Spring 2024

PMT, CBT, and Hybrid Models: Is Machine Learning the Future of Poverty Targeting?

Luca Orion Heidelberg
Bard College

PMT, CBT, and Hybrid Models: Is Machine Learning the Future of Poverty Targeting?

Senior Project Submitted to
The Division of Social Studies
of Bard College

by
Luca Heidelberg

Annandale-on-Hudson, New York
May 2024

Abstract

Addressing poverty in developing countries without relying on income data is becoming increasingly important, particularly since the Covid-19 pandemic has exacerbated the issue farther than anyone expected. This paper reviews literature on various poverty targeting models including Proxy Means Test, Community-Based Targeting, hybrid models, and machine learning-based models in hopes of finding the best method. The findings highlight the importance of model parameters, particularly in PMT, also revealing that the number of potential beneficiaries analyzed and number of indicators utilized can influence the targeting accuracy. CBT incorporates community involvement in the poverty targeting process at a lower cost than PMT, despite slightly lower accuracies. Hybrid models contain both benefits and problems with PMT and CBT, with many authors finding mixed results. The recent emergence of ML-based PMT models has shown promising results, particularly when applied to the feature selection phase. With targeting accuracies averaging 10% higher than PMT and cost efficiencies rivaling those from CBT, ML-based models might be the future of poverty targeting in developing countries.

Acknowledgements

I would like to thank my senior project advisor, Gautam Sethi, for assisting in the creation of this paper, and for providing invaluable resources and suggestions. In addition, I would like to thank Sanjay DeSilva for advising me throughout my college career, and lending his skills as a mentor and professor to help me succeed. This project would not have been possible without them.

Table of Contents

Chapter 1: Introduction

In developing countries, there has been an increasing demand for the alleviation of poverty, especially since COVID-19 has increased the headcount of the ultra-poor significantly. Developed countries have often employed the use of *means tests*, which created a threshold based on income data that determined who is eligible for benefits. However, this data is often lacking in developing countries because poverty might be defined by other measures (e.g. agricultural-based countries might find farmers to be wealthier, even if this is not reflected by income). Income might not always be the best indicator of social welfare. Researchers had to think of other characteristics to define poverty. At first, these definitions were defined on fairly one-dimensional space, in that they never really truly encapsulated the definition of poverty. These methods include but are not limited to self-targeting (a cutoff point for consumption is created) and geographic targeting (poorer areas were provided with more benefits). While these methods are useful when combined with new models, using them as a standalone program produces large targeting errors.

These targeting errors are how the effectiveness of a poverty targeting model is measured. More specifically, targeting errors are made up of inclusion and exclusion errors. Inclusion errors refer to individuals who should not have been included in the program, but the targeting model deemed them eligible. Exclusion errors refer to individuals who should have been included in the program, but the targeting model deemed them ineligible. These can also be referred to as leakage and undercoverage rates, respectively. The point of creating new poverty targeting methods is to reduce these errors, ideally identifying every household in need of benefits and excluding every other household. Going back to the example of geographic targeting, inclusion errors might occur because wealthy households might be located in a generally poor area. Similarly, exclusion errors might occur; poor households might be in a generally wealthier area.

In light of this, new methods have been developed that utilize the idea of Multi-dimensional Poverty (MDP), which "recognises people's breadth of needs and the different forms poverty can take that cut across different areas, or dimensions, of our lives" (Development Initiatives). The first modern method to emerge is an evolution of the means test mentioned before. Because means tests rely on income data, the idea of using proxy variables to represent income was introduced. Thus, the *Proxy Means Test* was born. There are no set variables to serve as a proxy, so it is up to the researcher to determine what characteristics best represent poverty relative to the region. Generally speaking, demographics of individuals, assets, and household characteristics are used to at least partially represent well-being. These variables are then run in a regression model against (normally) per-capita consumption. The coefficients are used to calculate the *PMT score*, which is a relative measure that allows researchers to create a cutoff point, in that only households that score below the cutoff point are eligible for benefits. The first PMT model to be conducted using these guidelines was in India by Revallion and Chao (1989). Due to a small budget and few variables, the authors found little significance in their results. The limits were recognized, however, and further research was encouraged by the authors. Literature on PMT models can be read in Section 2.1.

During this earlier period, the imperfections of proxy means tests led researchers to consider other methods to target poverty. The definition of poverty is the most important part of any poverty targeting model, so maybe it needs to be treated with a different perspective. In

PMT, administrators create the definition of poverty through the variables used and the weights assigned to them. In *community-based targeting* (CBT), administrators work with local community members to develop a definition based on the locals' understanding of poverty. The idea here is local community members possess information on what it means to be poor that is inaccessible through simple data analysis. Once community members create a list of criteria for what it means to be poor, the next step is often to have them also identify which households require benefits. The first CBT program to be conducted was in 1992 in Kenya during the drought emergency program, which required the government to understand what the information locals had to help alleviate the increase in water poverty. Literature on CBT can be read in Section 2.2.

Over the last couple of decades, the benefits and drawbacks of both these methods have been made clearer. For PMT, the definition of poverty created by the administrators might differ from what poverty means in the context of the area, which can lead to targeting errors. Still, if the proxy variables are chosen correctly, PMT is far more useful in targeting poverty than just self or geographic targeting. Data collection can also be costly, meaning it is most expensive to broaden the definition of poverty by including more proxy variables. Finally, the lack of involvement with the community can often lead to dissatisfaction from the target population, and an unwillingness to sign up for the program.

This is solved through CBT, where community involvement is the main focus. Because collecting data is done through an in-person group process, the cost of this method is significantly less. A key advantage that CBT has over PMT is the access of *knowledge*, specifically of social networks, which is provided by community members. However, the usability of this knowledge is up for debate. It has been found that community members might lack the specific information needed to correctly target the person as poor or non-poor (e.g. consumption data). Additionally, community members might act with *strategy* in mind, in that if the community is small enough, locals might pick and choose who receives benefits based on the scope of their social network. This leads into the idea of *elite capture*, which CBT is more prone to than PMT. Elite capture is when local leaders and local elites influence the targeting procedure in order to procure more benefits for themselves and their family/friends. CBT uses groups of community members to assist in targeting, which is easier to influence than any household survey or census.

Both of these methods have important strengths and weaknesses, which inspired researchers to attempt a stronger model by creating a *hybrid* of PMT and CBT. This can be done in a variety of ways, but is generally done using two methods. The first is when administrators choose a group of local community members to help identify what the local definition of poverty is. This definition is then applied to a PMT model using community-chosen variables and weights. The second is a more separated hybrid model, in that a CBT model is carried out in full, and the community's choices are then compared to the results of a PMT model (to verify accuracy and gain understanding of the differences). More literature on hybrid models can be read in Section 2.3.

While both have important strengths, the weaknesses of PMT and CBT have spurred updated models in recent years. These updated models have implemented the usage of *machine learning* to PMT models. For most of the literature, machine learning (ML) has been mostly used in the feature selection stage, which means it helps narrow down the variables and relative weights that are necessary to result in the highest targeting accuracy. There are a wide range of ML algorithms that could be applied, but the most common are regression trees, LASSO (Least Absolute Shrinkage and Selection Operator), and gradient boosting. [explain each here?]. The hope is that machine learning will both increase the speed and decrease the cost of targeting methods by quickly choosing the least amount of variables required. Literature on machine learning based models can be read in Section 2.4.

The results show PMT producing a targeting accuracy of 60% on average compared to CBT's 50%, with both containing similar amounts of inclusion and exclusion errors. PMT was found to have increased errors due to illogical coefficients being assigned based on the regression model, and CBT was prone to both a misunderstanding of poverty by the local population and elite capture. CBT was found to cost considerably less than PMT, leading to those authors promoting further research keeping community-based practices in mind. Hybrid models were found to perform either similarly to PMT or better. The authors attribute this to the combination of community involvement from CBT and the verifiability of PMT. Finally, machine learning based PMT models are analyzed, and are found to produce less errors than both PMT and hybrid models (around 70% targeting accuracy), particularly when using LASSO and regression forests. They also decrease the cost (relative to that of PMT) considerably because of the model's ability to choose the least amount of variables while maximizing the accuracy. ML-based models are promoted to be the most effective of all modern targeting methods.

Chapter 2: Literature Review

2.1 Literature that carries out and analyzes the effects of a PMT targeting method

Klasen and Lange use a Proxy Means Test to tackle the problem of targeting in a developing country that lacks expenditure or income data in *Targeting performance and poverty effects of proxy means-tested transfers.* Like other authors, Klasen and Lange are motivated to find the most effective targeting method in a country that does not have verifiable income data by maximizing poverty-removal benefits while minimizing costs on both the administrators and community members. The authors acknowledge alternative methods such as geographic targeting and community based targeting but choose to study PMT because it has not been implemented in Bolivia before, and there is readily available data to do so.

To carry out this PMT in Bolivia, Klasen and Lange use national household expenditure data from Encuesta de Hogares (2008 & 2011). This data contains information on food, education, rents, services, durable goods, and non food items. An Ordinary Least Squares regression model is then employed to study the relationship between the proxy variables and log real expenditure. The predictive model is then trained using the 2008 and a subset of the 2011 data (author refers to them as *calibration samples*, but a more common term is *training data*). The remaining households are then treated as the validation sample (*testing data*). The proxy variables used from the dataset to estimate poverty are then split into five groups: geographic, demographic, dwelling characteristics, electricity bill information, and assets. Following this, four different models are used to analyze the effects. The first model incorporates both geographic and demographic proxy variables, the second model adds dwelling characteristics and service use, and the third adds information on electricity services spending. The fourth and final model replaces spending on electricity with dummy variables representing ownership of durable goods (such as refrigerators, personal computers, TV sets, etc.). To evaluate the targeting error rate, areas under the curves are evaluated at poverty headcounts of 10, 25, and 50 percent of the population.

The results show varying targeting accuracy, depending on the imposed poverty headcount and the beneficiary proportion. Model four, taking into account a variety of proxy variables as well as durable goods, performs the best in terms of within sample and out of sample testing. That targeting accuracy for this model ranges between 61% and 65%, respectively. Models one through three have targeting accuracies of roughly 38%, 54%, and 58%, respectively. Overfitting, caused by leakage, is present in these models but mostly only affects those just above the poverty line, instead of incorrectly targeting the wealthy/elite. Klasen and Lange attribute the increase in accuracy to the broadening definition of poverty. Additionally, when the poverty headcount rises, the area under the curve increases, suggesting if you define more people as potential beneficiaries, more of the poor are correctly targeted. This observation is consistent with the findings of Alkire and Seth (2012), Schnitzer and Stoeffler (2021), Stoeffler et al (2016), and Grosh & Baker (1995).

Grosh and Baker (1995) run a simulation PMT designed to predict welfare through OLS models, utilizing data in Jamaica, Peru, and Bolivia. The authors want to discover which information is best used on a PMT and how it might improve the targeting accuracy. Latin America is a region of interest for the authors because Grosh (1994) previously showed the PMTs were the best solution for targeting in that area, since income data is not generally available. However, these studies were conducted with limited proxy variables and still produced significant errors. Grosh and Baker are therefore trying to discover if including more variables, in conjunction with minor tweaks, will improve the targeting accuracy of PMT.

The data for this study is pulled from the Living Standards Measurement Study (LSMS), which contains information on multiple socioeconomic and sociodemographic variables. These variables are then plotted against expenditure, and those that are most correlated are used in various Ordinary Least Squares regression models. The first model contains information on the location of the household, the second adds the housing type, the third adds family characteristics, and the fourth adds ownership of durable goods. Finally, a fifth model is included that consists of the most correlated variable from each of the other four models. This is a very similar setup to Klasen and Lange, in that the author is attempting to discover if more information is correlated with higher targeting accuracy. The authors also tested what would happen if only the poorest 50% of the population was analyzed through Model 4.

The results agree with Klasen and Lange, in that the more information is provided to the models, the better the accuracy (lower inclusion and exclusion errors). Grosh and Baker find that models three, four, and five all produce similar results, with models three and four performing slightly better (though durable goods did very little to improve the results, despite the opposite being found by Klasen and Lange). Specifically, model three had inclusion and exclusion rates of 33.7% and 41.3%, and 34.2% and 41% for model four. Grosh and Baker then move on to analyze the effects of excluding the wealthy and elite from targeting, focusing only on the poorest 50%. "For most social welfare programs, the recipient must take some action to apply. Members of the middle and upper class may not bother to do so" (Grosh and Baker et a 18). They found that the targeting accuracy of the model increased substantially. The exclusion errors decrease from 41% to 13% and Inclusion errors drop from 34.2% to 28.2%. This comes at the cost of increased program costs. Grosh and Baker sum up their results by stating that more information is better for proxy tests and cutting down the population to the poorest 50% improves exclusion and inclusion errors.

Proxy variables are the main ingredients when creating a PMT model. They represent poverty in ways that don't utilize income data. It's very important to consider what the proxy variables should be and how many should be included. Grosh & Baker and Klasen & Lange have both found that an increase in the number of proxies leads to an increase in targeting accuracy. Sebastian et al. (2018) aim to verify or challenge these results in *A Proxy Means Test for Sri Lanka*. The authors are motivated by the reformation needed by the Sri Lankan government regarding their targeting program, Samurdhi. Samurdhi currently uses self-targeting, which is subject to strategic moves by individuals to lie about assets in order to receive program benefits. In 2016, only 39% of cash transfers (monetary benefits) reached the poorest 20% of the population, so the authors will also analyze what percentage of the ultra-poor is targeted.

The dataset used is the HIES 2016, which is a household survey for Sri Lanka. When choosing the variables to use, Sebastian et al. outline three criteria; a small set to maintain cost effectiveness, easy to observe and measure, and difficult to manipulate. When choosing the variables, the authors use ANOVA (Analysis of Variance) to narrow down the criteria to only variables that explain variation in consumption by creating a partial sum of squares threshold. These variables are then utilized in an OLS regression model against the log per capita consumption and the coefficients are used to predict the poverty. The models get increasingly more restricted by decreasing the partial sum of squares threshold in three different models. The number of predictors are also reduced for each model, with the first having 36, the second having 27, and the third having 18. Sebastian et al. (2018) are hoping that this method of narrowing down variables will assist in reducing targeting errors.

The authors find model two performing the best, with model three behaving very similarly. Model three does have a slightly lower R-squared value due to the lower number of variables, but these variables have a significance that is not as present in the first model. However, “across all models, categorical household size variables are the largest contributors to the PMT score, with the score decreasing as household size increases” (Sebastian et al. 8). Additionally, the number of beneficiaries that would be considered ultra-poor increase dramatically. Under Samurdhi, only 39% of beneficiaries were in the bottom decile, but this number increases to about 65% for all three models. Finally, when these models are introduced to different eligibility cutoffs, Sebastian et al find that raising the eligibility cutoff both increases coverage and leakage rates, so a middle ground must be found. These results agree with those found by Klasen & Lange and Grosh & Baker in that more variables introduced to PMT reduce errors. But Sebastian et al. take it a step further by claiming that selection of variables is extremely important to consider. They conclude that Model three is the most realistic because it produces only slightly lower targeting efficiency than the other two, while using only a few variables to limit the potential cost.

In *Universal Basic Incomes versus Targeted Transfers: Anti-Poverty Programs in Developing Countries,* Hanna and Olken begin by focusing on one of the 17 Sustainable Development Goals presented by the U.N. which states the elimination of poverty by 2030. The recent economic growth of many countries has also exacerbated the problem of poverty, which means the need for proper targeting methods now are more important than ever. Due to the lack of income data in developing countries, proxy models and community-based models have been introduced, but still require further research. Universal Basic Income (UBI) is a poverty alleviation program that provides a fixed transfer to everyone, regardless of their incomes or assets. Hanna and Olken aim to compare (UBI) to PMT in order to see if the effect is worth using one over the other.

This study is conducted simultaneously in Peru and Indonesia, using the datasets from the Peruvian National Household Survey and the Indonesia National Socioeconomic Survey. The variables used are socioeconomic indicators that are used in the respective national targeting programs. For the proxy means test, the authors are basing the targeting method on consumption per capita, and subsequently split the data into training and testing sets. The per-capita consumption is then regressed on various indicator variables (on the training data), and then predicted based on the coefficients from the regression model. This is then compared to UBI in the sense of costs, errors, and benefits delivered to the poor.

For PMT, the R^2 value was determined to be between 0.53 and 0.66, meaning there is some explanatory power, but prediction errors are still present (exclusion errors at around 20% and inclusion at 22% ~ 31% for both Peru and Indonesia). This targeting accuracy is consistent with the findings of Klasen & Lange in *Targeting performance and poverty effects of proxy means-tested transfers,* a paper which Hanna & Olken claim is similarly carried out. When comparing this method to UBI, Hanna & Olken find PMT to provide more benefits to the poorest of the population, while also costing slightly less. There are also less inclusion errors present, meaning less benefits are wasted on those who are not in need. The authors note a key finding in that “narrowly targeted programs - those focused on distributing large benefits per capita to the poorest of the poor - appear to achieve much higher utility levels than less narrowly targeted programs, including but not limited to UBI” (Hanna & Olken 214). Utility levels refer to the usefulness and disability of the program. So, referring back to the main goal of eliminating poverty, Hanna & Olken find that PMT is more *useful* at helping the poorest of the poor than UBI.

Coady and Parker recognize, like many authors, the problem of inclusion and exclusion errors when targeting developing countries. Program benefits are often seen by those who are not in need, and missed by those who desperately need them. Coady points to one of his recent papers (Coady, Grosh, and Hoddinott 2004) that found very wide variations in targeting performances, but a correlation between number of methods used and targeting performance. One idea that has not been touched on is self-selection in conjunction with a proxy means test. This model is studied by Coady and Parker (2009).. Coady hopes that it will improve Mexico’s existing targeting program, *Oportunidades*, which uses PMT to focus on rural and urban areas, but suffers from considerable targeting errors.

The study is conducted using data from Mexico’s *Oportunidades* program, and uses socioeconomic variables like education, assets, and durable goods to create a PMT (weights were assigned to each variable, and a score was calculated that determined eligibility). Self-selection is combined with PMT in the following steps. The first, and one of the most important steps, is the process by which households learn about the program, in that they recognize the program may apply to them and go out of their way to sign up. Other important steps include whether or not they actually chose to apply and were subsequently accepted into the program. This entire process is through self-selection, in that it is up to each individual household to weigh the costs of applying for the program and the possible benefits it may provide. Households are then selected through the Proxy Means Test to determine who is eligible. In order to determine the success of the program, Coady used regressing the socioeconomic variables against welfare, instead of the more commonly used expenditure (as seen in Alatas et al. (2012), Dimri and Maniquet, Klasen and Lange).

On aggregate, Coady finds that 63% report knowing about the program, 47% report applying for the program, and 30% were deemed eligible. However, the author finds that for the bottom quintile, 87% of people who know about the program end up applying (compared to the top quintile, where only 45% apply). For those lower income households, nearly every household that applied was accepted, meaning that the distribution of knowledge is critical to reducing leakage to higher income households while not excluding the intended beneficiaries. Awareness of the program itself is something that other authors haven't studied as thoroughly as in this paper, despite it clearly playing a major role in the end result. However, the finding that focusing on the poorest improves targeting efficiency is shared by Hannah and Olken and Grosh and Baker.

2.2 Literature that carries out and analyzes the effects of a CBT targeting method

Before diving into various studies, Conning and Kevane (2002) provide important questions to be considered when creating a Community Based Targeting (CBT) model. "Will community mechanisms lower the costs of delivering benefits to a target population? What kind of distributions are likely to result after considering the need for incentives and the potential for elite capture? What determines the quantity and types of subsidies provided to those who were targeted?" (Conning and Kevane 377, 378). These questions should be kept in mind when analyzing the following case studies.

Conning and Kevane (2002) conduct an extensive literature review that studies the effect of CBT on multiple different communities. They move on to discussing the creation of a CBT, which involves certain categories of questions. They identify these as "Will community mechanisms lower the costs of delivering benefits to a target population? What kinds of distributions are likely to result when community-based targeting is employed after considering the need to provide incentives, program leakage, and the rents that intermediary agents might potentially capture? What determines the quantity and types of subsidies provided to those who were targeted" (Conning and Kevane 377-378). The authors then identify many advantages CBT has over other targeting methods, including lower administrative costs, better accountability, increased social capital, and more adaptability to local conditions. This amends many of the issues that the previous authors have with means tests and compliments the findings that CBT is an appropriate alternative. For example, Hillebrecht et al (2023) were concerned about PMTs ability to target the 'transitory poor.' This is another paper that stresses the importance of community involvement, and how crucial it is for CBT to outperform other targeting methods. Some disadvantages of CBT include minimal accounting, increased local division, and high opportunity costs. In order to design a CBT system, a demand driven approach must be taken that includes a diverse range of community agents from various boards and groups around the local area. They conclude with four key points regarding CBT: (a) communities have various ways to exchange information, influencing the cost saving advantages (b) local communities might have different levels of willingness to correctly target the poor; (c) the state of the national political economy can unexpectedly change the effects of programs, and (d) the funding and evaluation of CBT opens special conceptual and practical problems.

Savadogo et al. (2015) take into consideration the concept of Multidimensional Poverty, which is a redefined measure of poverty that takes education, health, and standard of living into consideration. This concept is also employed in a hybrid model by Alkire and Seth, analyzed later on. Savadogo et al. hope to understand the ever-broadening definition of poverty by utilizing local community information. Similar to Alkire in *Selecting a Targeting Method to Identify BPL Households in India,* Savadogo et al. find that traditional methods do not successfully target multidimensionally poor households. Community information can help administrators understand the multidimensional impacts on poverty, and is particularly useful because the definition of poverty is first defined by locals before the targeting exercise takes place.

To carry out this experiment, the authors conducted a study in the Nouna Health District located in northwest Burkina Faso. Participants were chosen from each village within that district, and asked them to participate in a focus group, where they would decide the criteria and perceptions of poverty in that region. The focus groups ended up defining poverty as: a scarcity of basic needs, vulnerability, powerlessness, voicelessness, indecent living conditions, absence of social capital, and a number of other criteria. Each household was then separated into a Poor group, Intermediate group, and Wealthy group by fellow community members.

After the criteria for poverty was defined by the focus group and the targeting exercise took place, the qualifications for each of the three groups were distinguished. For the poor group, most of the households were identified by not having sufficient food, health issues preventing their work, the inability to solve their own problems, or having nothing at all. The intermediate group was defined by households that had daily meals, had no need to beg for food or work, and could commute by their own means. The wealthy groups were able to solve their own problems, help others, own livestock, and be financially independent. Overall, the community-based targeting exercise was most accurate in defining the poor groups, which is consistent with the findings of Beaman et al. (2021) and Sebastien et al. (2018), who find the CBT aspect of their hybrid models is credited with targeting the ultra-poor accurately. Beauge et al. (2018) also look at the effectiveness of CBT on targeting the ultra-poor.

In *How much does community-based targeting of the ultra-poor in the health sector cost? Novel Evidence from Burkina Faso,* Beauge et al. (2018) conduct a cost-benefit analysis of a community based targeting model in Burkina Faso while keeping in mind previous literature has found PMT to perform better despite higher administrative costs. The authors will specifically analyze the accuracy of the ultra-poor, defining ultra-poor as “someone who is extremely socially and economically disadvantaged, unable to care for themselves and who is without internal or external resources" (Beauge et al. 3). The authors also compare the cost of CBT to Universal Basic Income, a program that provides transfers to every household regardless of consumption. This comparison is also conducted in Hanna and Olken's (2018) paper *Universal Basic Incomes vs. Targeted Transfers: Anti-Poverty Programs in Developing Countries*.

To conduct this, Beauge et al. (2018) analyze a previously conducted study launched by the Ministry of Health in 2014. This study contained a CBT model that involved community members ranking households based on socioeconomic and sociodemographic factors, with the poorest receiving free healthcare benefits. This model combined performance-based financing (PBF) with health-care coverage based subsidies for the poor. This means that health-care providers were paid based on a fixed price-per-services, and a lump sum is provided to reimburse the treatment of the ultra-poor. Costs for the government-side and community-side were traced for both the design of the program as well as the implementation, and subsequently compared to the cost of implementing UBI.

Beauge et al. find that the financial costs of the CBT interventional was $587,510, and the total cost of the program was $1,213,447. The selection process accounted for a significant portion at $392,060. The cost per person was determined to be $11.83 per ultra-poor person, and $5.73 per person for everyone else. Community costs incurred the highest proportion of funds, at 43%. The total cost of the program is still considerably less when compared to UBI (about half of the cost). Despite the community-level and implementation-level costs for CBT in Burkina Faso, Universal Basic Income would still cost considerably more, while providing a lower proportion of benefits to the ultra-poor. These results agree with those found by Hanna and Olken.

Alatas et al. (2016) begin by recognizing the growing importance of information possessed by local community members. However, Alatas et al. recognize that the type of information needed may not always be the type of information community members possess. “Agents often have to learn about a constantly evolving parameter (in our case, it is the wealth of others in their village) through the social network, and moreover, in assessing who is poorer than whom, they have to compare multiple bits of noisily-learned and potentially detailed information” (Alatas et al. 2016, p. 1664). This means that local community members might not understand the specific information needed to calculate who is in need.

To address this, Alatas et al. (2016) estimates a learning model that will utilize the previously understood information about how information spreads by involving an ever-changing state variable. This variable is estimating the well-being of some other household in the community, but changes over time as peoples’ situations change (specifically, the wealth of the household). This changes by agents receiving pieces of information and aggregating them. These pieces of information are each treated independently, which is “the same as full Bayesian learning, but on arbitrary networks, where people may receive the same piece of information through many different paths, this rule will not typically be Bayes optimal since it assumes that each piece of information is independent when in fact it is not” (Alatas et al 1665). This learning model is then applied to a network dataset on 631 Indonesian villages (previously used for a different CBT model) that contains information on social network data and the accuracy of descriptions of poverty status by households.

This paper stems from Alatas et al. (2012), which studies a PMT model also based on Indonesian data. The authors want to discover the implications of network knowledge in poverty targeting and poverty reduction. The results show centralized individuals are more likely to correctly rank other community members (especially if they are socially closer to the exact household they are ranking). Alatas et al. (2016) find the targeting accuracy to be 60%, which is on par with the results from other PMT tests."The findings show that the network structure and our learning model not only accurately predict how information spreads, but are also useful in understanding how real decisions are made using that information" (Alatas et al 2016, p. 1700). Using a learning model that aggregates over time provides key information on information networks, and how knowledge spreads within a community.

2.3 Literature that utilizes hybrid models and compares CBT and PMT

According to many previous studies, PMT has been found to perform better than CBT. The authors of these studies, however, have looked at the number of targeted households in the same year as the targeting exercise. This presents one of the problems of PMT, in that it might successfully identify persistently poor households, but exclude transitory poor households, or households that may be poor at the time of the targeting exercise but are poor in the following years (or vise versa). CBT helps solve this issue by utilizing the *information* that local community members possess about other members and giving out transfers based on that.

Hillebrecht et al. (2023), in *The dynamics of poverty targeting*, are motivated by the issue of the lag time between targeting exercises and program implementation.

To find the root of this problem, Hillebrecht et al. (2023) conduct a study in Nouna Village in Burkina Fast to determine the differences between Proxy Means Testing and Community Based Targeting. They do so by creating a simulated environment based on data from a demographic census of 6,148 households, which contains demographic characteristics of households that will be used to determine the PMT score. Weights for each of these characteristics (variables) are then obtained by looking at the regression coefficients when a regression is run on the per capita consumption of those variables. This is ultimately used to calculate the PMT score, which can serve as a threshold for poverty (those below a certain score are deemed 'poor'). One of the authors' goal for PMT is to avoid overfitting, which is a downward bias in the targeting error at baseline. A CBT exercise was also conducted, asking locals to create three 'poverty brackets,' and assign households to one of the poverty brackets. This was based on criteria that was determined in focus groups previous to the targeting exercise. The community subsequently elected three local 'key informants' to choose what bracket each household belonged to and who received benefits. In order to tackle the previously mentioned problem of lag time, the authors analyzed the effects of the targeting exercise over multiple years.

At baseline, Hillebrecht et al. (2023) find that PMT on average has eight and nine percent less targeting errors than CBT, outperforming CBT in both the rural and urban areas (statistically significant at the five percent level). However, this story changes dramatically when analyzing the effects twelve months later. PMT is only slightly more accurate when pooling the data, but CBT outperforms when looking at urban households. The exact same thing is found when analyzing the effects thirty months after baseline, in that PMT is slightly more accurate on aggregate, but CBT performs better in rural settings. This could loop back to the idea of 'strategic individuals' who might have bias towards households they know, in that the density of urban areas translates to the likelihood of knowing a particular household. The authors end by theorizing that, because of this difference in urban and rural settings, geographic targeting preceding CBT might have an even greater impact. Schnitzer and Steoffler's (2021) results agree with this theory in that geographic targeting has massive benefits if done prior to a targeting program.

Nssah (2018) begins by stating questions that should be asked when analyzing a targeting model. Did what was supposed to happen, in fact, happen? In other words, did the targeting model effectively target those who are in need, and did the programs employed help improve the quality of life of those individuals? "An intervention is relevant if it is the right thing to do for the target population, given the problem and the circumstances they face" (Nssah 3). For Nssah, the main goal of creating a hybrid model is to create a method that is fair, effective, and efficient. To do this, targeting errors must be reduced, and the two rules of equity must be followed: impartiality and consistency.

This is done using the priority method, a targeting mechanism that says those with the greatest claim to a resource are deserving of that resource. The dataset used from the Social Safety Nets Pilot Project (SSNPP) which contains socioeconomic information on previously carried out PMT and CBT programs in Cameroon. When adjusting the PMT model and CBT model, priority is based on those who are the most deprived of an indicator, and a list is generated from worst-off to best-off. A threshold is then implemented and identifies those who will receive benefits. Different poverty levels are then defined to understand what happens when a decreasing number of households are considered poor. When comparing the two models, Nssah focuses on analyzing the results keeping in mind what is *supposed* to happen (goes back to originally asked questions). This means comparing the targeting outcomes with "counterfactual outcomes from a neutral assignment mechanism which enforces equal treatment of all potential beneficiaries" (Nssah 30). If, when comparing the outcomes, the results vary significantly, the significance of the observed result must be considered.

Under the first scenario, where the poverty line defined 67.5% of the population as poor, the targeting success rate for PMT is 14% higher (67% vs 54%). The inclusion and exclusion rates were both found to be 9% lower for PMT as well. Under the second scenario, where 35% of the households are considered poor, PMT still outperforms CBT by 7% success rate (63% vs 56%) but the difference in error rates have decreased (PMT now only has 4% less exclusion errors and 8% less inclusion errors). The finding that this happens when the poverty line is decreased disagrees with the findings from Alkire and Seth (2010), Alatas et al. (2012), and Grosh and Baker (1995), who find the opposite occurs (when number potential beneficiaries are *increased*, CBT errors *decrease*). When comparing these results with the counterfactual, they find that PMT has a stronger relationship, possibly due to its targeting success rate despite the inclusion and exclusion errors in scenario 2. On aggregate, PMT performs better than CBT, but Nssah (2018) claims that the differences are still somewhat negligible, in that a deeper understanding of each is still needed.

Stoeffler et al. (2016) are interested in studying which targeting method (CBT or PMT) is appropriate for targeting. The authors want to shed light on the differences between the two methods since previous studies have had inconclusive results regarding what situation to use PMT vs CBT. For example, previous studies have shown CBT is more effective in targeting based on human capital and physical assets, whereas PMT is better at identifying low per capita consumption. CBT also often generates a lower efficiency (higher targeting error), but importantly results in greater satisfaction from community members. Recent literature (Handa et al. (2012), Sebates-Wheeler (2015), Hurrell (2018)) have suggested that, in Sub-Saharan Africa, including community members provides a slight advantage over PMT because the definition of poverty might change from region to region. Steoffler et al. (2016) are also aware of literature that has employed geographic targeting in conjunction with PMT to reduce errors (Nguetse-Tegoum, Mills 2015; World Bank), and are curious to see if a hybrid approach that includes community based targeting would further improve the efficiency.

This paper studies the performance of PMT, CBT, and a hybrid model by reconstructing a cash transfer program in Cameroon, specifically in an impoverished northern region of the country. The same data utilized in the cash transfer program is used by Steoffler et al. (2016), which contains consumption information on 2084 households (with variables such as household characteristics, housing conditions, and assets). For CBT, forums were set up with ‘Local Targeting Groups’ (GLC), which included random community members to identify roughly 70% of the village as poor. For PMT, the authors use the dataset mentioned before to score households based on weighted variables. A score is then calculated to represent the level of poverty, and the eligibility threshold is adjusted based on the ‘desired number of beneficiaries for the project’

(Steoffler et al. 8). The hybrid model will combine completed versions of both PMT and CBT. Households that are selected to receive benefits will be selected by the community *and* receive low PMT scores.

Steoffler et al. (2016) find that CBT performed worse than PMT when trying to target those with low per capita consumption but performed better in terms of low physical and human capital. The CBT model also had higher inclusion and exclusion rates, at 25.9% and 47%, respectively, when compared to that of PMT (16.7% exclusion errors and 21% inclusion errors). When poverty thresholds are adjusted for the specific village, CBT targeting errors decrease, but PMT still outperforms. The hybrid model produces similar results to PMT, with the exception of considerably higher exclusion errors. “This indicates that while using hybrid targeting may reduce PMT survey costs (if only households selected by the community are surveyed), exclusion errors are increased” (Stoeffler et al. (2016)). However, both methods perform better than CBT. The authors also find that the more potential beneficiaries selected the greater the accuracy (also found to be the case by Alkire and Seth (2012), Schnitzer and Stoeffler (2021), Grosh and Baker (1995)). One issue is the fact that CBT was carried out purely through random selection of community members, which can often lead to an increase in overall errors (Alatas et al 2016, Beaman et al. (2021)).

Schnitzer and Stoeffler (2021) are motivated by the lack of proper targeting methods despite the recent resurgence of cash transfer programs in *Targeting for Social Safety Nets: Evidence from Nine Programs in the Sahel*. Despite the increasing number of studies on this topic, the ‘best’ targeting method is still up for debate. These studies have often been misleading in terms of the concepts and methods used to create targeting performance which can slow down the ability to reach a consensus. One method that Schnitzer and Stoeffler claim is often not focused on by other studies is the idea of program coverage (share of the beneficiaries selected), which plays an extremely important role in targeting performance. The results of this study highlight the need for clearly defined objectives and methods. The authors focus on the Sahel region, which is considered to be one of the poorest regions in the world due to various economic crashes and shocks, which has only been exacerbated since COVID-19. These countries are in desperate need for a properly defined targeting program, because in the past, coverage rates for transfer programs were only 0.4%-1.6%, despite poverty rates ranging from 38% to 45%.

Schnitzer and Stoeffler (2021) use data from various censuses that included both demographic characteristics and information on previously employed targeting programs (which is how the authors know this area is in desperate need of a more defined program). The studies done using these datasets “indicate several issues in the implementation of the targeting operations (e.g., CBT thresholds deviating from the original target in Cameroon)” (Schnitzer and Stoeffler 12). The variables in the dataset, like Hillebrecht et al (2023), are analyzed through per capita consumption, which is a very common method for studying targeting performance. The methodology defined by the authors is meant to tackle comparing PMT and CBT both within databases and between databases. For within databases, the authors adjust the PMT threshold to the percentage of the population that CBT deems as poor. For between datasets, Schnitzer and Stoeffler (2016) create a harmonized selection threshold for PMT variables and set them all equal to 35%. This allows the authors to change the PMT threshold easily to allow for analysis across different databases. Additionally, it is important to note that geographical targeting took place before the collection of information in each of the datasets. This means the authors can focus on specific regions in the Sahel area that are the most in-need of transfer programs. This is exactly what Hillebrecht et al. (2023) suggested, despite Schnitzer and Stoeffler (2016) employing the idea 7 years earlier.

Utilizing CBT thresholds to adjust PMT thresholds, the authors find that PMT will select the households with the lowest per capita consumption relative to CBT. The targeting errors resulting from this for CBT and PMT are 50% and 39%, respectively. The authors notice an important variation that exists across databases. They find that when the selection threshold is higher (more beneficiaries chosen), the error rates for both CBT and PMT go down (with PMT still performing better). This finding regarding program coverage is consistent with findings from other authors (Alatas 2016, Alkire and Seth (2012)). Additionally, the results agree with the findings from Hillebrecht et al (2023), in that if targeting errors are analyzed during the same year that the targeting program took place, PMT performs better than CBT (slightly lower error rates).

Alatas et al. (2012) find the need for social safety net programs a necessity, particularly in developing countries where income data may be inaccessible. Ergo, targeting methods like PMT and CBT prove extremely useful because of their ability to utilize proxy variables and community knowledge (respectively) to predict and target households in need. The author is trying to weigh the differences between the two models and study if a hybrid outperforms either. These differences, according to Alatas, are mainly the “tradeoff between the better information that communities might have versus the risk of elite capture in the community process” (Alatas et al. 2). Elite capture is the idea that local elites will control who receives benefits due to their influence and power in the area. PMT can limit this by utilizing government consumption metrics that are out of the hands of local elites. The reason why Alatas et al. are conducting this experiment in Indonesia is due to the failure by the current beneficiary program (Bantuan Langsun Tunai, BLT). The World Bank estimates that around 45% of the funds in 2005-2006 were incorrectly distributed due to many oversights by the targeting system, indicating a dire need for an update.

Alatas et al. (2012) conduct a case study in 640 sub-villages in Indonesia in collaboration with the government. Prior to the exercises, geographic targeting took place which split Indonesian regions into 640 sub-villages, and villages were randomly assigned to one of the three models. One third of the villages is allocated for PMT, one third for CBT, and the remaining third for a hybrid model of the two. For the PMT, Alatas et al. analyze consumption data on a variety of demographic variables, and after running a regression, use the coefficients as weights to create the PMT score threshold. For CBT, the authors ask local community members to rank households from richest to poorest (prior to this exercise, the facilitators took time to go over what the community members considered when ranking households). In the hybrid model, communities would rank households as they did in CBT, but those rankings would limit the number of households that the government would conduct a survey on (PMT was then carried out using this survey data and determined eligibility). This hybrid model helps eliminate the potential for elite capture, because the community members are notified that households will be verified by government enumerators following the ranking exercise. Then to compare these three models, Alatas et al. create a baseline by collecting per capita expenditure, and every household that falls below PPP$2 per day is deemed poor.

The results show PMT outperforming both other methods when looking at the consumption-based error rate, with PMT incorrectly targeting households 30% of the time, and CBT and hybrid models incorrectly targeting about 33% of the households. As a result, PMT targets poorer individuals on average when compared to CBT and hybrid models. However, when comparing targeted households with daily per-capita consumption, CBT targets more households with a PPP of less than 1 dollar. Thus, in spite of the worse targeting rates, CBT seems to target more of the ultra poor than PMT. Alatas et al. also find that, similar to Schnitzer and Stoeffler, when analyzing the methods in terms of consumption-based error rates PMT performs better than both of the other models (CBT and hybrid). Because of its low cost and similar performance, Alatas et al. (2012) know that there is more to be understood about community targeting programs. This is why in Alatas et al. (2016), the same authors uncover more about how knowledge is spread in communities and how that affects decision making in the targeting process (this paper will be looked at in the CBT section). Both papers by Alatas et al. (2012, 2016) inspired Beaman et al. to look deeper into social networks and knowledge within communities.

Beaman et al. (2021) are motivated by the idea of social networks and how information flows within them. In countries, like Liberia, where administrative data on residents is either limited or nonexistent, policymakers often utilize community information to assist in targeting the poor. Beaman et al. are also driven by the findings in both Alatas et al. (2012) and Alatas et al. (2016). which show that CBT is more effective at targeting the ultra-poor and that social distance deteriorates the quality of knowledge. This leads the authors to the question: would asking community members who the most central individuals are increase the targeting performance of CBT?

This study took place in 13 neighborhoods in Monrovia, utilizing data from a 2018 household census. When taking the census, administrators asked households which 5 individuals they spend the most time with, which allows the authors to understand who is the most central in these networks. Community information is then analyzed as follows: random community members and local leaders were separately asked who should receive the benefits. Then, randomly selected community members were asked to nominate other community members to choose who receives the benefits. These three methods are then compared to each other in order to find out if nominated neighbors have better information than random community members. As a reference, Beaman et al. are also conducting a PMT on asset ownership.

The PMT shows that half the sample received frequent income shocks, particularly within the last year (COVID-19 had reached its peak the previous year in Monrovia). This, in conjunction with the low per capita daily consumption (PPP$2.51 per day) makes this population particularly vulnerable. During the census, 175 random community members, 345 community leaders, and 254 nominated neighbors were selected. Beaman et al. (2021) observed that leaders and nominated neighbors have very similar characteristics in terms of households, wealth, and assets, suggesting that leaders were often nominated by their neighbors. During the community knowledge interview, the authors find that leaders know the most households (41%), then nominated neighbors (36%), then random neighbors (32%). Using information from the PMT, the authors find that poor neighbors and non-poor neighbors are both equally known among all groups. Despite this, all groups inaccurately assess poverty of the known households. Again based on the PMT, the error rate was roughly 36% (64% of households identified correctly, but nominated neighbors performed slightly better). Ultimately, Beaman et al.'s (2021) findings disagreed with Alatas et al.'s (2016), in that there was no greater quality of knowledge from 'highly centralized' individuals. One main failure that I see in the study is the lack of actual centralized community members. During the census the administrators asked which 5 individuals that person socialized with the most, but these individuals were not used as the centralized entities. Instead, nominated individuals were those who were chosen to *identify households* by other neighbors (which ended up being mostly the leaders and the wealthy). I believe those neighbors that are the most socialized have the best quality of knowledge, and using them in the targeting process would improve the targeting accuracy.

Dimri and Maniquet (2019) introduce the idea of a Bayesian framework to tackle the issues with current poverty targeting schemes. This framework includes preferences towards essential goods, and how lack of a good in a certain area might result in seemingly well-off households leading an impoverished lifestyle based on consumption baskets. In other words, the preferences and income might be the same for two different households, but one household might be in a region where the particular preference has a higher price (thereby making the household with lower prices better off). If preferences are considered, then these prices can be better defined and the equivalent income can be calculated, theoretically reducing targeting errors.

Dimri and Maniquet (2019) use the 2004-2005 round of the Consumer Expenditure Survey to correctly adjust the income threshold, which is calculated after considering the price of a poverty line bundle in that region. The bundles, in this case, contained price data on cereal, vegetables, fuel, and clothing. To create variance within the poverty line bundle, the authors made the poverty threshold dependent on the prices of the bundles. "Then, the heterogeneity in the prices households face will be dealt with by looking at equivalent incomes, that is the incomes that leave households indifferent between their current situations and facing different prices" (Dimri and Maniquet 5). The poverty line bundle with the lower relative price was considered the reference price vector, and the preference groups were created using the CES data. These preference groups were chosen "such that the assumption is likely to be valid" (Dimri Maniquet 22) that the reference price vector is preferred.

They found that, nationally, 56% of the sample preferred the reference price vector, indicating poverty. By this standard, 68% are identified as poor in urban areas, and 41% in rural areas. They find that if preferences are considered, the poverty headcount rises, particularly in urban areas. This idea of 'painting a bigger picture' of poverty is exactly where the concept of Multidimensional Poverty stems from, a redefined measure of poverty that takes education, health, and standard of living into consideration. This concept is employed by Alkire and Seth (2012) to better understand poverty in India.

Alkire and Seth (2012) wanted to expand upon the definition of poverty by introducing Multidimensional Poverty (MDP) to poverty targeting. This is extremely significant in a country like India (country the study is conducted in), where the Indian government has not changed its poverty line once since 2002, and only changed its targeting methods once since then. In light of this, Alkire and Seth (2012) are motivated by the imperfections in the targeting programs conducted by the Indian government over the past three decades. Specifically, the self-reported income model of 1992, the consumption expenditure model of 1997, and the updated shift in 2002 to socio-economic indicators of well-being. Despite this recent shift, struggles persist with reducing targeting errors. A better definition of poverty is clearly needed, in that a large percentage of the population is struggling even above the poverty line.

As mentioned before, MDP further defines poverty by considering the dimensions of education, health, infrastructure, and standard of living. Each of these dimensions has several indicators (e.g. years of schooling for education, child mortality for health, access to electricity for standard of living). They then group people who are deprived of these indicators into one category. Those who are deprived of 33% or more of the indicators are grouped as BPL households.Recently, a Socio Economic Caste Census was implemented in India, which contained relevant variables useful for targeting MDP. Alkire and Seth (2012) use this data, in conjunction with the National Family Health Survey (NFSH) from 2005-2006, to create a robust tool for MDP by assigning weights to each of the variables in preparation for regression analysis. This tool is then compared to various poverty lines, starting with the current and moving up from there (35% to 80%). As the poverty line moves up, more individuals are selected for targeting and potentially receiving transfers. The authors then propose an additional solution through a ten-item binary scoring method. This method identifies households as deprived or not using 10 weighted indicators, and subsequently given a BPL score which represents the number of indicators that the household deprives. If the BPL deprivation score is higher than that of the poverty line cutoff, they are deemed poor. They find that, using this method, the MD-poor increase by about 20% (ranging between 37% and 75% depending on the chosen poverty line), and under-coverage and leakage rates drastically decrease. When more indicators are used and more potential beneficiaries are selected, more households are considered MD-poor.

When the poverty line is capped between 55% and 59%, 41% percent of the households are deemed poor, and 27% are deemed non-poor, with targeting error rates of 32%. However, when the poverty cap is moved up to 80%, the errors drastically decrease. This agrees with the idea of program coverage introduced by Schnitzer and Steoffler (2021) in that the more potential beneficiaries analyzed, the lower the targeting error rate. However, because a household needs to be deprived of three indicators in order to be considered poor, there are still households that are not selected but are still deprived of key indicators. "Given the high rate of deprivations in some of the indicators among households that are automatically excluded by the SECC 2011, are there households that are automatically excluded but are deprived in multiple dimensions? The answer is yes" (Alkire and Sethi 434). Pivoting to the ten-item binary scoring method, the MD-poor increase by about 20% (ranging between 37% and 75% depending on the chosen poverty line), and under-coverage and leakage rates drastically decrease. When more indicators are used and more potential beneficiaries are selected, more households are considered MD-poor. One complaint is with the SECC data, in that there is only information available on the household head, not on any other members. This could create errors in the targeting process that may be hard to identify without the proper data.

Follet and Henderson (2023) are interested in continuing to explore the implementation of a Bayesian framework into modern targeting methods. This is in light of the fact that these modern targeting methods (specifically, PMT and CBT), produce targeting errors that have significant social impact, and an updated model is in desperate need. The authors, in studying this hybrid model, are trying to keep the benefits of PMT and CBT and make up for the downfalls. For PMT, these downfalls include a lack of local information that may not capture certain features that are crucial in targeting the poor (no participatory involvement; the model's definition of poverty might differ from the local definition). CBT makes up for this in the participatory element, but lacks protection from elite capture (as stated in Alatas et al 2016).

This targeting program takes place in Indonesia and utilizes national census data for PMT and survey data containing sociodemographic characteristics and their respective covariates (taken from Alatas et al. 2012, and also used by Hillebrecht et al. (2023)). They identify two stages of PMT that they plan on keeping intact: "an estimation stage that calibrates a statistical model and a prediction stage that calculates the scores that determine program inclusion based on the statistical model" (Follet and Henderson 5). In terms of the Bayesian model, Follet and Henderson (2023) propose using a targeting model that resembles PMT, but utilizing local community members to help calculate the weights of the sociodemographic characteristics that are used in the model (these characteristics can include education, migration status, personal beliefs, and attitudes). To carry out the PMT stages while implementing the Bayesian framework, the authors first regress community preference rankings on household sociodemographic characteristics, then use the estimated weights to calculate scores used to determine program inclusion. Finally, elite capture is accounted for by noting which individuals hold leadership positions and correcting the impact of those decisions based on connections to the elite.

The results show a significant difference between the hybrid model and the baseline (PMT). Most of the gap in performance stems from the weights, which were calculated differently for each. For example, the hybrid model gave the variable regarding roof quality a positive coefficient whereas the PMT model calculated the coefficient as negative. This means the hybrid model is targeting those with a low-quality roof as poor and PMT is targeting them as non-poor (adding to the targeting error). The authors found that the hybrid model outperformed the standard PMT model with targeting error rates ranging from 12-16% lower. These results hold true when Follet and Henderson analyze the model under different poverty thresholds. This suggests that, when taking more encompassing preferences into account, variable weights differ from PMT in a more accurate way. This result agrees with those found by Alkire and Seth (2012) in that including more personal and behavioral based data will improve the targeting accuracy that modern models lack.

Leite (2014) explores various forms of poverty targeting and the advantages of each of them. The targeting methods studied were Means tests, Proxy means test, Community based targeting, Geographic targeting, and Self-targeting. Since the PMTs and CBTs were just described, the rest will be summarized. Means tests is simply analyzing actual consumption or income and comparing that numerical value to the poverty threshold. Geographic targeting is targeting by targeting by location, deeming certain areas as impoverished or not. Self-targeting is when the cost of transactions and programs are set so that only the households in need would enroll. Examples include Mexico and Kenya combining geographical targeting and PMT (Oportunidades & Orphans and Children program), Brazil using geographic targeting and means testing (Bolsa Familia), and Tanzania using geographic targeting along with PMT and CBT. The author also mentions Proxy means test plus, which is a form of PMT that attempts to increase the number of short-term poor households that are identified by accounting for the impact of major shocks like droughts and floods. This helps solve part of Hillebrecht's (2023) problem with PMT, which is the ability to target transition-poor, or short-term poor households. Leite (2014) also stresses the importance of a social registry, or household and individual-level data on the potential recipients of social assistance programs.

One thing to note is that the difference in targeting error rates between PMT and CBT seem to be varying depending on the study. The main differences seem to be the region in which they were conducted, which would explain the variability. The important piece of information is that PMT performs better than CBT across various types of regions or geography. Keep in mind that as many articles have said previously, this depends on the amount of community involvement, and is much more cost effective when conducting over multiple regions.

2.4 Literature that uses models involving machine learning

Mullainathan and Spiess (2017) look at the implications of machine learning on econometrics, and the variety of situations in which it can be applied. They start with a general overview of machine learning and how it can be implemented into econometrics. "The appeal of machine learning is that it manages to uncover generalizable patterns. In fact, the success of machine learning at intelligence tasks is largely due to its ability to discover complex structure that was not specified in advance" (Mullainathan and Spiess 88). In terms of regression models, the purpose of machine learning is to act more as $\hat{y}$ (what is being predicted) than $\boldsymbol{\beta}$ (the coefficients of the predictors). The authors note that in-sample performance can often exaggerate the potential of out-of-sample performance.

The authors then touch on some of the more common machine learning algorithms that are applied to econometrics: decision trees, random forests, or LASSO (Least Absolute Shrinkage and Selection Operator). LASSO uses regularization and applies some sort of penalty or correction every time overfitting is detected. Regression trees function as a tree with multiple nodes, each node is a variable determining the next two nodes, and a prediction is returned when a terminal node is reached. In random forests, multiple regression trees are presented with different amounts of noise for each, and the terminal nodes are then averaged. Random forests were found to perform the best out of the methods when compared to ordinary least squares (despite the overfitting), even when given less samples and covariates. There is still an issue of efficiency carrying over out of the sample, which is fixed using two methods: regularization and empirical tuning. In the context of regression trees, regularization is choosing the best tree that provides the most depth (multiple nodes). Empirical tuning utilizes overfitting to create an out-of-sample experiment within the sample.

Mullainathan and Spiess (2017) move on to the potential problems of machine learning in both unconventional data and poverty targeting. From a prediction standpoint, one of the main issues is the fact that much of the current data available is reflective of the past policies put in place. Behavioral tendencies are also a problem, in that we must consider the determinants behind faith in an algorithm and ask ourselves what factors lead to that faith. Mullainathan and Spiess (2017) ask future researchers to use economic insight and inference to start tackling these issues.

McBride and Nichols (2016) improve upon the standard proxy means test by implementing machine learning methods. In particular, the authors aim to use machine learning to improve out-of-sample performance, something that Mullainathan and Spiess (2017) noted as a key component of any model. The algorithms used include random forests, which are applied to PMT tools that have been developed by the IRIS Center at the University of Maryland. The purpose of this study is to create an improved version of the USAID Poverty Assessment Tool (PAT).

Using the same data, the authors then replicate the same process that IRIS took to develop its tool while using the following methods. McBride and Nichols (2016) utilize quantile regression forests, in which the entire conditional distribution of the response variable is estimated. “A quantile approach is particularly useful for the purposes of PMT tool development due to the fact that the very poor are often concentrated at one end of the conditional income distribution, far from the conditional mean” (McBride and Nichols 10). Similar to Mullainathan and Spiess (2017), the authors note that larger random forest trees are normally more beneficial for targeting methods, despite having a slightly higher variance. This variance is solved through bagging (bootstrapping with replacement), which is a process that takes regression trees with a high variance but low bias and averages across, reducing the variance. Lastly, the out of bag sample (OOB) is used to estimate the mean squared error of the prediction rate, using one third of the training data that has not been used in the regression tree. This is a similar method to empirical tuning, again used by Mullainathan and Spiess, that helps predict the potential performance out-of-sample.

Through the use of both linear regression forests and quantile regression forests, the authors expect the out-of-sample targeting accuracy to improve, under the assumption that the process of data generation does not change between the development and application. To quantify the improvements, the authors compare the out-of-sample bootstrap accuracy of the IRIS method and the out-of-sample accuracy generated by random forest regression and quantile forests. It was found that random forest methods do not improve the accuracy of the IRIS method greatly (improvements ranged from 2 percent to 8 percent, depending on the country), but parametric and nonparametric tests find that the improvements are highly statistically significant. Finally, the random forest generated methods have higher leakage rates than IRIS, but lower undercoverage rates. In other words, aid is reaching more intended and unintended beneficiaries of the program than in the model developed by IRIS (overfitting, an issue brought up by Mullainathan and other authors). Overall, machine learning, through regression forests, helps improve the targeting accuracy of IRIS substantially and specifically reduces leakage rates and undercoverage rates.

Poulin et al. (2022) analyze the improvement of targeting performance by creating a Machine Learning-based proxy means test (ML-based PMT). Pro-poor water subsidies are defined as subsidies that reach the poor and help increase their water accessibility. In parts of Asia and Sub-Saharan Africa, water subsidies are commonly put in place but rarely reach those who are truly in need of them. The authors claim that part of the problem is the government targeting based on water source, meaning that those who have a private connection don't have access to the same subsidies as someone who has a well. But this would then exclude those who have a proper water source but cannot afford to spend the money needed to get sufficient clean water. To help solve this problem, the authors used a ML-based PMT which predicted a household's official poverty status, as it was currently defined by the government.

LASSO is combined with PMT to create this model and helps improve the model by choosing 47 of the most relevant predictors from the GLSS7 dataset. LASSO would then rank the predictors by the size of their coefficients. A second model was then created for additional comparisons by taking the top 10 predictors ranked by LASSO and using them in a different PMT model. To analyze performance. These models are applied to the GLSS7 dataset, and then compared to other methods such as the Demographic and Health Survey (DHS, using PMT), the Poverty Probability Index (PPI, using PMT), CBT, and the LEAP (using CBT and PMT) program in Ghana.

The ML-based PMT performed better than every other PMT, particularly the pre-existing program PPI. It did, however, perform similarly to CBT, with both producing similar targeting accuracies. Poulin et al. (2022) also find very small overlap between the targeted poor of CBT and the targeted poor of PMT, which reveals "either a different poverty concept between communities and the statisticians developing PMT, or the possibility that communities do not always understand each others' condition" (Poulin et al 12). This is consistent with findings from Follet and Henderson (2023), who find the local community's definition of poverty entirely different from those conducting the PMT. The authors also find the ML-based model and CBT model to be better at targeting the chronically poor, which is consistent with the findings of Hillebrecht et al. (2023) (except they refer to chronic poverty as persistent poverty). The results also agree with those from McBride and Niels, in that PMT combined with machine learning is more accurate at targeting the poor than PMT as a standalone.

Aiken et al. (2023) take an interesting approach. Like many previous authors (Hillebrecht et al (2023), Dimri and Maniquet (2019), Leite (2014), and Stoefler (2016)), Aiken et al. (2023) uses consumption basket methods to identify the *ultra poor* in Afghanistan. Along with this, however, the authors implement machine learning into a asset-based PMT method, in the hopes that this will at least partially solve some of the drawbacks of PMT. More specifically, the authors are attempting to resolve the issues with Afghanistan's TUP (Targeting the Ultra Poor) program by implementing mobile cell phone data, which contains information about households that may help targeting accuracy.

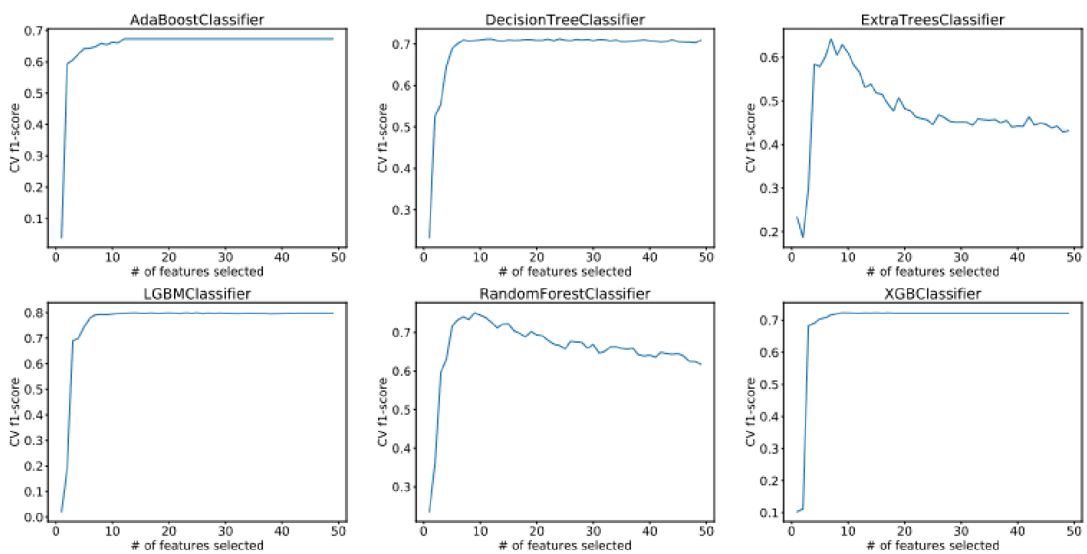
TUP is similar to the program that both Brazil and Tanzania employs in that they utilize geographic targeting, mentioned by Leite (2014). Using the data from this program, the authors carry out various models; a CDR-based method that applies machine learning to the data (specifically, gradient boosting, a sequential operation that constructs an accurate model by adding variables in descending order by strength). ; an asset-based wealth index using asset ownership to target poverty; a consumption basket method that serves as a benchmark. Along with this, Aiken utilizes data from call detail record (CDR) to gain understanding of how information flows within social networks. The machine learning analysis is boosted by a gradient model, which essentially assumes that the next best model, combined with the previous model(s) will produce the least amount of errors. The authors then follow up this method with a combined method that uses logistic regression to target using the predicted ultra-poor probability that the machine learning method gives us.

What Aiken et al. (2023) find is that the CDR method is as accurate as others that rely on consumption assets. The combined machine learning based CDR and wealth index method performs even better than the CDR-only method, with higher Area Under Curve rate (AUC 0.78), and lower error rates (~22%). The method that uses CDR, wealth index, and consumption data performs slightly worse, only because consumption data is difficult to gather when dealing with large populations. The main stipulation that these models expose is the reliance on individuals to own phones. If the model is performed in a country where a large percentage do not own phones, then the effectiveness drastically decreases. However, a massive advantage that this CDR and machine learning based model presents is the speed at which it can be employed, and the cost that it would take to do so (marginal costs decrease). Compared to PMT and CBT, this method costs significantly less and takes far less time to produce results. These results agree with Pouline et al. (2022) and McBride and Niels (2016) in that PMT's accuracy is improved when presented with machine learning algorithms. The speed of implementation could allow the transitory poor to receive benefits while they are still poor, which could partially account for the reduced targeting errors.

Alsharkawi et al. (2022) recognize the need for extreme poverty alleviation post COVID-19. The authors realize the drastic effects that COVID-19 has had on extreme poverty, particularly in Jordan. The expected number of people to reach extreme poverty globally is over one billion by 2030, the UN estimates. Alsharkawi (2022) claims that now is a point in time where poverty measuring strategies must be re-evaluated and updated. “Machine learning could dramatically change the game, making poverty measurement and identification as well as tracking and targeting cheaper and much closer to real-time” (Alsharkawi 86484). In order for machine learning to be effective, however, the model must be carefully constructed. Feature selection is one of the most important components of this model, because if irrelevant variables are selected for use, the predictive model could be extremely inefficient. This is done through an algorithm that searches every combination of features and outputs the most efficient one (in theory, however due to computational costs every possible combination is not explored).

To curb this cost, Alsharkawi (2022) utilizes a variety of filter methods to narrow down the combination of features through the Pearson correlation coefficient, mutual information, and a chi-square test. A particularly effective method is Recursive Feature Elimination (RFE), which introduces features in different iterations, and eliminates the two least important features to speed up the process. This feature elimination process is then implemented into three different machine learning models: LightGBM (gradient boosting machine), bagged decision trees, and random forests (mentioned by Mullainathan & Spiess (2017), McBride & Niels (2016). LightGBM is a learning method that uses gradient boosting, the same algorithm used as Aiken et al. Bagged Decision Trees are essentially a variation of the random forest algorithm, that takes multiple decision trees and multiple training sets, and averages the prediction results (multiple trees are used because using only one can cause overfitting). It’s important to note that the size of the dataset used is 63,211 households, which the author finds is more than sufficient to have an accurate predictive model (size beyond around 5000 households shows no change).

The author finds that both Bagged Decision Trees and LightGBM machine learning models were the most effective in conjunction with chi-square for the former, and RFE for the latter (RFE selected 41 important features out of the original 96). The following graph from Alsharkawi (2022) shows how each machine learning model performs in conjunction with RFE, and the number of features that that model selected.



Alsharkawi et al. (2022) then create an Artificial Neural Network (ANN) to compete with the final two models, and find that it performs worse than the LightGBM model. Therefore, the filtering method of RFE and the machine learning model of LightGBM is chosen as the most accurate model. The author concludes by promoting the LightGBM model as a tool with great power for predicting poverty, and encourages further research to verify generalizability. While this model is not compared to other modern targeting methods, it is impressive that each M.L.

model only required ~10 or less predictors to reach their maximum accuracy (as seen in the graphs above). This implies that M.L. models have the potential to retain the higher targeting accuracy of PMT, while maintaining the lower costs of CBT. Aiken et al. (2023) also found machine learning based models to have the potential to lower costs in addition to speeding up implementation time.

Chapter 3: Conclusions

This paper extensively reviews the literature on PMT, CBT, hybrid, and machine learning based poverty targeting models. The goal is to unpack the recent attempts at reducing poverty in developing countries without the use of income data. Despite the UN's goal of eliminating extreme poverty by 2030, COVID-19 has made problems worse than ever before. The need for an effective and efficient method for identifying the poor is crucial to improvements, but without a distinct non-monetary definition of poverty, it is proving to be a difficult task. In Section 2.1, case studies that utilize a PMT model are analyzed and the results are compared. In general, the papers choose variables based on a dataset to define poverty, assign weights, and use the PMT score to choose eligibility. They are conducted in a variety of developing countries, including Bolivia, Jamaica, Peru, Sri Lanka, India, and Indonesia. Generally, the authors found PMT to have a targeting accuracy of 60%, with most of the errors coming fairly evenly from inclusion and exclusion.

There were two particularly important findings by many of the authors. The first is that the size of the pool you select for targeting (the 'potential beneficiaries') matters, in that the more households you select, the greater your targeting accuracy (Steoffler et al. (2016), Alkire and Seth (2012), Schnitzer and Stoeffler (2021), Grosh and Baker (1995)). This is due to a reduction in the exclusion error, because you are allowing the model to distinguish between poor and non-poor instead of automatically deeming the household as non-poor (by not including them). The second important finding is the significance of the number of indicators (variables chosen from the dataset). Many authors found that the more indicators are used in the model, the greater the targeting accuracy (Grosh and Baker, Sebastien et al, Klasen and Lange). This could, in part, be due to the fact that the correct combination of indicators is rarely chosen, thus including more could give the model more to work with.

In Section 2.2, case studies of CBT models are analyzed, with community involvement taking place in the following ways. Some number of local community members are first selected (either randomly or not) to help set a predetermined definition of poverty for the village. They are then asked to rank households based on visual and descriptive aids from poorest to wealthiest. A threshold is then set based on either the rankings or the number of households the administrators intended to provide benefits. Some of the authors (Alkire and Seth (2012), Savadogo et al. (2015) introduce the idea of MDP to the CBT process. They try to broaden the definition of poverty by taking marital, religious, and other behavioral preferences into account. The results show the targeting accuracy for CBT to be roughly 10% lower than PMT, in that they receive slightly more inclusion and exclusion errors. This is offset by the fact that CBT costs considerably less than most PMT models, which is one of the reasons the authors give to keep CBT practices in mind during future research.

Alatas et al. (2016) had some key findings that allowed their CBT model to perform as well as most other PMT models. A learning model was created that involved a constantly evolving parameter, which represented the wealth of other households. Administrators would receive this information over periods of time from various households, and aggregate them to learn more about the social network. They then find that more socially centralized households were more accurate at ranking other households. Using these two methods, they conducted CBT and found the targeting accuracy to be around 60%. This presents a key finding: the more you understand about how a particular social network operates, the more effective community-based programs will be. However, the time it takes to study social networks enough to be able to pick out the socially central must be considered.

In Section 2.3, hybrid models and case studies that directly compare PMT and CBT are analyzed. In the literature that directly compares the two in the same region, the results hold that PMT performs about 10% better than CBT. However, Hillebrecht et al. (2023) make an important discovery when they analyze the results over the course of 18 months. CBT does perform about 10% worse than PMT if the results are analyzed soon after the targeting method takes place. But 12 months later, CBT is found to only perform about 2% worse on aggregate, and performs better in urban areas. This is a method that none of the other authors take, and leads to some important assumptions. First, it means CBT is better at differentiating between transitory poverty and chronic poverty. Second, it means because of PMT's inability to differentiate the two, the targeting errors presented at the baseline do not reflect the realistic performance of the model.

In the papers that create a new model containing elements from both CBT and PMT, the results were mixed. Some found hybrid models to perform better than PMT (Follet and Henderson (2023)), while others found the hybrid to perform the same (Stoeffler et al. (2021)) or worse (Alatas et al. 2012). This shows that just because some benefits are present in CBT and not in PMT, combining the two will not necessarily make a stronger model overall. Follet and Henderson (2023) did do one thing that could have improved the model, which is the utilization of community members to select the weights (or preferences) of certain sociodemographic characteristics. This Bayesian method of taking preferences into account could be attributed to the increase in performance over other hybrid models.

In Section 2.4, the most recent models are analyzed: machine learning-based PMT models. These models implemented machine learning in the feature selection phase, helping reduce the speed and cost of poverty targeting. The results show ML-based PMT models to perform about 10% better than PMT models, making them the most accurate at targeting poverty thus far. The ML algorithms used are regression forests, gradient boosting, and LASSO, with regression forests and gradient boosting performing the best. The results lead to the key assumption that the feature selection phase is a major contributor when it comes to the targeting errors in PMT. The features being selected, and relative weights, encapsulate poverty with a greater accuracy when presented with machine learning.

When the literature is analyzed to this extent, the results of various poverty targeting methods can be aggregated and averaged to promote suggestions for future research. First, the literature suggests that despite the potential utility of community knowledge and social networks, PMT still performs better on average (at baseline - keep in mind the findings of Hillebrecht et al. (2023)). Hybrid models have mixed results, meaning that a reusable structure has yet to be developed, or does not exist. If administrators take the time to understand the social network, as seen in Alatas et al. (2016), hybrid models have the potential for greater accuracy. The only models that have consistently performed better than PMT are ML-based PMT models, which produce targeting accuracies of 70%-80%. The important aspect of these ML-based models is that they employ various ML methods to the feature selection phase, which could be the most important part of any poverty targeting program. The authors of these studies promote ML-based

PMT models over the respective national programs. The results point to the fact that machine learning can drastically improve targeting accuracy and should be considered for any country that is suffering from significant poverty and lacks the data to make accurate assumptions.

References

Aiken, E. L., Bedoya, G., Blumenstock, J. E., & Coville, A. (2023). Program targeting with machine learning and mobile phone data: Evidence from an anti-poverty intervention in Afghanistan. *Journal of Development Economics*, *161*, 103016. https://doi.org/10.1016/j.jdeveco.2022.103016

Alatas, V., Banerjee, A., Chandrasekhar, A. G., Hanna, R., & Olken, B. A. (2016). Network structure and the aggregation of information: Theory and evidence from Indonesia. *American Economic Review*, *106*(7), 1663–1704. https://doi.org/10.1257/aer.20140705

Alatas, V., Banerjee, A., Hanna, R., Olken, B., & Tobias, J. (2010). *Targeting the Poor: Evidence from a Field Experiment in Indonesia*. https://doi.org/10.3386/w15980

Alkire, S., & Seth, S. (2012). Selecting a targeting method to identify BPL households in India. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2292942

Alsharkawi, A., Al-Fetyani, M., Dawas, M., Saadeh, H., & Alyaman, M. (2022). Improved Poverty Tracking and targeting in Jordan using feature selection and machine learning. *IEEE Access*, *10*, 86483–86497. https://doi.org/10.1109/access.2022.3198951

B. Essama-Nssah, 2018. "Assessing the performance of targeting mechanisms," Working Papers 457, ECINEQ, Society for the Study of Economic Inequality.

Beaman, L., Keleher, N., Magruder, J., & Trachtman, C. (2021). Urban Networks and targeting: Evidence from Liberia. *AEA Papers and Proceedings*, *111*, 572–576. https://doi.org/10.1257/pandp.20211061

Beaugé, Y., Koulidiati, J.-L., Ridde, V., Robyn, P. J., & De Allegri, M. (2018). How much does community-based targeting of the ultra-poor in the health sector cost? novel evidence from Burkina Faso. *Health Economics Review*, *8*(1). https://doi.org/10.1186/s13561-018-0205-7

Coady, D. P., & Parker, S. W. (2009). Targeting performance under self‐selection and administrative targeting methods. *Economic Development and Cultural Change*, *57*(3), 559–587. https://doi.org/10.1086/596615

Conning, J., & Kevane, M. (2002). Community-based targeting mechanisms for social safety nets: A critical review. *World Development*, *30*(3), 375–394. https://doi.org/10.1016/s0305-750x(01)00119-x

Dimri, A., & Maniquet, F. (2019). Income poverty measurement in India: Defining group-specific poverty lines or taking preferences into account? *The Journal of Economic Inequality*, *18*(2), 137–156. https://doi.org/10.1007/s10888-019-09434-6

Follett, L., & Henderson, H. (2023). A hybrid approach to targeting social assistance. *Journal of Development Economics*, *160*, 103002. https://doi.org/10.1016/j.jdeveco.2022.103002

Grosh, M. E., & Baker, J. L. (1995). *Proxy Means Tests for Targeting Social Programs*. https://doi.org/10.1596/0-8213-3313-5

Hillebrecht, M., Klonner, S., & Pacere, N. A. (2023). The dynamics of poverty targeting. *Journal of Development Economics*, *161*, 103033. https://doi.org/10.1016/j.jdeveco.2022.103033

Hanna, R., & Olken, B. (2018). *Universal Basic Incomes vs. Targeted Transfers: Anti-Poverty Programs in Developing Countries*. https://doi.org/10.3386/w24939

Handa, S., Huang, C., Hypher, N., Teixeira, C., Soares, F. V., & Davis, B. (2012). Targeting effectiveness of social cash transfer programmes in three African countries. Journal of development effectiveness, 4(1), 78-108.

McBride, L., & Nichols, A. (2016). *Improved Poverty Targeting through Machine Learning: An Application to the USAID Poverty Assessment Tools*. https://doi.org/10.1596/1813-9450-7849

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87–106. https://doi.org/10.1257/jep.31.2.87

Pereira Guimaraes Leite,Phillippe George. *Effective Targeting for the Poor and Vulnerable (English).* Social protection and labor technical note ; no. 6 Washington, D.C. : World Bank Group. http://documents.worldbank.org/curated/en/747591468125685125/Effective-Targeting-for-the-Poor-and-Vulnerable

Poulin, C., Trimmer, J., Press-Williams, J., Yachori, B., Khush, R., Peletz, R., & Delaire, C. (2022). Performance of a novel machine learning-based proxy means test in comparison to other methods for targeting pro-poor water subsidies in Ghana. *Development Engineering*, *7*, 100098. https://doi.org/10.1016/j.deveng.2022.100098

Savadogo, G., Souarès, A., Sié, A., Parmar, D., Bibeau, G., & Sauerborn, R. (2015). Using a community-based definition of poverty for targeting poor households for premium subsidies in the context of a community health insurance in Burkina Faso. *BMC Public Health*, *15*(1). https://doi.org/10.1186/s12889-014-1335-4

Schnitzer, P., & Stoeffler, Q. (2021). Targeting for social safety nets: Evidence from nine programs in the sahel. *Policy Research Working Papers*. https://doi.org/10.1596/1813-9450-9816

Sebastian, A., Shivakumaran, S., Silwal, A. R., Newhouse, D., Walker, T., & Yoshida, N. (2018). *A Proxy Means Test for Sri Lanka*. https://doi.org/10.1596/1813-9450-8605

Sabates-Wheeler, R., Hurrell, A., & Devereux, S. (2014). Targeting social transfer programmes: Comparing design and implementation errors across alternative mechanisms: WIDER Working Paper.

Stoeffler, Q., Mills, B., & del Ninno, C. (2016). *Reaching the Poor: Cash Transfer Program Targeting in Cameroon*. https://doi.org/10.1596/24255

Stephan Klasen & Simon Lange, 2015. "Targeting Performance and Poverty Effects of Proxy Means-Tested Transfers: Trade-offs and Challenges," Ibero America Institute for Econ. Research (IAI) Discussion Papers 231, Ibero-America Institute for Economic Research.

What is multidimensional poverty?. Development Initiatives. (n.d.). https://devinit.org/resources/what-multidimensional-poverty/