Journal of Development Economics 148 (2021) 102564

Contents lists available at ScienceDirect

Journal of Development Economics

journal homepage: www.elsevier.com/locate/devec

Regular Article

Targeting humanitarian aid using administrative data: Model design and validation

Onur Altındağ[a], Stephen D. O'Connell[b,f,*], Aytuğ Şaşmaz[c], Zeynep Balcıoğlu[d], Paola Cadoni[e], Matilda Jerneck[e], Aimee Kunze Foong[e]

[a] Bentley University Department of Economics and Economic Research Forum, USA
[b] Emory University Department of Economics, USA
[c] Harvard University Department of Government, USA
[d] Northeastern University Department of Political Science, USA
[e] UNHCR Lebanon, Lebanon
[f] IZA Institute of Labor Economics, Germany

ARTICLE INFO

JEL classification:
I39
I32
O12

Keywords:
Poverty targeting
Proxy means test
Cash transfers
Refugees
Forced displacement
Lebanon

ABSTRACT

We develop and assess the performance of an econometric prediction model that relies on administrative data held by international agencies to target over $380 million annually in unconditional cash transfers to Syrian refugees in Lebanon. Standard metrics of prediction accuracy suggest targeting using administrative data is comparable to a short-form Proxy Means Test, which requires a survey of the entire target population. We show that small differences in accuracy across approaches are largely attributable to a few data fields. These results are robust to a blind validation test performed on a random sample collected after the model derivation, as well as the type of estimator used for prediction. We discuss relative costs, which are likely to feature prominently when alternative approaches are considered in practice.

1. Introduction

"A refugee used to be a person driven to seek refuge because of some act committed or some political opinion held … With us the meaning of the term 'refugee' has changed. Now 'refugees' are those of us who have been so unfortunate as to arrive in a new country without means and have to be helped by Refugee Committees." - Hannah Arendt, *We Refugees*, 1943.

Governments and aid organizations face persistent challenges in targeting social welfare programs to accurately identify and reach intended beneficiaries. In the case of unconditional cash transfers, which are popular in many low- and middle-income countries, accurate targeting is often complicated by limited institutional capacity and reliable data. Available aid is thus allocated by any number of proxy mechanisms, including simple approaches such as geographic or demographic targeting, as well as more sophisticated allocation mechanisms such as self- or community targeting, or proxy means tests (PMTs). The performance of these methods exhibits substantial variation across implementations and contexts, with one review showing that, in practice, "*a quarter of programs' … [targeting] performed worse than a random allocation of resources*" (Coady et al., 2004).

Among such alternatives, PMTs are the most common to target the poor. They rely on existing survey data to choose a small set of predictors to collect in a short survey that is administered, in principle, to the

* Corresponding author. Emory University Department of Economics, USA.
E-mail addresses: oaltindag@bentley.edu (O. Altındağ), soconnell@emory.edu (S.D. O'Connell), sasmaz@g.harvard.edu (A. Şaşmaz), balcioglu.z@northeastern.edu (Z. Balcıoğlu), cadoni@unhcr.org (P. Cadoni), jerneck@unhcr.org (M. Jerneck), kunzefoo@unhcr.org (A.K. Foong).

https://doi.org/10.1016/j.jdeveco.2020.102564
Received 12 February 2020; Received in revised form 26 July 2020; Accepted 11 August 2020
Available online 15 October 2020

entire potentially eligible population (Basurto et al., 2017; Kshirsagar et al., 2017; Schreiner, 2010).[1] The popularity of PMTs is likely to increase in the future thanks to the recent developments in econometric targeting approaches that prioritize out-of-sample prediction performance (McBride and Nichols, 2018; Kshirsagar et al., 2017).[2] This econometric approach typically uses consumption or expenditure data from a representative household survey as a proxy for poverty, and derives a model that assigns weights to factors used to predict poverty in the broader population of the potentially eligible. The predictors in a standard PMT model comprise a set of household assets and demographics that are easily verifiable, which eschews measurement error and misreporting and diminishes the cost of assessing households' assistance eligibility. The methodology to choose measures that predict consumption thus becomes the key step in targeting the eligible population (Brown et al., 2018).

In this study, we present the design and validation of an econometric targeting model that uses routinely collected administrative data to target over $380 million per year of unconditional cash and in-kind assistance to Syrian refugees in Lebanon. Our study is motivated by three contemporaneous phenomena influencing the practice of modern poverty targeting. The first is an increasing degree of administrative data availability, integration, and interoperability, which can allow governments, international organizations, and other entities the ability to securely access, use, and analyze data to provide better programs, policies, and services (Fantuzzo and Culhane, 2015). The second concerns perennial limitations in financial and administrative capacity to implement a successful PMT, which requires hiring and training a substantial workforce and coordinating logistics for carrying out household visits of the entire potentially eligible population to gather a census of verifiable assets in order to target the program to the poorest. The accuracy of any PMT thus heavily depends on the capacity of the implementing agency, and this can exhibit substantial variation from one program to another (Coady et al., 2004). This is particularly salient in the context of humanitarian assistance, which is relatively new to unconditional cash-based interventions. Furthermore, chronic underfunding – as is the case with the vast majority of humanitarian programs – thus becomes a major factor determining the effectiveness of a PMT. Finally, if trends in recent decades continue, the future is likely to see increases in the frequency, intensity, and duration of forced international migration – whether driven by conflict or climate change (CARE, 2019); providing alternative and cost-effective tools to improve program targeting is thus crucial for the effective deployment of aid resources to vulnerable populations.

We combine a nationally representative expenditure survey with routinely collected administrative data and cross-validated regularized linear estimators to generate a prediction model for household per capita expenditure. We then compare the prediction accuracy of a set of models relying on administrative data to a simulation of the short-form survey PMT approach,[3] which would rely on data on household characteristics and verifiable assets collected by survey. While there is no specific expectation that existing administrative data would be well-suited for the targeting of humanitarian aid, we show that the use of basic demographic information from typical administrative records held by aid organizations and governments is approximately as accurate in targeting the poor compared to a short-form PMT that requires a household survey of the entire population. While the survey-based approach yields decreases in inclusion and exclusion error of about two percentage points, these differences are not statistically significant. Furthermore, we are able to isolate a small number of fields in the survey data that provide additional predictive power. All models we present also perform around the median of the 85 targeted interventions in various developing countries reviewed by Coady et al. (2004), suggesting that (a) targeting accuracy among refugee populations is not meaningfully different from those of other populations around the world, and (b) differences in accuracy rates across methods are relatively minor.

Finally, we exploit a unique opportunity to conduct a contemporaneous out-of-sample validation using data from households that were not included in the model derivation sample and were surveyed after the model development process. The out-of-sample validation survey was carried out under the same survey protocol by the same organizations and enumerators involved in collecting the survey data that provided household expenditures for the training data — an important feature for independent data sets to yield meaningful comparisons (Heckman and Smith, 1995). The fact that the validation survey was available only after the model development stage ensures zero degree of discretion regarding the components in the prediction model, and lends additional insight into intertemporal reliability as the validation sample was collected closer to the date of program implementation than the training (survey) data used to develop the model. This out-of-sample test suggests targeting accuracy is comparable to cross-validated error rates from the data used to develop the model.

Our primary contribution is the development and validation of an administrative-data-based econometric targeting model for a large-scale, ongoing, unconditional cash transfer program. We show that such an approach can be used to generate targeting models whose performance compares equivalently to a traditional survey-based approach that is often too costly, too cumbersome, or limited by logistical constraints for antipoverty programs of even moderate scale. Given that there is not a strong conceptual reason to expect existing administrative data to be apt for this purpose, this finding adds a new approach to the toolbox of aid targeting. The main logistical advantages of the proposed tool compared to a typical PMT approach are in avoiding (i) non-response or reachability issues related to a population level short-form survey, and (ii) the well-documented problems with the misreporting of assets or household structure during short-form surveys (Banerjee et al., 2018; Camacho and Conover, 2011).

The structure of the study is as follows. In Section 2, we briefly review the existing literature on the targeting of anti-poverty programs, then describe the background and context of humanitarian assistance in Lebanon. In Section 3, we discuss the data we used to develop the targeting model, the methodology applied, and the resulting model and its prediction properties. We then discuss the sampling and survey methodology for the out-of-sample validation exercise, and present the results from the analysis of those data within the same section. Section 4 concludes with a brief discussion on the future of scalable econometric targeting methods in similar contexts.

2. Background and literature review

2.1. Proxy targeting of anti-poverty programs

The PMT approach is a popular tool for targeting anti-poverty programs (Coady et al., 2004; Brown et al., 2018). Typically, a representative household expenditure survey provides data to determine the relative importance of predictors of household consumption. The model building process then results in assigning weights to demographic variables that are observed for the population to generate a metric for program eligibility. The two main advantages of PMT are: (i) the ease of implementation due to the short surveys that collect information on verifiable assets, and (ii) the ability to account for informal economic

[1] Recent work has shown some benefit to self- and community targeting over proxy means tests: self-targeting mechanisms can increase targeting efficiency (Alatas et al., 2016), while community targeting does not perform better than PMTs, although it may increase beneficiary satisfaction with aid programs (Alatas et al., 2012; Schüring, 2014).

[2] See Devereux et al. (2015) for a cross-country review of recent PMT-based programs; see Sharp (2015) for a detailed review of cash and food assistance for refugees in Lebanon, Jordan, and Egypt.

[3] Short-form PMT surveys are also often referred to as "scorecards" (Kshirsagar et al., 2017; Schreiner, 2010).

activity (Basurto et al., 2017; Kshirsagar et al., 2017; Schreiner, 2010).

There is, however, a well-documented substantial variation in exclusion and inclusion error rates across implementations of PMTs.[4] The existing evidence suggests that better targeting is associated with stronger administrative capacity, larger variation in poverty, reliable up-to-date survey and administrative data, and the availability of proxies that are strongly correlated with poverty (Coady et al., 2004; Devereux et al., 2015; Kidd et al., 2017).[5] Even in ideal circumstances, however, a PMT is usually only partially successful in accurately targeting the poor, and the more homogeneous in observables is the target population, the larger the proportion that will be incorrectly excluded (Brown et al., 2018).

While the main goal of targeting is to accurately predict welfare in a population for which the data on the outcome of interest is not available, assessments of a program's targeting accuracy often rely on in-sample prediction performance. Only more recently have there been meaningful strides in analyzing the out-of-sample prediction performance of various econometric targeting tools. McBride and Nichols (2018) show that overfitting the prediction sample yields poor out-of-sample performance, and prediction tools that are designed to minimize out-of-sample error can likely increase targeting accuracy.

Our study contributes to the literature assessing the performance of various approaches for econometric targeting of social or aid programs. This includes, but is not limited to, Andini et al. (2018) for Italy's national tax rebate program, Sohnesen and Stender (2017) for predicting poverty in several African countries, Baird et al. (2013) for poverty in Tanzania, and Kilic et al. (2014) for farm input subsidies in Malawi. Perhaps the most pertinent studies are McBride and Nichols (2018) and Brown et al. (2018), who evaluate the impact of different methodological tools on targeting effectiveness. Using the United States Agency for International Development (USAID) poverty assessment tool from several countries, McBride and Nichols (2018) show that approaches that prioritize out-of-sample accuracy perform substantially better in accurately identifying the poor population compared to a standard PMT approach relying only on in-sample fitting. Brown et al. (2018) show that simple demographic surveys do as well, or nearly as well, as econometric PMT methods across nine African countries. Our study adds to this literature by showing that routinely collected administrative data can offer a potentially equally reliable, and less costly, alternative to existing PMT strategies.

2.2. Basic needs assistance to refugees in Lebanon

Worldwide, more than 61 percent of 25.9 million refugees live in non-camp settings in developing countries under the mandate of the United Nations High Commissioner for Refugees (UNHCR, 2017).[6] The primary destinations for displaced populations are neighboring countries, which often have constrained economic and operational resources to host these populations.[7] As a result, international organizations, in partnership with governmental and non-governmental organizations (NGOs), have been a primary source of cash and in-kind assistance to displaced individuals in conflict regions.[8]

Since 2011, the Syrian Civil War forcibly displaced more than 5.6 million people internationally. This number includes non-Syrians too. Per UNHCR website there are 879,598 Syrian refugees in Lebanon at the moment. Lebanon hosts over 1.5 million refugees, resulting in the highest per capita population share of refugees in the world. Following the beginning of refugee outflows from Syria in 2012, several separate cash transfer and voucher programs have been implemented in Lebanon by organizations including UNHCR, the United Nations Children's Fund (UNICEF), and the World Food Programme (WFP), among others. UNHCR and WFP provided four cash assistance programs for Syrian refugees in Lebanon. As of 2018, the Multi-Purpose Cash Assistance Program (MCAP), operated by UNHCR, assists around 33,000 severely vulnerable refugee families every year. Supported families receive $175 every month for a year. Assistance is provided through an ATM card, allowing families withdraw cash from ATMs across the country. WFP also operates three other cash assistance programs targeting Syrian refugees in Lebanon. The Multi-Purpose Cash Program (MPC) started in October 2017 and assists approximately 23,000 severely vulnerable Syrian refugee families. In this program, the beneficiaries have the choice either to redeem their assistance at WFP-contracted shops or to withdraw cash from ATMs across the country. The Cash for Food program started in 2017 and provides food assistance to 170,000 Syrian refugees, either as complementary food assistance to UNHCR's MCAP or as food assistance only, and scaled up to reach 220,000 Syrian refugees by late 2018. Finally, the Food e-Card started in 2013 and currently targets 345,000 Syrian refugees; similar to the Cash for Food program, this assistance modality provides either food assistance alone or as a complement to food assistance through UNHCR's MCAP. UNHCR additionally provides winter assistance to 162,000 families (in 2018) through a lump-sum payment of $375 per household.

Targeting welfare programs is challenging in the context of forced displacement: refugees from a conflict zone typically constitute the very poorest and most vulnerable segment of the host country population, having lost or left assets in their home country. This induces the population to become more observably homogeneous and poorer, reducing the potential predictive capacity of typical econometric approaches that use verifiable household assets as a proxy for economic well-being. Moreover, data quickly become outdated due to displaced populations' ongoing movements both within the host country and between the host and home countries. While targeting limited assistance resources to such populations is crucial to achieve the typical goals of humanitarian organizations, little is known about the performance of PMTs, or their alternatives, in such contexts.

Importantly, eligibility for these transfer programs is based on a common, unified scoring system. Since comprehensive data on consumption and expenditure do not exist, program targeting has had to rely on the use of information available in administrative records held by the humanitarian agencies and in nationally representative surveys. Since2016, UNHCR and WFP have used a regression-based approach to determine the predictors of per capita consumption from a nationally representative[9] household expenditure survey called the Vulnerability Assessment of Syrian Refugees (VASyR). The model coefficients are then

[4] Type I and exclusion errors are interchangeable terms, both indicating a poor family that is wrongly excluded from the program. Type II/inclusion error accordingly reflects a non-poor family that is wrongly included within the program eligible population due to prediction error.

[5] See Coady et al. (2004); Devereux et al. (2015); Kidd et al. (2017) for reviews of targeting effectiveness in various welfare transfer programs around the world.

[6] As of 2016, 19.9 million refugees globally were living under the mandate of the UNHCR (UNHCR, 2017).

[7] According to World Bank (2018), in 2015, 80 percent of the internationally forcibly displaced population took shelter in a neighboring country, and those who moved to non-neighboring countries tended to be more skilled.

[8] Contextually related to our work, Verme et al. (2016) are the first to provide a detailed welfare assessment of refugees in Jordan and Lebanon. Using household survey data collected between 2013 and 2014, the authors provide a comprehensive description of poverty among the first waves of refugee populations in both countries. Combining administrative data and a large survey from Jordan, the study also investigates the observable characteristics of the registered refugee population that predict welfare (as measured by expenditure per capita). In a follow-up study, Verme and Gigliarano (2019) offer a methodology to optimize the under-coverage and leakage under a budget constraint using an index-based simulation exercise.

[9] "Nationally representative," as used throughout this work, refers to representativity of the Syrian refugee population in Lebanon.

used to predict expenditure per capita in the population using refugee households' background and demographic information collected comprehensively by aid agencies. The model and household scores have historically been updated annually, a process that typically occurred over the months of July and August; the newly generated scores were then used to determine assistance receipt from November to the following October. For the purposes of this paper, we are concerned with the classification of a household as "severely vulnerable," defined as a household with per capita expenditure below $87 per month,[10] which reflects the subsistence level of consumption for a typical family as determined by the Lebanese government.[11]

Two points are worth addressing with respect to the reliability of expenditure per capita as the outcome measure to target the vulnerable population. The first is whether expenditure per capita is a relevant measure of poverty for refugees when other assessment measures such as multidimensional indices, principal components, or ad-hoc categorization by aid agencies are available. For this paper, we took as given the institutional decision to target cash assistance based on expenditure per capita. This measure, however, parallels the nature of a program that provides additional cash assistance to families to increase expenditure levels that fall short of fulfilling basic needs. The same agencies also run alternatively targeted programs that provide assistance to address needs for education, health services, and heating/shelter needs.

The second is whether the absolute (level) or the relative (ranking) measure of poverty is better in accuracy assessment of the prediction model given that the former could potentially underestimate poverty. The question is empirically testable and we show that using the nominal values of expenditure per capita yields very similar predicted and actual estimates of poverty rates for the refugee population. Thus, reported inclusion and exclusion errors rely on expenditure per capita expressed in absolute terms.

3. Targeting model

3.1. Data

We develop the model and validation analysis using three data sources. The first is nationally representative survey data from the Vulnerability Assessment of Syrian Refugees in Lebanon (VASyR) 2018, which collected detailed information on households and expenditure patterns. The sample includes information on 4364 households across 26 districts in Lebanon. We construct expenditure per capita (in USD) for each household as of the survey date, which spanned three weeks of April and May 2018.[12] Unique household and individual identifiers allow us to link the survey records to the administrative databases described below.

The second data source is the UNHCR database, which is an administrative data set that contains information on the demographic background of all Syrian refugees in Lebanon known to UNHCR. As is typical in many contexts, Syrian refugees in Lebanon must make humanitarian agencies aware of their presence in order to receive humanitarian aid. Undertaking this process also provides refugees a proof of identity that can protect against forced return or arbitrary arrest, and eases family unification and resettlement efforts. Refugees thus have strong incentives to make their presence and situation known to humanitarian agencies and to be included in the administrative data.

Information in the UNHCR database is updated on a regular basis through mobile phone and in-person communication with refugee families. Individual-specific information includes the individual's arrival date, the governorate and district of origin (in Syria), a self-reported education level, age (in years), relationship to the household head, gender, and a series of other indicators reflecting specific vulnerabilities or protection concerns.[13] For the targeting model, we construct household-level analogues of these variables (typically in terms of the share of household members with a given characteristic), along with additional measures of household structure. Our modeling and analysis uses a snapshot of the database as of June 2018. Importantly, the UNHCR database serves as the sampling frame for both the VASyR and the validation surveys, and is also the data to which the model is ultimately applied in practice.

The third data source is the Refugee Assistance Information System (RAIS), which includes up-to-date information on all refugee families who receive assistance in Lebanon from any of the major international organizations or their partners. Our data were current as of June 2018 and include details on the type(s) of assistance (cash and/or food) currently being received. Unique family identifiers in the RAIS allow this data set to be merged with both the administrative and survey datasets described above.

Table 1, Panel A shows summary statistics of the individual demographic background information from the UNHCR database, which includes Syrian refugees in Lebanon known to UNHCR. The refugee population is young, balanced by gender, and relatively uneducated. The initial interview includes a question about refugees' most recent profession prior to displacement. Responses to this question are recorded without strict categorization; we aggregated them into six categories: none, unknown, housekeeper, labor, services, and student. Occupational patterns are in line with the education distribution, and indicate a relatively low-skill labor force. Table 1 Panel B shows the constructed measures by household. High fertility is seen alongside a high share of dependents; working age males constitute only 23% of the individuals in the average household.[14]

Fig. 1 contains a conceptual mapping of our model-building and validation process and the various data sources used therein. We first merge the UNHCR database, RAIS, and VASyR data sets to create our training sample, which includes only families for whom we have information on household expenditure. Summary statistics on expenditure are shown in Table 2, which indicates a right-skewed distribution of consumption with a mean and median of $114 and $87 per capita, respectively.[15] We then use the estimated coefficients derived from the training sample to predict expenditure per capita in the population. In the final stage of the analysis, consumption and expenditure data are collected from a random sample of 521 households that were not interviewed in VASyR 2018 to assess the out-of-sample discrepancy between actual and predicted expenditure.

[10] See Verme et al. (2016) for a comprehensive discussion of concepts related to economic welfare of refugees and a detailed welfare and vulnerability assessment of refugees in Jordan and Lebanon. Verme et al. (2016) make a distinction between welfare and vulnerability, suggesting the latter refers to the ability of households to respond to future shocks and the risk of remaining in or falling into poverty in near future. In line with the operationalization of the concept by international organizations in the context in which we conducted this research, we use the terms welfare, vulnerability, and deprivation interchangeably, with all three terms conveying a concept of socio-economic welfare.

[11] In the Lebanese government's official poverty line calculation, the typical family is assumed to be composed of two adults, one child over five years of age, and two children under five years of age. The calculation is then based on a survival-level minimum food expenditure basket; amount of rent for an informal tent settlement; and minimum water, clothes, communication and transportation costs. A full description can be found at UNHCR UNICEF WFP (2017).

[12] VASyR survey instruments, as well as the summary report, are available at https://data2.unhcr.org/en/documents/details/66669 and https://data2.unhcr.org/en/documents/details/67380, respectively.

[13] Due to data sensitivity, we are unable to report some of the questions that are asked to refugees during the initial interview. These include questions about specific medical conditions, children's daily activities, and relationships among family members, among others.

[14] We define the dependency ratio as the total number of household members over 60 and under 15 divided by the total number of household members.

[15] The Lebanese pound was pegged to the US dollar during the study period with an exchange rate of approximately 1 USD = 1500 LBP. All currency values referred to throughout the paper are in USD.

Table 1
Summary statistics, UNCHR database.

Panel A: Individual records	Mean	Std. Dev.
Age	20.41	16.42
Female	0.52	0.50
Disabled	0.03	0.18
No education	0.24	0.43
Less than primary school	0.16	0.37
Primary school	0.23	0.42
Secondary school	0.16	0.37
High school and above	0.08	0.26
Education Unknown	0.12	0.33
Housekeeper	0.15	0.36
Service	0.04	0.20
Student	0.01	0.12
Laborer/Other	0.11	0.32
None	0.07	0.25
Profession Unknown	0.07	0.26
Panel B: Constructed household records	Mean	Std. Dev.
Size	4.20	2.25
Head's age	36.99	12.46
Head female	0.31	0.46
% members aged 0-5	0.19	0.21
% members aged 6-10	0.12	0.17
% members aged 11-17	0.12	0.19
% male members aged 18-50	0.23	0.27
% female members aged 18-50	0.24	0.21
% members aged 60+	0.04	0.16
Dependency ratio	0.48	0.28
% members with no education	0.14	0.30
% members with less than primary school education (%)	0.04	0.16
% members with primary school education (%)	0.33	0.39
% members with secondary school education (%)	0.29	0.37
% members with high school education and above (%)	0.17	0.32
% members who worked in service (%)	0.10	0.24
% members who worked as a housekeeper (%)	0.34	0.33
% members who were students (%)	0.03	0.13
% members who worked as a laborer/other profession (%)	0.25	0.30
% members who were not working (%)	0.14	0.29

Note: This table shows the mean and the standard deviation of the demographic characteristics of the Syrian refugee population in Lebanon based on the UNHCR database. Panel A reports the individual level data whereas Panel B shows the household level characteristics. Summary statistics represent the cross-section of active cases as of June 2018. Due to data sensitivity, we are unable to report sample sizes and the details of some of the questions that are asked to refugees during intake. These include questions about specific medical conditions, children's daily activities, and relationships among family members, among others. For approximate sample size of the individual- and household-level records, we refer the reader to publicly available sources such as UNHCR (2018), which report more than 976,000 Syrian refugees registered in Lebanon as of 31 July 2018.

For the prediction model, the observational unit is a "case," which is typically a nuclear family or a household that registered together with the UNHCR. The survey information is based on household visits, and only in rare cases does a household include multiple cases who live together. We assigned the same outcome for multiple families who live in the same household given that expenditure can only be observed by household.

3.2. Prediction model

3.2.1. Regression framework

We use the following linear specification:

$$log(y_i) = \pi_0 + \sum_{j=1}^{k} x_{ij}\pi_j + \varepsilon_i \tag{1}$$

where y_i is the log per capita expenditure for case *i*, which is predicted by k independent variables, ε is the unknown error term and π_0 denotes a common intercept. As recently shown by McBride and Nichols (2018), approaches using in-sample validation — such as the standard implementation of Ordinary Least Squares — are likely to overfit in a prediction exercise. Instead, tools and methods designed for out-of-sample prediction, such as cross-validated penalized linear regression, should be used for the out-of-sample prediction exercise that PMTs comprise.

To estimate the coefficients $\pi_0, \pi_1, \ldots, \pi_k$, our primary approach relies on a least absolute shrinkage and selection operator (lasso) regression, which, when combined with cross-validation to choose hyperparameters, has been shown to consistently perform well across various out-of-sample prediction settings (Abadie and Kasy, 2019). Cross-validated lasso is now commonly used to predict outcomes for which acquiring direct information on the outcome is costly or impossible.[16]

[16] Some examples of machine-learning tools that have recently been applied include the prediction of economic activity, productivity, or growth with nighttime lights (Jean et al., 2016; Donaldson and Storeygard, 2016; Henderson et al., 2012; Chen and Nordhaus, 2011), wealth and poverty using mobile phone logs (Blumenstock et al., 2015; Blumenstock, 2016), food security and resilience (Knippenberg et al., 2018), and community poverty (Abelson et al., 2014; Sohnesen and Stender, 2017, among others).

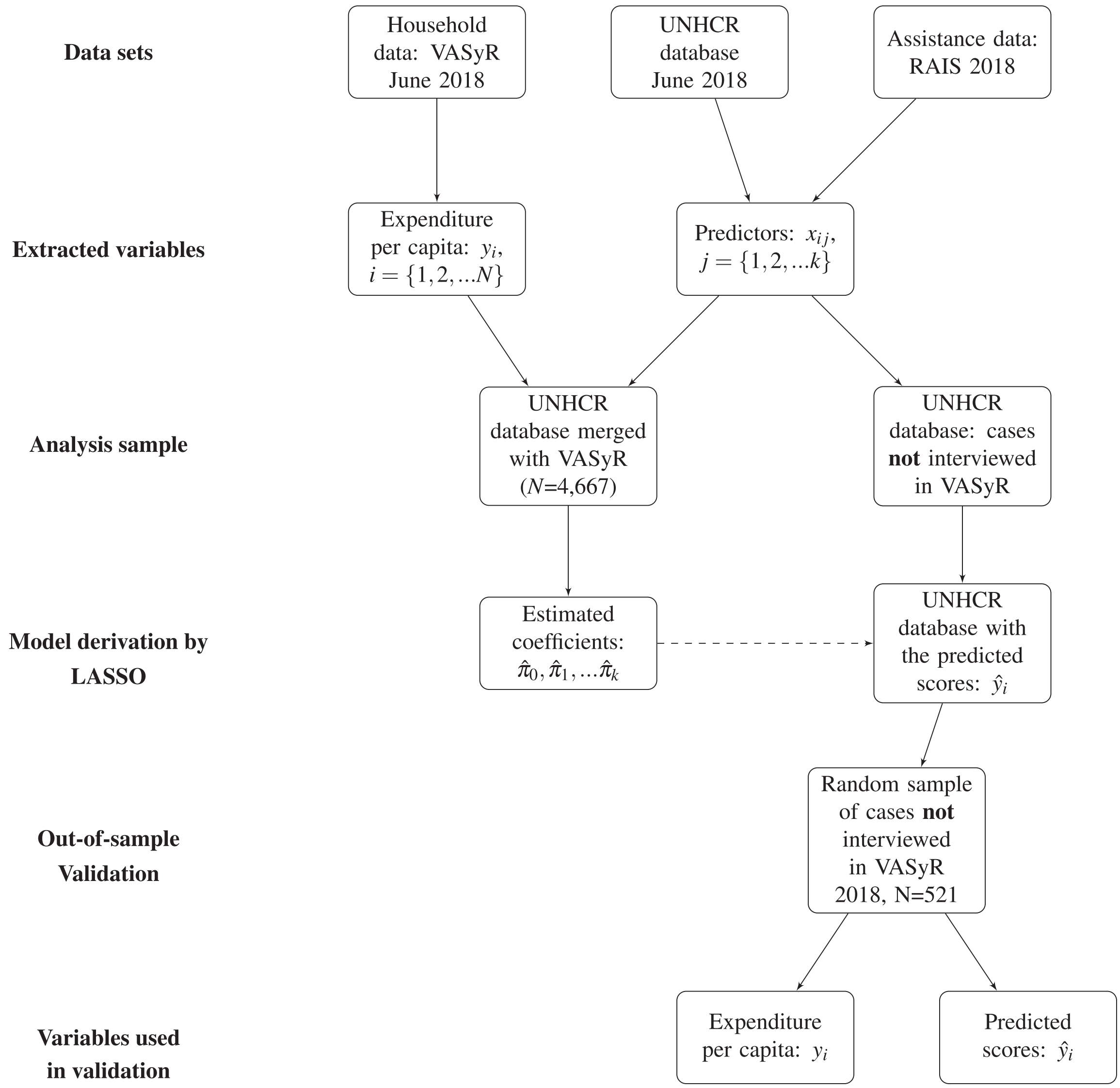


Fig. 1. Conceptual mapping of datasets used.

It solves the following optimization problem:

$$\underset{\pi_0,\pi_1,\ldots\pi_k}{\arg\min} = \sum_{i=1}^{N} (y_i - \pi_0 - \sum_{j=1}^{k} x_{ij}\pi_j)^2 \text{such that} \sum_{j=1}^{k} |\pi_j| \le \lambda \qquad (2)$$

where the constraint denotes the $L1$ norm of the regression coefficients and λ is a hyperparameter for coefficient regularization. We calculated the latter through a K-fold cross-validation process with $K = 10$, and chose a regularization parameter that yields the model with the fewest number of predictors that is within one standard error of the estimated minimum error rate (Hastie et al., 2009). Alternative models adjust the penalty parameter to estimate ridge and elastic net regressions, as well as perform random forest regression. For benchmarking, all models are compared to a forward selection algorithm. All of our results from these models use only administrative data in the vector of predictors; they are then compared to the results of an approach using an expansive vector of household characteristics and verifiable assets that would be available to develop a survey-based PMT, as described below.

3.2.2. Outcome and prediction variables

We model and predict a continuous measure of the natural log of expenditure per capita so that the targeting score can be used flexibly by humanitarian agencies in the form of a categorical classification, a ranking, or directly as predicted expenditure per capita. As described above, we construct the training sample by combining household expenditure per capita from the 2018 VASyR (survey) data and the household-level demographic variables from the June 2018 UNHCR (administrative) database. Only the dependent variable of the prediction model (log per capita expenditure) is taken from the 2018 VASyR survey, and candidate predictors come from the administrative data. This ensures consistency in the information used to model and predict per capita expenditure by reducing the discrepancies in the data sources across the two uses.[17]

The independent variables in the prediction model are based on the administrative records of household characteristics stored in the UNHCR database. We include the basic demographic variables in addi-

[17] For example, for a family who was surveyed in VASyR 2018, the education information was available in both administrative and survey data. We used the education information from the administrative data, which is more likely to be missing. While this can be expected to reduce in-sample prediction power, it ensures that the differential measurement error will have no impact when predicting the majority of the population for whom the same information is only available in the administrative data set.

Table 2
Summary statistics, VASyR 2018– household expenditure per capita.

Statistic	Mean	Std. Dev.	Median	N
Expenditure per capita (USD)	113.862	113.115	86.667	4364
ln (Expenditure per capita)	4.524	2.161	4	4364
Is severely vulnerable	0.502	0.500	1	4364

Note: This table shows summary statistics based on VASyR 2018.

Table 3
Characteristics of households at quantiles of predicted expenditure.

Quantile	Household size	Female Head HH	Disabled Dependent	HH Head Disabled	Share working age males
10	5.54	0.46	0.09	0.07	0.10
30	4.84	0.35	0.08	0.07	0.14
50	4.46	0.31	0.06	0.05	0.16
70	3.54	0.26	0.05	0.05	0.19
90	2.15	0.16	0.01	0.03	0.46
	Dependency ratio	Share no occ.	Share service sector occ.	Share below primary ed.	Share post-secondary ed.
10	0.65	0.22	0.04	0.24	0.05
30	0.58	0.18	0.06	0.19	0.10
50	0.55	0.13	0.09	0.12	0.12
70	0.50	0.07	0.13	0.08	0.20
90	0.15	0.06	0.19	0.05	0.35

Note: This table reports the demographic characteristics of households by quantiles of predicted expenditure based on lasso regression.

tion to measures of adults' previous occupations (in Syria) and education levels, the governorate of origin of the household head, the district of residence, and other specific medical issues or vulnerability measures. We also explicitly create a category for the share of records with missing data in any categorical variable so that all households can receive a predicted score. Appendix Table 1 contains a listing of the candidate variables used in the model-building process.

A potentially important issue is that existing transfers might contaminate the outcome of interest (expenditure per capita) through their effect on household expenditures. Because we are able to observe accurately which sample households are receiving assistance as of the survey date – another advantage of the administrative data – we train the model using an unadjusted measure of expenditure and include indicators for assistance receipt in the vector of candidate predictors. Put another way, we avoid making any assumptions about the marginal propensity to consume (or expend) cash transfers and do not adjust our outcome measure *ex ante*. Instead, we allow the model selection algorithm to provide a non-zero weight for the prediction step should assistance be sufficiently strongly linked to changes in expenditure. Therefore, both predicted and actual expenditure per capita account for the existing transfers that the families are receiving.[18]

3.2.3. Population characteristics at percentiles of predicted expenditure

We present characteristics of households in different percentiles of predicted expenditure per capita in Table 3. Overall, families who are predicted to be poor tend to be larger, have a higher share of disabled members, are substantially more likely to be female-headed, are less likely to have a working-age male, and have a higher share of dependents. Education and former occupation also follow an expected pattern, in which the model is more likely to target households with lower education and with a larger share of members who had no previous occupation before their arrival to Lebanon.

3.3. Model assessment

3.3.1. Prediction performance: administrative vs. short-form proxy-means

Table 4 contains the definition of our various measures of prediction performance. We first present a standard confusion matrix in Panel A, which classifies the four types of possible prediction outcomes based on true and predicted expenditure relative to our targeting eligibility cutoff. Panel B then defines inclusion and exclusion error, which are standard in the literature. We additionally use the Coady-Grosh-Hoddinott (CGH) Ratio, from Coady et al. (2004), which is the ratio of total benefits distributed to the targeted population to the ratio of total benefits that the same population would receive in the case of random or universal allocation at a given percentile of the distribution. For example, the CGH-40 ratio for Mexico's successful conditional cash transfer program, PROGRESA, is 1.56 — meaning that the households in the bottom 40 percent of the expenditure per capita distribution receive 62.4 percent of the resources in the PROGRESA program (62.4/40 = 1.56). Because of its flexibility in assessing targeting accuracy across different segments of the population, the CGH metric gives a more robust characterization of prediction performance across the distribution of targeted households and allows us to compare our findings to those documented across the 122 interventions reviewed and analyzed in Coady et al. (2004).[19]

Fig. 2 and Table 5 contain the above metrics of prediction performance across modeling approaches, along with 95% empirical confidence intervals based on 1,000 bootstrap replications. We begin with the benchmarked forward selection model using administrative data, which exhibits inclusion and exclusion errors of 35.7% and 26.1%, respectively, in Panel A. The lasso model provides substantial improvements, with inclusion error of 30.9% and exclusion error of 26.7%. We then show that the choice of the form of the penalty function makes little difference to prediction performance, with inclusion/exclusion errors from an elastic net regression (31.0%/26.7%) or ridge regression (30.5%/26.8%) being highly similar to those of lasso. Random forest regression yields slightly lower exclusion error (23.1%) at substantial cost of inclusion error (39.1%) – and with a substantially higher degree of variability than the regularized regressions.

[18] For program implementation, we manually set to zero any weights on indicators for current assistance, in order to predict a form of counterfactual expenditure per capita in the absence of any cash transfer. This allows us to avoid penalizing, in the new round of targeting, households who exhibit higher expenditure due to current receipt of transfers. However, for the purpose of calculating accuracy metrics for this paper, we make no such adjustment as it is not necessary for accurate representations of prediction performance.

[19] See Coady et al. (2004) for the details of the ranking methodology and the list of countries and programs that are included in the ranking list.

Table 4
Confusion matrix and targeting performance measures.

Panel A: Confusion Matrix		
	$1\{y_i < \$87\} = 1$	$1\{y_i > \$87\} = 1$
$1\{\hat{y}_i < \$87\} = 1$	True Positive (TP)	False Positive (FP)
$1\{\hat{y}_i > \$87\} = 1$	False Negative (FN)	True Negative (TN)

Panel B: Performance measure definitions	
Inclusion error (Leakage)	$\frac{FP}{TP+FP}$
Exclusion error (Undercoverage)	$\frac{FN}{TP+FN}$
Coady-Grosh-Hoddinott (CGH) Ratio	$\frac{\text{share of benefits reaching the poorest } x \text{ percentile}}{x}$

Note: Definitions of inclusion and exclusion error are presented as standard in the literature. The Coady-Grosh-Hoddinott (CGH) ratio is described in (Coady et al., 2004) and relates the ratio of the share of aid potentially disbursed under a given targeting scheme at a given percentile of the poverty distribution to the share of aid disbursed under a neutral (random) allocation scheme. For example, if the bottom 20 percent of the poverty distribution receive 50 percent of the aid disbursed, this ratio is 2.5. A higher value is associated with better targeting performance. Assuming homogeneous benefits and that total aid would reach all of the truly eligible, the CGH ratio can formally be expressed as $\frac{TP_x}{TP+FN} \div \frac{TP_x+FP_x+FN_x+TN_x}{TP+FP+FN+TN}$, where the latter fraction represents a universal, neutral, random assignment of aid — which by construction evaluates to x, the fractional percentile for which the CGH ratio is being calculated. Subscripted terms represent the cumulative sum of types at the xth percentile of the true poverty distribution, and nonsubscripted terms represent the total sum of types in the population.

Finally, the short-form survey approach (using lasso regression) considers a large vector of candidate features from measures of household characteristics and verifiable assets in survey data, which are listed in Appendix Table 2. In the short-form model derivation, both inclusion exclusion errors are slightly lower than those of lasso model, although as the overlapping 95% confidence intervals indicate, these differences are not statistically meaningful. The short-form survey yields an inclusion error of 28.6% (CI: [27.2%, 30.2%]) and an exclusion error of 24.3% (CI: [22.8%, 25.8%]), while the corresponding metrics for the lasso model are 30.9% (CI: [29.2, 32.4]) and 26.7% (CI: [25.2, 28.3]), respectively.

Panel B of Fig. 2 reports CGH metrics at the 10th, 20th, and 40th percentiles of the expenditure per capita distribution along with 95% confidence intervals from 1,000 bootstrap replications. In terms of targeting accuracy across the distribution of households, the lasso model allocates households below the 10th, 20th, and 40th percentile of the (true) expenditure per capita distribution to receive 15.3, 30.0, and 57.6 percent of available assistance. (In a program in which 50% of the population will receive assistance, the theoretical maxima of these percentages is 20, 40, and 80.) Dividing by the share of the population at which that statistic is evaluated yields CGH ratios of 1.53, 1.50, and 1.44, respectively, with a theoretical maximum of two at any percentile. As with the error rates in Panel A, the performance of any of the regularized linear models is roughly comparable; forward selection and OLS models yield a higher variance, and there are performance losses to the use of random forest models. The survey-based approach yields no gain at the 10th percentile (1.51), and only modest improvements over lasso at the 20th and 40th percentiles (1.55 and 1.48, respectively).

Taking these results together, we draw two major conclusions: first, there is no substantive difference in the capacity of administrative data – which includes no information on assets – to predict poverty, in this context, relative to traditional survey-based methods. Second, the short-form survey approach yields only small reductions in inclusion and exclusion error - in our context, only up to two percentage points. These improvements are not particularly surprising, as information on assets (or lack thereof) is likely to provide useful additional explanatory power in predicting the capacity of households to fulfill basic needs. The magnitude of performance gain, however, is important to quantify – especially in consideration of the relative costs of each approach. Our results suggest that the primary consideration when considering these alternative methods (survey vs. administrative data) should be one which weighs the higher cost of the survey-based approach with an expectation of only slightly higher errors when using administrative data, with the strong assumption that targeting institutions are capable of reaching the full population for assessment. In Section 3.7 below, we discuss cost considerations.

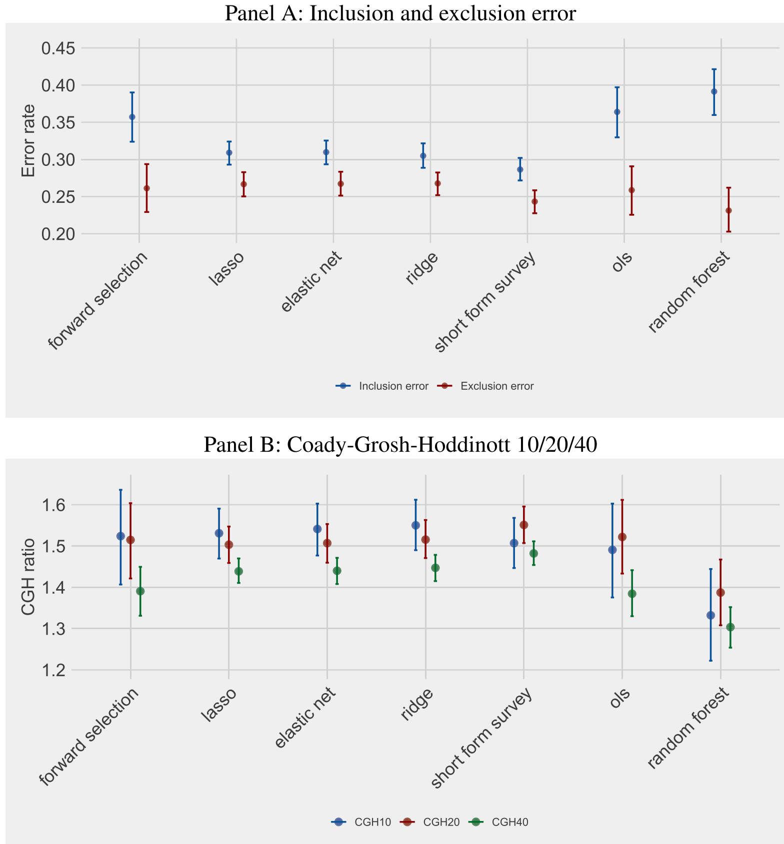
3.4. Tracing the source of prediction accuracy differences

We next undertake an analysis in which we explore the specific factors that generate differences in error rates between the survey-based and administrative data approaches. That is, are there a small number of features that could be selectively added to the administrative data to achieve survey-level error rates, and if so, what would these be? To do this, we augment our administrative data-based lasso models iteratively with the features from a single survey question to estimate the feature-wise contribution to targeting accuracy. Fig. 3 contains the result of this analysis, plotting the marginal change to net inclusion and exclusion error rates by feature added to the model; Appendix Fig. 1 plots effects on inclusion and exclusion error separately.

In Fig. 3, we see a clear pattern of prediction performance gains attributable to additional knowledge about the type of housing inhabited by the household,[20] providing improvement jointly in inclusion and exclusion error. A small number of basic household furniture questions (beds, refridgerator) provide modest improvements in overall error; the vast majority of the other features provide trivial improvements in error rates, with some features increasing model error.[21]

[20] This variable can take 14 values, including: Active construction site, Agricultural/engine/pump room, Apartment/house, Concierge's room in residential building, Factory, Farm, Garage, Hotel room, Prefab unit, School, Shop, Tent, Warehouse, or Workshop. Additional analyses suggest the most predictive individual values of this question are whether the household resides in an apartment, or a tent.

[21] Whether such features could be accurately captured through intake interviews that take place on UNHCR premises and not the dwelling of the refugee is an open question; this analysis is used to illustrate the principle behind one way to identify additional accuracy-enhancing features.



Note: Figure presents prediction performance metrics across methods and data sources. Statistics presented are the mean error rate and a 95% empirical confidence interval from 1,000 bootstrap replications.

Fig. 2. Inclusion and exclusion error, by prediction methodology.

Table 5
Prediction performance metrics, by prediction and adjustment methodology.

	Inclusion error	Exclusion error	Share of transfers to bottom: 10%	20%	40%
Panel A: Models using administrative data					
forward selection	0.357	0.261	0.152	0.303	0.556
	[0.324, 0.39]	[0.229, 0.294]	[0.141, 0.164]	[0.284, 0.321]	[0.532, 0.58]
lasso	0.309	0.267	0.153	0.301	0.576
	[0.293, 0.325]	[0.252, 0.283]	[0.147, 0.16]	[0.292, 0.31]	[0.564, 0.587]
elastic net	0.31	0.267	0.154	0.301	0.576
	[0.293, 0.325]	[0.251, 0.283]	[0.148, 0.16]	[0.292, 0.311]	[0.563, 0.589]
ridge	0.305	0.268	0.155	0.303	0.579
	[0.289, 0.322]	[0.252, 0.282]	[0.149, 0.161]	[0.294, 0.313]	[0.566, 0.591]
OLS	0.364	0.259	0.149	0.304	0.554
	[0.33, 0.397]	[0.225, 0.291]	[0.138, 0.16]	[0.287, 0.322]	[0.532, 0.577]
random forest	0.391	0.231	0.133	0.277	0.521
	[0.36, 0.421]	[0.203, 0.262]	[0.122, 0.144]	[0.262, 0.293]	[0.502, 0.541]
Panel B: Short form verifiable assets survey					
lasso	0.286	0.243	0.151	0.31	0.593
	[0.272, 0.302]	[0.228, 0.258]	[0.145, 0.157]	[0.301, 0.319]	[0.582, 0.605]

Note: Figure presents prediction performance metrics across methods and data sources. Statistics presented are mean error rate and a 95% empirical confidence interval from 1000 bootstrap replications.

Adding a single type of housing question to the administrative database would improve targeting accuracy around two percentage points, which would fully compensate for the targeting differences across models reported in Fig. 2.

3.5. Prediction performance: out-of-sample validation

For out-of-sample validation, we use data collected by UNHCR and WFP in July 2018 from 521 randomly selected households that were not

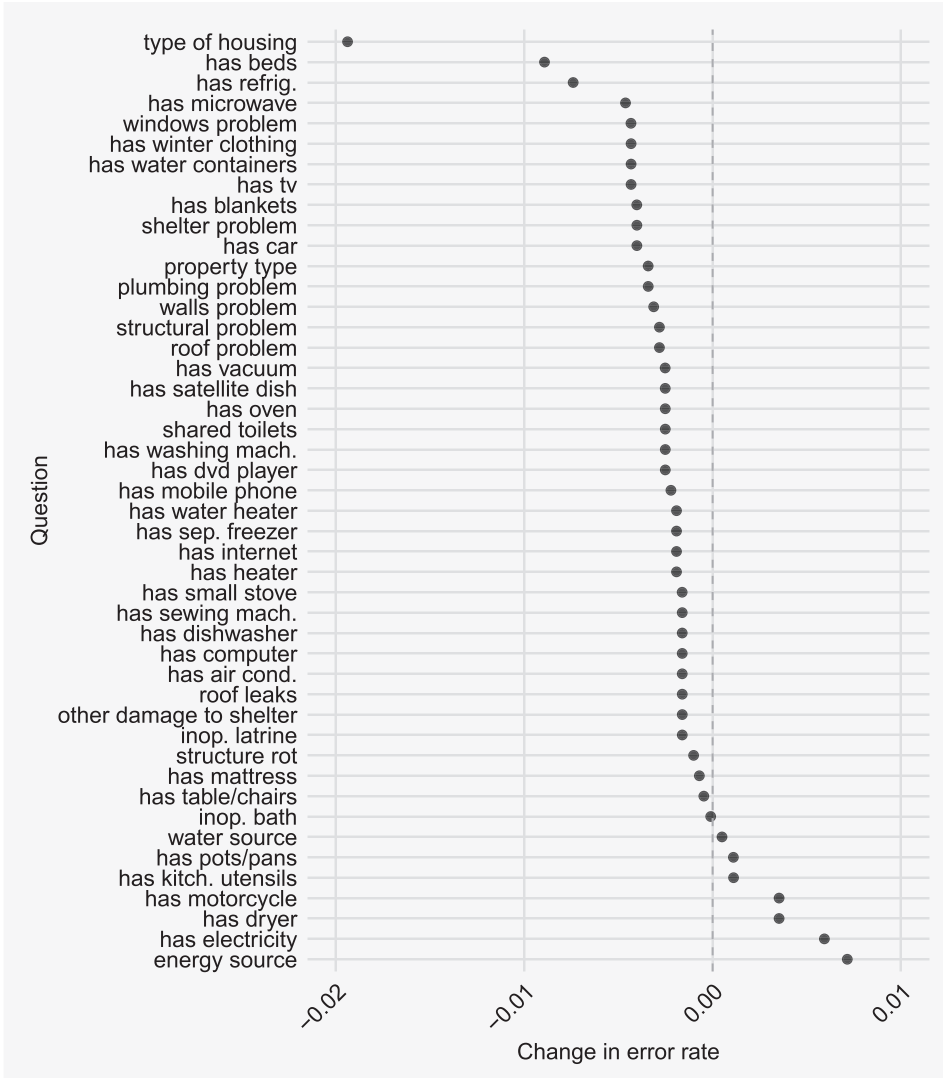
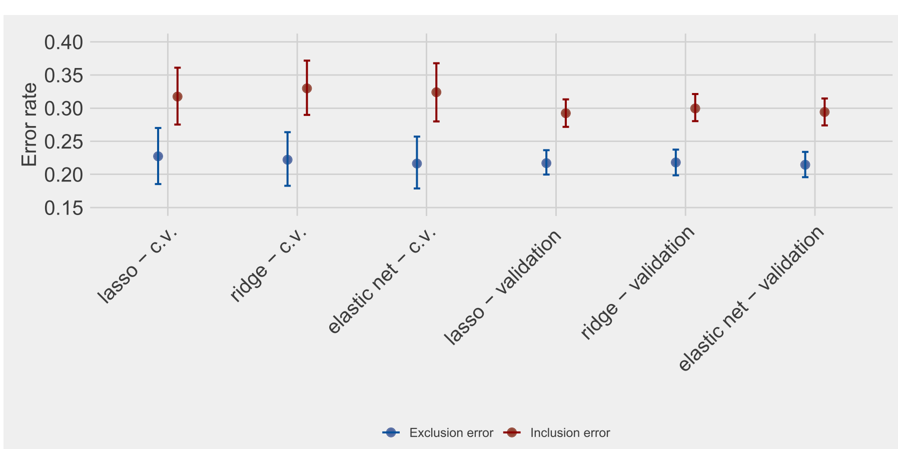


Fig. 3. Reduction in targeting error from single survey question added to administrative-data-based lasso model.



Note: Figure plots cross-validated error rates and out-of-sample error rates from the additional validation data collected after model development.

Fig. 4. Model cross-validation error rates versus blind out-of-sample validation.

part of the 2018 VASyR sample.[22] While such validation is commonly (and more easily) done by splitting the initial sample into training, testing, and (blinded) validation subsets, the choice to collect an additional sample was made to assess both validity and reliability of the model, as the data would be collected several months later than the training sample. Given the volatility of refugee households' situations as well as high levels of mobility, this approach lends additional insight into whether the model is intertemporally accurate – specifically whether accuracy changes substantially between the time the survey data were collected and the implementation of the program (using more recent data) several months later.

In the validation survey, the expenditure module was the same as that used in the 2018 VASyR survey, allowing us to recover a measure of expenditure per capita equivalent to that used in the modeling process. Furthermore, the same enumerators who collected the VASyR data also collected the validation survey. The sample was constructed to exclude households in the training sample and was collected after, but blind to the outcome of, the targeting model's prediction(s). To gauge and reduce measurement error, each household was visited and assessed by two enumerators.[23]

Fig. 4 contains inclusion and exclusion error rates based on model cross-validation as well as the blinded out-of-sample validation sample for households in the set of districts that surveyed in the validation sample. Overall, this test yields highly comparable prediction performance relative to the cross-validated error rates: in this sample, lasso, ridge, and elastic net models all generate cross-validated inclusion and exclusion error rates between 31 and 32 percent; in the blinded validation data, these same models yield error rates slightly lower (between 29 and 30 percent). This analysis not only confirms the cross validated error rates, as expected, but provides evidence of intertemporal reliability of the model, at least to the degree to which there is a time gap in the data used for model derivation and program implementation using more recent administrative records.[24]

3.6. Accuracy of targeting a refugee population

A remaining question is whether poverty targeting has the potential to be accurate among refugees relative to targeting the native poor. Given the unique characteristics of refugee populations relative to the poor in the host community, it is not obvious as to whether policymakers should hold similar expectations over the ability to accurately target poverty in humanitarian crises independent of the targeting method. To provide some insights, Fig. 5 overlays the performance of the survey-based approach (in blue) and our regularized regression models (in red) in the distribution of accuracy rates of the 85 targeted social welfare programs implemented in developing countries and are reviewed by Coady et al. (2004). First, the targeting of poverty among refugees, even in the survey-based approach, performed slightly above the median of other known programs – implying that targeting poverty among refugees is neither substantially more difficult nor substantially easier compared to other antipoverty programs and their respective beneficiaries. Second, the use of survey or administrative data does not meaningfully change the relative performance of the program in view of the distribution of accuracy across programs worldwide.

3.7. Cost of targeting

As discussed above, the main consideration between alternative approaches is one of relative cost: targeting based on survey data incurs a substantial cost per household visit, even for short-form surveys. Assuming an all-inclusive cost of \$25 per household visit[25] and a program scale (using the figures reported in Table 1 of nearly one million people of an average size of five, yielding 200,000 households), the survey-based approach would cost around \$5 million. This amount would allow the program to include more than 2300 additional families in the cash program – approximately equal to the implied number of wrongly excluded families by the lasso model using administrative data relative to the short-form survey PMT (110,000 × [0.266–.243]). This calculation, of course, assumes that logistical capacity would be available for a program of this size, that costs would not grow with the scale of the operation, and that all households could be found, reached, and scored for targeting; in practice, these factors may or may not be present in any given context.

3.8. Further considerations

We now discuss additional considerations related to the application of the above approach to poverty targeting. The first is whether accuracy is maintained when targeting smaller shares of the population. Up to this point, the analysis presents error rates when targeting approximately 50% of the refugee population; Fig. 6 contains inclusion/exclusion error rates when targeting from one to 99 percent of the population.[26] By construction, error rates approach 100% as the targeted share of the population approaches zero. Using non-regularized linear regression appears to improve accuracy when the objective is to predict a small number of households into the lower tail of the distribution, although programs of such small scale are often more conducive to categorical targeting or a case management approach in which individual cases are assessed through interviews and household visits.

Second, an important feature of the error rates above is that they average across subpopulations that differ in prediction accuracy. Recognizing this, implementing agencies may be interested in assessing for which populations predictive models generate higher versus lower prediction error; approaches to addressing this issue are only recently emerging in the poverty targeting literature.[27] While a detailed subgroup analysis would be required to fully understand the extent of such

[22] Due to logistical constraints, the sample was drawn from 11 of 26 districts in Lebanon, of which 9 were randomly selected. See Section A2 for more details of the sampling design.

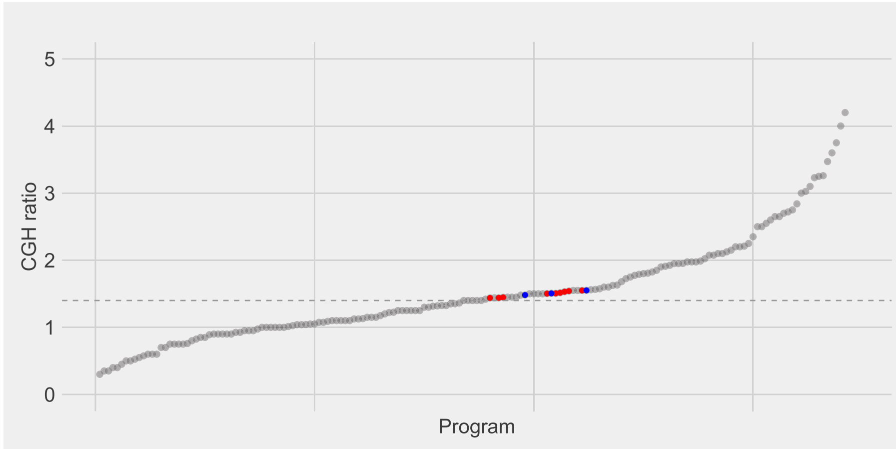
[23] Appendix Table 3 contains summary statistics of the variables collected in the validation sample, and compares them to comparable measures from the VASyR both for the same district sample and the whole sample. As expected, the households in the validation sample are highly similar in basic measures of welfare and demographics to the comparable VASyR sample: across Panel A and B, we see that the median household expenditure per capita was about \$81.79 in the validation sample, compared to \$79.24 in the VASyR. Household size is similar (approximately five in both samples), as is the share of severely vulnerable households (53 percent, compared to 56 percent).

[24] For practical purposes, the choice to collect an additional sample for validation also preserved a larger sample size for training the data, and was useful in building consensus as to the value of using a method that otherwise eliminates face-to-face contact between field workers and refugee households for targeting. The process we follow made predictions about families that field staff have never visited before, and then they had the opportunity to verify the predictions via household visits – somewhat easing the skepticisms of field staff on the capacity of this type of approach to accurately predict vulnerability.

[25] \$25 per household visit is an average cost for a medium-sized survey in the region. Experts we consulted suggest that the unit price would increase substantially if a very large number of households were to be surveyed in a relatively short period (under four weeks), and preparations for such an operation would require additional time and fiscal resources, with the potential for costs to reach up to \$40 per household visit; the estimates we present are thus conservative.

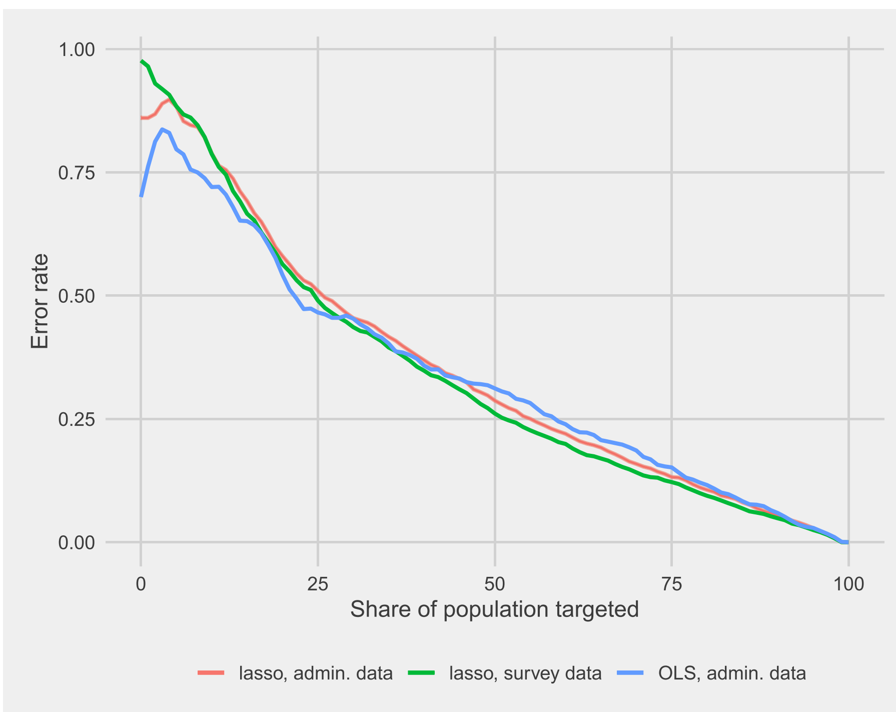
[26] Because this calculation is based on targeting shares of the population, inclusion and exclusion errors are equivalent conditional on share targeted and method due to the fact that there is a one-for-one replacement of false positives for false negatives. Inclusion and exclusion error rates will diverge when the error is calculated based on an absolute value, as in all other analysis presented in the paper.

[27] See Noriega-Campero et al. (2020), for example.



Note: Figure presents prediction performance metrics across programs reviewed by Coady et al. (2004). Each position on the x-axis is one program CGH ratio; some programs report multiple. The red series shows the CGH ratios for the lasso, elastic net, and ridge regression models described above; each model reports three ratios (CGH10, 20 and 40). The blue series plots analogous CGH ratios for the simulated survey-based approach.

Fig. 5. Prediction performance metrics, Lebanon cash targeting compared to programs analyzed in Coady et al. (2004).

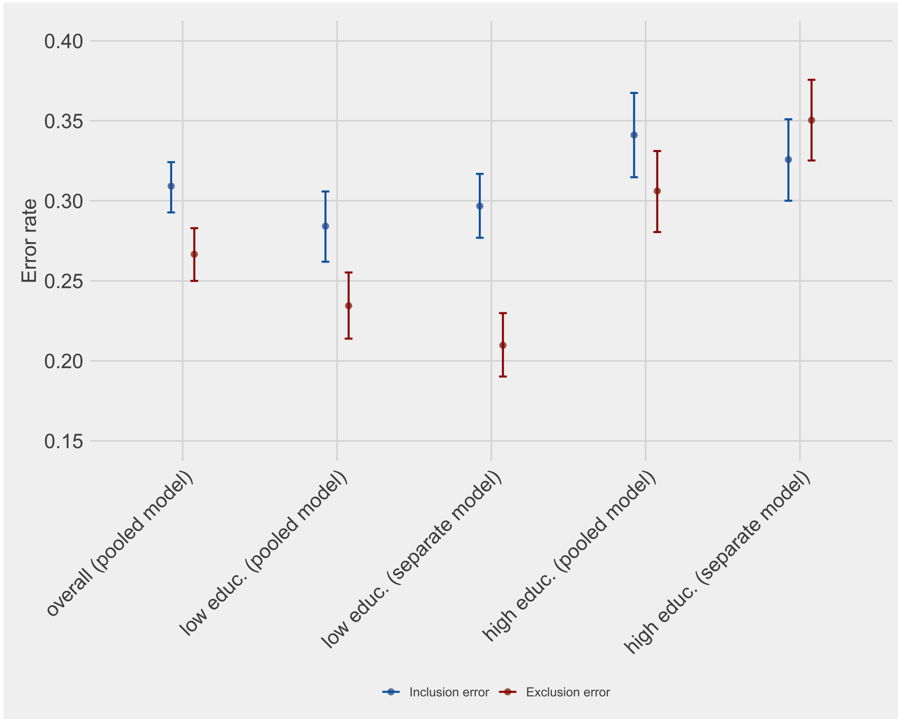


Note: Figure presents prediction performance metrics across methodology and differing shares of population targeted.

Fig. 6. Inclusion/exclusion error by share of population targeted.

heterogeneity, which we leave for future work, we inspect one dimension – education levels – for the purpose of illustration in this context.

We first show heterogeneity in error rates from a single lasso model across more- versus less-educated households: Fig. 7 shows that less-educated households are subject to lower rates of inclusion and exclusion error than more-educated households. We then test whether targeting these two groups separately would reduce this differential targeting accuracy, and find that subgroup targeting does not necessarily yield uniform reductions in targeting error. As shown in Fig. 7, less-educated households see exclusion error reduced alongside a slight increase in inclusion error; more-educated households, on the other hand, see higher exclusion error and slightly lower inclusion error than in the pooled model. Understanding the causes of varying error rates across subgroups thus likely requires additional data collection, such as



Note: Figure presents prediction performance metrics across education level subsamples and pooled versus split sample model training. "High education" is defined as those households that have at least one member with some secondary education; "low education" households are those with no members with at least some secondary education.

Fig. 7. Inclusion/exclusion error by education level and split sample vs. pooled modeling.

detailed followup surveys among groups subject to high rates of error.

Finally, we offer reflections on the frequency with which targeting models should be updated. In an environment with large population in- or out-flows or a rapidly changing policy environment, frequent re-assessment is likely to be beneficial to targeting accuracy. In more stable, protracted situations, collecting a new sample survey and reestimating the model annually might only yield minimal benefit to targeting accuracy. As to the frequency of updates to administrative data, while up-to-date administrative records are ideal, the strongest predictors of economic vulnerability are often structural and persistent – such as education and the ability to supply labor. Therefore, fields that hold promise for improving targeting but are often not found in administrative data include accurate measurements of human capital and labor supply capacity. Future work should test the value to targeting of various measures of human capital and labor supply capacity, among others, with a view towards their potential incorporation into administrative records at the registration stage.

4. Conclusion

An econometric targeting model for unconditional cash transfers based on limited information captured in typical administrative records held by humanitarian agencies performs approximately as well as a traditional PMT requiring a short-form survey of the entire potentially eligible population. These findings have implications for the understanding of the prerequisites for successful targeting of large scale cash and food assistance programs, especially in the context of humanitarian aid. The use of administrative data, which captures structural predictors of poverty, reduces the concern over misreporting of household composition or assets in annually-repeated targeting surveys (Banerjee et al., 2018; Camacho and Conover, 2011). A small reduction in targeting error comes at substantial costs, and could be alternatively achieved through the addition of a small number of fields to the administrative records. Our findings suggest that policymakers should consider targeting methods that maximize the use of existing data, and should additionally provide scope to include new features into administrative data should they be identified as beneficial for program targeting.

Acknowledgements

Altındağ acknowledges financial support from Bentley University' 2019 Summer Research Grant. Altındağ, O'Connell, Şaşmaz, and Balcıoğlu consulted for UNHCR Lebanon from 2018 to 2020 on topics related to the econometric targeting of cash-based interventions, and thank UNHCR Lebanon for providing the data used in this study. Several colleagues provided useful input and feedback on this work, including Cinzia Papavero, Simon Renk, Catherine Saiid, and Pablo Vizcaino. We would also like to thank Quy-Toan Do, Thomas Ginn, Linden McBride, Paolo Verme, Philip Verwimp, three anonymous referees, the editor, and participants at the Joint Data Center on Forced Displacement Research Conference 2020 for valuable comments. The views and conclusions found herein pertain to the context studied, and are those of the authors and not any organization with which they are or may be associated.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jdeveco.2020.102564.

References

Abadie, A., Kasy, M., 2019. Choosing among regularized estimators in empirical economics: the risk of machine learning. Rev. Econ. Stat. 10 (5), 743–762.

Abelson, B., Varshney, K.R., Sun, J., 2014. Targeting direct cash transfers to the extremely poor. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1563–1572.

Alatas, V., Banerjee, A., Hanna, R., Olken, B.A., Tobias, J., 2012. Targeting the poor: evidence from a field experiment in Indonesia. Am. Econ. Rev. 102 (4), 1206–1240.

Alatas, V., Banerjee, A., Hanna, R., Olken, B.A., Purnamasari, R., Wai-Poi, M., 2016. Self-targeting: evidence from a field experiment in Indonesia. J. Polit. Econ. 124 (2), 371–427.

Andini, M., Ciani, E., de Blasio, G., D'Ignazio, A., Salvestrini, V., 2018. Targeting with machine learning: an application to a tax rebate program in Italy. J. Econ. Behav. Organ. 156, 86–102.

Arendt, H., 1943. We Refugees. Menorah J. 31 (1), 69–77.

Baird, S., McIntosh, C., Özler, B., 2013. The regressive demands of demand-driven development. J. Publ. Econ. 106, 27–41.

Banerjee, A., Hanna, R., Olken, B.A., Sumarto, S., 2018. The (Lack of) Distortionary Effects of Proxy-Means Tests: Results from a Nationwide Experiment in Indonesia. Working Paper 25362. National Bureau of Economic Research.

Basurto, P., Dupas, P., Robinson, J., 2017. Decentralization and Efficiency of Subsidy Targeting: Evidence from Chiefs in Rural Malawi. NBER Working Paper No. 23383.

Blumenstock, J.E., 2016. Fighting poverty with data. Science 353 (6301), 753–754.

Blumenstock, J., Cadamuro, G., On, R., 2015. Predicting poverty and wealth from mobile phone metadata. Science 350, 1073–1076.

Brown, C., Ravallion, M., van de Walle, D., 2018. A poor means test? Econometric targeting in Africa. J. Dev. Econ. 134, 109–124.

Camacho, A., Conover, E., 2011. Manipulation of social program eligibility. Am. Econ. J. Econ. Pol. 3 (2), 41–65.

CARE, 2019. Suffering in Silence Iii: the 10 Most Under-reported Humanitarian Crises of 2018. (CARE.org).

Chen, X., Nordhaus, W.D., 2011. Using luminosity data as a proxy for economic statistics. Proc. Natl. Acad. Sci. Unit. States Am. 108 (21), 8589–8594.

Coady, D., Grosh, M., Hoddinott, J., 2004. Targeting of Transfers in Developing Countries: Review of Lessons and Experience. The World Bank.

Devereux, S., Masset, E., Sabates-Wheeler, R., Samson, M., te Lintelo, D., Rivas, A.-M., 2015. Evaluating the Targeting Effectiveness of Social Transfers: a Literature Review.

Donaldson, D., Storeygard, A., 2016. The view from above: applications of satellite data in economics. J. Econ. Perspect. 30 (4), 171–198.

Fantuzzo, J., Culhane, D.P. (Eds.), 2015. Actionable Intelligence: Using Integrated Data Systems to Achieve a More Effective, Efficient, and Ethical Government. Palgrave Macmillan.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer Series in Statistics, second ed. Springer.

Heckman, J.J., Smith, J.A., 1995. Assessing the case for social experiments. J. Econ. Perspect. 9 (2), 85–110.

Henderson, J.V., Storeygard, A., Weil, D.N., 2012. Measuring economic growth from outer space. Am. Econ. Rev. 102 (2), 994–1028.

Jean, N., Burke, M., Xie, M., Davis, W.M., Lobell, D.B., Ermon, S., 2016. Combining satellite imagery and machine learning to predict poverty. Science 353 (6301), 790–794.

Kidd, S., Gelders, B., Bailey-Athias, D., 2017. Exclusion by Design : an Assessment of the Effectiveness of the Proxy Means Test Poverty Targeting Mechanism. ILO working papers. International Labour Organization.

Kilic, T., Whitney, E., Winters, P., 2014. Decentralised beneficiary targeting in large-scale development programmes: insights from the Malawi farm input subsidy programme. J. Afr. Econ. 24 (1), 26–56.

Knippenberg, E., Jensen, N., Constas, M., 2018. Resilience, Shocks, and the Dynamics of Food Insecurity: Evidence from Malawi. (Working paper).

Kshirsagar, V., Wieczorek, J., Ramanathan, S., Wells, R., 2017. Household Poverty Classification in Data-Scarce Environments: a Machine Learning Approach. arXiv preprint arXiv:1711.06813.

McBride, L., Nichols, A., 2018. Retooling poverty targeting using out-of-sample validation and machine learning. World Bank Econ. Rev. 32 (3), 531–550.

Noriega-Campero, A., Garcia-Bulle, B., Cantu, L.F., Bakker, M.A., Tejerina, L., Pentland, A., 2020. Algorithmic targeting of social policies: fairness, accuracy, and distributed governance. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, FAT∗ '20. Association for Computing Machinery, New York, NY, USA, pp. 241–251.

Schreiner, M., 2010. Seven extremely simple poverty scorecards. Enterp. Dev. Microfinance 21 (2), 118–136.

Schüring, E., 2014. Preferences for community-based targeting - field experimental evidence from Zambia. World Dev. 54, 360–373.

Sharp, K., 2015. Review of Targeting of Cash and Food Assistance for Syrian Refugees in Lebanon, Jordan and Egypt. Report. United Nations High Commissioner for Refugees and the World Food Programme.

Sohnesen, T.P., Stender, N., 2017. Is random forest a superior methodology for predicting poverty? An empirical assessment. Poverty & Public Policy 9 (1), 118–133.

UNHCR, 2017. The 2016 Statistical Yearbook. Report. United Nations High Commissioner for Refugees.

UNHCR, 2018. Map of Registered Syrian Refugees by District in lebanon - 31/07/2018 Technical report. UNHCR.

UNHCR, UNICEF, WFP, 2017. The Vulnerability Assessment for Syrian Refugees in Lebanon. Report, United Nations High Commissioner for Refugees, United Nations Children's Fund, the World Food Programme.

Verme, P., Gigliarano, C., 2019. Optimal targeting under budget constraints in a humanitarian context. World Dev. 119, 224–233.

Verme, P., Gigliarano, C., Wieser, C., Hedlund, K., Petzoldt, M., Santacroce, M., 2016. The Welfare of Syrian Refugees: Evidence from Jordan and Lebanon. The World Bank, Washington, DC.

World Bank, 2018. Moving for Prosperity: Global Migration and Labor Markets. The World Bank.