ECONOMIC RESEARCH FORUM

منتدى البحوث الاقتصادية

Working Paper Series

ERF 25TH ANNIVERSARY

TARGETING HUMANITARIAN AID USING ADMINISTRATIVE DATA: MODEL DESIGN AND VALIDATION

Onur Altındağ, Stephen D. O'Connell, Aytuğ Şaşmaz, Zeynep Balcıoğlu, Paola Cadoni, Matilda Jerneck and Aimee Kunze Foong

Working Paper No. 1343

TARGETING HUMANITARIAN AID USING ADMINISTRATIVE DATA: MODEL DESIGN AND VALIDATION

Onur Altındağ[1], Stephen D. O'Connell[2], Aytuğ Şaşmaz[3], Zeynep Balcıoğlu[4], Paola Cadoni[5], Matilda Jerneck[6], and Aimee Kunze Foong[7]

Working Paper No. 1343

September 2019

Several colleagues provided useful input and feedback on this work, including Cinzia Papavero, Simon Renk, Catherine Saiid, and Pablo Vizcaino. We would like to thank Linden McBride 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.

Send correspondence to:
Onur Altındağ
Bentley University Department of Economics
oaltindag@bentley.edu

[1] Bentley University Department of Economics and Economic Research Forum.
[2] Emory University Department of Economics and IZA Institute of Labor Economics, soconnell@emory.edu.
[3] Harvard University Department of Government, sasmaz@g.harvard.edu.
[4] Northeastern University Department of Political Science, balcioglu.z@husky.neu.edu.
[5] UNHCR Lebanon, cadoni@unhcr.org.
[6] UNHCR Lebanon, jerneck@unhcr.org.
[7] UNHCR Lebanon, kunzefoa@unhcr.org.

First published in 2019 by
The Economic Research Forum (ERF)
21 Al-Sad Al-Aaly Street
Dokki, Giza
Egypt
www.erf.org.eg

Abstract
We develop and assess the performance of an econometric targeting model for a large scale humanitarian aid program providing unconditional cash and food assistance to refugees in Lebanon. We use regularized linear regression to derive a prediction model for household expenditure based on demographic and background characteristics; from administrative data that are routinely collected by humanitarian agencies. Standard metrics of prediction accuracy suggest this approach compares favorably to the commonly used “scorecard” Proxy Means Test, which requires a survey of the entire target population. We confirm these results through a blind validation test performed on a random sample collected after the model derivation.
Keywords: poverty targeting, proxy means test, cash transfers, refugees, forced displacement, Lebanon.
JEL Classifications: I39, I32, O12

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-targeting, community decisions, 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 widely used to target the poor. They are relatively easy to implement — typically relying on existing survey data to choose a set of predictors to collect in a short survey (typically referred to as a "scorecard") that is administered, in principle, to the potentially eligible population (Basurto et al., 2017; Kshirsagar et al., 2017; Schreiner, 2010).[1] Recent developments in econometric targeting approaches that prioritize out-of-sample prediction performance are also likely to increase the popularity of PMTs (McBride and Nichols, 2018; Kshirsagar et al., 2017).[2] The econometric approach to PMTs typically uses consumption or expenditure data from a representative household survey as a proxy for poverty, and derives a model, typically using forward stepwise regression, that assigns weights to factors used to predict poverty in the broader eligible population. 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 the 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 currently used to target over 300 million USD per year of unconditional cash and in-kind

[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; Schring, 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.

assistance to Syrian refugees in Lebanon. We combine a nationally representative expenditure survey with routinely collected administrative data to predict household expenditure by a commonly used regularized linear regression. We then compare the prediction accuracy of this model to that of a PMT "scorecard" approach, which would rely on data on household characteristics and verifiable assets collected via survey.

Our results show that this use of basic demographic information from typical administrative records held by aid organizations and governments is at least as accurate in targeting the poor compared to the traditional "scorecard" PMT that requires a household survey of the entire population. Our targeting model performs around the median of the 42 targeting interventions that use either a means or proxy-means testing approach in various developing countries reviewed by Coady et al. (2004). On average, targeting in this manner using administrative data allocates 40 to 50 percent more resources to the poor than a random allocation. We also show that using the same data source for model derivation and prediction has implications for targeting accuracy: the use of survey data for model derivation, as is common, performs less effectively than the use of the equivalent administrative records when only administrative data are available for predicting poverty among the full population.

We then 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 to the researchers only after the model development stage ensures zero degree of researcher discretion regarding the components in the final prediction model. This out-of-sample test yields targeting accuracy comparable to our preferred approach.

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. The model's performance compares favorably to a traditional "scorecard" approach which is often too costly, too cumbersome, or limited by logistical constraints for antipoverty programs of even moderate scale. Second, we perform a contemporaneous, out-of-sample validation of the prediction model. While the targeting model is derived using in-sample cross-validation, we take advantage of a rare availability of additional data collected for a contemporaneous out-of-sample test. The results are robust to this "blind" validation exercise.

The structure of the study is as follows. In Section 2, we briefly review the existing literature on targeting of anti-poverty programs, then describe the background and context of humanitarian assistance in the context of 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 presents the results from the analysis of those data within the same section. Section 4 concludes with thoughts 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 an increasingly 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) ease of implementation due to the short-form "scorecard" surveys that collect information on easily verifiable assets, and (ii) the ability to account for informal economic activity (Basurto et al., 2017; Kshirsagar et al., 2017; Schreiner, 2010).

There is, however, well-documented substantial variation in exclusion and inclusion error rates across implementations of PMTs.[3] The existing evidence suggests that better targeting is associated with stronger administrative capacity, larger variation in poverty, and the availability of reliable survey and administrative data, as well as the availability of proxies that are strongly correlated with poverty (Coady et al., 2004; Devereux et al., 2015; Kidd et al., 2017).[4] Even in ideal circumstances, however, a PMT is usually only partially successful in accurately targeting the poor, and the more homogeneously poor 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 the 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.

[3] 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.

[4] 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.

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 scorecards 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, 2018).[5] The primary destinations for displaced populations are neighboring countries, which often have constrained economic and operational resources to host these populations.[6] 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.[7]

Since 2011, the Syrian Civil War forcibly displaced more than 5.5 million people internationally. Lebanon hosts over 1.5 million refugees, resulting in the highest per capita population share of refugees in the world. Following the beginning of the refugee outflows from Syria in 2012, several separate cash transfer and voucher programs have been implemented in Lebanon by organizations including UNHCR, UNICEF, and WFP, among others. UNHCR and WFP provided four cash assistance programs for Syrian refugees in Lebanon.

[5] As of 2016, 19.9 million refugees globally were living under the mandate of the UNHCR (UNHCR, 2018).

[6] 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.

[7] 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.

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 USD 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 220,0000 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 through a lump-sum payment of 375 USD per household.

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. Since 2016, UNHCR and WFP have used a regression-based approach to determine the predictors of per capita consumption from a nationally representative[8] household expenditure survey called the Vulnerability Assessment of Syrian Refugees (VASyR). The model coefficients are then 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,[9] which reflects the subsistence level of consumption for a typical family as determined by

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

[9] 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.

the Lebanese government.[10]

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 a homogeneity in the population while also reducing the potential prediction power 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.

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 4,667 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/May 2018.[11] 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 who declared their refugee status and presence in Lebanon with the 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. Given these incentives, nearly all refugees make their presence and situation known to humanitarian agencies and thus are captured in the administrative data.

Information in the UNHCR database is updated on a regular basis through mobile phone

[10] 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, and communication and transportation costs. A full description can be found at UNHCR, UNICEF, WFP (2017).

[11] 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.

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.[12] 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.

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 the 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 with 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 all displaced populations known to UNHCR in Lebanon. 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.[13] Importantly, 33% of the households receive some form of assistance; of the available assistance programs, WFP's Cash for Food Assistance Program has the largest share of recipients.

Figure 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 $106 and $86 per capita, respectively.[14] We then use the estimated coefficients derived from the training sample to predict expenditure per capita in the population. In the

[12] 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.

[13] 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.

[14] The Lebanese pound was pegged to the US dollar during the study period with an exchange rate of approximately 1 USD = 1,500 LBP. All currency values referred to throughout the paper are in USD.

Table 1: Summary statistics, refugee records

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.

final stage of the analysis, consumption and expenditure data are collected from a random sample of 521 households that were not interviewed in the original VASyR sample to assess the out-of-sample overlap between actual and predicted expenditure (see Figure 1).

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

Statistic	Mean	Std. Dev.	Median	N
Expenditure per capita (USD)	105.598	75.680	86.000	4,659
ln(Expenditure per capita)	4.431	0.709	4.454	4,659
Household size	4.931	1.971	5	4,659

Note: This table shows summary statistics of household consumption per capita based on VASyR 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

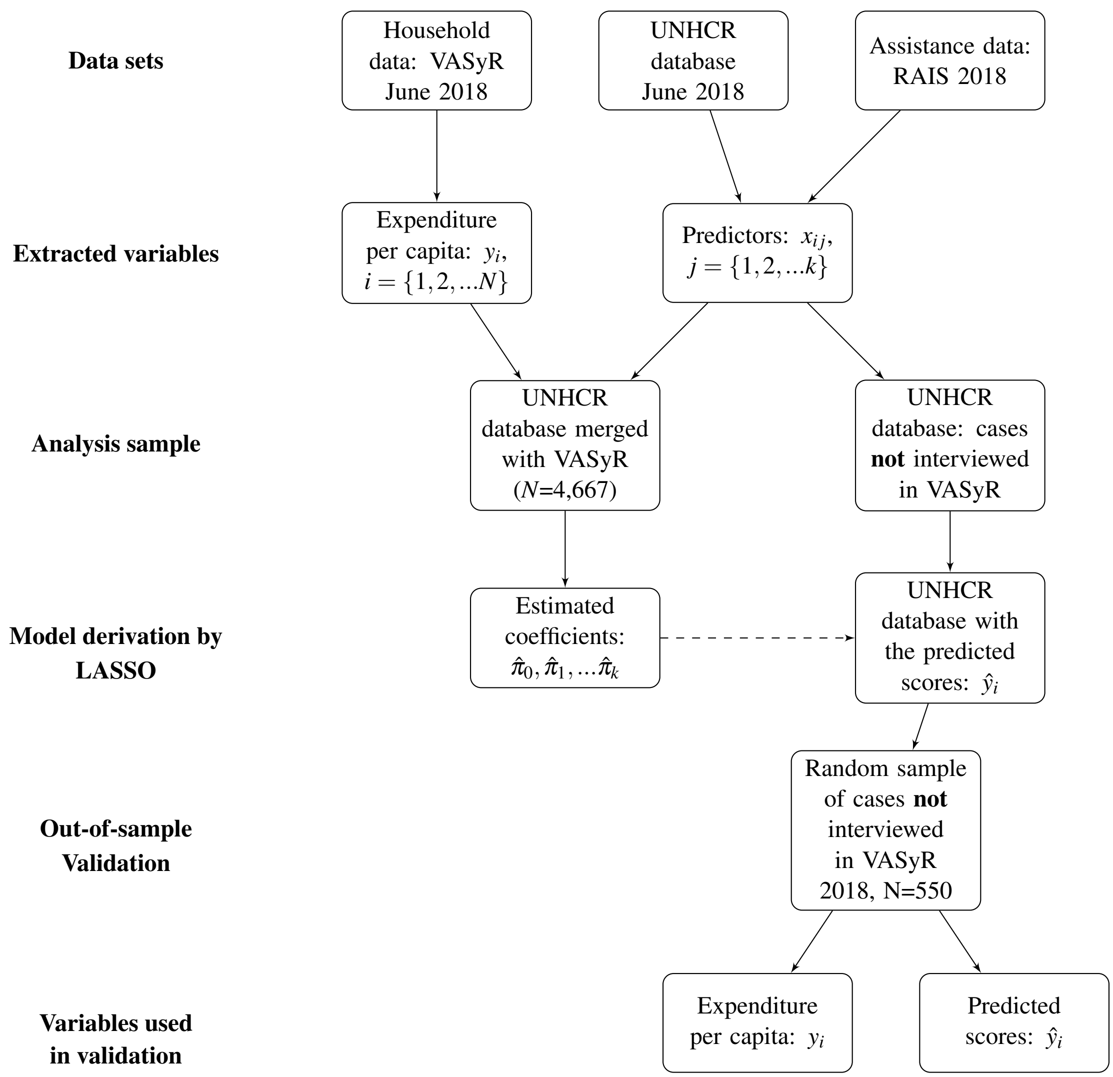
We use the following linear specification:

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

where y_i is the log per capita expenditure for case i, which is predicted by k independent variables, e 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 $\pi_0, \pi_1, ..., \pi_k$, we rely 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

Figure 1: Conceptual mapping of datasets used



Note: This figure shows the merging process of different data sets that are used for the targeting exercise. The validation sample was randomly drawn from the population that was not interviewed in VASyR 2018 and was performed after the targeting model was estimated.

or impossible.[15] It solves the following optimization problem:

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

where the constraint denotes the $L1$ norm of the regression coefficients and λ is a hyper-parameter for coefficient regularization. We calculated the latter through a K-fold cross-validation process, 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).

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. We show below that this conceptual change has implications for targeting accuracy.[16]

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 addition 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. Importantly, we include indicator

[15] 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).

[16] 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.

variables for each type of assistance that the family is currently receiving to account for any association between existing program participation and expenditure.

In our preferred approach, we predict expenditure net of existing cash transfers. Since three (of seven) indicators for assistance receipt are positively weighted in the LASSO model, predicting the outcome net of current assistance has meaningful implications for households currently receiving aid. In other words, when the LASSO model generates nonzero weights on measures of current assistance as predictors of per capita expenditure, we manually set these weights to zero in order to predict 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. We refer to results implementing this method of transfer adjustments in the tables below as "model-adjusted LASSO."

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.

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	5.82	0.43	0.08	0.06	0.10
30	4.84	0.37	0.08	0.07	0.13
50	4.44	0.32	0.06	0.05	0.16
70	3.58	0.27	0.05	0.05	0.21
90	2.09	0.16	0.01	0.03	0.49
	Dependency ratio	**Share no occ.**	**Share service sector occ.**	**Share below primary ed.**	**Share post-secondary ed.**
10	0.66	0.23	0.04	0.23	0.05
30	0.59	0.17	0.06	0.18	0.11
50	0.55	0.11	0.08	0.14	0.12
70	0.45	0.08	0.14	0.10	0.19
90	0.14	0.08	0.17	0.04	0.36

Note: This table reports the demographic characteristics of households by quantiles of predicted expenditure.

3.3 Model assessment

3.3.1 Prediction performance: administrative vs. scorecard 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 famous and 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 distribution, 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).[17]

Table 5 contains these metrics of prediction performance across modeling approaches. In Panel A, we compare LASSO regression with administrative data predictors to a classic forward selection stepwise regression ''scorecard" approach based on verifiable assets in survey data. This latter methodology involves selecting the set of verifiable assets and demographic information from the household survey (VASyR) and implementing a forward selection stepwise regression algorithm to determine the set of predictor variables in the final model. In order to generate out-of-sample prediction metrics, we split the survey data into equal-sized training and test samples with which we derive and validate the model, respectively. We iterate this process 1,000 times on independent equal splits of the data into test/train samples, and we report the median out-of-sample performance metric. Appendix Table 2 contains a listing of the candidate variables used in the model-building process for the "scorecard" model.

The cross-validated LASSO model yields an inclusion error of 32.2 percent and an exclusion error of 25.5 percent. In the scorecard approach, the median inclusion error is lower than LASSO, at 26.0 percent, while exclusion error is higher at 30 percent. In terms of targeting accuracy across the distribution of households, the model-adjusted LASSO scores target the population at the bottom 20th, 40th, and 60th percentile of the (true) expenditure per capita distribution to receive 29, 56, and 78 percent of the available assistance,

[17] 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.

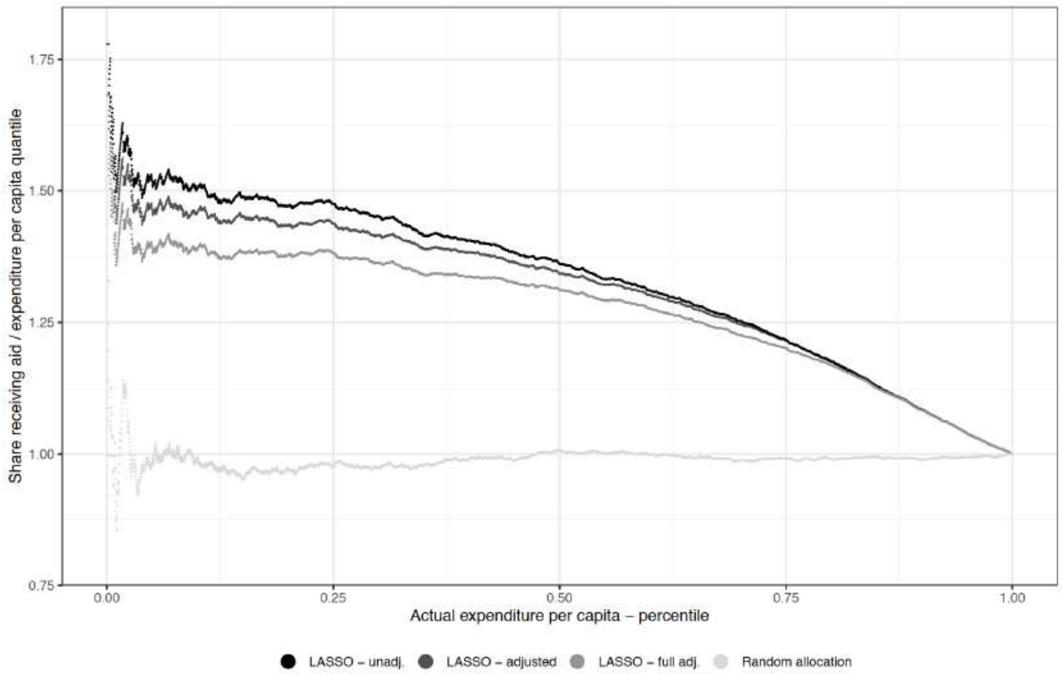
respectively. These results suggest a relatively favorable performance compared to more than 100 interventions studied by Coady et al. (2004). Our preferred methodology would place around the median of the 42 means and proxy means targeting methods analyzed by Coady et al. (2004). These ratios are highly similar in the scorecard method, especially for hypothetical programs of lower scale (*i.e.*, those targeting a smaller share of the population). In the third row, we include comparable statistics for a simulated random allocation of benefits among the population, which (by construction) exhibits approximately 50 percent inclusion and exclusion errors and an exactly proportionate CGH ratio at any point in the distribution.

In Panel B, we compare the performance of the LASSO model based on predictors derived from the administrative data (as in Panel A) and alternatively based on survey data. (Note that to ensure comparability, we re-estimate the model using only the variables that are commonly available in both sources, which excludes schooling, previous occupation, and governorate of origin). When estimating the model using survey-based predictor variables, exclusion error is 3.4 percentage points higher, increasing from 26.4 percent to 29.8 percent, and inclusion error decreases only slightly. We thus conclude that in situations where only administrative data are available to generate targeting model predictions, there are likely performance gains from using the same administrative data when such data are linkable to the poverty proxy measure used as an outcome in the model training step.

Panel C compares performance of the model-adjusted LASSO (described above) to two alternative treatments of existing transfers: one in which we performed no adjustment for current transfer receipt ("LASSO — unadjusted") and another in which we directly subtracted the per capita value of cash assistance from the predicted expenditure level ("LASSO — full adjustment"). Had we not adjusted the predicted expenditure for transfers received, the model would have resulted in slightly lower inclusion error (31.9 percent) and slightly higher exclusion error (26.6 percent), indicating that the choice between a model adjustment and no adjustment at all has only a minor impact on targeting accuracy. However, when we impose the strong assumption that $1 of cash assistance increases expenditure by exactly $1, and then remove the exact amount of assistance currently being received from the predicted score (indicated by — fully adjusted-LASSO), exclusion error reduces substantially but at a substantial cost of greater inclusion error, which reaches 34 percent. This comes also with poorer distributional performance, as can be seen in the CGH metrics. The full (or direct) adjustment for current transfers received thus over-includes some families, who then displace those slightly more needy but not currently receiving transfers.

Figure 2 calculates the CGH targeting ratio across the distribution of households in the targeting sample, compared to the same model variants in Table 5. The primary point to see here is the fact that differences in targeting accuracy between traditional scorecard methods are the largest when the beneficiary share of the population is the smallest. This is perhaps somewhat mechanical, since all series will naturally converge to 1 when the eligibility cutoff is high enough to cover the entire population. Another important implication of Figure 2 is that proxy targeting is only partially effective independent of the method used or the share of the poor targeted in the program — a finding that is repeatedly documented in existing evaluations of non-refugee populations (Coady et al., 2004; Devereux et al., 2015; Brown et al., 2018; Alatas et al., 2012).

Figure 2: Share of population receiving aid conditional on expenditure percentile



Note: Figure plots the CGH ratio across model variants and a single draw of a simulated random allocation of aid across households.

3.3.2 Prediction performance: out-of-sample validation of administrative LASSO

For out-of-sample validation, we use data collected by UNHCR and WFP in July 2018 from 521 randomly selected households that were not part of the 2018 VASyR sample.[18] 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

[18]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.

by two enumerators.[19]

Panel D of Table 5 contains inclusion and exclusion error rates and CGH 10/20/40 ratios based on the out-of-sample validation data, as well as (in-sample) LASSO error rates based on the VASyR data that include the same districts that were surveyed in the validation sample. Overall, this out-of-sample test yields highly comparable prediction performance relative to the previously reported in-sample metrics based on the VASyR survey, which are robust to a blind out-of-sample testing. In the validation sample that includes 11 districts, LASSO out-of-sample inclusion and exclusion errors are 33 percent and 18 percent, respectively, which closely correspond to analogous in-sample error rates (using the VASyR sample) of 31 and 20 percent, respectively.

3.3.3 Empirical validation of expenditure as a proxy in contexts of high vulnerability

In the validation survey, we also collected information on household coping mechanisms.[20] Specifically, after the household interview, enumerators were asked to rate, on a scale of 0-3 in increasing severity, the coping mechanisms they observed or were told that the household exhibited.[21] We then investigate the empirical association between the severity of coping mechanisms and predicted log expenditure per capita, finding that the severity of coping mechanisms engaged in is positively correlated with predicted expenditure per capita. That is, we provide direct evidence that families who adopt more severe forms of undesirable coping strategies are predicted to have higher expenditures by the targeting model (r = 0.133, p-value = 0.002) despite that fact that the prediction model does not include coping mechanisms as predictor variables. Thus, we empirically confirm the notion that an econometric targeting model based on expenditure per capita should, in principle, be complemented by alternative, complementary structures that provide the opportunity for such households to be considered for humanitarian aid.

[19] 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).

[20] "Coping mechanisms" are generally defined as activities or actions taken by the extremely poor or vulnerable to maintain subsistence-level consumption. These can include stress coping mechanisms (lowest intensity; includes spending savings, selling household goods, buying goods on credit, or incurring debt), crisis coping mechanisms (moderate intensity; includes selling productive assets or means of transport, withdrawing children from school, reducing non-food expenses, marrying children under 18, or engaging in survival sex), and emergency coping mechanisms (highest intensity; includes involving children in income activities, begging, accepting high-risk jobs, selling house or land).

[21] The Web Appendix contains the survey instrument.

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

Panel A: Administrative LASSO vs. survey scorecard approach	Inclusion error	Exclusion error	Share of transfers to bottom: 10%	20%	40%
LASSO (model-adjusted)	0.322	0.255	0.148	0.289	0.557
Forward selection (verifiable assets X)	0.260	0.300	0.152	0.307	0.584
Random allocation	0.492	0.492	0.103	0.200	0.405
Panel B: LASSO performance using predictors from UNHCR database vs. survey data					
LASSO (model-adjusted, UNHCR X)	0.320	0.264	0.146	0.292	0.558
LASSO (model-adjusted, survey X)	0.313	0.298	0.147	0.293	0.564
Panel C: LASSO performance under various treatments of existing transfers					
LASSO (unadjusted)	0.319	0.266	0.149	0.293	0.561
LASSO (fully adjusted)	0.340	0.225	0.140	0.276	0.536
Panel D: Out-of-sample validation (11 districts)					
LASSO (model-adjusted, validation)	0.332	0.183	0.119	0.253	0.504
LASSO (model-adjusted, VASyR)	0.307	0.201	0.132	0.262	0.507

Note: Panel A forward selection models present the median value of the prediction measure across 1000 split-sample model derivations. Panel B shows the differential targeting accuracy when using administrative versus survey-derived predictors (using a common vector of candidate predictors available across both sources). Panel C presents measures of prediction accuracy of the LASSO model subject to different methods for adjusting for current transfers received. Panel D presents measures of prediction accuracy in the out-of-sample test as well as comparable measures for a similar district sample in the VASyR data.

4 Conclusion

An econometric targeting model for unconditional cash transfers based on limited information captured in typical administrative records held by humanitarian agencies performs as well as, if not better than, "scorecard"-style PMTs requiring household surveys 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. Furthermore, the use of administrative data, which captures structural predictors of poverty, reduces the concern over misreporting of household assets in annually-repeated targeting surveys (Banerjee et al., 2018). Our findings suggest that policymakers consider the cost and scalability of such targeting approaches. Finally, we show that the data and adjustment choices by the program designer have significant implications on targeting accuracy, and a blinded out-of-sample validation test can be used to better assess the implications of such choices.

References

Abadie, A. and Kasy, M. (2019). Choosing among regularized estimators in empirical economics: The risk of machine learning. *The Review of Economics and Statistics*, (forthcoming).

Abelson, B., Varshney, K. R., and Sun, J. (2014). Targeting direct cash transfers to the extremely poor. *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1563–1572.

Alatas, V., Banerjee, A., Hanna, R., Olken, B. A., Purnamasari, R., and Wai-Poi, M. (2016). Self-targeting: Evidence from a field experiment in Indonesia. *Journal of Political Economy*, 124(2):371–427.

Alatas, V., Banerjee, A., Hanna, R., Olken, B. A., and Tobias, J. (2012). Targeting the poor: Evidence from a field experiment in Indonesia. *American Economic Review*, 102(4):1206–40.

Andini, M., Ciani, E., de Blasio, G., D'Ignazio, A., and Salvestrini, V. (2018). Targeting with machine learning: An application to a tax rebate program in Italy. *Journal of Economic Behavior & Organization*, 156:86 – 102.

Baird, S., McIntosh, C., and zler, B. (2013). The regressive demands of demand-driven development. *Journal of Public Economics*, 106:27 – 41.

Banerjee, A., Hanna, R., Olken, B. A., and 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., and Robinson, J. (2017). Decentralization and efficiency of subsidy targeting: Evidence from chiefs in rural Malawi. *NBER Working Paper No. 23383*.

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

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

Brown, C., Ravallion, M., and van de Walle, D. (2018). A poor means test? Econometric targeting in Africa. *Journal of Development Economics*, 134:109 – 124.

Chen, X. and Nordhaus, W. D. (2011). Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences*, 108(21):8589–8594.

Coady, D., Grosh, M., and 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., and Rivas, A.-M. (2015). Evaluating the targeting effectiveness of social transfers: a literature review.

Donaldson, D. and Storeygard, A. (2016). The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives*, 30(4):171–98.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics)*. Springer.

Heckman, J. J. and Smith, J. A. (1995). Assessing the case for social experiments. *Journal of Economic Perspectives*, 9(2):85–110.

Henderson, J. V., Storeygard, A., and Weil, D. N. (2012). Measuring economic growth from outer space. *American Economic Review*, 102(2):994–1028.

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

Kidd, S., Gelders, B., and 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., and Winters, P. (2014). Decentralised Beneficiary Targeting in Large-Scale Development Programmes: Insights from the Malawi Farm Input Subsidy Programme. *Journal of African Economies*, 24(1):26–56.

Knippenberg, E., Jensen, N., and Constas, M. (2018). Resilience, shocks, and the dynamics of food insecurity: Evidence from Malawi. *Working paper*.

Kshirsagar, V., Wieczorek, J., Ramanathan, S., and Wells, R. (2017). Household poverty classification in data-scarce environments: a machine learning approach. *arXiv preprint arXiv:1711.06813*.

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

Schreiner, M. (2010). Seven extremely simple poverty scorecards. *Enterprise Development and Microfinance*, 21(2):118136.

Schring, E. (2014). Preferences for community-based targeting - field experimental evidence from Zambia. *World Development*, 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. and Stender, N. (2017). Is random forest a superior methodology for predicting poverty? An empirical assessment. *Poverty & Public Policy*, 9(1):118–133.

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.

World Bank (2018). *Moving for Prosperity: Global Migration and Labor Markets*. The World Bank.

UNHCR (2018). The 2016 Statistical Yearbook. Report, United Nations High Commissioner for Refugees.

Verme, P. and Gigliarano, C. (2019). Optimal targeting under budget constraints in a humanitarian context. *World Development*, 119:224 – 233.

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

A1 Appendix Tables and Figures

Appendix Table 1: Candidate variables list

Demographic Variables	Type	*Year of Arrival*	Type
Household Size	Positive integer	2011 and before	Indicator [0/1]
Household Size Squared	Positive integer	2012	Indicator [0/1]
Dependency ratio	Continuous [0,1]	2013	Indicator [0/1]
Age of the HoH	Continuous [0,100]	2014	Indicator [0/1]
Female HoH	Indicator [0/1]	2015	Indicator [0/1]
Frac. of HH members aged 0-5	Continuous [0,1]	2016	Indicator [0/1]
Frac. of HH members aged 6-10	Continuous [0,1]	2017 and after	Indicator [0/1]
Frac. of HH members aged 11-17	Continuous [0,1]		
Frac. of male members aged 18-50	Continuous [0,1]	*Protection Measures*	
Frac. of female members aged 18-50	Continuous [0,1]	Frac. of HH members with a disability	Continuous [0,1]
Frac. of members older than 60	Continuous [0,1]	Frac. of HH members with a medical condition	Continuous [0,1]
Education		Disabled HoH	Indicator [0/1]
Frac. of HH members education Unknown	Continuous [0,1]	Existence of a disabled dependent member	Indicator [0/1]
Frac. of HH members no education	Continuous [0,1]	Existence of child at risk	Vector of indicators
Frac. of HH members some education below primary	Continuous [0,1]	Single Parent	Indicator [0/1]
Frac. of HH members with primary education	Continuous [0,1]	Existence of older person at risk	Indicator [0/1]
Frac. of HH members secondary education	Continuous [0,1]		
Frac. of HH members above secondary education	Continuous [0,1]		
Occupation			
Frac. of HH members with Unknown occupation	Continuous [0,1]	*Assistance received*	
Frac. of HH members Housekeepers	Continuous [0,1]	7 types of assistance received	Vector of indicators
Frac. of HH members with No Occupation	Continuous [0,1]	*Location of Arrival*	
Frac. of HH members in Service jobs	Continuous [0,1]	26 Districts of Lebanon	Vector of indicators
Frac. of HH members Laborer	Continuous [0,1]	*Location of Origin*	
Frac. of HH members Students	Continuous [0,1]	10 Main Governorates in Syria	Vector of indicators

Note: There are three types of variables we use: integer, continuous, and indicator. Integer variables take positive integer values only. Continuous variables can (but do not necessarily) take any value on the closed interval indicator. Indicator variables capture whether the household exhibits the characteristic indicated.

Appendix Table 2: Scorecard approach verifiable demographics, conditions, and assets candidate variables list

Home size and meals	*Type*
# rooms	Positive integer
M2 of dwelling	Positive integer
Household size	Positive integer
M2 per person	Positive integer
Bathrooms	Positive integer
Meals per day - Adults	Positive integer
Meals per day - Children	Positive integer

Shelter characteristics	*Type*
Has Air conditioning	Indicator [0/1]
Have Beds	Indicator [0/1]
Have Blankets	Indicator [0/1]
Have Light Vehicle	Indicator [0/1]
Have Computer	Indicator [0/1]
Have Dishwasher	Indicator [0/1]
Have Dryer	Indicator [0/1]
Have DVD Player	Indicator [0/1]
Have Heater	Indicator [0/1]
Have Internet	Indicator [0/1]
Have Kitchen Utensils	Indicator [0/1]
Have Mattresses	Indicator [0/1]
Have Microwave	Indicator [0/1]
Have Mobile Phone	Indicator [0/1]
Have Motorcycle	Indicator [0/1]
Have Oven	Indicator [0/1]
Have Pots & Pans	Indicator [0/1]
Have Refridgerator	Indicator [0/1]
Have Satellite Dish	Indicator [0/1]
Have Separate Freezer	Indicator [0/1]
Have Small Gas Stove	Indicator [0/1]
Have Sewing Machine	Indicator [0/1]
Have Tables+Chairs	Indicator [0/1]
Have TV	Indicator [0/1]
Have Vacuum Cleaner	Indicator [0/1]
Have Washing Machine	Indicator [0/1]
Have Water Containers	Indicator [0/1]
Have Water Heater	Indicator [0/1]
Have Winter Clothing	Indicator [0/1]

Structure/shelter conditions	*Type*
Structure in Dangerous Condition	Indicator [0/1]
Structure in Substandard	Indicator [0/1]
Shelter may collapse	Indicator [0/1]
Shelter has damaged roof	Indicator [0/1]
Shelter structurally weak	Indicator [0/1]
Shelter had unsealed windows	Indicator [0/1]
Shelter has leaking roof	Indicator [0/1]
Structure has rot	Indicator [0/1]
Shelter has damaged walls	Indicator [0/1]
Shelter has faulty plumbing	Indicator [0/1]
Shelter has unuseable latrine	Indicator [0/1]
Shelter has unuseable bath	Indicator [0/1]
Shelter lacks electricity	Indicator [0/1]
Shelter had other damage	Indicator [0/1]
Type of housing (14 categories)	Vector of indicators
Shelter type (Perm./Non-perm./Resid.)	Indicator [0/1]
Household shares toilets	Indicator [0/1]
Wastewater destination (7 categories)	Vector of indicators
Energy source (15 categories)	Vector of indicators

Demographics	*Type*
Men	Positive integer
Women	Positive integer
HH members <5	Positive integer
HH members 6-10	Positive integer
HH members 11-17	Positive integer
HH members 18-60	Positive integer
HH members 61+	Positive integer
Working-age men	Positive integer
Working-age women	Positive integer
Medical condition	Positive integer
Disability	Positive integer
60+ with medical cond.	Positive integer
Age of HH head	Positive integer
Female Head of Household	Indicator [0/1]
<18 Head of Household	Indicator [0/1]
60+ Head of Household	Indicator [0/1]
Disabled Head of Household	Indicator [0/1]
Head of Household has Med. Cond.	Indicator [0/1]
Dependents	Positive integer
Working-age adults	Positive integer
Disabled dependents	Positive integer
Members with serious medical cond.	Positive integer
60+ unable to work	Positive integer
60+ caregivers	Positive integer

Note: See Appendix Figure 1

Appendix Table 3: Summary statistics, Validation sample

Statistic	Mean	St. Dev.	Median	N
Panel A: Summary statistics, Validation sample (11 districts)				
Expenditure per capita (USD)	107.098	100.237	81.792	1,042
Household size	5.094	2.335	5	1,042
Is vulnerable	0.534	0.499	1	1,042
PBVS (0-110; 110=most vulnerable)	58.023	17.813	58.333	1,042
Panel B: Summary statistics, VASyR (11 districts)				
Expenditure per capita (USD)	96.843	71.568	79.238	2,389
Household size	4.980	1.987	5	2,389
Is vulnerable	0.563	0.496	1	2,389
Panel C: Summary statistics, VASyR (All districts)				
Expenditure per capita (USD)	105.598	75.680	86.000	4,659
Household size	4.931	1.971	5	4,659
Is vulnerable	0.508	0.500	1	4,659

A2 Data Collection for Model Validation

A2.1 Survey Instrument

The data collection instrument aimed to (i) collect basic data about the household, (ii) open the conversation between the interviewers and the family to observe the welfare of the household, (iii) collect expenditure and food consumption data in a similar way to VASyR to undertake an out-of-sample prediction test and (iv) collect the subjective assessments of the interviewers on fulfillments of basic needs in the household. The data collection instrument included the following components:

- General information on the household composition
- Dimensions of vulnerability (food, shelter conditions, access to healthcare, income and debt, coping strategies, general subjective assessment)
- Monthly expenditure
- Weekly food consumption

The sections on household composition, expenditure and food consumption followed the VASyR data collection instrument closely so that the data can be comparable. Each section in the "dimensions of vulnerability component" was composed of background questions through which the interviewer can open the conversation with the household and receive information. Each of 12 subjective assessment questions answered by the interviewers had options that are equivalent to vulnerability categories used by UNHCR (not vulnerable, mildly vulnerable, highly vulnerable, severely vulnerable).[22]

A2.2 Sampling Design and Data Collection

The data collection for the validation exercise was based on a random sample of 521 households from 11 districts in Lebanon, nine of which were randomly selected from the administrative data. Two other districts were imposed on the sample to ensure that governorates with only one district (Akkar and Beirut) are represented in the sample. To ensure sample size was maintained, we prepared a list of 2750 additional households to the field offices as potential sample replacements in the case of non-response among the households sampled. Note that the validation exercise sample is not representative of the population of concern in Lebanon due to administrative constraints in conducting the survey.[23] However, we did not expect this to have a substantial impact on the validation exercise. The main aim of this exercise was to assess the correlation between subjective vulnerability assessment and expenditure per capita and predicted scores. The requirements of this aim was fulfilled

[22] The full survey instrument is available in the Web Appendix.

[23] These included time constraints (the survey had to be fielded in one week) and additional authorizations required in some regions of the country.

given the high level of variation in vulnerability across and within 11 districts in which the survey was conducted. Appendix Table 4 provides information on the sampling structure.

Appendix Table 4: Sampling design for the validation exercise data collection

Region	Number of randomly selected districts	Imposed districts	Total districts	Number of randomly selected villages	Number of randomly selected households
North	2	Akkar	3	30	150
South	2		2	20	100
Bekaa	3		3	30	150
Beirut and Mount Lebanon	2	Beirut	3	30	150
Total	9		11	110	550

During each household visit, two interviewers collected data together but separately provided answers to the same questionnaire. This was to ensure that the subjective nature of perception-based assessment is at least partially mitigated by collecting two subjective assessment about the same household. The collected raw data included 1216 observations including both Syrian and non-Syrian populations of concern. The final sample excluded non-Syrian households, those with a missing case ID[24] as well as families for which we could only obtain one survey response.[25] The main analysis is based on a sample of 1042 observations (521 households), and Appendix Table 5 shows the distribution of the sample by governorate.

[24] Households in which the data collection team were not able to collect a correct case ID.

[25] Data from non-Syrian households were collected by the same data collection teams and with the same data collection instrument, as UNHCR wanted to conduct a similar validation exercise for the formula they use for distributing cash assistance to non-Syrian refugee families. They were excluded from the sample for these analyses because the formulae used for Syrian refugees and non-Syrian refugees are different.

Appendix Table 5: Realized sample and distribution of observations across governorates

Region	Governorate	Number of observations in the realized sample
North	North	198 (99)
	Akkar	100 (50)
South	South	98 (49)
	Nabatieh	100 (50)
	Baalbek-Hermel	100 (50)
Bekaa	Bekaa	200 (100)
Beirut and Mount Lebanon	Beirut	74 (37)
Beirut	Mount Lebanon	172 (86)
Total		1042 (521)