A Machine Learning Approach to Targeting Humanitarian Assistance among Forcibly Displaced Populations

Angela C. Lyons, Alejandro Montoya Castano, Josephine Kass-Hanna, Yifang Zhang, Aiman Soliman

2023
May 4 - 6,
Cairo Egypt

ECONOMIC RESEARCH FORUM
منتدى البحوث الاقتصادية

ERF 29th Annual Conference

A machine learning approach to targeting humanitarian assistance among forcibly displaced populations

Angela C. Lyons, University of Illinois at Urbana-Champaign[1]
Alejandro Montoya Castano, University of Illinois at Urbana-Champaign[2]
Josephine Kass-Hanna, IÉSEG School of Management[3]
Yifang Zhang, University of Illinois at Urbana-Champaign[4]
Aiman Soliman, University of Illinois at Urbana-Champaign[5]

December 14, 2022

Abstract

Increasing trends in forced displacement and poverty are expected to intensify in coming years. Data science approaches can be useful for governments and humanitarian organizations in designing more robust and effective targeting mechanisms. This study applies machine learning techniques to data collected from Syrian refugees in Lebanon over the last four years to help develop more robust and operationalizable targeting strategies. Our findings highlight the importance of a comprehensive and flexible framework that captures various poverty dimensions along with the commonly used expenditure metric, while also allowing for regular updates to keep up with (rapidly) changing contexts over time. The analysis also points to geographical heterogeneities that are likely to affect the effectiveness of targeting strategies. The insights from this project have important implications for agencies seeking to reduce the inclusion and exclusion errors, especially with shrinking humanitarian funding.

JEL classification: I3, I32, I38, O1, O53, R23, H1

Key words: poverty, forced displacement, refugees, humanitarian assistance, machine learning

[1] Corresponding author: Angela C. Lyons, Associate Professor, University of Illinois at Urbana-Champaign, Department of Agricultural and Consumer Economics, 440 Mumford Hall, 1301 W. Gregory Drive, Urbana, IL 61801 USA. Phone: +1 (217) 244-2612. Email: anglyons@illinois.edu
[2] Alejandro Montoya Castano, PhD Candidate, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, USA. Email: am26@illinois.edu
[3] Josephine Kass-Hanna, Assistant Professor, IÉSEG School of Management, Finance Department, Paris – La Défense, France. Email: j.kass-hanna@ieseg.fr
[4] Yifang Zhang, Data Analyst, National Center for Supercomputing Applications (NCSA), University of Illinois at Urbana-Champaign, USA. Email: zhang303@illinois.edu
[5] Aiman Soliman, Research Scientist, National Center for Supercomputing Applications (NCSA) and Research Assistant Professor, Department of Urban and Regional Planning, University of Illinois at Urbana-Champaign, USA. Email: asoliman@illinois.edu

A machine learning approach to targeting humanitarian assistance among forcibly displaced populations

Abstract

Increasing trends in forced displacement and poverty are expected to intensify in coming years. Data science approaches can be useful for governments and humanitarian organizations in designing more robust and effective targeting mechanisms. This study applies machine learning techniques to data collected from Syrian refugees in Lebanon over the last four years to help develop more robust and operationalizable targeting strategies. Our findings highlight the importance of a comprehensive and flexible framework that captures various poverty dimensions along with the commonly used expenditure metric, while also allowing for regular updates to keep up with (rapidly) changing contexts over time. The analysis also points to geographical heterogeneities that are likely to affect the effectiveness of targeting strategies. The insights from this project have important implications for agencies seeking to reduce the inclusion and exclusion errors, especially with shrinking humanitarian funding.

JEL classification: I3, I32, I38, O1, O53, R23, H1

Key words: poverty, forced displacement, refugees, humanitarian assistance, machine learning

1. Introduction

The United Nations High Commissioner for Refugees (UNHCR) estimates that there are more than 84 million forcibly displaced persons (FDPs) worldwide (UNHCR, 2022a, 2022b). At least 70% of these are living in conditions of extreme poverty, without access to food, water, and basic services (e.g., InfoMigrants, 2021). Humanitarian support is primarily provided by international agencies such as UNHCR, the World Food Programme (WFP), and UNICEF to address these basic needs. However, increasingly protracted forced displacement, and humanitarian response focused on cash-based assistance is reaching its limits and current levels of support are no longer sufficient (Lyons, Kass-Hanna, Molena, 2021; Lyons, Kass-Hanna, Montoya Castano, 2021). As a result, agencies are needing to redesign and implement cash-based assistance programs that more efficiently identify and target the most vulnerable families.

Targeting models used by government and humanitarian agencies mostly rely on a proxy means testing (PMT) approach, where support programs target families whose estimated consumption falls below a certain threshold (e.g., Altındağ et al, 2021; Brown et al., 2018; Chaaban et al., 2018; Chaaban et al., 2020; Lyons et al., 2021; Moussa et al., 2021; Schnitzer, 2019; Verme & Gigliarano, 2019). Advocates contend that the PMT method provides a relatively accurate and cost-effective tool to target the poor, that is suitable for large-scale assistance programs and is less prone to manipulation (Mills et al., 2015). However, these models have some important limitations. First, PMTs are based on expenditures, which are highly susceptible to variations in prices, especially in countries with hyperinflation or where the prices of goods and services differ considerably across regions. Second, PMTs are the best available tools if poverty is only based on a monetary measure such as income, or more commonly, consumption expenditures. However, there may be segments of the population who are not consumption poor, but who are nevertheless poor. Such populations may experience deprivations in other dimensions. They may, for instance, experience hardships such as food insecurity, inadequate housing, and a lack of employment opportunities. They could also have little or no access to essential services and resources, including healthcare, education, sanitation, and clean water. If a more comprehensive definition of poverty is used (e.g., multidimensional measure of social welfare), then PMTs are not very accurate predictors of poverty. Third, even if the best metric for defining poverty is expenditure and the PMT is the best tool to measure it, the predictions are not very accurate. Some have even suggested that PMT methods are largely arbitrary and akin to a “lottery” (Kidd et al., 2017; Kidd & Wylde, 2011).

The main objective of this paper is to develop effective targeting strategies that: (1) are not so susceptible to price changes, (2) do not rely on a measure that is solely based on expenditures, and (3) help to improve the accuracy of the targeting mechanisms. To construct and test our methodology, we use data from Syrian refugees in Lebanon. The Syrian refugee crisis is one of the largest mass displacements in recent years, and one of the worst humanitarian crises of our time. For over a decade, Lebanon has hosted an estimated 1.5 million Syrian refugees. The majority live in precarious conditions in the most impoverished areas of Lebanon where they represent more than 20% of the population – the highest per capita proportion of refugees in the world (Chaaban et al., 2020; Government of Lebanon & United Nations, 2020). The prolonged nature of the conflict, coupled with the COVID-19 pandemic and Lebanon’s dire economic conditions and political crises, have resulted in deteriorated living conditions for the refugees. According to UNHCR, WFP and UNICEF (2021), more than 90% of refugee households live in extreme poverty, below the Survival Minimum Expenditures Basket (SMEB).

In recent years, researchers have been increasingly devoting attention and resources to Lebanon’s case and are striving to improve the targeting mechanisms used by humanitarian agencies (Altındağ et al, 2021; Chaaban, Gattas, Irani & Thomas, 2018; Lyons et al., 2021). Research has focused on the selection of variables to better predict poverty, as well as the inclusion of analytical tools and criteria to classify those households who are most in need of assistance. For instance, Altındağ et al. (2021) proposed a low-cost methodology that used limited administrative data and machine learning (ML) techniques to predict household expenditures with accuracy comparable to that of survey-based models that have used PMT. Verme and Gigliarano (2019) used data from Syrian refugees in Jordan, a neighboring country of Lebanon that has experienced a similar influx of refugees, and proposed that researchers use ROC (Receiver Operating Characteristics) curves to define the optimal poverty cutoffs that reduce leakage and increase coverage. Despite the valuable contributions of these studies, they measure poverty only in terms of expenditure per capita. Other approaches, such as Chaaban, Ghattas, Irani, and Thomas (2018) move beyond monetary measures to include non-monetary measures such as food security. However, even in these cases, metrics that focus on the lack of material resources are not always good proxies of living standards, because individuals often have different needs and face different costs in trying to achieve the same living standards. Lyons et al. (2021) have proposed one of the most comprehensive multidimensional approaches. They constructed a multidimensional poverty index (MPI) akin to that of Alkire and Foster (2011) and Alkire and Santos (2014) to classify poor households that were deprived in several dimensions of human life, including health, food security, education, living standards, employment, personal security, and social inclusion. While the approach identifies more precisely which households and geographical locations are vulnerable to experiencing protracted poverty, the index requires data that is not readily available to agencies and costly to collect, which make it challenging to operationalize.

This study contributes to the literature and improves upon current targeting mechanisms in four key respects. First, we substitute expenditures per capita as the variable to measure poverty with a multidimensional metric based on three variables: expenditure, food security, and coping strategies. Our measure brings some elements from the multidimensional poverty literature (Alkire & Foster, 2011; Alkire & Santos, 2014; Lyons et al., 2021), as we measure poverty beyond the traditional expenditure approach, acknowledging that households face different costs to achieve the same standards of living and that higher expenditures are not an exact indicator for satisfying all basic needs. We then use a distance function to measure the distance between each household and the "poorest profile" and show that the classification of households who are poor is highly sensitive to the definition of poverty; and similarly, so are the exclusion and inclusion errors.

Second, we follow a rigorous methodology using machine learning (ML) techniques to better predict which households are more likely to be classified as poor based on a set of sociodemographic characteristics. We show some of the potential problems that arise in the definition of and use of socioeconomic variables at predicting PMT scores. We expect that these steps will help future research and international organizations in the calculation of PMT scores. Data science approaches have recently been used to address poverty and economic vulnerability in general (e.g., Abdul Rahman et al., 2021; Coromaldi & Drago, 2017; Yoder et al., 2021). However, relatively few studies have applied ML techniques to assist in the targeting of humanitarian assistance for forcibly displaced populations (Altındağ et al., 2021). ML analysis improves upon traditional econometric methods, as it does not require strong assumptions about the distribution of the data. At the same time, it enables the interaction of the variables that explain poverty (to create clusters) in flexible ways, which is not possible with linear methods.

Third, we include geospatial covariates in our ML models. The work of Lyons, Kass-Hanna, and Montoya Castano (2021) identified the importance of taking into consideration not only people-based poverty, but place-based poverty as well. Lebanon is a country that exhibits considerable heterogeneities across geographical locations in terms of land use, climate, employment opportunities, economic growth, etc. There are clearly geographical heterogeneities that need to be taken into consideration in the construction of current and future targeting algorithms. We are among the first to include geospatial indicators as predictors of poverty in ML models for refugee populations.

Fourth, and finally, we are also among the first to consider time trends. Humanitarian crises exist in conditions that are dynamic and in states of constant turmoil and flux. As such, humanitarian organizations must regularly update their targeting mechanisms, typically on an annual basis. Most previous studies have only investigated the PMT method using a single year of data. None, to our knowledge, have conducted a rigorous comparison of the effectiveness of these targeting mechanisms to consistently predict poverty over time, especially in a country such as Lebanon that has experienced hyperinflation and one crisis after another in rapid succession. We are among the first to compare our results over time to assess the stability of PMT and distance formulas over a four-year period that includes a pre- and post-COVID timeframe.

In the end, the ultimate goal of this study is to show in a systematic way how ML techniques can be combined with a multidimensional approach to improve traditional PMT targeting mechanisms. Increasing trends in poverty and displacement are expected to intensify in coming years due to population growth, climate change, economic inequality, and increased conflicts. Our findings show how governments and humanitarian organizations might use data science approaches to design more robust and effective targeting mechanisms in the face of increasing poverty and displacement, along with limited assistance resources. This work is particularly timely given the current Russo-Ukrainian crisis, where more than 4.2 million refugees have fled the country and over 6.5 million have been displaced inside Ukraine (UNHCR, 2022). The results from this study using data on Syrian refugees in Lebanon can help to inform resource allocation decisions related to this and other forced displacement crises in the future.

The remainder of this paper is structured as follows.[6] The next section describes the data. The third section presents our methods for constructing our multidimensional poverty measure using the distance formula and for using machine learning to generate our poverty predictions. The fourth section presents the results from the various comparisons of the traditional PMT using expenditure with the distance formula using our multidimensional PMT. The final section summarizes the key findings and highlights implications for humanitarian and development organizations seeking to improve current targeting mechanisms, especially given increasing poverty and displacement and limited humanitarian funding.

2. Data

We use survey data taken from the *Vulnerability Assessment of Syrian Refugees (VASyR)* jointly gathered by the UNHCR, WFP, and UNICEF for the years 2018, 2019, 2020, and 2021. The *VASyR* is a nationally representative survey of Syrian refugee households in Lebanon that includes detailed information on: (1) individual and household demographics, including work and schooling; (2) shelter, utility, sanitation, and settlement conditions; (3) income, expenditures, assets and debts; (4) food consumption and dietary diversity; (5) health and safety; and (6) coping strategies (UNHCR, UNICEF, & WFP, 2018, 2019, 2020, 2021). The UN

[6] Note that this paper is a work in progress. We plan to release a more complete version of the paper at the end of January and prior to the ERF Annual Conference.

agencies use the results from this annual survey to inform the distribution of humanitarian assistance and other interventions.[7]

We supplement the *VASyR* data with official data on the types, amounts, and duration of cash and non-cash assistance provided to refugee families who are registered with the humanitarian agencies. We also merge the PMT scores generated internally by the humanitarian agencies for each refugee household with the most updated administrative data, using a proprietary algorithm that predicts households' consumption expenditures. Those with PMT scores below a certain expenditure level (usually the Minimum Expenditure Basket - MEB) are classified as poor by the UNHCR. This score largely determines a refugee family's eligibility for assistance and other interventions.

Our analysis was conducted at the household level and by survey year and governorate. The initial sample size included 18,551 refugee households for all four survey years (4,434 in 2018, 4,670 in 2019, 4,480 in 2020, and 4,967 in 2021). Households with heads less than 15 years old or who had missing information about their educational attainment and other key variables included in this study were excluded from the sample. The final sample consisted of 17,642 refugee households (4,141 in 2018, 4,377 in 2019, 4,288 in 2020, and 4,836 in 2021).

We used the merged data to construct the standard PMT score that approximates expenditures per capita (the traditional PMT) and our multidimensional PMT score based on the 3 key dimensions: (1) expenditures per capita, (2) food consumption score (FCS), and (3) reduced coping strategies (rCSI). These 3 factors are used most often by UNHCR and WFP to measure vulnerability among the refugees. Other variables used in this study account for the household's structure in terms of household size, dependency ratio, proportion of female-headed and single-parent households, and the fraction of household members by age, gender, education, employment status, health and disability, and residency status. In addition, we include variables that identify households that are receiving cash for food and/or multipurpose cash assistance. Variables that capture other dimensions of vulnerability to poverty and identify household deprivations related to basic living standards and social welfare are also included. See Table 1 for a complete listing of all the variables included in our study and how they were constructed. Table 2 presents the descriptive statistics for these variables by survey year; p-values are reported to identify which variables differed significantly across the years.

[INSERT TABLE 1]

[7] In each survey year, data were collected from Syrian refugee households who were randomly selected from the 26 administrative districts across the eight governorates of Lebanon. To ensure representativeness at the district and governorate levels, sampling was based on a two-stage cluster approach whereby clusters (villages, neighborhoods, or towns) were selected within each district, and then refugee cases were randomly selected within each cluster. Specifically, probability proportionate to size (PPS) methodology was used, where clusters with larger concentrations of refugees were more likely to be selected. Weights were also constructed at the district level based on the refugee population in each district. See UNHCR, UNICEF and WFP (2018, 2019, 2020, 2021) for more details about the sampling and survey methodology.

[INSERT TABLE 2]

Geospatial indicators are likely to also be significant predictors of poverty and were key covariates included in our models. As previously noted, we are among the first researchers to include such an extensive and unique set of geospatial variables. The extraction of the geospatial attributes was conducted using the district administrative units. The geographic boundaries of twenty-six districts were used. We extracted the fraction area coverage of five different land cover types, namely built-up area, crops, permanent water area, seasonal water, and area covered in snow using the annual 100m global land cover maps in raster format available from the Copernicus Global Land Service (CGLS) portal.[8] The fractions of land area coverage for the five land cover types were calculated for each of the 26 districts. In addition, the total population count was extracted from the WorldPop rasters, where the population counts were adjusted to match the UN population estimates.[9] Also, included was the Normalized Difference Vegetation Index (NDVI), a standardized measure of healthy vegetation and how sensitive the vegetation in a particular area may be to drought. The average Normalized Difference Vegetation Index (NDVI) per district was extracted from the CGLS, which includes the NDVI Collection 300m (Versions 1 and 2). We also estimated the average elevation and its standard deviation for each district using the USGS/NGA Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010) at the resolution of 30 arc-seconds (approximately 1 km at the equator).[10] Finally, we extracted the monthly nighttime light intensity using the Visible Infrared Imaging Radiometer Suite (VIIRS) V10 produced by the Earth Observation Group (EOG) at the resolution of 15 arc second (approximately 500m at the equator).[11] Light intensity was averaged over each district to provide a proxy for economic development.

3. Methodology

3.1. Distance functions as our measure of poverty

We use distance functions to approximate multidimensional poverty. Particularly, we compare the distance of each household to the poorest profile in our sample. The distance is calculated as the weighted average of the distances to the poorest profiles across three dimensions: expenditures (Exp), food consumption score (FCS), and reduced coping strategies (rCSI). The poorest profile is constructed as the average of the fifth percentile in each dimension. In the context of Syrian refugees, households that fall in the fifth percentile in the three dimensions are identified as being deprived or poor in these dimensions. We prefer the Manhattan distance function[12], as its values better resemble a normal distribution (see Figure A1 in the Appendix). It is

[8] https://land.copernicus.eu/global/
[9] https://www.worldpop.org/
[10] https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-global-multi-resolution-terrain-elevation
[11] https://eogdata.mines.edu/products/vnl/
[12] As a robustness check, we also analyzed the Euclidian and Minkowski distance functions. See Figure A1 in the Appendix for a comparison of the distributions of the Manhattan, Euclidean, and Minkowski distance formulas.

important to note that the distance function, as any other measure of multidimensional poverty, requires the assignment of weights to each dimension of poverty. For comparison purposes, and as a robustness check, we estimate two models: one in which each component is given equal weight in the distance function, and another that assigns 50% to expenditure and 25% to the other two components. Equation (1) shows the estimated distance ($DIST_i$) for the i^{th} household, where W_x denotes the weights of each component and $\overline{X_{5\%}}$ is the average value of dimension X for the fifth percentile:

$$DIST_i = W_{Exp}|Exp_i - \overline{Exp_{5\%}}| + W_{FCS}|FCS_i - \overline{FCS_{5\%}}| + W_{rCSI}|rCSI_i - \overline{rCSI_{5\%}}|. \tag{1}$$

As a final step, we determine our multidimensional poverty score by adjusting the distances such that households who are closer to the poorest profile receive the highest value and households who are not considered as poor receive a value that is closer to 0. Equations (2) and (3) show the inverse distance of each household to the poorest profile, such that higher scores are associated with households who are more in need of humanitarian assistance:

$$MPscore_i = Max(DIST_i) - DIST_i \tag{2}$$

$$MPscore_i = Max(DIST_i) - (W_{Exp}|Exp_i - \overline{Exp_{5\%}}| + W_{FCS}|FCS_i - \overline{FCS_{5\%}}| + W_{rCSI}|rCSI_i - \overline{rCSI_{5\%}}|). \tag{3}$$

Distance functions have been used for estimating which households fall into multidimensional poverty. To our knowledge, there are two main applications in the literature. The first application follows the concept of Sen's functioning or capabilities, where households seek to guarantee a certain level of capabilities, such as good housing conditions or a high level of education. These capabilities can be achieved by the use of inputs, such as income, savings, or assets. In this sense, this first approach uses distance functions in the form of stochastic production frontiers, by measuring the number of inputs necessary to achieve a certain level of capabilities (Deutsch & Silber, 2005; Ramos, 2008). The second approach is the use of cluster analysis to classify households into poverty levels or predict which households are more likely to be poor based on a set of socioeconomic characteristics (Otoiu, Titan, & Dumitrescu, 2014; Sani et al, 2018; Usmanova, Rakhmonov, & Osamy, 2022). Implicitly, clustering algorithms use distance functions to measure the dissimilarity between observations and to classify them into clusters. Our methodology differs from both approaches, as our ultimate goal is not the classification of households into clusters, but the use of distances to create an indicator of poverty that resembles the standard PMT approach. To this end, we use a metric that quantifies how far a certain household is from not being poor and eligible to assistance, that is the inverse of each household's distance to the poorest profile. In this sense, higher scores are associated with households who are more in need of humanitarian assistance. Furthermore, we use inputs (i.e., socioeconomic characteristics) to predict out the likelihood of households to be poor, just as the standard PMT approach does. Nevertheless, our approach is more flexible than the traditional PMT based solely on expenditures, because it allows including other dimensions of poverty.[13]

Note that the distance approach does not address the fact that the main problem with multidimensional measures of poverty is that they are hard to operationalize. They require data that are often unavailable or costly to obtain, and they require a clear definition of the weighting scheme for each dimension, etc. (Lyons et al., 2021). Also, it is not certain whether the assistance provided by humanitarian organizations can help to reduce all deprivations included in standard multidimensional poverty indices. For example, it is unlikely that more humanitarian assistance will increase the level of education of adult members in the household, an indicator commonly used by the multidimensional poverty index designed by Alkire and Foster (2011), among others. To address these concerns, we opted for a measure of poverty that includes a small set of variables which can be influenced by the humanitarian assistance. As such, our measure can be easily calculated and predicted with available information and can, thus, be more practically operationalized by humanitarian organizations.

3.2. Machine learning and poverty predictions

We predicted poverty for all refugee households using machines learning (ML) techniques. Our training protocol can be described as follows. We compared the performance of three ML models: Lasso Regression (Lasso), Random Forest (RF), and Gradient Boosting (GB). The models were trained to predict the poverty rankings using both the traditional PMT and then our multidimensional poverty score based on the distance from the poorest five percent. Models were fitted using R statistical language version (4.2). The predictors of poverty included in our models were described in the data section. These variables were selected, because they are included in some form in UNHCR's official administrative data collected for all refugees and so are readily available to humanitarian organizations. This approach follows that of Altındağ et al. (2021). We added to the models the set of geospatial indicators not previously used by other researchers, but that are also available for all refugees.

We used a repeated K-fold cross-validation strategy to evaluate the performance of our models. We divided the data into five equal folds and trained the models using four partitions and then tested the models using the remaining partition. This process was repeated three times. Modeling and cross validation were implemented using the R package Caret version (6.0).

The models were calibrated to identify the best model at predicting our distance-based multidimensional poverty (MP) score, as well as the traditional PMT. Five accuracy metrics were used to evaluate the performance of the three models – namely, the absolute error (Abs. Error), the Pearson correlation coefficient (Correlation), the Mean Squared Error (MSE), the Root Mean Squared Error (MSE), and the R-squared.

[13] It is also possible to estimate the traditional PMT based solely on expenditures by not assigning any weight to the other components in the distance function.

$$\text{Abs. Error} = \boldsymbol{\Sigma}|y(i) - \hat{y}| \tag{4}$$

$$\text{Mean Squared Error (MSE)} = \boldsymbol{\Sigma}(y(i) - \hat{y})^2 / N \tag{5}$$

$$\text{Root Mean Squared Error (MSE)} = \text{sqrt}(\boldsymbol{\Sigma}(y(i) - \hat{y})^2 / N) \tag{6}$$

$$\text{R-squared} = 1- (\text{sum of squares of residuals / total sum of squares}). \tag{7}$$

Table 3 presents the results for the five metrics for the three models estimated for our distance-based multidimensional poverty score and the traditional PMT score (based on expenditure). The models were estimated separately for each year. In comparing the five metrics, the results were similar for all three models. However, Gradient Boosting (GB) performed slightly better than the Lasso and Random Forest and so we focused on refining the calibration of this model.[14] This better performance of the Gradient Boosting method is mainly attributed to accounting for non-linear relationships between variables and optimizing the model weights via the gradient descent method, although the results are comparable. We also see that applying the same ML models to predict poverty as a function of expenditure only resulted in slightly better prediction results, as the ML model did not need to account for interactions between the three-dimensional poverty measure, which were considered when predicting the poorest households using the distance-based MP score.

[INSERT TABLE 3]

We performed a parameter grid search to obtain the best values of the gradient boosting model parameters. The parameters search was done for interaction depths of 1, 5 and 10, and for the number of trees from 10 to 200 tress with a step of 10. Table 4 presents the GB results for our distance-based MP score and the traditional PMT across the four years. The table reveals the covariates that were the top predictors of poverty and the importance of each in explaining the prediction of the models. Three key findings are worth noting. First, the key predictors of poverty vary considerably when comparing the results for both approaches. For example, for the model that predicted expenditure only, we observed that the most important predictors indicated by fitting the gradient boosting were predominately from the survey covariates. Some geographical covariates were identified as important as well. For instance, the fraction area covered with crops and the NDVI, which are proxies for the level of agriculture activities within an administrative unit, were found to be important predictors. Similarly, the mean and standard deviation of land elevation, which are proxies for geographic accessibility and nighttime light intensity and are well-documented proxies for economic performance, were also identified as important predictors.

Second, when looking at the predictors of the distance-based MP score across time, we see that they vary considerably across years, with the top predictors tending to be related to the geospatial indicators, including

[14] The results for the Lasso and Random Forest models were similar and are available upon request.

governorate 7 which was North Lebanon. Interestingly, for the traditional PMT, the top predictors also varied considerably across time, but they tended to be more related to the socio-demographic factors.

Third, when it comes to predicting poverty, location clearly matters, whether as covariates at the governorate level or geospatial attributes at the district level. There are place-based elements to poverty that previous research has not been able to adequately capture. For example, when the model predicted the distance from the poorest, the covariate importance list featured a number of dummy variables that were used to encode the geographic location at the district and governorate levels. These results lead us to believe that when predicting a more complex definition of poverty that is not solely expenditure-based, the model clearly identifies the spatial correlation between the observations at different spatial scales (e.g., governorate, district, etc.) and benefits from the dummy geographic variables to account for such spatial effects indirectly.

Combined, these three findings point to the fact that the amount of heterogeneity between the two models and across time varies significantly, and hint to an even greater need for humanitarian organizations to be regularly updating their definition of poverty and the algorithms used to predict it for the entire refugee population.

[INSERT TABLE 4]

Figure A2 in the Appendix presents the distribution of the real and predicted values for the distance-based MP score by year and governorate. For the most part, we see that the distributions tend to be fairly similar over time. Figure A3 in the Appendix presents the distributions for each of the three dimensions in the distance-based MP score. We see that the distributions for expenditures per capita and the food consumption look similar across the years. However, some anomalies for reduced coping strategies (rCSI) can be observed. In particular, we see that for governorate 7 (North Lebanon) a larger share of the distribution is using reduced food coping strategies compared to the other governorates. This indicates higher levels of food insecurity for this specific governorate, which could be surprising at first sight, given that other governorates are known for having higher poverty rates among both host and refugee communities (i.e., Bekaa, Baalbek-El Hermel, and Akkar). However, a closer look into the heterogeneities across Lebanon's geographic locations and the ensuing economic dynamics help to provide a plausible explanation. In fact, despite being among the poorest places, these governorates include large agricultural areas unlike North Lebanon. Moreover, refugees tend to have employment opportunities in the agriculture sector which make them less prone to food insecurity. On the other hand, refugees in the North Lebanon governorate, which is mostly composed of urban areas and mountainous regions, face more serious challenges in accessing food. This likely also explains why governorate 7 was found to be a key predictor of the distance-based MP score in the GB model.

4. Results: A Comparison of the Traditional PMT with the Distance-based MP score

We used a few different methods to compare the PMT results to those for our distance-based MP score. First, we compared the overlap between refugees who were identified as the poorest households using our distance-based method to those who were identified as the poorest using the traditional PMT. For this purpose, we took the 10th, 20th, 30th, 40th, and 50th percentiles for both methods, and identified among the households who belonged to the X^{th} percentile in the distance function, the percentage that was also classified in the X^{th} percentile for expenditure. Our results showed that depending on the metric chosen for measuring poverty, the households that were classified as the poorest ones differed significantly (compare the overlap between the two measures for both the real and predicted values in Table 5). Similarly, there was high volatility in the overlap by year. For example, for 2019 and 2021, the households that were predicted to be the poorest ones (10th percentile) were completely different. In fact, there appears to be little, if any, overlap between those who were identified as the poorest using the traditional PMT versus the distance-based MP method. This result is unfortunate, as a correct targeting scheme should be able to at least identify accurately the extremes of the distribution: the households who are worse and better off. We would expect that differences between the two approaches to emerge at the middle of the distribution, where it is not clear which households are worse off. Additionally, we observed that food insecurity (measured through FCS and rCSI) was not necessarily correlated with expenditures. When comparing the real values reported in Table 5, we found that for the bottom 10th percentile, the overlap between the two approaches for 2019, 2020, and 2021 was less than 30%.

[INSERT TABLE 5]

Second, we analyzed whether our distance-based MP score was correlated with other measures of poverty. We classified as poor those households in the 30th percentile for both the PMT and distance-based methods and calculated the proportion of those households who were also deprived in different dimensions of poverty or social welfare. For instance, in the first row of Table 6, we estimated the percentage of households in the 30th percentile using both methods who had: (1) expenditures below the Survival Minimum Expenditure Basket (SMEB), (2) an rCSI score that was equal to or higher than 19 (or Phase 2), (3) an FCS below acceptable (lower than 42), or (4) who were deprived in all three dimensions at the same time. As a comparison point, we calculated the percentage of the total refugee population that was deprived in that particular dimension of poverty or social welfare (the “Total” columns in Table 6). Table 6 shows that the predictions using the PMT method did not classify correctly the households deprived in rCSI and FCS. In fact, the percentage of households deprived in these two characteristics, who were classified as poor according to expenditure (bottom 30% of the distribution), was lower than the average for the whole population for almost all years.

To complement this analysis, we also included deprivations related to living standard (e.g., having children of school age who were not attending school, cooking only with dung or charcoal, not having electricity for more than 16 hours per day, having less than 4.5m^2 per person, not having access to adequate sanitation, not having access to drinking water, and facing insecurity issues (robbery, kidnapping, harassment, etc.)). Contrary to our previous results, the PMT-based method was better than the distance-based MP method at predicting most of the deprivations for the living standards. However, the differences between the two approaches were generally small in most cases. Also, the percentage of households deprived in the living standards, who were also classified as poor, was sometimes not significantly different of that for the population as a whole. This result suggests there is room to improve the distance-based MP approach, by incorporating other dimensions of poverty. Nevertheless, it still captures other dimensions of poverty (i.e., food security and reduced coping strategies) that are not usually accounted for in the traditional PMT method that only uses expenditures per capita.

[INSERT TABLE 6]

Third, we extended our analysis to compare the characteristics of refugee households predicted to be poor using both the distance-based MP score and the traditional PMT. In Table 7, we split the sample into 4 groups: those that were classified as poor in both expenditures and multidimensional poverty (Both), those who were not classified as poor by either method (None), and those who were classified as poor according to only one of the two methods. For this comparison, we expanded the classification of poverty to the 40th percentile to increase the sample size of the overlapping group that was classified as poor using both methods. We found significant differences in some key variables. For example, household size and dependency ratio were important predictors of poverty in terms of the PMT, but their relevance decreased when the MP score was used. Perhaps this difference could by accounted for by introducing economies of scale in the calculation of expenditures per capita – for example, the marginal cost of food decreases with the number of household members, particularly with children. We also found that households classified only as poor using the MP score were more likely to be female headed or have a disabled head, although the differences compared to those who were only classified as poor using the PMT were small.

Interestingly, the main differences in the characteristics between the two methods were related to the location of the households and whether they were receiving humanitarian assistance (cash for food and multi-purpose cash). Households located in Baalbek-El Hermel and Bekaa were highly likely to be classified as poor using the traditional PMT, but not when the MP score was used, whereas the opposite was true for North Lebanon. Additionally, households located in Mount Lebanon were usually not classified as poor according to the PMT, but a significant percentage of the households in this governorate were classified as poor according to the distance-based MP score. The latter result is of particular interest, as the cost of living in Mount Lebanon is relatively higher compared to the rest of the country, and consequently, households require higher levels of expenditure. However, households in Mount Lebanon can experience other forms of poverty, such as food insecurity. The proportion of households who were receiving assistance (both MPC and Cash for food) was higher in the group that was only classified as poor in expenditure only than the group classified as poor according to both measures. In fact, the number of households who were receiving assistance and were classified as poor using the MP score only was particularly low and almost as low as the households who were not classified as poor by either of the two measures. These findings, again, hint to the importance of the geospatial component of poverty.

[INSERT TABLE 7]

Finally, we compared the exclusion errors using the predicted values for the distance-based MP score and the traditional PMT. To do this, we estimated the exclusion errors (households who were predicted to be non-poor when in reality they were actually poor) for different percentiles using both methods. The upper panel in Table 8A presents the exclusion errors using our distance-based MP score, while the lower panel in Table 8A presents the exclusion errors using the PMT. As expected, when the metric for poverty was based on expenditures only, the predictions for the standard PMT resulted in lower exclusion errors; and when the metric was based on multidimensional poverty, the predictions using the distance-based MP score proved to do better. However, when we compared the best predictors (see the numbers in italics), the exclusion errors were relatively similar for both models. In this sense, our methodology behaves similarly to the standard PMT regarding exclusion errors when the metric for defining poverty includes other dimensions of poverty apart from expenditure.

[INSERT TABLE 8A]

It is important to point out that the predictions presented in Table 8A assigned equal weights to the three dimensions included in the distance-based MP model. As one would expect, the traditional PMT assigned 100% of the weight to expenditure. For comparison purposes, and as a robustness check, we wanted to see how the exclusion errors might change if more weight was placed on expenditures in the MP model. To this end, we re-estimated our models assigning 50% weight to expenditure and 25% each to the food consumption score (FCS) and the reduced coping strategies index (rCSI). The results are presented in Table 8B. In comparing the findings for both Tables 8A and 8B, we found that the exclusion errors do, in fact, vary depending on: (1) how poverty is measured, (2) how that measure of poverty changes over time, and (3) how the weights are assigned to the different dimensions of poverty. As these assumptions change, exclusion errors can also change, with some assumptions resulting in less volatility over time and reducing the errors by more than others.[15]

[INSERT TABLE 8B]

[15] Currently, we are working to run additional robustness checks to check the sensitivity of the exclusion and inclusion errors to various weighting schemes. These results and related discussion will be included in the revised paper.

5. Conclusions

This paper uses data from Syrian refugees in Lebanon to construct a convenient multidimensional poverty score using the weighted average of the distances to the poorest profiles across three dimensions: expenditures, the food consumption score, and the reduced coping strategies index. Models were estimated, first, by assigning equal weights to each dimension and then by assigning 50%, 25% and 25% weights to the dimensions, respectively. The results were compared with those generated by the PMT method used traditionally by humanitarian agencies to predict households' consumption expenditures. Essentially, the latter represents the results obtained if we were to assume a 100% weight for expenditure. Such comparisons underscore the sensitivity of classifications to the definition of poverty. As such, adopting unidimensional definitions and measures of poverty versus multidimensional ones is a key decision that shapes any targeting strategy and determines to what extent it can reduce the exclusion and inclusion errors.

From this standpoint, our approach is more flexible as it allows researchers and organizations to include multiple poverty dimensions, and thus capture heterogeneities that the PMT formula focused only on expenditure cannot capture. In this study, we identified what profiles of households tend to be excluded from receiving assistance based on the scores generated by the traditional PMT method used by humanitarian organizations compared to our proposed multidimensional approach. Further, we examined the results across multiple years, which reflect not only the time element, but also the changing socio-economic conditions, given that the time frame covered by our analysis includes two years before major shocks (COVID and Lebanon's compounded crises) and two years post-shocks. This highlights the consistency of our methodology over time and across systemic events. Our approach is particularly less volatile at predicting which households are more in need of assistance in periods of hyperinflation. These findings have important implications for government and international agencies seeking to develop robust and effective targeting mechanisms, in a world where it has become increasingly hard to predict how things will be changing. In such a context, any targeting strategy needs to be designed to be flexible enough to allow for regular changes and updates.

The findings from this study also emphasize the importance of the geographical element. They suggest that geographical heterogeneities need to be considered in the construction of any future algorithms that classify who is poor and thus eligible for assistance. That said, a better understanding of the geographical aspects and their relationship with the inclusion and exclusion errors is still needed. There are place-based elements to poverty that this study and previous research have yet to further explore.

In addition, a few limitations of our work need to be acknowledged. Our models and predictions are only as good as the available data. We do not have access to longitudinal data to track the refugees over time, which would allow us to better capture how poverty is changing over time and within and across refugee households. Also, for security purposes, we do not have access to the geo-coordinates for the households. At present, we only know where the refugees are located at the governorate and district levels within Lebanon. This makes it difficult for us to conduct more rigorous geospatial analysis. We also do not have access to the administrative data or specific algorithms that are used by the humanitarian organizations to generate the PMT scores using traditional methods. These data restrictions limit our ability to test and refine our models so that the predictions are more accurate and consistent with internal methods currently being used by the humanitarian organizations to rank the refugee households and to identify priority needs in the face of limited resources.

Regardless, we have access to several years of data and can link these data to humanitarian assistance received, the actual PMT scores generated internally by UNCHR, as well as a robust set of geospatial attributes at the district level. Further, we are among the first to apply a data-science based approach to developing and testing a more comprehensive, yet flexible, targeting mechanism that goes beyond expenditure and can be adapted to various definitions and priorities associated with poverty. In this respect, we view our paper as a "road map" that can be used by other researchers and humanitarian organizations to more rigorously design, test, and update current targeting methods. As we mentioned at the beginning of the paper, this work is particularly timely given the current Russo-Ukrainian crisis, as our findings have general applicability to other crises situations, where humanitarian organizations need to make critical decisions on how to allocate limited resources among forcibly displaced populations worldwide.

References

Aaberge, R., & Brandolini, A. (2015). Multidimensional poverty and inequality. In *Handbook of income distribution* (Vol. 2, pp. 141-216). Elsevier.

Abdul Rahman, M., Sani, N. S., Hamdan, R., Ali Othman, Z., & Abu Bakar, A. (2021). A clustering approach to identify multidimensional poverty indicators for the bottom 40 percent group. *Plos One*, *16*(8), e0255312.

Aiken, E., Bellue, S., Karlan, D., Udry, C. R., & Blumenstock, J. (2021). *Machine learning and mobile phone data can improve the targeting of humanitarian assistance* (No. w29070). National Bureau of Economic Research. https://www.nber.org/papers/w29070

Aiken, E. L., Bedoya, G., Blumenstock, J. E., & Coville, A. Program Targeting with Machine Learning and Mobile Phone Data: Evidence from an Anti-Poverty Intervention in Afghanistan. https://jblumenstock.com/files/papers/jblumenstock_ultra-poor.pdf

Alkire, S., & Foster, J. (2011). Counting and multidimensional poverty measurement. *Journal of Public Economics*, *95*(7-8), 476-487.

Alkire, S., & Santos, M. E. (2014). Measuring acute poverty in the developing world: Robustness and scope of the multidimensional poverty index. *World Development*, *59*, 251-274.

Altındağ, O., & O'Connell, S. D. (2020). Unconditional cash-based assistance to the poor: What do at-scale programs achieve? *Available at SSRN 3719946*.

Altındağ, O., O'Connell, S. D., Şaşmaz, A., Balcıoğlu, Z., Cadoni, P., Jerneck, M., & Foong, A. K. (2021). Targeting humanitarian aid using administrative data: model design and validation. *Journal of Development Economics*, *148*, 102564.

Bienvenido-Huertas, D., Pulido-Arcas, J. A., Rubio-Bellido, C., & Pérez-Fargallo, A. (2021). Prediction of Fuel Poverty Potential Risk Index Using Six Regression Algorithms: A Case-Study of Chilean Social Dwellings. *Sustainability*, *13*(5), 2426.

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

Chaaban, J., Ghattas, H., Irani, A., & Thomas, A. (2018). Targeting mechanisms for cash transfers using regional aggregates. *Food Security, 10*(2), 457-472.

Chaaban, J., Salti, N., Ghattas, H., Moussa, W., Irani, A., Jamaluddine, Z., & Al-Mokdad, R. (2020). Multi-purpose cash assistance in Lebanon: Impact evaluation on the well-being of Syrian refugees. American University of Beirut Press. https://reliefweb.int/sites/reliefweb.int/files/resources/77377.pdf

Chi, G., Fang, H., Chatterjee, S., & Blumenstock, J. E. (2021). Micro-Estimates of Wealth for all Low-and Middle-Income Countries. *arXiv preprint arXiv:2104.07761*.

Coromaldi, M., & Drago, C. (2017). An analysis of multidimensional poverty: Evidence from Italy. In R. White (Ed.), *Measuring Multidimensional Poverty and Deprivation* (pp. 69-86). Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-319-58368-6_4

Deutsch, J., & Silber, J. (2005). Measuring multidimensional poverty: An empirical comparison of various approaches. *Review of Income and wealth*, *51*(1), 145-174.

Egger, D., Haushofer, J., Miguel, E., Niehaus, P., & Walker, M. W. (2019). *General equilibrium effects of cash transfers: Experimental evidence from Kenya* (No. w26600). National Bureau of Economic Research.

Ferreira, F. H., & Lugo, M. A. (2013). Multidimensional poverty analysis: Looking for a middle ground. *The World Bank Research Observer*, *28*(2), 220-235.

Government of Lebanon & United Nations. (2020). *Lebanon crisis response plan: 2017-2020 (2020 update)*. https://reliefweb.int/sites/reliefweb.int/files/resources/74641.pdf

InfoMigrants. (2021, March 16). UNHCR: 5.5 million Syrian refugees, 70% in poverty. https://www.infomigrants.net/en/post/30887/unhcr-55-million-syrian-refugees-70-in-poverty

Inter-Agency Coordination Lebanon. (2021, May 31). Inter-Agency Lebanon Crisis Response Plan (LCRP) Situation Update: Operational Environment in Lebanon (January-May 2021). https://data2.unhcr.org/en/documents/details/87896

Kidd, S., Gelders, B., & Bailey-Athias, D. (2017). Exclusion by design. *Extension of Social Security (ESS) Working Paper No. 56*. Geneva, Switzerland: International Labour Organization and Development Pathways. https://www.alnap.org/system/files/content/resource/files/main/Exlcusion%20by%20design.pdf

Kidd, S., & Wylde, E. (2011). Targeting the poorest: An assessment of the proxy means test methodology. *AusAID Research Paper*. Canberra, Australia: Australian Agency for International Development.

Lyons, A. C., Kass-Hanna, J., & Molena, E. (2021). A multidimensional approach to poverty that strengthens the humanitarian-development nexus. *T20 Italy Policy Brief*. Prepared for 2021 G20 Summit by T20 Italy, Task Force 5: 2030 Agenda and Development Cooperation. https://www.t20italy.org/2021/09/21/a-multidimensional-approach-to-poverty-that-strengthens-the-humanitarian-development-nexus/

Lyons, A. C., Kass-Hanna, J., & Montoya Castano, A. (2021). A multidimensional approach to measuring vulnerability to poverty of Syrian refugees in Lebanon. *Economic Research Forum's Working Paper Series No. 1472*. Cairo, Egypt: Economic Research Forum.

Mills, B., del Ninno, C., & Leite, P. (2015). Effective targeting mechanisms in Africa: Existing and new methods. In C. del Ninno & B. Mills (Eds.), *Safety Nets in Africa: Effective Mechanisms to Reach the Poor and Most Vulnerable*, 19-37.

Morel, R., & Chowdhury, R. (2015). Reaching the ultra-poor: Adapting targeting strategy in the context of South Sudan. *Journal of International Development*, *27*(7), 987-1011.

Moussa, W., Irani, A., Salti, N., Al Mokdad, R., Jamaluddine, Z., Chaaban, J., & Ghattas, H. (2021, February). The Impact of Cash Transfers on Syrian Refugee Children in Lebanon. In *Economic Research Forum's Working Paper Series, No. 1457*. Cairo, Egypt: Economic Research Forum.

Otoiu, A., Titan, E., & Dumitrescu, R. (2014). Are the variables used in building composite indicators of well-being relevant? Validating composite indexes of well-being. *Ecological indicators*, *46*, 575-585.

Ramos, X. (2008). Using efficiency analysis to measure individual well-being with an illustration for Catalonia. In Quantitative approaches to multidimensional poverty measurement (pp. 155-175). Palgrave Macmillan, London.

Ravallion, M. (2011). On multidimensional indices of poverty. *The Journal of Economic Inequality*, *9*(2), 235-248.

Sani, N. S., Rahman, M. A., Bakar, A. A., Sahran, S., & Sarim, H. M. (2018). Machine learning approach for bottom 40 percent households (B40) poverty classification. *International Journal on Advanced Science, Engineering and Information Technology*, *8*(4-2), 1698.

Schnitzer, P. (2019). How to target households in adaptive social protection systems? Evidence from humanitarian and development approaches in Niger. *The Journal of Development Studies*, *55*(sup1), 75-90.

United Nations High Commissioner for Refugees (UNHCR). (n.d.). Operational data portal: Ukraine refugee situation. https://data2.unhcr.org/en/situations/ukraine

United Nations High Commissioner for Refugees (UNHCR). (2022). Refugee data finder. https://www.unhcr.org/refugee-statistics/

United Nations High Commissioner for Refugees (UNHCR). (2022). Refugee statistics: Global trends at-a-glance. https://www.unrefugees.org/refugee-facts/statistics/

United Nations High Commissioner for Refugees (UNHCR), the United Nations World Food Programme (WFP), & the United Nations Children's Fund (UNICEF). (2018). *VASyR 2018: Vulnerability assessment of Syrian refugees in Lebanon*. Beirut, Lebanon. Retrieved from https://www.unhcr.org/lb/wp-content/uploads/sites/16/2018/12/VASyR-2018.pdf

United Nations High Commissioner for Refugees (UNHCR), the United Nations World Food Programme (WFP), & the United Nations Children's Fund (UNICEF). (2019). *VASyR 2019: Vulnerability assessment of Syrian refugees in Lebanon*. Beirut, Lebanon. https://data.unhcr.org/en/documents/details/73118

United Nations High Commissioner for Refugees (UNHCR), the United Nations World Food Programme (WFP), & the United Nations Children's Fund (UNICEF). (2021). *VASyR 2020: Vulnerability assessment of Syrian refugees in Lebanon*. Beirut, Lebanon. https://data.unhcr.org/en/documents/details/85002

United Nations High Commissioner for Refugees (UNHCR), the United Nations World Food Programme (WFP), & the United Nations Children's Fund (UNICEF). (2022). *VASyR 2021: Vulnerability assessment of Syrian refugees in Lebanon*. Beirut, Lebanon. https://data.unhcr.org/en/documents/details/90589

Usmanova, A., Aziz, A., Rakhmonov, D., & Osamy, W. (2022). Utilities of Artificial Intelligence in Poverty Prediction: A Review. *Sustainability*, *14*(21), 14238.

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

Vyas, S., & Kumaranayake, L. (2006). Constructing socio-economic status indices: how to use principal components analysis. *Health policy and planning*, *21*(6), 459-468.

World Bank, UNDP, & UNICEF. (2021). A roadmap for countries measuring multidimensional poverty. *Equitable Growth, Finance and Institutions Insights*. World Bank: Washington, DC. https://openknowledge.worldbank.org/handle/10986/35808

Yoder Clark, A., Blumenfeld, N., Lal, E., Darbari, S., Northwood, S., & Wadpey, A. (2021). Using K-Means Cluster Analysis and Decision Trees to Highlight Significant Factors Leading to Homelessness. *Mathematics*, *9*(17), 2045.

Table 1. Variable definitions

Variables	Definitions
Main variables	
Exp per capita	Expenditure per capita in Lebanese Pounds (LBP)
rCSI	Reduced food coping strategies index. Measures the strategies that households use to cope with the lack of food and the severity of these strategies to compare the hardship faced by households due to shortage of food. Ranges from 0 (no coping strategies) to 56 (severe).
FCS	The Food Consumption Score measures the diversity and frequency of households' diets in the week prior to the survey. The index ranges from 0 to 112.
Covariates	
Household size	Number of household members
Household size squared	
Dependency ratio	Ratio of dependent household members (aged below 15 or above 60) relative to total household members.
Female Head	=1 if female headed household
Frac. of HH members aged 0-4	Percentage of children aged 0 to 4 in each household
Frac. of HH members aged 5-9	Percentage of children aged 5 to 9 in each household
Frac. of HH members aged 10-19	Percentage of household members aged 10 to 19 in each household
Frac. of male members aged 20-49	Percentage of male adults aged 20 to 49 in the household
Frac. of female members aged 20-49	Percentage of female adults aged 20 to 49 in the household
Frac. of members older than 60	Percentage of household members aged 60 and above
Frac. of HH members education unknown	Percentage of household members who do not report any educational level
Frac. of HH members no education	Percentage of household members who did not go to school
Frac. of HH members some education below primary	Percentage of household members who did not complete primary education
Frac. of HH members with primary education	Percentage of household members who completed primary education
Frac. of HH members secondary education	Percentage of household members who completed secondary education
Frac. of HH members above secondary education	Percentage of household members with high school, technical, or college diploma
Frac. of HH members working	Percentage of household members who are working
Frac. of HH members unemployed	Percentage of household members who are unemployed
Frac. of HH members inactive	Percentage of household members who are inactive
Frac. of HH members studying	Percentage of household members who are receiving education online or going to the school/university or both
Frac. of HH members with a disability	Percentage of household members with any disability (seeing, hearing, walking, etc.)
Frac. of HH members with a medical condition	Percentage of household members with a chronic illness or unable to care for themselves
Disabled Head	=1 if the head has a disability
Existence of a disabled dependent member	=1 if at least one member of the household other than the head has a disability
Single Parent	=1 if the household head is a single parent
Frac. Illegal residency	Percentage of household members aged 15 or older who do not have legal residency in Lebanon
Received MPC	=1 if the household received multi-purpose cash in the 6 months prior to the survey
Received WFP cash for food	=1 if the household received cash for food in the 6 months prior to the survey
Received any other monetary assistance	=1 if the household received any other cash assistance in the 6 months prior to the survey
Governorate	Fixed effects for the 8 governorates in Lebanon
District	Fixed effects for the 26 districts in Lebanon
Other variables	
Survival Minimum Basket - SMEB	=1 if the monthly expenditures per capita fell below the survival minimum basket cutoff. This cutoff varies by year. For 2018 to 2019 it was equivalent to 87 USD; for 2020 308,722 LBP, for 2021 490,028 LBP.
rCSI Phase 2	=1 if the rCSI is higher or equal than 19
FCS not acceptable	=1 if FCS is below 42
Not attending school	=1 if household has a child in school age (5 to 14) who is not attending school
Cooking fuel	=1 if household did not have access to electric or gas stove and cooks only with dung, wood, or charcoal.
Electricity	=1 household does not have access to electricity or has access for less than 16 hours

Improved sanitation	=1 if household did not have access to basic sanitation (i.e., no access to flushed toilets or improved pit latrines with a cement slab, and was not sharing the toilets with other households).
Crowdedness of shelter	=1 if household was living in an overcrowded shelter with less than 4.5m^2 per person
Water	=1 if household did not have access to clean drinking water
Insecurity	=1 if a member of the household experienced any form of insecurity (robbery, extortion, harassment, kidnapping, etc.)
Geographic variables	
Dem mean	Mean district elevation
Dem std	Standard deviation of elevations in the district
Built_cf_count	Fraction cover of built-up area in the district in years 2018 and 2019
Built_cf_mean	(Average fraction cover and count number of pixels)
Crop_cf_count	Fraction cover of crop covered area in the district in years 2018 and 2019
Crop_cf_mean	(Average fraction cover and count number of pixels)
Perm_Water_cf_count	Fraction cover of permanent water area in the district in years 2018 and 2019
Perm_Water_cf_mean	(Average fraction cover and count number of pixels)
Seas_Water_cf_count	Fraction cover of seasonal water area in the district in years 2018 and 2019
Seas_Water_cf_mean	(Average fraction cover and count number of pixels)
Snow_cf_count	Fraction cover of snow area in the district in years 2018 and 2019
Snow_cf_mean	(Average fraction cover and count number of pixels)
Pop_sum	Sum of the total population per district in years 2018, 2019, and 202 based on the population counts taken from the WorldPop adjusted to match the UN estimation count.
Night_Lights_count	Nighttime lights in the district for the years 2018, 2019, 2020, 2021
Night_Lights_mean	
NDVI_count	Healthy vegetation (Agriculture)
NDVI_mean	

Sources: 2018, 2019, 2020, 2021 Vulnerability Assessment of Syrian Refugees (VASyR).

Table 2. Descriptive statistics by year

Variables	2018 (N=4,141)	2019 (N=4,377)	2020 (N=4,288)	2021 (N=4,836)	p-value
Expenditure (LBP)	157,519 (115,380)	148,831 (106,531)	196,960 (152,471)	341,497 (260,928)	0.000
rCSI	17.8 (14.3)	18.5 (15.0)	17.0 (13.5)	19.8 (14.2)	<0.001
FCS	53.3 (20.2)	55.3 (18.9)	45.3 (18.3)	48.5 (18.2)	<0.001
Household size	4.93 (2.23)	5.13 (2.41)	5.05 (2.19)	5.04 (2.18)	0.001
Dependency ratio	0.46 (0.23)	0.45 (0.24)	0.46 (0.24)	0.46 (0.23)	0.728
Female Head	0.16 (0.37)	0.16 (0.37)	0.17 (0.37)	0.16 (0.36)	0.394
Frac. of HH members aged 0-4	0.17 (0.18)	0.17 (0.18)	0.18 (0.19)	0.18 (0.18)	0.058
Frac. of HH members aged 5-9	0.15 (0.17)	0.15 (0.16)	0.14 (0.16)	0.14 (0.16)	0.005
Frac. of HH members aged 10-19	0.18 (0.21)	0.18 (0.21)	0.19 (0.22)	0.19 (0.21)	0.240
Frac. of members older than 60	0.03 (0.13)	0.03 (0.13)	0.03 (0.13)	0.03 (0.13)	0.656
Frac. of male members aged 20-49	0.21 (0.19)	0.21 (0.20)	0.21 (0.20)	0.22 (0.20)	0.202
Frac. of female members aged 20-49	0.21 (0.15)	0.18 (0.14)	0.20 (0.15)	0.20 (0.14)	<0.001
Frac. of HH members education Unknown	0.36 (0.25)	0.24 (0.22)	0.25 (0.23)	0.11 (0.14)	0.000
Frac. of HH members no education	0.12 (0.21)	0.07 (0.16)	0.07 (0.17)	0.37 (0.33)	0.000
Frac. of HH members some education below primary	0.15 (0.21)	0.29 (0.26)	0.28 (0.26)	0.22 (0.24)	<0.001
Frac. of HH members secondary education	0.09 (0.17)	0.09 (0.17)	0.11 (0.19)	0.09 (0.17)	<0.001
Frac. of HH members above secondary education	0.06 (0.16)	0.07 (0.18)	0.06 (0.16)	0.05 (0.15)	<0.001
Frac. of HH members inactive	0.09 (0.15)	0.34 (0.24)	0.30 (0.23)	0.25 (0.21)	0.000
Frac. of HH members studying	0.01 (0.06)	0.02 (0.07)	0.02 (0.07)	0.00 (0.03)	<0.001
Frac. of HH members working	0.17 (0.20)	0.16 (0.21)	0.16 (0.21)	0.20 (0.21)	<0.001
Frac. of HH members unemployed	0.10 (0.19)	0.07 (0.16)	0.09 (0.17)	0.07 (0.15)	<0.001
Frac. of HH members with a disability	0.03 (0.09)	0.06 (0.15)	0.06 (0.15)	0.07 (0.16)	<0.001
Frac. of HH members with a medical condition	0.16 (0.23)	0.15 (0.22)	0.15 (0.23)	0.15 (0.23)	0.061
Disabled Head	0.04 (0.19)	0.08 (0.27)	0.10 (0.30)	0.11 (0.31)	<0.001
Existence of a disabled dependent member	0.06 (0.23)	0.10 (0.30)	0.11 (0.31)	0.10 (0.30)	<0.001
Single Parent	0.05 (0.22)	0.06 (0.23)	0.05 (0.22)	0.05 (0.23)	0.218
Frac. Illegal residency	0.66 (0.41)	0.71 (0.38)	0.63 (0.42)	0.67 (0.40)	<0.001
Governorate					<0.001
Governorate 1: Akkar	426 (10.3%)	474 (10.8%)	483 (11.3%)	522 (10.8%)	
Governorate 2: Baalbek-El Hermel	333 (8.04%)	434 (9.92%)	485 (11.3%)	485 (10.0%)	
Governorate 3: Beirut	377 (9.10%)	412 (9.41%)	322 (7.51%)	471 (9.74%)	
Governorate 4: Bekaa	502 (12.1%)	477 (10.9%)	480 (11.2%)	481 (9.95%)	
Governorate 5: El Nabatieh	554 (13.4%)	537 (12.3%)	627 (14.6%)	646 (13.4%)	
Governorate 6: Mount Lebanon	847 (20.5%)	874 (20.0%)	768 (17.9%)	1009 (20.9%)	
Governorate 7: North Lebanon	683 (16.5%)	702 (16.0%)	677 (15.8%)	759 (15.7%)	
Governorate 8: South Lebanon	419 (10.1%)	467 (10.7%)	446 (10.4%)	463 (9.57%)	
Received WFP cash for food	0.65 (0.48)	0.66 (0.47)	0.65 (0.48)	0.51 (0.50)	<0.001
Received MPC	0.47 (0.50)	0.49 (0.50)	0.50 (0.50)	0.50 (0.50)	0.033
Received any other monetary assistance	0.76 (0.43)	0.74 (0.44)	0.74 (0.44)	0.28 (0.45)	0.000
Not attending school	0.28 (0.45)	0.30 (0.46)	0.29 (0.45)	0.13 (0.34)	<0.001
Cooking fuel	0.06 (0.24)	0.13 (0.34)	0.18 (0.38)	0.14 (0.35)	<0.001
Electricity	0.40 (0.49)	0.27 (0.45)	0.38 (0.48)	0.39 (0.49)	<0.001
Crowdedness of shelter	0.33 (0.47)	0.29 (0.45)	0.23 (0.42)	0.20 (0.40)	<0.001
Improved sanitation	0.31 (0.46)	0.27 (0.44)	0.24 (0.43)	0.24 (0.43)	<0.001
Water	0.12 (0.32)	0.13 (0.34)	0.14 (0.35)	0.12 (0.33)	0.002
Insecurity	0.03 (0.18)	0.13 (0.34)	0.09 (0.29)	0.15 (0.36)	<0.001

Notes: Standard errors in parentheses.

Table 3. Comparison of machines learning results across models (Lasso, Random Forest, Gradient Boosting)

a. Distance-based multidimensional poverty score

Method	Stats	2018	2019	2020	2021
Random Forest	Abs. Error	0.258	0.237	0.224	0.221
	Correlation	0.438	0.502	0.433	0.508
	MSE	0.107	0.090	0.078	0.076
	RMSE	0.326	0.300	0.279	0.276
	R-squared	0.191	0.252	0.186	0.258
LASSO	Abs. Error	0.259	0.239	0.223	0.222
	Correlation	0.444	0.500	0.443	0.499
	MSE	0.106	0.090	0.077	0.077
	RMSE	0.326	0.301	0.277	0.278
	R-squared	0.196	0.247	0.196	0.248
Gradient Boosting	Abs. Error	0.258	0.239	0.223	0.221
	Correlation	0.451	0.497	0.446	0.505
	MSE	0.105	0.090	0.077	0.077
	RMSE	0.324	0.301	0.277	0.277
	R-squared	0.203	0.246	0.198	0.254
	BEST	GB	RF	GB	RF

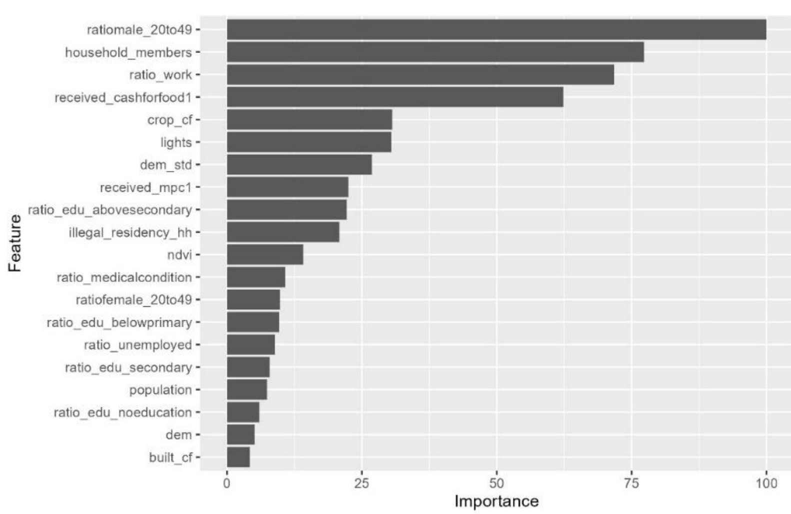
b. Expenditure-based traditional PMT

Method	Stats	2018	2019	2020	2021
Random Forest	Abs. Error	42.2	41.7	75853.8	125009.0
	Correlation	0.592	0.467	0.556	0.577
	MSE	3869.1	3465.4	1.63E+10	4.566E+10
	RMSE	62.2	58.9	127492.7	213680.6
	R-squared	0.349	0.213	0.309	0.332
LASSO	Abs. Error	43.3	41.8	80805.8	128165.3
	Correlation	0.594	0.463	0.529	0.572
	MSE	3852.0	3473.8	1.7E+10	4.608E+10
	RMSE	62.1	58.9	130531.7	214656.5
	R-squared	0.352	0.211	0.275	0.326
Gradient Boosting	Abs. Error	42.5	42.0	75902.1	124949.5
	Correlation	0.601	0.462	0.527	0.577
	MSE	3800.6	3479.6	1.7E+10	4.558E+10
	RMSE	61.6	59.0	130500.4	213504.5
	R-squared	0.361	0.210	0.276	0.333
	BEST	GB	RF	RF	GB

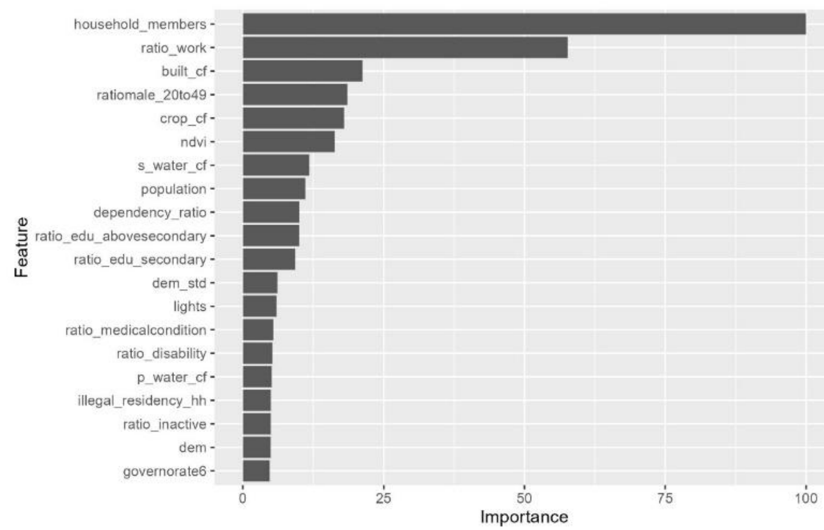
Table 4. Poverty predictors using gradient boosting for traditional PMT versus distance-based multidimensional score by year

Expenditure-based traditional PMT:

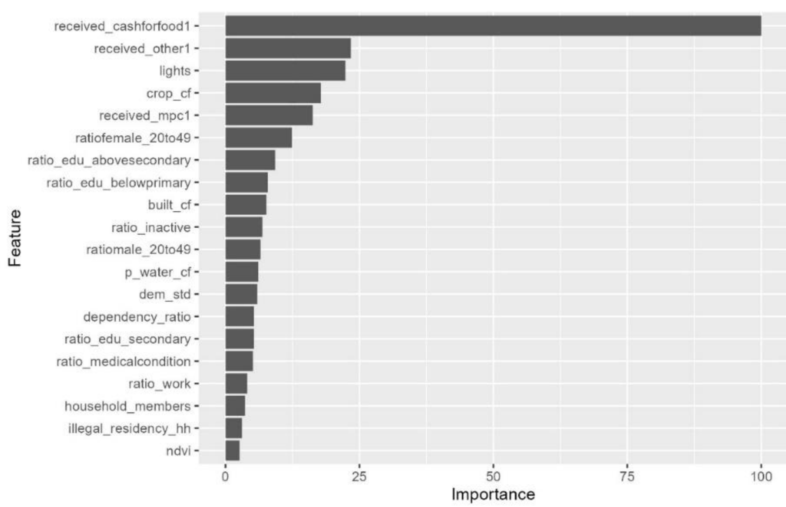
2018



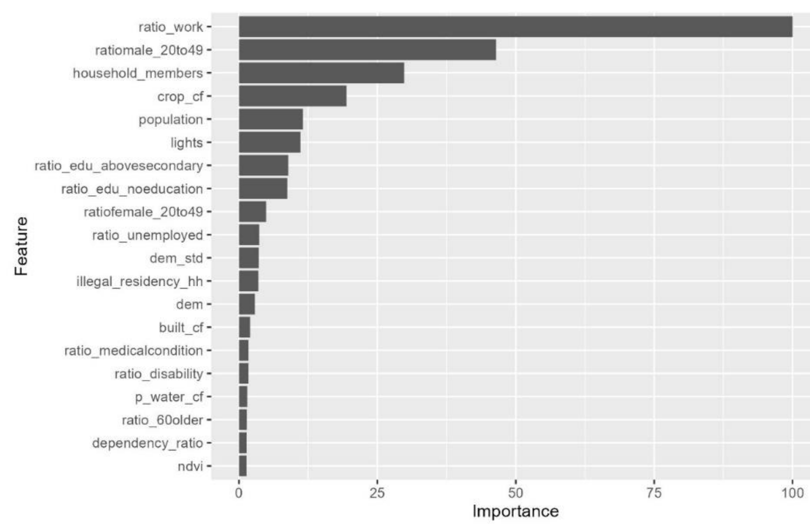
2020



2019

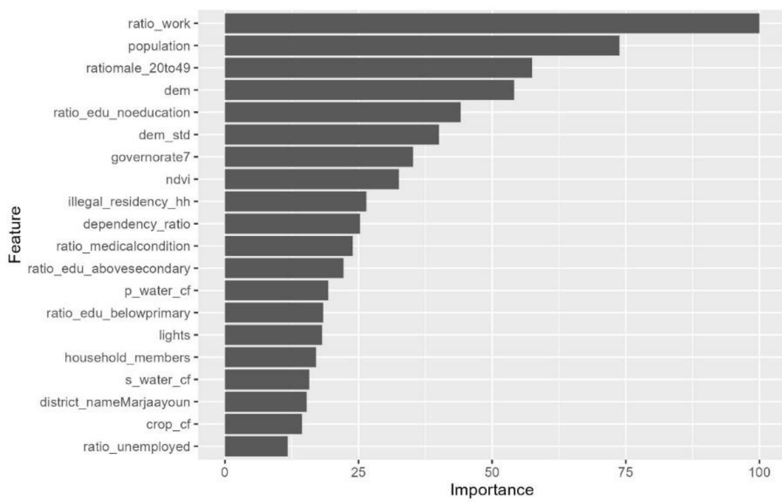


2021

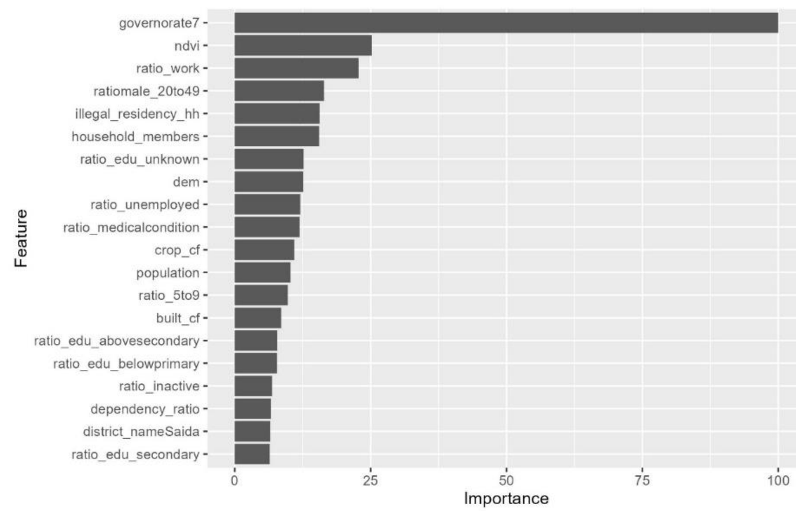


Distance-based multidimensional poverty score:

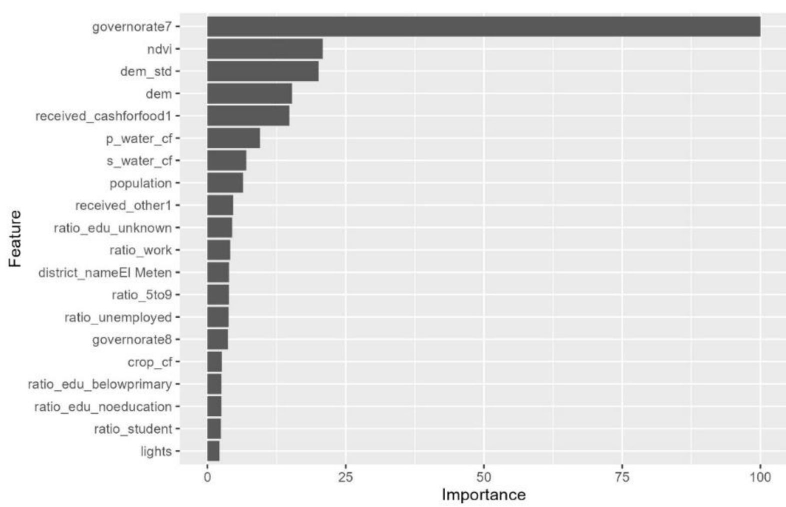
2018



2020



2019



2021

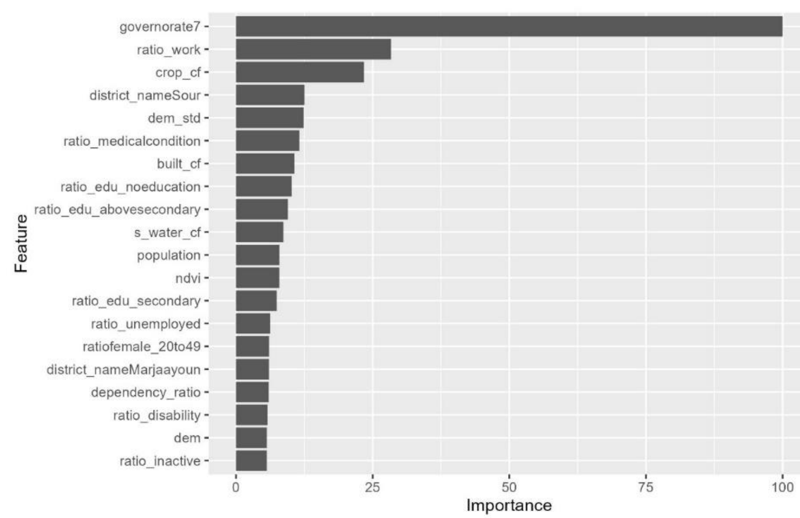


Table 5. Overlap between traditional PMT and distance-based multidimensional poverty scores

Real values

Year	Perc 10	Perc 20	Perc 30	Perc 40	Perc 50
2018	30.9%	41.5%	51.6%	58.6%	66.4%
2019	18.0%	31.5%	42.4%	51.9%	61.5%
2020	24.5%	39.0%	46.4%	54.7%	62.5%
2021	17.4%	29.9%	41.1%	51.0%	59.5%

Predicted values

Year	Perc 10	Perc 20	Perc 30	Perc 40	Perc 50
2018	17.9%	35.1%	49.1%	59.1%	67.3%
2019	0.0%	16.6%	38.7%	49.8%	65.5%
2020	45.2%	46.2%	47.9%	51.8%	58.9%
2021	0.0%	7.5%	22.3%	36.6%	49.5%

Notes: Estimations are based on a distance function that assigns equal weights to each dimension.

Table 6. Relative efficiency of the distance-based MP and traditional PMT methods at capturing other forms of deprivation

	SMEB			rCSI			FCS			All (SMEB, rCSI, FCS)		
Year	Dist	Exp	Total	Dist	Exp	Total	Dist	Exp	Total	Dist	Exp	Total
2018	65.6%	79.7%	51.0%	67.4%	41.1%	40.2%	40.1%	29.3%	32.1%	22.6%	16.1%	10.4%
2019	62.3%	80.4%	54.9%	73.3%	34.7%	41.8%	29.5%	17.8%	25.4%	14.5%	6.9%	7.2%
2020	92.6%	98.6%	87.6%	59.6%	37.6%	38.6%	67.0%	55.7%	47.6%	36.5%	23.5%	18.3%
2021	86.0%	97.5%	84.5%	77.9%	36.6%	46.6%	53.1%	38.5%	42.0%	35.8%	14.1%	18.2%

	School attendance			Cooking Fuel			Electricity			Shelter Crowdedness		
Year	Dist	Exp	Total	Dist	Exp	Total	Dist	Exp	Total	Dist	Exp	Total
2018	34.2%	42.5%	28.3%	6.4%	3.3%	6.1%	44.8%	38.6%	39.6%	35.1%	41.4%	32.6%
2019	29.6%	33.2%	29.8%	13.3%	16.4%	13.3%	30.4%	33.9%	27.5%	27.9%	36.6%	29.2%
2020	34.4%	43.1%	29.2%	21.5%	23.9%	17.6%	52.2%	46.4%	37.6%	22.5%	29.6%	23.5%
2021	16.2%	18.5%	13.0%	12.7%	17.5%	13.9%	44.6%	41.5%	38.8%	18.9%	29.6%	20.2%

	Sanitation			Water			Security		
Year	Dist	Exp	Total	Dist	Exp	Total	Dist	Exp	Total
2018	40.3%	39.0%	31.2%	13.7%	11.5%	11.8%	2.3%	1.7%	3.3%
2019	33.4%	34.2%	27.1%	17.5%	10.5%	13.0%	10.1%	12.2%	13.2%
2020	24.8%	23.5%	24.3%	19.7%	16.8%	14.3%	10.1%	7.8%	9.2%
2021	22.4%	29.8%	24.3%	14.6%	10.8%	12.1%	17.4%	12.1%	14.8%

Notes: Households were classified as poor if their scores were in the lowest 30% of the distribution for either the distance-based MP score or the traditional PMT. The column "Total" measures the proportion of the whole population of refugees that was deprived in that particular indicator of poverty or social welfare. Thus, it is expected that the poorest households (using both methods) have higher deprivation rates than the average population ("Total").

Table 7. Comparison of the characteristics of refugee households predicted to be poor according to the distance-based MP score and the traditional PMT.

Variables	Both (n=3,445)	None (n=6,976)	Only Distance (n=3,611)	Only Expenditure (n=3,610)	p-value
rCSI	24.0 (14.8)	14.6 (12.5)	26.5 (14.9)	12.1 (10.1)	0.000
FCS	47.7 (19.0)	52.7 (20.2)	45.7 (18.6)	53.8 (17.1)	<0.001
Household size	5.85 (2.04)	4.35 (2.17)	4.39 (1.91)	6.25 (2.15)	0.000
Dependency ratio	0.53 (0.20)	0.39 (0.25)	0.43 (0.23)	0.54 (0.20)	<0.001
Female Head	0.20 (0.40)	0.13 (0.34)	0.19 (0.39)	0.16 (0.36)	<0.001
Frac. of HH members aged 0-4	0.19 (0.18)	0.16 (0.19)	0.18 (0.19)	0.18 (0.17)	<0.001
Frac. of HH members aged 5-9	0.19 (0.17)	0.11 (0.15)	0.13 (0.17)	0.18 (0.15)	<0.001
Frac. of HH members aged 10-19	0.21 (0.21)	0.16 (0.21)	0.15 (0.21)	0.24 (0.21)	<0.001
Frac. of members older than 60	0.02 (0.09)	0.04 (0.15)	0.04 (0.15)	0.03 (0.10)	<0.001
Frac. of male members aged 20-49	0.15 (0.12)	0.27 (0.25)	0.21 (0.18)	0.15 (0.12)	<0.001
Frac. of female members aged 20-49	0.19 (0.11)	0.20 (0.16)	0.22 (0.17)	0.18 (0.10)	<0.001
Frac. of HH members education Unknown	0.28 (0.24)	0.21 (0.23)	0.22 (0.22)	0.24 (0.23)	<0.001
Frac. of HH members no education	0.18 (0.26)	0.13 (0.25)	0.18 (0.27)	0.22 (0.29)	<0.001
Frac. of HH members some education below primary	0.26 (0.24)	0.21 (0.25)	0.22 (0.25)	0.27 (0.24)	<0.001
Frac. of HH members secondary education	0.06 (0.12)	0.12 (0.21)	0.10 (0.18)	0.06 (0.12)	<0.001
Frac. of HH members above secondary education	0.03 (0.09)	0.09 (0.21)	0.06 (0.15)	0.03 (0.10)	<0.001
Frac. of HH members inactive	0.22 (0.21)	0.25 (0.24)	0.27 (0.25)	0.24 (0.20)	<0.001
Frac. of HH members studying	0.01 (0.05)	0.02 (0.07)	0.01 (0.05)	0.01 (0.06)	<0.001
Frac. of HH members working	0.10 (0.13)	0.25 (0.25)	0.18 (0.19)	0.11 (0.13)	0.000
Frac. of HH members unemployed	0.09 (0.15)	0.08 (0.19)	0.10 (0.19)	0.07 (0.11)	<0.001
Frac. of HH members with a disability	0.06 (0.12)	0.05 (0.15)	0.07 (0.17)	0.05 (0.11)	<0.001
Frac. of HH members with a medical condition	0.15 (0.19)	0.15 (0.25)	0.21 (0.26)	0.13 (0.18)	<0.001
Disabled Head	0.10 (0.31)	0.07 (0.25)	0.11 (0.31)	0.07 (0.26)	<0.001
Existence of a disabled dependent member	0.11 (0.31)	0.07 (0.26)	0.10 (0.30)	0.11 (0.31)	<0.001
Single Parent	0.02 (0.15)	0.09 (0.28)	0.05 (0.21)	0.02 (0.14)	<0.001
Frac. Illegal residency	0.72 (0.38)	0.62 (0.42)	0.68 (0.40)	0.69 (0.38)	<0.001
Governorate					0.000
Akkar	987 (28.7%)	230 (3.30%)	338 (9.36%)	350 (9.70%)	
Baalbek-El Hermel	206 (5.98%)	401 (5.75%)	11 (0.30%)	1119 (31.0%)	
Beirut	98 (2.84%)	1234 (17.7%)	209 (5.79%)	41 (1.14%)	
Bekaa	275 (7.98%)	518 (7.43%)	38 (1.05%)	1109 (30.7%)	
El Nabatieh	331 (9.61%)	1251 (17.9%)	365 (10.1%)	417 (11.6%)	
Mount Lebanon	355 (10.3%)	2127 (30.5%)	815 (22.6%)	201 (5.57%)	
North Lebanon	1026 (29.8%)	190 (2.72%)	1596 (44.2%)	9 (0.25%)	
South Lebanon	167 (4.85%)	1025 (14.7%)	239 (6.62%)	364 (10.1%)	
Received WFP cash for food	0.84 (0.36)	0.44 (0.50)	0.48 (0.50)	0.88 (0.33)	0.000
Received MPC	0.72 (0.45)	0.31 (0.46)	0.36 (0.48)	0.75 (0.44)	0.000
Received any other monetary assistance	0.76 (0.43)	0.54 (0.50)	0.57 (0.50)	0.68 (0.47)	<0.001
Not attending school	0.36 (0.48)	0.18 (0.38)	0.22 (0.41)	0.31 (0.46)	<0.001
Cooking fuel	0.15 (0.36)	0.10 (0.31)	0.12 (0.33)	0.16 (0.37)	<0.001
Electricity	0.43 (0.50)	0.31 (0.46)	0.41 (0.49)	0.35 (0.48)	<0.001
Crowdedness of shelter	0.32 (0.47)	0.22 (0.41)	0.21 (0.41)	0.34 (0.47)	<0.001
Improved sanitation	0.32 (0.47)	0.23 (0.42)	0.27 (0.44)	0.28 (0.45)	<0.001
Water	0.14 (0.35)	0.11 (0.32)	0.16 (0.37)	0.11 (0.31)	<0.001
Insecurity	0.09 (0.29)	0.11 (0.31)	0.12 (0.32)	0.10 (0.29)	0.05

Notes: This table compares the characteristics of the households that fall into four groups: both (meaning that they were classified as poor using both methods to predict poverty; none (meaning that they are not classified as poor according to the predictions of either method); and and those who were classified as poor according to only one of the two methods.

Table 8A. A comparison of exclusion errors using the predicted values for the distance-based MP score assuming equal weights versus and expenditure-based* traditional PMT methods

Exclusion errors (Distance-based MP score)						
	Distance-based MP score			Expenditure-based traditional PMT		
Year	Perc 30	Perc 40	Perc 50	Perc 30	Perc 40	Perc 50
2018	*18.4%*	*25.8%*	*33.6%*	27.2%	35.4%	43.2%
2019	*18.1%*	*25.2%*	*34.2%*	31.0%	39.3%	46.3%
2020	*20.1%*	*25.9%*	*32.0%*	27.3%	37.0%	45.9%
2021	*17.6%*	*24.6%*	*30.5%*	32.0%	41.2%	49.5%

Exclusion errors (Expenditure-based traditional PMT)						
	Distance-based MP score			Expenditure-based traditional PMT		
Year	Perc 30	Perc 40	Perc 50	Perc 30	Perc 40	Perc 50
2018	24.4%	31.2%	37.0%	*19.6%*	*25.4%*	*29.4%*
2019	28.5%	36.7%	43.2%	*21.1%*	*26.2%*	*32.3%*
2020	24.0%	33.5%	43.0%	*21.4%*	*26.7%*	*31.4%*
2021	30.0%	39.4%	47.5%	*21.7%*	*25.7%*	*31.1%*

* At its core, the expenditure-based traditional PMT assumes 100% weight to expenditure.

Table 8B. A comparison of exclusion errors using the predicted values for the distance-based MP score assuming unequal weights versus and traditional PMT methods

Exclusion errors (Distance-based MP score)						
	Distance-based MP score			Expenditure-based traditional PMT		
Year	Perc 30	Perc 40	Perc 50	Perc 30	Perc 40	Perc 50
2018	*19.2%*	*26.2%*	*33.1%*	25.9%	32.9%	39.6%
2019	*19.1%*	*27.3%*	*34.8%*	30.2%	36.7%	42.4%
2020	*20.2%*	*26.7%*	*32.6%*	26.0%	34.9%	43.8%
2021	*18.3%*	*24.1%*	*31.0%*	30.4%	38.8%	46.9%

Exclusion errors (Expenditure-based traditional PMT)						
	Distance-based MP score			Expenditure-based traditional PMT		
Year	Perc 30	Perc 40	Perc 50	Perc 30	Perc 40	Perc 50
2018	22.0%	28.8%	34.1%	*19.6%*	*25.4%*	*29.4%*
2019	28.4%	33.4%	39.6%	*21.1%*	*26.2%*	*32.3%*
2020	23.5%	31.3%	38.9%	*21.4%*	*26.7%*	*31.4%*
2021	28.4%	36.0%	44.4%	*21.7%*	*25.7%*	*31.2%*

*Note*s: The models were re-estimated by assigning 50% weights to expenditure and 25% to each of the food consumption score and the reduced coping strategies index. Recall that the expenditure-based traditional PMT assumes 100% weight to expenditure.

APPENDIX

Figure A1. Distribution of Euclidean, Manhattan, and Minkowski distance formulas

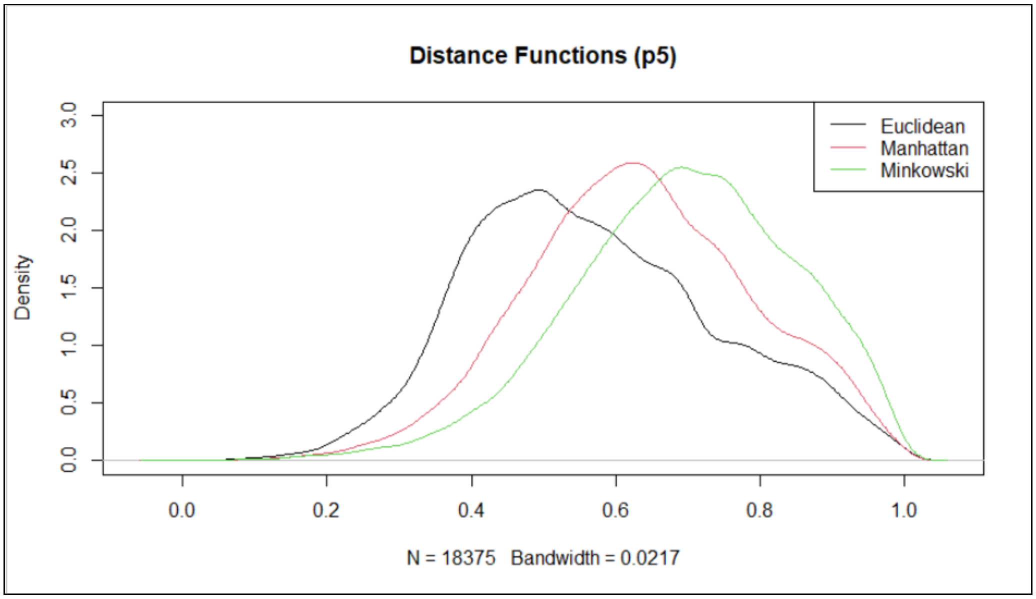
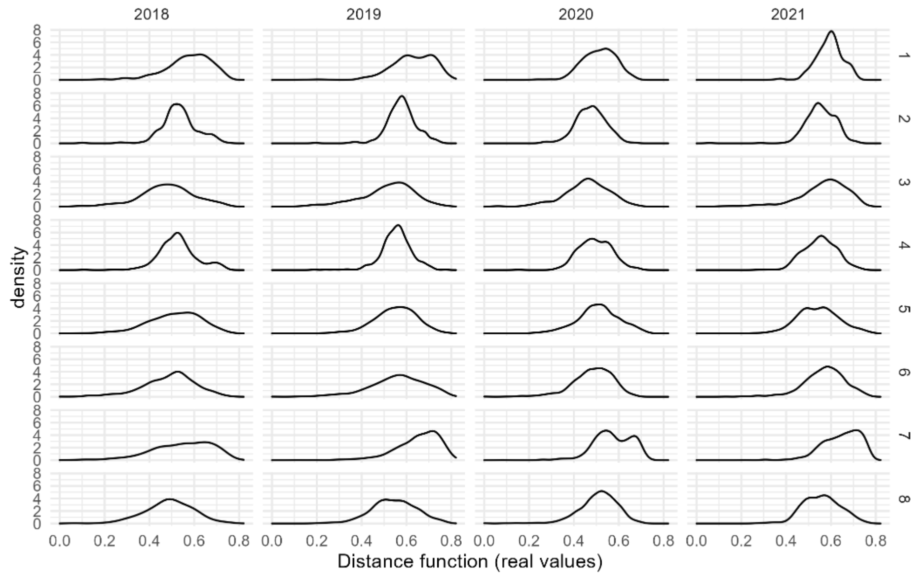


Figure A2. Distribution of the distance-based MP score by year and governorate (real vs. predicted values)



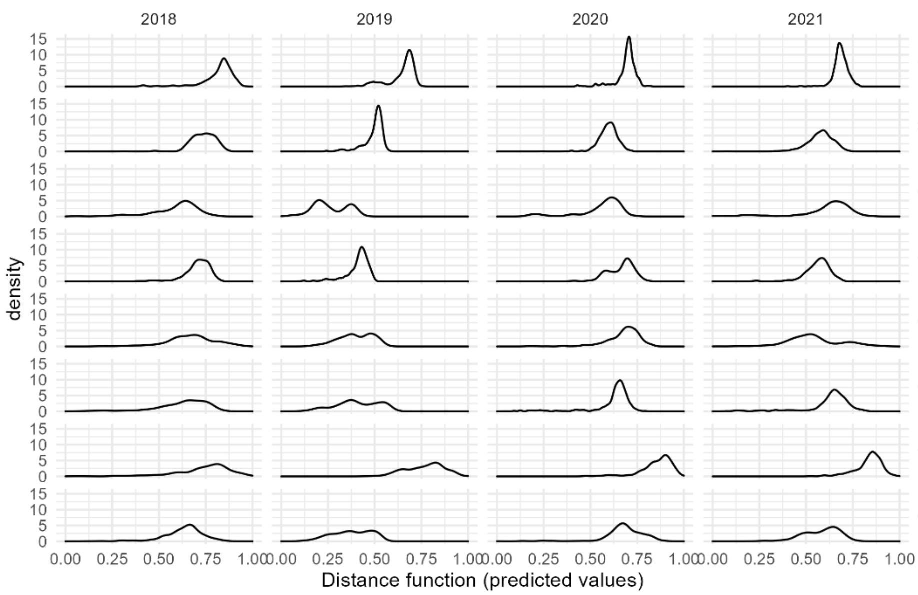


Figure A3. Distribution for the variables used to construct the distance function by year and governorate