contrast, the official poverty statistics require national consumption surveys conducted by the national statistical office with samples that are nationally representative. At the time of writing this paper, these national data were not available. This will provide another opportunity to validate the cross-survey imputation method proposed in this paper.

The paper is organized as follows. Section II outlines the country context. Section III presents the data and analytical framework. The estimation results are presented in section IV, and Section V evaluates the targeting strategy used in Chad and our targeting method in light of the global experience. Section VI discusses the limitations of the study, and Section VII concludes.

II. Country Context

Chad is one of the poorest countries in the world. According to the household consumption survey administered in 2011, 29% of the population fall below the food poverty line and 47% fall below the national poverty line (World Bank 2018). The past decade has seen much instability for Chad with negative consequences on household well-being. Per capita Gross Domestic Product (GDP) decreased by 15% between 2015 and 2017, from an average of US\$963 in 2015 to US\$823 in 2017 (in 2010 purchasing power parity (PPP)). In terms of overall development, Chad ranks 187th of 189 countries on the Human Development Index (World Bank 2019). Due to these challenges, the country struggled to meet many of the Millennium Development Goals in 2015. Barring unforeseen economic growth or large increases in official development assistance, the country appears unlikely to meet many of the SDG objectives set for 2030.

Despite the current negative economic downturn, Chad continues to host a high number of refugees and is among the countries that top the world's list in this respect (Table A1). Chad is the 10th largest host country for refugees in the world and the fifth largest host country in Sub-Saharan Africa (after Ethiopia, Kenya, Uganda, and the Democratic Republic of Congo). Chad's refugee population represents a significant portion of the national population, about 3%. The number of forcibly displaced persons increased from 474,478 in 2015 to 667,586 as of March 2019, of which about 69% were refugees or asylum seekers.⁴ Of the 459,809 current refugees and asylum seekers, the majority (74%) are Sudanese refugees living in the eastern part of Chad, 21% are Central African Republic refugees living in southern Chad, and a much smaller number of Nigerian refugees (about 2%) are living in the Lake Chad Basin. The situation is further complicated by the

⁴ UNHCR uses the term “people of concern” to describe those who are forcibly uprooted from their homes, including asylum-seekers, refugees, stateless persons, the internally displaced, and returnees.

large population of IDPs in the Lake Chad region, which is estimated at 165,313 at the end of 2018 (UNHCR 2018). Map 1 shows the locations of the refugee camps in Chad.

III. Analytical Framework and Data

III.1. Analytical Framework

The methodology used in this paper relies on the cross-survey imputation framework, which was first introduced by Elbers et al. (2003) to generate poverty maps.⁵ Most recently, Dang et al. (2017) built on this literature to propose a model that imposes fewer restrictive assumptions and offers an explicit formula for estimating the poverty rate and its variance. Three new contributions introduced by this study are: (i) it offers a simple variance formula, which is in line with the recent statistical literature; (ii) it can accommodate complex design sampling; and (iii) the framework remains applicable to two surveys with different designs (such as imputing from a household consumption survey into a labor force survey). Finally, the approach allows for different modeling methods, including the standard linear regression model, its variant with a flexible specification of the empirical distribution of error terms, a logit model, and/or a probit model.

Formally, let x_j be a vector of characteristics that are commonly observed between two surveys, where j indicates survey type, with 1 and 2 being respectively the base survey and the target survey. Assume that the welfare indicator is a function of household and individual characteristics (x_j):

$$y_j = \beta_j x_j + \mu_{cj} + \varepsilon_j$$

where y_j is the welfare indicator (consumption per capita per month), β_j is a vector of parameters, μ_{cj} is cluster (c) random effects, and ε_j is the idiosyncratic error term.

⁵ See also Tarozzi (2007) and Mathiassen (2009) for further improvements and adaptation of this approach (e.g., by estimating the standard errors in a different way). Newhouse et al. (2014), Dang et al. (2019), and Dang (2020) offer recent reviews of previous imputation studies that discuss the main advantages and different approaches of welfare imputation practices as well as provide useful insights into the imputation process. There is also an established statistical literature on imputation methods to address this issue; see Little and Rubin (2019) for a recent review.

This imputation framework is based on two assumptions. The first assumption (Assumption 1), which is critical for poverty imputation, states that measurement of household characteristics in each sample of data is a consistent measure of the characteristics of the whole population. In other words, it stipulates that the surveys considered are representative of the same target population. In our context, the two surveys represent the same population of refugees, and they were conducted approximatively at the same time. Therefore, we do not expect major issues with this first assumption. However, we will conduct means difference tests on the observed overlapping variables between the target data and base data to ensure that this is the case. The second assumption states that changes in x_j between the data collection periods of the two data sets can capture the change in welfare over the period (Assumption 2). Since data collection for the two data sets we use were collected in the same year, there is no need to test Assumption 2. Under these two assumptions, the imputed welfare is

$$y_2^1 = \beta_1' x_2 + u_1 + \varepsilon_1. \quad (1)$$

Dang et al. (2017) propose different imputation methods for parameter estimation. The first method relies on the assumption of the normal distribution for the two error terms (μ_{cj} and ε_j are uncorrelated and $\mu_{cj} / x_j \sim \mathcal{N}(0, \sigma_{\mu_{cj}})$ and $\varepsilon_j / x_j \sim \mathcal{N}(0, \sigma_{\varepsilon_j})$). Hereafter, this method is referred to as the normal linear regression model. An alternative method proposed is the empirical error method, which assumes no functional form for these error terms and uses instead the empirical distribution to estimate the parameters. This study also proposes two other alternative methods—the Probit Model and the Logit Model—which are more restrictive and model poverty status (poor and nonpoor) instead of consumption expenditure.

Once the parameters are estimated, the welfare indicator, which is household consumption per capita, is obtained as follows:

$$\hat{y}_{2,s}^1 = \hat{\beta}_1' x_2 + \tilde{u}_{1,s} + \tilde{\varepsilon}_{1,s}. \quad (2)$$

The imputed poverty rate and its variance are then estimated as:

$$\text{i) } \widehat{P}_2 = \frac{1}{S} \sum_{s=1}^S P(\widehat{Y}_{2,s}^1 \leq z_1) \quad (3)$$

$$\text{ii) } V(\widehat{P}_2) = \frac{1}{S} \sum_{s=1}^S V(\widehat{P}_{2,s}|x_2) + V\left(\frac{1}{S} \sum_{s=1}^S \widehat{P}_{2,s}|x_2\right) \quad (4)$$

These poverty estimators offer consistent estimates of the parameters of interest. Furthermore, in terms of prediction accuracy, these estimators outperform the traditional proxy means testing technique, which typically omits the error terms $u_1 + \varepsilon_1$ and results in biased estimates of the welfare indicator (Dang et al., 2019).

III.2. Data

As part of its mandate to protect displaced persons in host countries, the UNHCR collects data to track refugees and other populations of interest, monitor these populations, and deliver assistance and services. In the framework of this study, we use three sets of data collected by the UNHCR (Table A2). The first one is the ProGres data set, which is the UNHCR's registration system covering all refugees or asylum seekers requiring assistance. The ProGres data set is a live instrument that is continuously updated as new refugees/asylum seekers arrive, or existing refugees contact the UNHCR. The data that we use were extracted at the end of December 2017. This set of data contains socioeconomic variables (such as household size, marital status, gender, age, country of origin, and region of residence) but no consumption or expenditure data. This data set can be considered as the "census" of refugees.

The second set of data, the Targeting data set, is also a census-like data set for refugees living in Chad. The main objectives of this data set are to fill knowledge gaps on refugee livelihoods and the levels and differences of vulnerability in refugee households, and to categorize refugees into different income levels for assistance regarding cash and food. Besides these objectives, the Targeting data set aims to identify factors that can enable self-reliance. This data set is based on a mixed methods approach, including qualitative and quantitative methods. The first step is the use of focus groups with refugee leaders, women's organizations, and youth associations, to identify the wealth characteristics and key challenges specific to age and gender. Next, a sample survey is carried out across camps to confirm the wealth characteristics that were identified by refugees in

the first step. Based on the outcomes of the first two steps, a quantitative survey that is designed to capture wealth characteristics is administered to all the refugee households.

The Targeting data set includes all Sudanese, Central African, and Nigerian refugees living in Chad. The data were collected between June 17, 2017 and July 15, 2017 and cover 19 refugee sites and refugees living in nine host villages. After the data are collected, a statistical model, which takes into account household welfare, is used to classify households into four socioeconomic groups (very poor, poor, average, and better off). For the variables that are relevant for this study, this data set contains demographic variables (household size, gender, age, country of origin, and region of residence), variables for asset and animal ownership, and variables reflecting shock-coping strategies. Similar to the ProGres data set, the Targeting data set does not collect information on consumption or expenditure. However, it collects information on wealth.

The last set of data is the Post-Distribution Monitoring (PDM) data set, which is from a sample survey that covers similar themes as the Targeting data set. The PDM data set, which was collected in 2017 by the World Food Program (WFP), aims to provide a better understanding of how refugees use food assistance and contains data on consumption and expenditure. The sampling design is a two-stage stratified random sample, where the first stage includes the selection of camps and the second stage the selection of households. Beforehand, the different camps are stratified in three zones: (i) North East (Ourecassoni, Amnaback, Iridimi Touloum), (ii) Centre-East (Goz Amir, Djabal, Gaga, Tegaine, Bredjing, Farchana), and (iii) South (Amboko, Dossey, Gondjé, Belom, Moyo) (Map 1). In addition, the sampling takes into account the kind of humanitarian assistance that is provided to refugees (in-kind, food voucher, or cash). The survey includes two consumption aggregates measuring monthly total consumption and monthly food consumption, using retrospective questions with varying recall periods depending on the item considered (from seven days to one year). The consumption aggregate is compiled by aggregating the different food and non-food items, including expenditures on education, health, durable assets, and rent. For this study, we consider two welfare indicators from the PDM data set. The first is the household total

consumption expenditure per capita per month, and the second is the household food consumption per capita per month.⁶

For poverty imputation purposes, three data sets are constructed from the ProGres, Targeting, and PDM data sets. The first, which we refer to as “ProGres 2” is obtained by appending the ProGres data to the end of the PDM data. As the ProGres and PDM data share only demographic variables, ProGres 2 contains demographic variables for all observations, although only the observations from the PDM data have consumption expenditure. So “ProGres 2” dataset allows us to calibrate the welfare model using the PDM data and subsequently use this model to estimate household consumption using the proGres dataset. The second set of data constructed, “Targeting 2,” comes from appending the Targeting data to the end of the PDM data. Therefore, Targeting 2 contains demographic variables, asset and animal ownership, and coping strategies variables as well as consumption data. The last set of data, “ProGres Targeting,” is obtained by first merging the ProGres and Targeting data (matching 72% of the observations) and then appending these data to the end of PDM data. This set of data is the most complete data set in terms of variables.

The motivation behind constructing these three sets of data is to check whether the different sources of data as well as the different sets of variables generate different poverty figures, such that we can determine the set that best predicts poverty. To ensure comparability across the three data sets, we restrict the analysis to 16 of the 19 refugee sites, because the PDM data cover only 16 sites. Consequently, this study covers the refugees in Chad that come from the Central African Republic and Sudan only.

IV. Estimation Results

As a first step, we check whether our data sets are representative of the same underlying population (Assumption 1) by performing means difference tests across key predictors. Given that the PDM data is a subsample of the Targeting or ProGres data sets, we use a test that is suitable

⁶ The aim of the paper is not to measure consumption accurately or estimate nationally representative poverty figures for refugees in Chad. The purpose of the paper is only to test the cross-survey imputation methodology using a sample of refugee data. In this respect, our only concern is that the poverty predictions are close to the poverty rates calculated with the consumption data. Whether our consumption data are accurate or not, this is less relevant for us. We should expect the cross-survey methodology to produce even better results if the quality of the consumption aggregate improves.

for partially overlapping samples. Here we use the same test proposed by Verme et al. (2015) applied to ProGres data on refugees in Jordan. Table 1 provides the results. It shows that all the variables are not significantly different in terms of means, indicating that the two samples are representative of the same population.

To evaluate the performance of the welfare estimation model, we consider three models. Model 1 includes demographic and geographic variables (region of residence and country of origin). This is the most parsimonious model and uses the variables that are readily available in the ProGres data set. Model 2 adds to Model 1 variables related to animal and asset ownership. Model 2 is richer than Model 1, but it is more demanding in terms of the control variables, which may also be less reliable or more likely to be missing in the census data. Model 3 adds to Model 2 variables measuring coping strategies. To test for multicollinearity, Table 2 reports the variance inflation factor (VIF) for the different models. It shows that no variable has a VIF that is over 5, which is far lower than the rule-of-thumb value of 10 given for harmful collinearity by Kennedy (2008). We conclude that multicollinearity is not an issue for any of the models considered.

Next, we test the out-of-sample performance and possible overfitting of the three models, using the PDM data and the root mean square error (RMSE) and mean absolute error (MAE) as performance functions. To do so, the data set is split into five equal folds. In the first iteration, the first fold is used to test the model, and the rest are used to train the model. In the second iteration, the second fold is used as the testing set, while the rest serve as the training set. This process is repeated until each of the five folds has been used as the testing set. The performance function is obtained as the mean across the five iterations.

For the food consumption aggregate, the three models have similar measures of goodness-of-fit for both indicators (Table 3). Model 1's RMSE is 0.55, Models 2 and 3's RMSE is 0.54. For the MAE, Models 1 and 3 have a value of 0.42, whereas Model 2 has an RMSE of 0.41. When we turn to the overall consumption aggregate, we note differences between the three models. The RMSE values range from 0.53 to 0.58, with Model 3 and Model 1 having the smallest and highest RMSE, respectively. The MAE is quite similar across the three models, within a range from 0.39

(Model 3) to 0.43 (Model 1). These results suggest that no model consistently outperforms the other models.

Table 4 applies the model to the three data sets described in the data section (ProGres 2, Targeting 2, and ProGres Targeting data), using the normal linear regression model and the empirical error model. We also shows the results using two poverty lines in this table: (i) a poverty line of US\$1.9 a day (PPP), which represents the international poverty line for extreme poverty (panel A), and; (ii) the national poverty line, which corresponds to around US\$2.6 (World Bank 2013) (panel B). Table 5 further presents the results for the national food poverty line of US\$1.8 (PPP). These three poverty lines are among the set of the arbitrary poverty lines considered in the general simulation above to evaluate the quality of the prediction.⁷ With one exception, the predicted poverty rates are not statistically different from the poverty rates obtained from the survey consumption data (henceforth, “survey estimates”). For the case of the food and international poverty lines, this is partly due to the large standard errors of the prediction estimates, but these findings hold for the national poverty line where the standard errors of the predicted values are much smaller.⁸

Figure 1 repeats the exercise of Table 5 for the ProGres Targeting data and all the poverty lines between the 66th and 99th percentiles of the consumption distribution. Panels A and B offer estimation results using the normal linear model and the empirical error model, respectively. The results suggest that Models 1 and 2 predict the poverty rates for different poverty lines well. The predicted poverty rates are within the 95% confidence interval for all the arbitrary poverty lines considered. The predictions are also very similar across the normal and empirical error models. However, Model 3 overestimates poverty, and the predicted poverty rates provided are outside the 95% confidence interval of the survey-based rates for the set of different poverty lines considered. As Model 3 adds variables related to coping strategies, it might be that households do not accurately report these strategies, for example, by overestimating the frequency of using these strategies to receive more assistance from humanitarian organizations.

⁷ See also Table A3 in the annex for the full base model.

⁸ The national and international poverty lines are used for illustrative purposes only. As the consumption aggregate used in this paper is not comparable to a full consumption aggregate as usually derived from nationally representative household surveys, the resulting poverty statistics are not comparable to national or international poverty statistics. In our case, these lines are simply alternative thresholds used to test the sensitivity of poverty estimates to different poverty lines.

Figure 2 shows the predicted welfare rates for the set of different poverty lines for all three models, but with a focus on food security. Welfare based on food security is defined in humanitarian settings as the inability to afford the minimum expenditure basket required to purchase a food basket (to satisfy basic needs). The minimum expenditure basket is defined by the WFP “as what a household requires in order to meet their essential needs, on a regular or seasonal basis, and its average cost” (WFP 2018). The results are similar to the overall welfare results displayed in Figure 1. The results indicate that Models 1 and 2 predict the actual welfare rates well based on food security for different poverty lines and are within the 95% confidence interval for all the arbitrary poverty lines considered. The predictions are also similar for the two different estimation models of error terms, the normal linear model and empirical error model. Again, Model 3 overestimates the poverty rates, as the predicted welfare rates are outside the confidence interval of the survey-based rates.⁹

In summary, Figures 1 and 2, corresponding to Models 1 and 2, underestimate welfare for low poverty lines and overestimate for high poverty lines but are still within the 95% confidence intervals. Model 3 always overestimates poverty for smaller poverty lines and its predictions are outside the 95% confidence interval. In general, Models 1 and 2 predict poverty and food security poverty well for different arbitrary poverty lines. Based on these results, we conclude that these two models provide fairly accurate poverty estimates for refugees living in Chad, and that the variables currently available in the ProGres UNHCR registration system can be combined with other survey data to predict the poverty rates of refugees efficiently.¹⁰

⁹ To check for possible heterogeneity, we split the sample with respect to country of origin. The results were similar except larger estimate variances (less precision), which might be due to sample size for refugees from the Central African Republic.

¹⁰ Our results are also consistent with the findings in previous studies (e.g., Dang et al. (2017), Dang et al. (2018), and Dang and Verme (2019)) that emphasize the importance of selecting few key predictor variables rather than too many predictors (to avoid model overfitting). In particular, Model 1 contains a few demographic variables only are in line with the empirical evidence offered in previous studies and the theoretical evidence in Luca et al. (2018). Previous empirical studies also highlight that adding household assets helps to improve on poverty estimates, and Model 2, which adds asset and animal ownership to Model 1, is consistent with this evidence. However, adding more variables may lead to overfitting, resulting in less accurate welfare estimates. The results of Model 3 could be placed in this context.

V. Targeting Performance

The imputed welfare estimates can be useful in evaluating ex-post the inclusion/exclusion errors of the food assistance programs administered by government and humanitarian organizations during 2016/17. The targeting strategy for food assistance was agreed to and implemented by the UNHCR, WFP, and the Chadian government agency responsible for refugees, the CNARR. We perform an analysis to show how accurately the current targeting strategy identifies poor households in terms of inclusion (leakage) and exclusion (undercoverage) errors.

The inclusion error is defined as the proportion of households that the targeting method considers as poor despite not being poor. This is expressed as $\frac{FP}{TP+FP}$, where FP (false poor) is the number of non-poor households incorrectly considered by the targeting method and TP (true poor) is the number of poor households correctly reported poor. The exclusion error is defined as the proportion of households in poverty that the targeting method considers as non-poor. It is defined as $\frac{FN}{TP+FN}$, where FN (false non-poor) is the number of poor households incorrectly considered non-poor by the targeting method.

Both error types are important from different perspectives. The inclusion error matters primarily from a budget perspective, as it represents a waste of resources. The exclusion error summarizes the program's failure to cover households in need. Another targeting indicator used is the Coady-Grosh-Hoddinott (CGH) Ratio (Coady, Grosh, and Hoddinott 2004). Coady et al. (2004) report an index of targeting performance obtained by dividing the proportion of beneficiaries falling within the target population by the proportion of beneficiaries that would result from a random allocation. For example, if the bottom 40% of the income distribution receives 60% of the funding, the performance indicator is 1.5 (=60/40). The higher the indicator, the greater is the performance of the targeting strategy (see Table 6 for more details).

The current UNHCR/WFP/CNARR targeting approach relies on the Food Consumption Score (FCS) generated by WFP's PDM surveys, which is a composite score based on dietary diversity, food consumption frequency, and the relative nutritional importance of different food items. As is the case for any index, the FCS is contingent on the selection of the food group weights as well as the food item thresholds, which are based on inherently subjective choices. Survey-to-survey

methods have been shown to outperform these types of index approaches, whereas the proposed cross-imputation method has been shown to perform better than the proxy means testing also in refugee contexts (Dang and Verme 2021).¹¹

In light of the previous findings, we empirically evaluate how the UNHCR/WFP/CNARR targeting strategy performs relative to the targeting method based on predicted consumption and relative to the available international evidence. Table 7 shows the undercoverage and leakage rates for the different approaches. The method we propose (panel B) outperforms the targeting method currently used in Chad (panel A) for all the poverty lines except the 25th percentile poverty line. The errors are not low overall, with the UNHCR/WFP/CNARR undercoverage rates ranging from 9% to 32% and the leakage rates from 12% to 36%, and our model-based undercoverage rates from 6% to 40% and the leakage rates from 9% to 41%. However, these methods perform relatively well when compared with international evidence. For example, Skoufias et al. (2001) find that the undercoverage and leakage rates for the PROGRESA program in Mexico were 7% and 70%, respectively, for a poverty rate of 25%. These figures represents slightly better performance on the undercoverage rate but much worse performance on the leakage rate compared with those for Chad.

The estimated targeting rates for Chad are also better than the median performance of similar scores for programs across the world. Table 8 reports Coady-Grosh-Hoddinott (CGH) Ratio for the 85 programs considered by Coady et al. (2004) (A), the UNHCR/WFP/CNARR targeting program (B), and our proposed methodology (C). The UNHCR/WFP/CNARR program does not perform poorly when compared with the international evidence. Notably, our methodology outperforms the UNHCR/WFP/CNARR targeting program and the median value of the programs covered by Coady et al. (2004).

For an additional and more general test, we empirically evaluate how the UNHCR/WFP/CNARR targeting strategy performs relative to the proposed targeting method based on predicted consumption for different poverty lines varying from 38% to 99% of the consumption distribution (Figure 3). The results suggest that the method we propose outperforms

¹¹ On optimal targeting in humanitarian contexts, see also Verme and Gigliarano (2019).

the targeting method currently used in Chad for all the poverty lines between 38% and 99%. In other words, our proposed method would more accurately identify and reach intended beneficiaries than the targeting method currently used in Chad for any welfare programs targeting poor refugees.

VI. Limitations

The objective of this work is to test how cross-survey imputation methods perform in estimating poverty for refugee populations, using Chad as a case study. Although the results of our cross-survey imputation exercise show that key demographic variables from ProGres predict reasonably well the welfare measure offered in the PDM, additional work is needed to assess the performance of this methodology in other refugee contexts, particularly in poor countries and data-scarce environments. To do this, data sets should ideally contain more detailed information on consumption. For efficient analysis, different data sets should be linked at the individual level using the IDs available in the ProGres registration data.

Furthermore, the PDM data measure consumption using relatively fewer variables than those in round 4 of the Chadian Household Consumption and Informal Sector Surveys (ECOSIT4) and the data used by this paper only cover a subset of refugees in Chad. Our main objective is to test a cross-survey methodology, and, for this purpose, we used a subsample of UNHCR refugee data that are not nationally representative of the refugee population in Chad. By contrast, official poverty statistics require national consumption surveys conducted by the national statistical office, with samples that are nationally representative. Therefore, the poverty estimates presented in this paper do not reflect the official poverty estimates monitored by the government and the international community and poverty estimates based on the methodology proposed can be improved once ECOSIT4 data are made available.

The data we analyze did not cover refugees who live outside camps. As these refugees live in different environments, predicting their welfare may require different sets of variables. And measuring consumption among refugees who rely on a combination of handouts and informal incomes is uncharted territory. Existing survey instruments may need to be adapted, and the meanings of concepts such as utility and capabilities among refugees may need to be reconsidered.

VII. Conclusion

Tracking the progress made toward SDG Goal 1 of eradicating poverty for all requires the availability of high-quality household consumption surveys. However, the majority of countries across the world, especially developing countries, face challenges in collecting poverty data. High-quality consumption surveys that are comparable for forcibly displaced persons and their hosts are and will remain in limited supply, given the cost and challenges associated with these types of surveys. In the meantime, cross-survey imputation methods can provide a second-best alternative that can potentially save time and resources.

This study combined survey and census-type data on refugees to estimate welfare for refugees in Chad. We showed how different sets of variables as well as different sources of data perform in the identification of poor households, in particular how well the set of variables available in the ProGres database can predict poverty. In a second step, the paper estimated the accuracy of the current UNHCR/WFP/CNARR targeting strategy and compared it with the targeting strategy based on imputed consumption.

The results suggest that the set of variables available in ProGres accurately predicts the welfare rates for different poverty lines. Adding variables related to asset and animal ownership provides predictions that are very close to the ones with only the variables available in the ProGres data set. Since the UNHCR ProGres data are available in most refugee locations where the UNHCR runs the registration system, these methods may be replicable in many settings of forcibly displaced persons.

The current targeting strategy that is used for food, livelihoods, and cash-based assistance, despite its simplicity, is rather accurate when compared with the existing international evidence. The targeting errors resulting from the current UNHCR/WFP/CNARR targeting strategy for a poverty rate of 25% are in the same error range as other targeting methods around the world, as reported in Coady et al. (2004). Yet, we also showed that the existing targeting method can be improved by using the imputation method proposed in this paper.

References

- Altindag, Onur, O'Connell, Stephen, Şaşmaz, Aytuğ, Balcioğlu, Zeynep, Cadoni, Paola, Jerneck, Matilda and Foong, Aimee Kunze, (2021), Targeting humanitarian aid using administrative data: Model design and validation, *Journal of Development Economics*, 148: 102564.
- Beegle, Kathleen, Luc Christiaensen, Andrew Dabalen, and Isis Gaddis. (2016). *Poverty in a Rising Africa*. Washington, DC: The World Bank.
- Coady, David, Margaret Grosh, and John Hoddinott. 2004. *Targeting of Transfers in Developing Countries: Review of Lessons and Experience*.
- Dang, Hai-Anh H. 2020. "To Impute or Not to Impute, and How? A Review of Poverty-Estimation Methods in the Absence of Consumption Data." *Development Policy Review*. Doi: <https://doi.org/10.1111/dpr.12495>
- Dang, Hai-Anh H, and Peter F Lanjouw. 2018. "Poverty Dynamics in India between 2004 and 2012: Insights from Longitudinal Analysis Using Synthetic Panel Data," *Economic Development and Cultural Change*, 67(1), 131-170.
- Dang, Hai-Anh H., Peter F. Lanjouw, and Umar Serajuddin. 2017. "Updating Poverty Estimates in the Absence of Regular and Comparable Consumption Data: Methods and Illustration with Reference to a Middle-Income Country." *Oxford Economic Papers*, 69(4): 939-962.
- Dang, Hai-Anh, Dean Jolliffe, and Calogero Carletto. 2019. "Data Gaps, Data Incomparability, and Data Imputation: A Review of Poverty Measurement Methods for Data-Scarce Environments". *Journal of Economic Surveys*, 33(3): 757-797.
- Dang, Hai-Anh, and Paolo Verme. 2021. "Estimating Poverty for Refugee Populations: Can Cross-Survey Imputation Methods Substitute for Data Scarcity?" *ECINEQ Working Paper 2021-578*.
- Doudich, M., A. Ezzrari, R. van der Weide, and P. Verme. 2016. "Estimating Quarterly Poverty Rates Using Labor Force Surveys." *World Bank Economic Review*, 30(3): 475-500.
- Elbers, Chris, Jean O. Lanjouw, and Peter Lanjouw. 2003. "Micro-Level Estimation of Poverty and Inequality." *Econometrica* 71 (1): 355-64.
- "Global_POVEQ_TCD.Pdf." n.d. Accessed December 2, 2018. http://databank.worldbank.org/data/download/poverty/33EF03BB-9722-4AE2-ABC7-AA2972D68AFE/Archives-2018/Global_POVEQ_TCD.pdf.
- Kennedy, Peter. 2008. *A Guide to Econometrics*. John Wiley & Sons.
- Little, Roderick JA, and Donald B. Rubin. 2019. *Statistical Analysis with Missing Data*. Vol. 793. John Wiley & Sons.
- Luca, Giuseppe De, Jan R. Magnus, and Franco Peracchi. 2018. "Balanced Variable Addition in Linear Models." *Journal of Economic Surveys* 32 (4): 1183-1200.
- Mathiassen, Astrid. 2009. "A Model Based Approach for Predicting Annual Poverty Rates without Expenditure Data." *Journal of Economic Inequality* 7 (2): 117-35.
- Newhouse, David, S. Shivakumaran, Shinya Takamatsu, and Nobuo Yoshida. 2014. *How Survey-to-Survey Imputation Can Fail*. The World Bank.

- Skoufias, Emmanuel, Benjamin Davis, and Sergio de la Vega. 2001. "Targeting the Poor in Mexico: An Evaluation of the Selection of Households into PROGRESA." *World Development* 29 (10): 1769–84.
- Tarozzi, Alessandro. 2007. "Calculating Comparable Statistics From Incomparable Surveys, With an Application to Poverty in India." *Journal of Business & Economic Statistics* 25 (3): 314–36.
- UNHCR. 2018. "UNHCR - Figures at a Glance." 2018. <http://www.unhcr.org/figures-at-a-glance.html>.
- . 2020. "GlobalTrends: Forced Displacement in 2019."
- Verme, Paolo, and Chiara Gigliarano. 2019. "Optimal Targeting under Budget Constraints in a Humanitarian Context." *World Development* 119 (July): 224–33.
- Verme, Paolo, Chiara Gigliarano, Christina Wieser, Kerren Hedlund, Marc Petzoldt, and Marco Santacroce. 2015. *The Welfare of Syrian Refugees: Evidence from Jordan and Lebanon*. The World Bank.
- WFP. 2018. "World Food Programme, 2018. Minimum Expenditure Basket: Interim Guidance Note. WFP. <https://docs.wfp.org/api/documents/WFP-0000074198/download/>."
- World Bank. 2013. "Dynamics of Poverty and Inequality Following the Rise of the Oil Sector."

Table 1 : Means Difference Tests

	PDM			ProGres Targeting			Two-sided p-value
	Mean	Std. Dev.	N Obs.	Mean	Std. Dev.	N Obs.	t-test for overlapping groups
Demographics and geographical variable							
HH size	4.76	2.96	1440	4.11	2.53	65943	0.83
Gender	0.65	0.48	1440	0.69	0.46	65943	0.96
Age of HH head	42.34	14.01	1441	42.19	14.70	65943	0.97
Education							
No Education	0.63	0.48	1440	0.55	0.50	56838	0.81
Koranic School	0.15	0.36	1440	0.19	0.39	56838	0.99
Primary	0.12	0.33	1440	0.15	0.36	56838	0.90
Secondary	0.09	0.28	1440	0.10	0.30	56838	0.92
Higher	0.01	0.07	1440	0.01	0.08	56838	0.99
Marital status							
Married	0.08	0.28	1440	0.10	0.30	65934	0.91
Divorced	0.08	0.28	1440	0.10	0.30	65934	0.90
Widowed	0.18	0.38	1440	0.11	0.31	65934	0.96
Single	0.05	0.22	1440	0.08	0.27	65934	0.90
Occupation is agriculture	0.49	0.50	1439	0.789	0.408	65943	0.99
Origin	0.467	0.499	1441	2.79	0.41	65943	0.61
Asset and animal ownership							
HH has phone	0.15	0.36	1440	0.17	0.38	65943	0.99
HH has carts	0.02	0.14	1440	0.03	0.18	65943	0.92
HH has bike	0.05	0.23	1440	0.02	0.15	65943	0.98
HH has moto	0.02	0.12	1440	0.02	0.13	65943	0.98
HH has radio	0.06	0.23	1440	0.08	0.27	65943	0.91
HH has cattle	0.02	0.15	1441	0.02	0.13	65943	0.98
HH has donkeys	0.07	0.25	1441	0.44	0.50	65942	0.87
HH has sheep	0.04	0.19	1441	0.09	0.29	65943	0.16
HH Has goats	0.06	0.24	1441	0.14	0.35	65943	0.18
HH Has horses	0.06	0.23	1441	0.04	0.19	65942	0.87
HH Has poultry	0.09	0.28	1441	0.17	0.38	65943	0.19
Coping strategies							
Consume seeds	0.17	0.38	1104	0.17	0.38	65943	0.92
Sell assets	0.01	0.07	1104	0.06	0.24	65943	0.80
Send children for Begging	0.03	0.16	1104	0.05	0.21	65943	0.99
Sell last breeding Female	0.01	0.10	1104	0.05	0.22	65943	0.87
Region of residence							
Region 1	4.74	2.19	1441	5.08	2.02	65943	0.97
Region 2	0.13	0.32	1441	0.101	0.30	65943	0.95
Region 3	0.19	0.39	1441	0.12	0.32		0.88
Region 4	0.18	0.38	1441	0.07	0.25	65943	0.80
Region 5	0.09	0.30	1441	0.18	0.48	65993	0.61
Region 6	0.16	0.37	1441	0.03	0.17	65943	0.74
Region 7	0.09	0.29	1441	0.18	0.38	65943	0.86
Region 7	0.15	0.36	1441	0.13	0.338	65943	0.96

Source: Authors' calculations.

Table 2: Collinearity Tests

	Model 1		Model 2		Model 3	
	VIF	1/VIF	VIF	1/VIF	VIF	1/VIF
HH size	1.15	0.87				
Age of head of HH	1.29	0.77	1.33	0.75	1.29	0.77
HH is Farming	1.37	0.73	1.4	0.71	1.34	0.75
Head of HH has primary education	1.25	0.80	1.27	0.79	1.27	0.78
Head of HH attended Islamic School	1.14	0.87	1.2	0.83	1.23	0.81
Head of HH has Secondary education	1.16	0.86	1.19	0.84	1.36	0.74
Head of HH has Higher education	1.02	0.98	1.03	0.97	1.06	0.95
HH is Female	1.41	0.71	1.5	0.66	1.42	0.70
Head of HH is divorced	1.16	0.86	1.19	0.84	1.14	0.88
Head of HH is widowed	1.35	0.74	1.4	0.71	1.49	0.67
Head of HH is single	1.14	0.87	1.19	0.84	1.14	0.88
Country origin is Sudan	3.71	0.27	4.94	0.20	3.46	0.29
Region 2	2.12	0.47	2.69	0.37	1.91	0.52
Region 3	2.3	0.43	3.01	0.33	2.3	0.43
Region 4	1.9	0.53	1.96	0.51	1.5	0.66
Region 5	2.08	0.48				
Region 6	2.07	0.48	2.21	0.45	1.71	0.59
Region 7	1.79	0.56	1.81	0.55	1.66	0.60
HH has Phone			1.21	0.83	1.23	0.81
HH has Carts			1.15	0.87	1.14	0.87
HH has Bikes			1.21	0.83	1.4	0.71
HH has Moto			1.05	0.95	1.08	0.93
HH has Radio			1.17	0.85	1.22	0.82
HH has Cattle			1.06	0.94	1.07	0.94
HH has Horses			1.30	0.77	1.31	0.76
HH consumes seeds as coping strategies					1.15	0.87

Table 3: Models Out of Sample Performance, Individual Level

	Food Consumption aggregate			Overall Consumption aggregate		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
RMSE	0.55	0.54	0.54	0.58	0.53	0.54
MAE	0.42	0.41	0.42	0.43	0.41	0.39

Note:* The sample size of PDM dataset that 1441 is divided into five parts. Performances functions (RMSE and MAE) are obtained as the mean across the five iterations.

Source: Authors' calculations.

Table 4: Predicted Poverty Rates Using the International and National Poverty Lines*

	<i>ProGres 2</i>		<i>Targeting 2</i>		<i>ProGres Targeting</i>		
	Model 1 (1)	Model 1 (2)	Model 2 (3)	Model 3 (4)	Model 1 (5)	Model 2 (6)	Model 3 (7)
Panel A: Poverty rates using international standard							
Normal linear regression model	80.9 (4.2)	76.4 (3.42)	76.0 (3.4)	75.4 (4.2)	78.1 (3.5)	79.7 (3.6)	79.8 (4.2)
Empirical error model	81.5 (4.2)	77.0 (3.6)	76.55 (3.6)	75.5 (4.4)	79.0 (3.5)	80.5 (3.6)	80.4 (4.3)
Survey-based poverty estimate		78.8 (1.9)					
Panel B: Poverty rates using national standard							
Normal linear regression model	90.9 (2.4)	88.0 (2.1)	87.5 (2.7)	87.0 (2.7)	88.6 (2.2)	90.0 (2.1)	90.0 (2.6)
Empirical error model	91.6 (2.1)	89.0 (2.1)	88.4 (2.1)	87.4 (2.7)	89.7 (2.1)	90.9 (2.0)	90.1 (2.5)
Survey-based poverty estimate		89.7 (1.5)					
Control Variables							
Demographics & Employment	Y	Y	Y	Y	Y	Y	Y
Asset and animal ownership	N	N	Y	Y	N	Y	Y
Coping Strategies	N	N	N	Y	N	N	Y
R ² adjusted	0.57	0.55	0.57	0.66	0.52	0.55	0.62
Observations (N)	65242	82468	82467	82467	56830	56829	56829

Note:*The international total poverty line is \$1.88 PPP per person per day while the most recent national total (Food) poverty line in Chad is \$2.60 (\$1.88) per person per day. Robust standard errors in parentheses are clustered at the camp level. We use 1,000 simulations for each model run.

Source: Authors' calculations.

Table 5: Food Poverty

	<i>ProGres Targeting</i>		
	Model 1 (1)	Model 2 (3)	Model 3 (3)
Food Poverty rates			
Normal linear regression model	82.1 (3.5)	83.4 (3.2)	81.6 (4.0)
Empirical error model	82.8 (3.3)	83.2 (3.3)	82.0 (4.2)
Survey-based poverty estimate		80.1 (1.9)	

Table 6: Targeting performance measures

Predicted Poverty Status	True Poverty Status	
	Poor	Non-Poor
Poor	True Positive (TP)	False Positive (FP)
Non-poor	False Negative (FN)	True Negative (TN)
Inclusion error (Leakage)	$\frac{FP}{TP + FP}$	
Exclusion error (Undercoverage)	$\frac{FN}{TP + FN}$	
Coady-Grosh-Hoddinott (CGH) Ratio	$\frac{\text{share of beneficts reaching the } x \text{ poorest percentile}}{x}$	

Table 7: Comparison of Coverage and Leakage Rates (%)

	Poverty lines			
	25 th Percentile	50 th Percentile	80 th Percentile	90 th Percentile
A. Current UNHCR/WFP/CNARR targeting strategy approach				
Undercoverage Rate	32	32	19	9
Leakage Rate	36	36	22	12
B. Our predicted consumption-based targeting				
Undercoverage Rate	40	26	12	6
Leakage Rate	41	28	14	9
C. PROGRESA's method targeting				
Undercoverage Rate	7	10	16	
Leakage Rate	70	43	16	

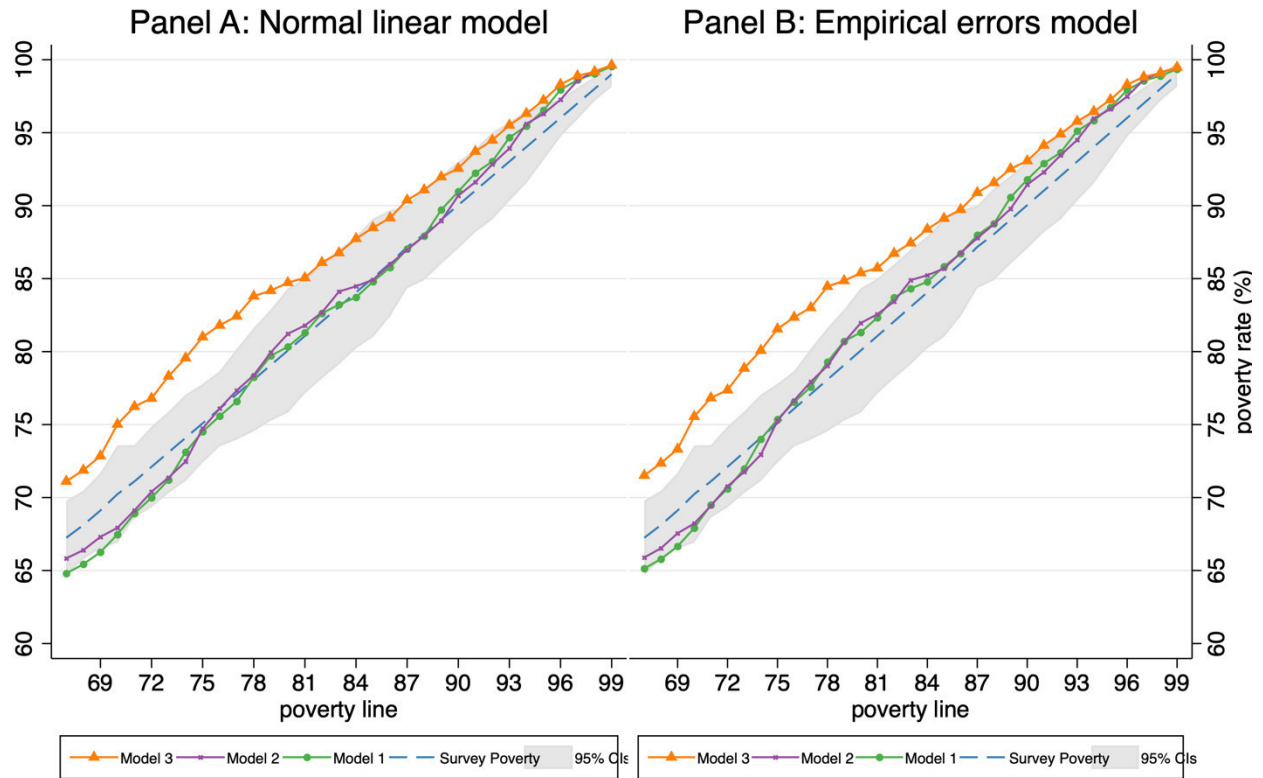
Source: Authors' calculations for UNHCR/WFP/CNARR Targeting Strategy and *Skoufias et al. (2001)*.

Table 8: Targeting Performance of sample of programs, current UNHCR/WFP/CNARR, and our imputed consumption-based Targeting

	Poverty lines								
	10th Percentile			20th Percentile			40th Percentile		
	Median	Min	Max	Median	Min	Max	Median	Min	Max
A. All 85 programs in Coady et al. (2004).	2.8	0.8	7.5	2.2	0.7	4.3	1.5	1.0	2.1
B. UHNCR Targeting	4			3.1			1.6		
C. Imputed Consumption based Targeting	5.5			3.3			1.9		

Source: Authors' calculations for UNHCR/WFP/CNARR Targeting Strategy and compilations based on data in *Coady et al. (2004)*.

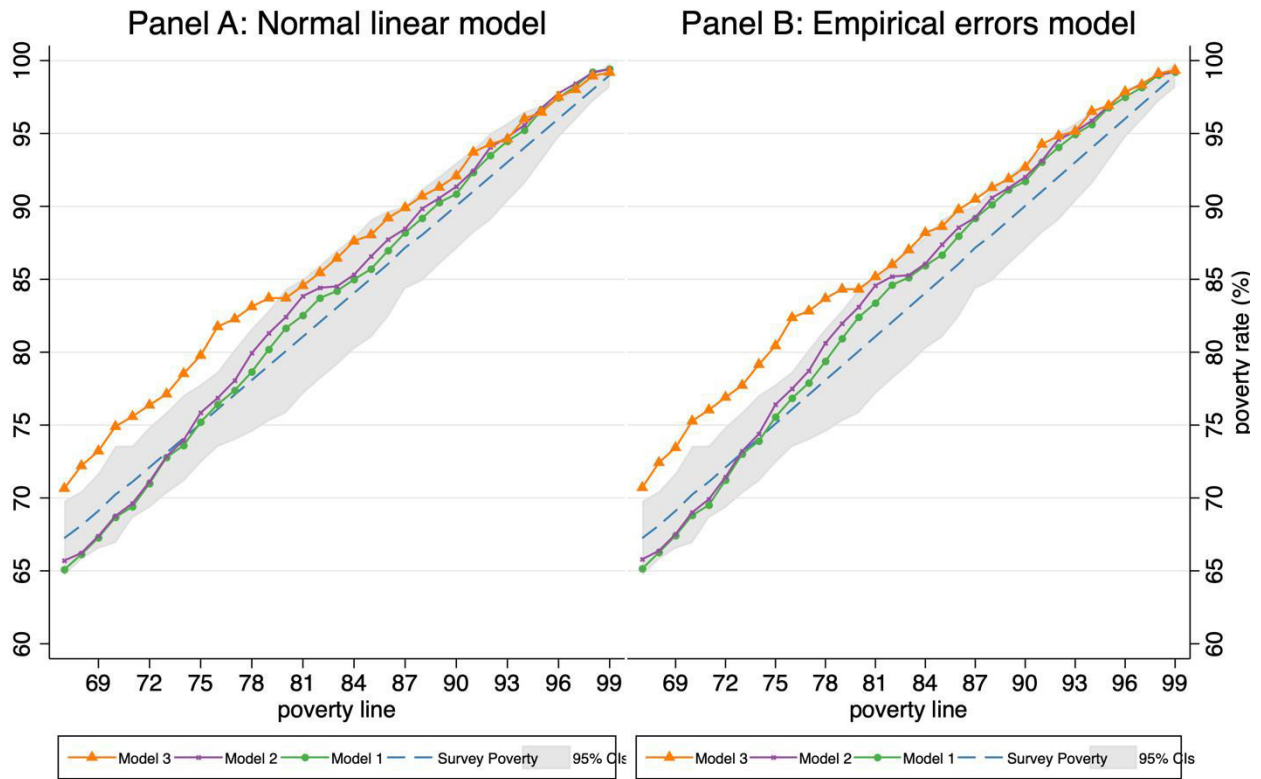
Figure 1: Predicted welfare and Survey based welfare for different poverty lines, ProGRES Targeting



Note: Comparison of models

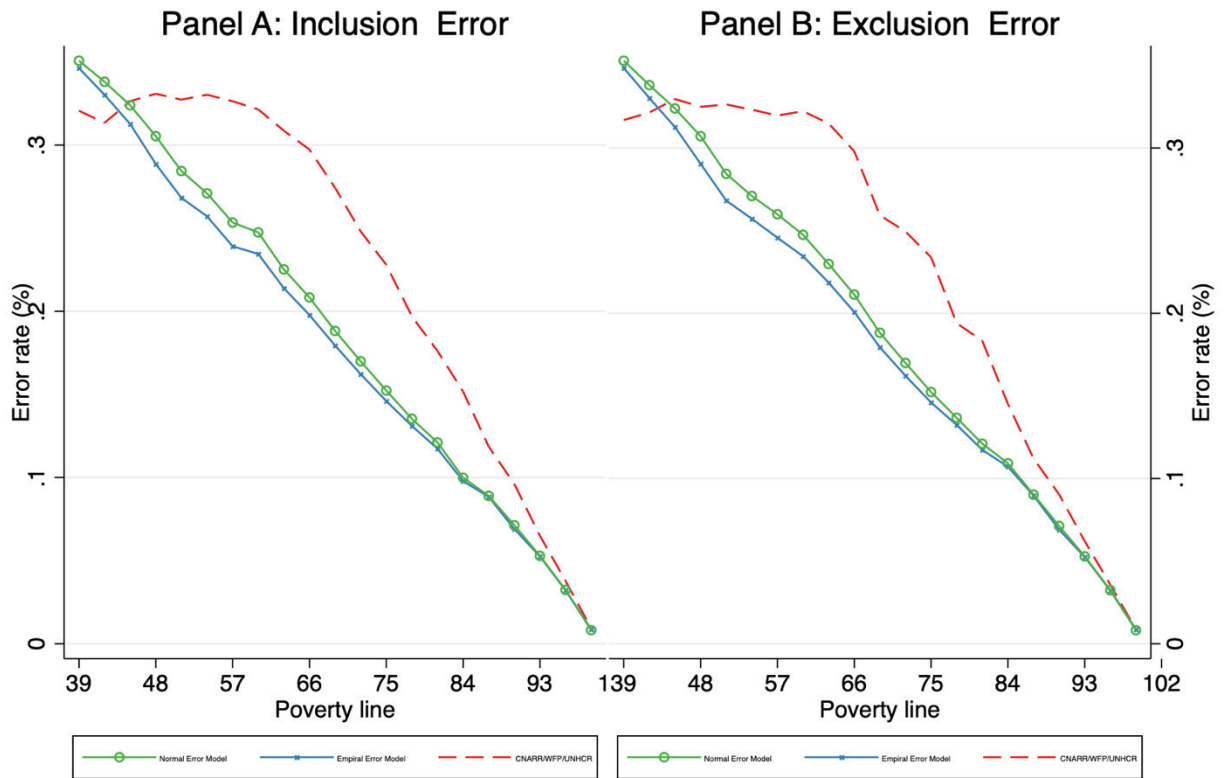
Note: The blue dashed curve presents the actual poverty rates derived from the PDM observations in the ProGRES Targeting, meaning that the blue dashed curve presents poverty rates derived from observed consumption of the PDM. The green solid curve with circle symbol represents the predicted poverty rates from Model 1 with observations from Merged ProGRES Targeting. The indigo solid curve with symbol “x” represents the predicted poverty rates from Model 2 with the Merged ProGRES Targeting observations while the orange solid curve with the triangle symbol represents the predicted poverty rates from Model 3 with the Merged ProGRES Targeting observations.

Figure 2 : Predicted welfare and survey based welfare based on food security for different poverty lines, ProGres Targeting



Note: Comparison of models

Figure 3: Comparison of targeting performances of different targeting methods



Note: Comparison of targeting performances

Annex: Additional tables and figures

Table A1: Distribution of persons of concern by Group in Chad

Type	Number	Proportion
Refugee and Asylum seeker	459809	68.9
Returnees	5746	0.9
IDPs	165313	24.8
Chadian Returnees from CAR	16718	2.5
Others	20000	3.0
Total	667586	100.0

Source: Authors' calculations, ProGres.

Table A2: Summary of data

Number	Dataset	Overview	Date	Implementing Agency	Existence of Consumption expenditure information	Relevant Variables to poverty imputation available
Panel A: Data available						
1	UNHCR Registration Data (ProGres)-	Census for all refugee households	June 2017	UNHCR	No	1. Demographics
2	Targeting Database 2017- All Chad	Census for all refugee households	June-July 2017	UNHCR/WFP and CNARR	No	1. Demographics 2. Asset and animal ownership 3. Coping strategies
3	Post Distribution Monitoring 2017-	Sub-Sample of refugees	2017	WFP	Yes	1. Demographics 2. Asset and animal ownership 3. Coping strategies 4. Consumption expenditure
Panel B: Data constructed for poverty imputation						
1	ProGres 2	CAR and Sudanese refugees living in regions covered by PDM	-	Constructed by authors	-	1. Demographics 2. Consumption expenditure for observations from PDM
2	Targeting 2	CAR and Sudanese refugees living in regions covered by PDM	-	Constructed by authors	-	1. Demographics 2. Asset and animal ownership 3. Coping strategies 4. Consumption for observations from PDM
3	ProGres Targeting	CAR and Sudanese refugees living in regions covered by PDM	-	Constructed by authors	-	1. Demographics 2. Asset and animal ownership 3. Coping strategies 4. Consumption for observations from PDM

Source: Authors calculations

Table A3: Estimation Model

	(1) Model 1	(2) Model 2	(3) Model 3
HH size	-0.16** (0.01)	-0.15** (0.01)	-0.14** (0.01)
Age of head of HH	-0.00 (0.00)	0.00 (0.00)	-0.00 (0.00)
HH is Farming	0.08* (0.05)	0.11** (0.04)	0.18*** (0.05)
Head of HH has primary education	-0.06 (0.06)	-0.01 (0.05)	-0.06 (0.07)
Head of HH attended Islamic School	-0.00 (0.08)	-0.09 (0.07)	-0.18** (0.09)
Head of HH has Secondary education	0.07 (0.09)	-0.06 (0.09)	-0.11 (0.10)
Head of HH has Higher education	-0.23 (0.33)	-0.25 (0.31)	-0.55* (0.30)
HH is Female	-0.18*** (0.05)	-0.12** (0.05)	-0.21*** (0.06)
Head of HH is divorced	0.00 (0.09)	0.01 (0.08)	0.11 (0.13)
Head of HH is widowed	-0.19*** (0.07)	-0.13* (0.07)	-0.25*** (0.08)
Head of HH is single	0.04 (0.12)	0.05 (0.10)	0.11 (0.16)
Country origin is Sudan	0.42*** (0.10)	0.59*** (0.11)	0.84*** (0.11)
Region 2	-0.60*** (0.11)	-0.30** (0.12)	-0.45*** (0.14)
Region 3	-1.13*** (0.12)	-0.94*** (0.12)	-0.77*** (0.14)
Region 4	-0.44*** (0.07)	-0.51*** (0.07)	-0.66*** (0.09)
Region 5	0.00 (.)	0.00 (.)	0.00 (.)
Region 6	-0.38*** (0.08)	-0.45*** (0.07)	-0.63*** (0.10)
Region 7	0.03 (0.06)	0.01 (0.06)	-0.03 (0.06)
HH has Phone		0.07 (0.06)	0.09 (0.07)
HH has Carts		0.33** (0.14)	0.39* (0.20)
HH has Bikes		0.11 (0.13)	0.15 (0.16)
HH has Moto		0.35* (0.21)	0.32 (0.21)
HH has Radio		0.20** (0.09)	0.15 (0.11)
HH has Cattle		-0.05 (0.12)	-0.11 (0.14)
HH has Horses		0.08 (0.07)	0.09 (0.11)
HH consumes seeds as coping strategies			0.01 (0.07)
_cons	6.43*** (0.13)	6.06*** (0.13)	5.99*** (0.15)
N	803	803	503
R ² adjusted	0.52	0.55	0.62

Source: Authors' calculations, PDM survey.

Note: Standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the log of household expenditure per capita and results obtained from the PDM survey alone.