ARTICLE IN PRESS

World Development xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

World Development

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

Optimal targeting under budget constraints in a humanitarian context

Paolo Verme *, Chiara Gigliarano

The World Bank, 1818 H St., NW Washington DC 20433, USA
Department of Economics, Università dell'Insubria, via Monte Generoso 71, 20100 Varese, Italy

ARTICLE INFO

Article history:
Available online xxxx

JEL classification:
I31
I32
I38

Keywords:
Food vouchers
Refugees
Receiver Operating Characteristics curves
Targeting

ABSTRACT

The paper uses Receiver Operating Characteristic (ROC) curves and related indexes to determine the optimal targeting strategy of a food voucher program for refugees. Estimations focus on the 2014 food voucher administered by the World Food Program to Syrian refugees in Jordan using data collected by the United Nations High Commissioner for Refugees. The paper shows how to use ROC curves to optimize targeting using coverage rates, budgets, or poverty lines as guiding principles to increase the overall efficiency of a program. As humanitarian organizations operate under increasing budget constraints and increasing demands for efficiency, the proposed approach addresses both concerns.

Public Disclosure Authorized

1. Introduction

The United Nations has estimated the number of forcibly displaced people in the world at 65.6 million in 2016, most of them located in the Middle East and Sub-Saharan Africa.[1] Forced displacement is largely the consequence of conflict as shown by countries such as Nigeria, South Sudan, Somalia and the Republic of Yemen. Northern Nigeria has over 2 million Internally Displaced Persons (IDPs) as a direct consequence of the conflict with Boko Haram; South Sudan may now have up to 80 percent of its population either displaced or hosting displaced people as a consequence of the civil war; Somalia has generated millions of refugees over decades of instability who settled in neighboring countries or moved to third countries; the Republic of Yemen has almost 3 million people counted as either refugees or IDPs as a consequence of the civil war.[2]

In these countries, conflict has been the primary driver of forced displacement but conflict also contributes to other negative outcomes such as food insecurity and poor health, factors that can increase displacement in their own right.[3] Moreover, the same countries that experienced mass forced displacement in recent years are also the same countries that are experiencing persistent droughts. The UN estimated that up to 20 million people may be at risk of famine in Sub-Saharan Africa and the Middle East by the end of 2017 with countries such as Nigeria, South Sudan, Somalia and the Republic of Yemen being highly affected.[4] This factor reinforces the negative consequences of conflict including forced migration.

Because of this combination of factors, humanitarian organizations have been facing severe budget constraints that made targeting more compelling. Donors find themselves pulled between the necessity to mitigate starvation and save lives and provide for shel-

* Corresponding author.
E-mail addresses: pverme@worldbank.org (P. Verme), chiara.gigliarano@uninsubria.it (C. Gigliarano).

[1] http://www.unhcr.org/en-us/figures-at-a-glance.html.1.

[2] For more details of these crises, see www.unhcr.org (country updates) and www.iom.int (particularly the Displacement Tracking Matrix – DTM – at http://www.globaldtm.info/).

[3] See, among others, Brück and d'Errico (2017) for a discussion on the relationships between food insecurity and various forms of violent conflict, insecurity and fragility. Mercier, Ngenzebukez, and Verwimpx (2017) show that Burundi households living in localities exposed to the 1993 civil war exhibit a significantly higher level of deprivation than non-exposed households. See Shemyakina (2017) on the health consequences of conflict. See also Martin-Shields and Stojetz (2017) for a recent review of the literature on conflict and food security, van Weezel (2017) for a global study of the impact of conflict on food security covering 106 countries and Brück, d'Errico, and Pietrelli (2017) for a study of the impact of conflict on food security in Gaza. For a discussion about the uncertainty that conflicts cause on agricultural production, see Arias, Ibáñez, and Zambrano (2017).

[4] http://www.cnn.com/2017/03/11/africa/un-famine-starvation-aid/4.

https://doi.org/10.1016/j.worlddev.2017.12.012
0305-750X/© 2017 The Author(s). Published by Elsevier Ltd.

Please cite this article in press as: Verme, P., & Gigliarano, C. Optimal targeting under budget constraints in a humanitarian context. *World Development* (2018), https://doi.org/10.1016/j.worlddev.2017.12.012

ter, services and any other needs of displaced populations in countries where host governments have scarce means of their own to face these challenges. Budgets are very stretched and humanitarian organizations face tough choices on the ground. Universal coverage of assistance programs such as cash or food assistance becomes the exception rather than the rule and humanitarian organizations are forced to ration and target resources. This is where targeting becomes a very relevant activity for donors and humanitarian organizations alike.

This paper is a contribution in the direction of making targeting more effective when budgets are stretched. It exploits Receiver Operating Characteristics (ROC) curves and related indices to devise a relatively simple methodology for optimizing coverage, poverty reduction and leakage in the presence of budget constraints. We therefore focus on optimizing outcomes of social protection programs based on cash transfers.

ROC curves are one of the most common statistical tools employed to assess the performance of a diagnostic rule based on a predictive model (Lusted, 1971). They are generated by plotting the fraction of true positives out of the positives (true positive rate) versus the fraction of false positives out of the negatives (false positive rate), at various probability thresholds. If we are assessing a poverty reduction program, this would correspond to plotting the fraction of poor persons who benefit from the program (coverage rate) versus the fraction of non-poor persons who benefit from the program (leakage rate). This curve can be used to determine the probability cut point that optimizes coverage and leakage.

For example, using a poverty prediction model, one can estimate from microdata the probability of an individual to be poor. To classify these same individuals as poor or nonpoor based on these predictions, one must then decide the probability threshold to use as cut point between poor and nonpoor. The standard approach is to consider as poor those individuals who have a probability of being poor above 50 percent and as non poor those individuals who have a probability equal or below 50 percent. However, this practice is arbitrary and may result in coverage and leakage rates that are not optimal. By changing the probability cut point, one can plot the curve that describes all possible combinations of coverage and leakage rates (the ROC curve) and choose the probability that correspond to the best combination of the two outcomes.

The main applications of the ROC curves are in medicine (see, e.g., Hand, 2010) or more generally in diagnostics (see, e.g., Hand & Anagnostopoulos, 2013) and in credit risk analysis (see, e.g. Gigliarano, Figini, & Muliere, 2014; Thomas, 2009), and the Area Under the Curve (AUC) is a popular measure to evaluate the discriminative power of a predictive model (see, e.g. Hand, 2009, 2012; Krzanowski & Hand, 2009). The use of these curves in economics has been less frequent (see Wodon, 1997 for an early application) and, to our knowledge, these curves have not been used in the context of humanitarian programs, with the exception of our previous contribution on which this paper builds (Verme et al., 2016).

In the context of a welfare program such as cash assistance or food vouchers, the approach proposed can be used when policy makers work with a coverage, poverty or budget target. [5] It answers questions such as: *i) What is the budget required to reduce poverty by X percent? ii) What is the budget required to increase household coverage by Y percent? iii) What is the coverage or poverty reduction we can obtain with a given Z budget? iv) Can targeting be improved by shifting the poverty line?* The answers to these questions can help donors make funding decisions and humanitarian organizations make targeting choices. We show that this approach can be applied to existing programs such as food voucher programs administered by the World Food Program (WFP) or cash programs administered by the United Nations High Commissioner for Refugees (UNHCR).

These are important questions considering the experience with targeting programs in development contexts. There are different methodologies based on microdata that are used for targeting such as simple means tests, proxy means test, community based or geographical targeting, categorical targeting or various types of self-selection methods. In a classic study on targeting performances of these methods that covered 122 interventions from 48 countries, Coady, Grosh, and Hoddinott (2004) found that targeting performances varied very significantly across programs worldwide. The median program was found to transfer to the poor only 25 percent more resources than a random allocation of transfers whereas 21 of the 85 programs evaluated were regressive, meaning that they performed worse than a random allocation. We are not aware of a comparable study in humanitarian contexts but it is evident that improving the targeting capacity of programs is essential.

This paper uses the same modelling used for proxy means testing, which is one of the best performing methodology according to Coady et al. (2004). Our contribution is to improve on this methodology by adopting a more rigorous system to select the optimal probability threshold for classifying households as poor or nonpoor. The methodology proposed is not a substitute for a good welfare or poverty model, which remains the cornerstone for a good targeting performance, but helps to improve on targeting after the model has been defined. We show how to operationalize such approach in the context of welfare improving humanitarian operations using relatively simple visual devices suitable for policy makers. In this respect, the paper is a complement to the relatively new literature on targeting food and cash programs in humanitarian contexts (Coll-Black et al., 2012; Maxwell, Young, Jaspars, Burns, & Frize, 2011; Morris, Levin, Armar-Klemesu, Maxwell, & Ruel, 1999; Tranchant et al., 2017).

The paper is organized as follows. The next section describes the food voucher program that we use to illustrate the methodology proposed. Section 3 describes the data, Section 4 outlines the models employed to caliber the methodology, Section 5 outlines results and Section 6 concludes.

2. The 2014 food voucher program for refugees in Jordan

As a case study, the paper uses the 2014 food voucher program administered by WFP to Syrian refugees in Jordan. It should be noted upfront that Jordan is a middle-income country and has a very different capacity to provide for displaced populations as compared to countries such as South Sudan or Somalia. The funding per refugee available in this country is higher than what is generally available in poorer countries and the quality and quantity of refugee data is superior to those available in Sub Saharan Africa. For example, the estimated UNHCR expenditure for the Middle East and North Africa region in 2015 was higher than the estimated expenditure for Sub Saharan Africa, in spite of the fact that the projected needs and caseload was higher in Sub Saharan Africa.[6] The infrastructure and the education level of the local staff in a middle income country like Jordan also contributed to improve the quality of data collection, management and use. However, we will show that the approach proposed can be implemented using micro data that the WFP and UNHCR routinely collect and that this is a method not beyond the existing capacity of these organizations.

The material assistance provided to refugees in Jordan in 2014 was essentially structured in two programs: i) a cash program

[5] The technology proposed in this paper requires basic econometric skills and trained econometricians but the resulting graphs presented in the results section are designed for policy makers and do not require any particular training.

[6] http://www.unhcr.org/56d7f9884.pdf6.

Please cite this article in press as: Verme, P., & Gigliarano, C. Optimal targeting under budget constraints in a humanitarian context. *World Development* (2018), https://doi.org/10.1016/j.worlddev.2017.12.012

administered by the UNHCR and ii) a food voucher program administered by the WFP. In 2014, the UNHCR cash program provided 50 Jordanian Dinars (JD)[7] per month to cases[8] including one or two members, 100 JD to cases with 3–5 members and 120 JD to cases with more than five members. The WFP program included two bi-weekly vouchers for a total value of 24 JD per person per month. The voucher was provided to the principal applicant and it could be spent via a network of 652 stores distributed across all governorates of Jordan. In this paper, we focus on the food voucher program but the approach proposed can be equally used for any cash program. Note that the two programs are not substitutes but complements. Refugees can receive both programs and neither of the programs can exclude households from the other program.

By the end of 2013, the food voucher program reached almost universal coverage and was assessed positively by both its administrator and external organizations. In addition to the benefits accrued to the direct beneficiaries, the WFP found that "*The program has also led to some US$2.5 million investment in physical infrastructure by the participating retailers; created over 350 jobs in the food retail sector; and generated about US$6 million in additional tax receipts for the Jordanian government. In terms of indirect effects, this study finds a predictive multiplier ranging from 1.019 to 1.234. In other words, WFP's plan to distribute US$250 million in vouchers during 2014 would lead to some US$255-US$308 million of indirect benefits for the Jordanian economy.*" (WFP, 2014, p. 1) The World Bank and the UNHCR jointly evaluated the poverty reduction of the program and found it to reduce the poverty headcount from 69.2 to 32.3 percent of the refugee population living outside camps (Verme et al., 2016).

Towards the end of 2014, the humanitarian community started to face budget shortages that forced the WFP to scale down its food voucher program. The organization found itself with the difficult choice of how to prioritize beneficiaries and opted to follow a welfare approach whereby income or consumption would be used to assess welfare and target households. Targeting based on welfare modeling – where the dependent variable is a measure of wellbeing such as income, expenditure or consumption and the independent variables are a set of individual and household socioeconomic characteristics –, is a well-developed methodology in economics and has produced important tools like Proxy Means Targeting (PMT) that help organizations such as the World Bank or governments to target cash-based programs. However, this is not an exact science and it is known that this approach invariably results in sizeable leakage and undercoverage errors.

3. Data

The paper uses a combination of the UNHCR proGres (PG) registry data and the UNHCR Home Visits (HV) data. The PG database is the official global registration system of the UNHCR. Much of the published statistics on refugees derive from this database and any survey or home visit targeting refugees is usually based on this database. ProGres contains information collected from refugees at different stages starting from the first brief interview administered when refugees cross the border to more extended interviews carried out when refugees are settled. Data are updated on a continuous basis and they are principally used to identify beneficiaries of the various protection and cash programs that the UNHCR and other organizations administer to refugees. It is, in short, the refugee "census" and contains individual and household (case) socioeconomic information but does not contain information on welfare.

The second data set is the Home Visits database (HV), which has been administered in Jordan in successive rounds starting from 2013 for the purpose of targeting the cash assistance program. The HV questionnaire results in almost 200 variables that can be used for analysis and includes questions on income and expenditure, which we use to construct our welfare aggregate. For the purpose of this paper, we use the second round implemented between November 1, 2013 and September 30, 2014.

The home visits survey is administered to refugees living outside camps, which is the focus group of this paper. In 2014, less than 18 percent of refugees were hosted in camps with almost 14 percent residing in the Zaatari camp alone. Syrian refugees living in Jordan at the time were not allowed to work although a few hundred refugees obtained work permits through special arrangements granted to selected employers. Despite this restriction, about 44 percent of refugees living outside camps reported to have some form of income suggesting that many worked informally. Reported income was very low, below or close the minimum wage of Jordan and also typically occasional (Verme et al., 2016).

The unit of observation used by the study is the "case". The UNHCR abides to the principle of family unity whereby a case should include family members but family members may have been separated or several families may be living together. This implies that a case may correspond to a set of related individuals, a family or a household made of several families.[9] The case identifier is present in both PG and HV data and for this reason it was possible to merge the PG and HV data into one dataset. The final sample used for this paper included over 43,000 observations, about a third of all cases registered in Jordan in 2014. This sample is not a random sample of the refugee population but Verme et al. (2016) showed that it is the closest approximation to a random sample that one could have with existing data.

The welfare aggregate was constructed using two questions on expenditure present in the HV data. The HV data contain questions on income and expenditure. The income question was clearly underreported and, as customary when measuring welfare in poor and middle-income countries, expenditure or consumption is preferred as a measure of welfare. The two questions on expenditure were both based on a 30 days' recall period but one had answers organized in six items and the second in ten items.[10] The two questions were aggregated by selecting the highest value between the two questions for items common to both questions and preserving the rest of the items from both questions. This allowed to reduce underreporting to a minimum. The resulting welfare aggregate showed common properties to consumption aggregates in terms of distribution.[11] We had also knowledge of the households (cases) that received assistance via the UNHCR cash program. For these households, the value of the programs was subtracted from the aggregate to obtain expenditure net of social assistance. Most households were also receiving food vouchers but interviewers and respondents to the home visits questionnaire were instructed to exclude these expenditures when filling the questionnaire.[12] Finally and for the purpose of poverty modeling, the welfare aggregate was transformed into per capita basis.

[7] A Jordanian Dinar in 2014 was roughly equivalent to one Euro.

[8] In the UNHCR jargon, a "case" is the household or family unit used to register refugees. See also the data section.

[9] See UNHCR, Operational Guidance Note on Resettlement Case Composition, June 2011. See also: http://www.unhcr.org/3d464ee37.pdf9.

[10] The questionnaire contained two questions on expenditure because one was designed for measuring general expenditure and the other for better understanding food expenditure.

[11] For more details on the welfare aggregate and its properties see Verme et al. (2016, p. 60).

[12] Excluding the value of the food voucher from the expenditure aggregate was essential to make sure that cash assistance could be cumulated with food assistance. In other words, the UNHCR provided cash assistance to cases whose expenditure net of food assistance was found to be below the poverty line. This is also essential for our analysis where we evaluate the pre- and post-transfers poverty of beneficiaries.

Please cite this article in press as: Verme, P., & Gigliarano, C. Optimal targeting under budget constraints in a humanitarian context. *World Development* (2018), https://doi.org/10.1016/j.worlddev.2017.12.012

Table 1
Summary statistics of welfare aggregates monthly).

Variable	Mean (USD)	Mean (JD)	Std. Dev. (JD)	Min (JD)	Max (JD)
Income per capita	49.63	34.95	64.41	0	3000
Income per capita with no zeroes	93.01	65.50	75.99	0.5	3000
Expenditure per capita	110.65	77.92	74.72	1	1675
Expenditure per capita net of UNHCR cash assistance	100.25	70.60	76.29	0	1675

Source: Authors' estimations based on Jordan UNHCR data.

Table 1 provides the summary statistics for income and expenditure. It is clearly visible that income per capita is underreported when compared to expenditure. It is also evident that the UNHCR cash program contributes, on average and per capita, with about 7 JD per month, which is in line with what we should expect given the coverage of the program in 2014. Fig. 1 also shows how the shape of the welfare aggregate of choice (expenditure per capita net of UNHCR cash) exhibits the expected characteristic of a bell-shaped distribution.

The poverty line adopted for the study is 50 JD per capita per month. This threshold was selected because it was the poverty line adopted by the UNHCR for providing cash assistance. At the end of 2013, this line was equivalent to 71 USD and 160 USD at Purchasing Power Parity (PPP) per capita per month. This latter amount corresponded to 5.25 USD PPP per day, a significantly higher amount than the international poverty line of 1.9 USD PPP recommended for poor countries but a lower amount as compared to the poverty line adopted by Jordan at the same time (8.2 USD/day/PPP). Considering that Jordan is a middle-income country and that most refugees live in urban areas, this poverty line is reasonable.[13] Based on the sample used for this study, a poverty line of 50 JD/capita/month results in a poverty rate of 52.5 percent for cases and 69.2 percent for the population. In other words, 7 in 10 refugees in 2014 were estimated to be poor based on a 50 JD per person poverty line.

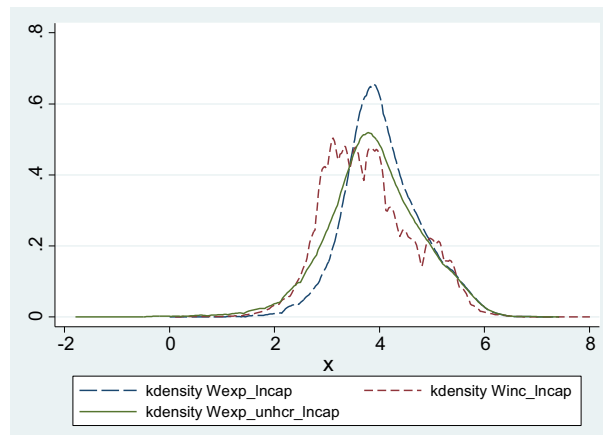

Fig. 1. Income and expenditure distributions. Source: Authors' estimations based on Jordan UNHCR data. On the x-axis is the log of income or expenditure per capita. On the y-axis is the density (percentage of population).

4. Models

The underlying model of the predictions and simulations that will follow is a simple poverty model described as follows:

$$P_i = \alpha + \gamma X_i + \varepsilon_i \tag{1}$$

where $P_i = 1$ if the case is under the poverty line and $P_i = 0$ if the case is on the poverty line or above; X_i = vector of case characteristics that derive from merging the UNHCR proGres registry data and the home visits data (see Table A in annex for the full list of variables); ε_i = normally distributed error term with zero mean; and i = case.

The model is estimated with a probit function and the main objective is to maximize its poverty prediction capacity. To achieve this objective, we followed a systematic procedure to optimize the construction of the independent variables and the explanatory power of the model. For the variables construction, we constructed case indicators and turned some of the categorical variables into dichotomous variables.[14] This avoided the question of discontinuity between categorical variables. To maximize the explanatory power of the model, we first explored the explanatory capacity of each individual variable with binomial models to prepare a first shortlist of variables and then used this shortlist with a backward and forward stepwise selection method to identify the best model.

The optimal prediction model is then used to predict poverty as follows:

$$\hat{P}_i = \hat{\alpha} + \hat{\gamma} X_i \tag{2}$$

where $\hat{P}_i$ can be interpreted as the expected probability of being poor given a set of X characteristics and based on the estimated parameters $\hat{\alpha}$ and $\hat{\gamma}$ from Eq. (1).

To determine whether a household is expected (predicted) to be poor or not, the standard approach is to assign a prediction of "i = non-poor" to households with $0 \leqslant \hat{P}_i \leqslant 0.5$ and "i = poor" to households with $0.5 < \hat{P}_i \leqslant 1$. This is not particularly efficient because a cut point of 0.5 may not be the cut point that maximizes the coverage rate (i.e. the fraction of poor correctly predicted as poor) and minimizes the leakage rate (i.e. the fraction of nonpoor predicted as poor), when these are the objectives of the program's administrators. One approach to solve this problem is to use indices that promise to optimize coverage and leakage. Two popular indices are the Youden Index (YI) and the Distance Index (DI) described as follows:

$$YI = max\left(\frac{P_t}{P} - \frac{NP_t}{NP}\right) \tag{3}$$

$$DI = \sqrt{\left(\frac{P_{nt}}{P}\right)^2 + \left(\frac{NP_t}{NP}\right)^2} \tag{4}$$

where N = population; P = number of poor; NP = Number of non-poor; t = targeted; nt = non-targeted. Consequently: coverage rate = $\frac{P_t}{P}$; undercoverage rate = $\frac{P_{nt}}{P}$; and leakage rate = $\frac{NP_t}{NP}$.

[13] See also the discussion on poverty line in Verme et al. (2016, p. 64).

[14] The categorical variables that have been turned into dichotomous variables are: *House: electricity*, *House: sanitary*, *House: concrete house*, *House: for rent or owned*, *Receiving any kind of NFIs*. This transformation, although affecting the variance of these explanatory variables, was due to the fact that some modalities of the original variables had very small frequencies, thus making the estimation of the regression coefficients unreliable.

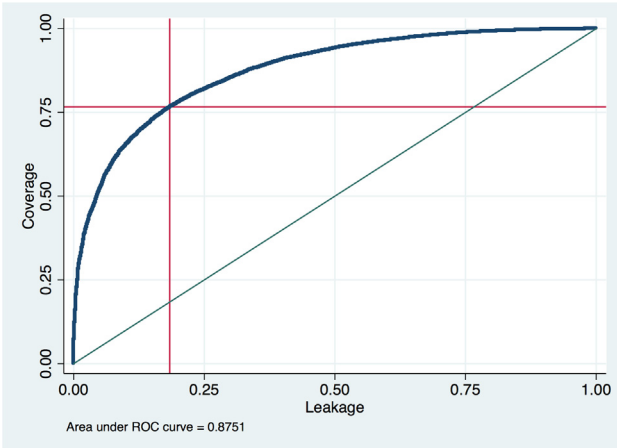

Fig. 2. Coverage and leakage rates with different Probability Thresholds (ROC curve). Source: Authors' estimations based on Jordan UNHCR data.

Similarly, one can draw the ROC curve, which plots for each possible cut point between 0 and 1 the corresponding leakage rate (*x-axis*) and coverage rate (*y-axis*), thus determining the optimal trade-off between these two indicators (see also Wodon, 1997). This is represented in Fig. 2 below. In such a graph, the diagonal represents an equal probability of coverage and leakage, which is what one would obtain if targeting was random (blind assignment of the program). All the points on the left of the diagonal show the performance of models that do better than random draws. The Area Under the ROC Curve (AUC) becomes therefore an indicator of the predicting capability of the model and it can be used in conjunction with the R-Squared statistics of a model to assess its predicting capacity. Differently from the R-squared, the AUC value varies between 0.5 and 1 where 1 represents perfect prediction capacity.

The example illustrated in Fig. 2 shows that the model does better than a random assignment. The AUC value is 0.8751, which indicates that the model is capable of predicting poverty correctly 87.5 percent of the time, while a random assignment would predict correctly 50 percent of the time. The curve also shows that there is one point where the distance between the diagonal (random assignment) and the curve (the model) is maximized (vertical line in the figure). This is the point that offers the best coverage while minimizing leakage, given the selected model. To this point corresponds a probability threshold of 0.703, which is what can be used to maximize the targeting performance of the model. This also corresponds to the Youden index, which can be graphically represented as the longest vertical distance between the ROC curve and the diagonal. The Distance index, instead, corresponds to the shortest Euclidean distance between errorless classification (the point (0,1)) and the ROC curve. Therefore, in the example depicted in Fig. 2, the probability threshold that provides optimal targeting is well above 50 percent according to both indexes and the ROC curve.

This is precisely what ROC curves can help to do. Each threshold corresponds to a specific coverage rate and leakage rate and there is clearly a different trade-off between coverage and leakage in correspondence of each probability threshold. Changing the threshold may improve on one of the two rates while making the other rate worse. By knowing preferences for minimizing leakage (save money) or maximizing coverage (reduce poverty) and the corresponding trade-offs depicted by the ROC curves, policy makers can evaluate and compare alternative outcomes in anticipation of implementing a program.

The methodology that we propose has some limitations. First of all, it considers as equally important the misclassifications due to false negative and those due to false positive. Other approaches have been proposed in the literature, which assume, instead, that it is possible to quantify the different costs incurred in misclassifying individuals. These methods combine benefits and costs with the coverage and leakage in order to find the value on the ROC curve that minimizes the average cost (or maximizes the average benefit) of a poverty assessment. Another drawback of the method is that comparisons over time and across countries may be problematic, since different ROC curves will provide different optimal probability thresholds.

5. Results

As in any other econometric effort, the key to a successful implementation of the methodology proposed is to find a poverty model with good explanatory power. Table A in the annex shows the results of alternative poverty models based on the 2014 UNHCR Jordan data. The dependent variable is the dummy variable that indicates whether a case's expenditure per capita is below the poverty line (equal to 50 JD),[15] whereas the independent variables are divided into groups of variables. In Model 1 we consider only socio-demographic variables such as case size, proportion of children, age, education, employment and marital status of the Principal Applicant (PA),[16] as well as place of destination and origin of the household. In Model 2 we also add variables related to characteristics of the house and of the WASH[17] system, and Model 3 includes also variables related to coping strategies implemented by the refugees as well as information on whether the case (household) receives humanitarian assistance.

In Table A, the three models are compared in terms of pseudo R-squared and Area Under the ROC Curve (AUC). These indicators show that Model 3 shows the best performance followed by models 2 and 1 in this order. Regressors related to housing and WASH included in models 2 and 3 seem to make the difference with model 1 and contribute significantly in explaining the probability of being poor. Fig. 3 shows graphically the same results using the ROC curves. The AUC is almost the same in models 2 and 3 and the corresponding ROC curves practically overlap (the green line in Fig. 3), whereas the ROC curve corresponding to model 1 (the blue line in Fig. 3) is clearly below.

Considering Model 3 as our best poverty model, we can now test its capacity to predict poverty correctly. The exercise consists of using the parameters estimated by the models to predict the dependent variable of the model (poor/non-poor) as if we did not have information on this variable. Table 2 shows the results using different probability thresholds. For any given cut point between 0 and 1, we can evaluate the corresponding coverage rate (that is the probability of correctly classifying the poor) and leakage rate (that is the probability of classifying as poor a non-poor). The table compares coverage and leakage rates for three arbitrary cut point values (0.3, 0.5 and 0.6), revealing that in correspondence to a low cut point we obtain a very high coverage rate but also a high leakage rate.

In particular, the central panel of Table 2 shows that, using a poverty line of 50 JD and a 50 percent threshold, the model would be able to predict correctly if an individual is poor 90.4 percent of the time, which implies that 9.6 percent of the time the model would predict poor individuals as non-poor (under-coverage rate

[15] See Verme et al. (2016) for a sensitivity analysis based on different poverty lines.

[16] The principal applicant is the reference case person for the UNHCR, similarly to head of household in household budget surveys.

[17] WASH is an acronym frequently used by humanitarian organizations to describe the Water, Sanitation and Hygiene sectors.

Please cite this article in press as: Verme, P., & Gigliarano, C. Optimal targeting under budget constraints in a humanitarian context. *World Development* (2018), https://doi.org/10.1016/j.worlddev.2017.12.012

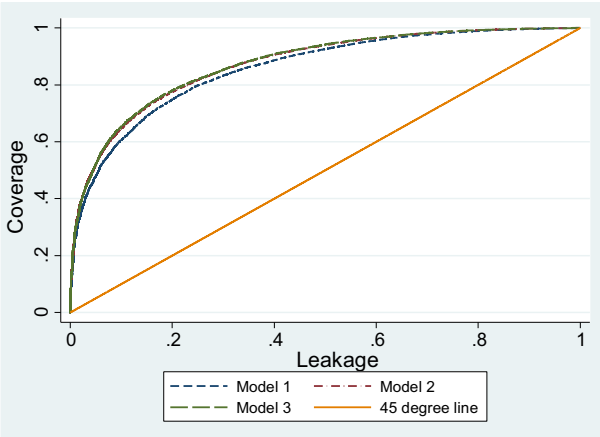

Fig. 3. Comparing poverty models' ROC curves. Source: Authors' estimations based on Jordan UNHCR data.

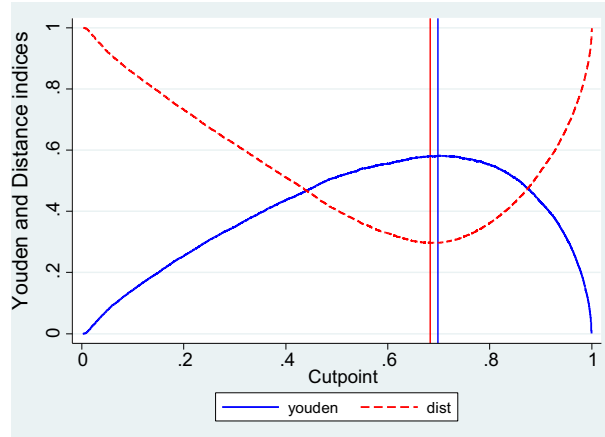

Fig. 4. Youden and distance indices for different cut point. Source: Authors' estimations based on Jordan UNHCR data.

Table 2
Coverage and leakage rates for different choices of cut point.

		Observed poor		
		No	Yes	Total
Predicted poor		*Cut point 30%*		
	No	37.64	2.92	13.59
	Yes	**62.36**	**97.08**	86.41
		Cut point 50%		
	No	61.01	9.59	25.39
	Yes	**38.99**	**90.41**	74.61
		Cut point 60%		
	No	71.20	15.63	32.71
	Yes	**28.80**	**84.37**	67.29

Source: Authors' estimations based on Jordan UNHCR data.
Bold values refer to coverage and leakage rates as referred in the text where the table is discussed.

or exclusion error). The model would also predict correctly if an individual is non-poor 61 percent of the time, which means that 39 percent of the time the model would predict as poor nonpoor individuals (leakage rate or inclusion error). The first type of error (under-coverage) is more problematic from a policy and welfare perspective, whereas the second type of error (leakage) is more problematic from a budget perspective.

The first and the last panels repeat the exercise with a probability threshold of 30 percent and 60 percent, respectively. In the former case, the under-coverage rate improves (2.9 percent) while the leakage rate worsens (62.4 percent). In the latter case, the under-coverage rate worsens (15.6 percent) while the leakage rate decreases considerably (28.8 percent). Clearly, changing probability threshold affects targeting results. Hence, it is important to fine-tune the parameter related to the probability threshold to obtain results the closest as possible to the error we want to minimize (under-coverage or leakage).

We now determine the optimal probability threshold that maximizes coverage and minimize leakage, by computing the Youden and the Distance indices (Fig. 4). In general, the two indicators may suggest different optimal thresholds (indicated in Fig. 4 by the two vertical lines). In our case, we find the two values rather close: according to the Youden index, the optimal probability threshold should be 0.703 (corresponding coverage, 0.766; corresponding leakage, 0.184), while according to the Distance index the probability cutoff should be 0.685 (corresponding coverage, 0.782; corresponding leakage, 0.201). Also noteworthy and from a graphical perspective, one could consider as an optimal choice, the maximum vertical distance between the Youden and Distance indices curves. This falls indeed within the two optimal thresholds indicated respectively by the two indices and could solve the issue of choosing between two optimal values.

As an alternative, it is possible to select the optimal probability threshold using the available budget for the food voucher program as baseline. The WFP would normally know the budget available for targeting and this is what frequently drives targeting decisions. The WFP or the UNHCR do not have large regular annual budgets and they rely on donors' contributions pledged during humanitarian crises. Based on the effective contributions received following the pledges, these organizations take targeting decisions accordingly. Therefore, the targeting decisions are largely based on budget rather than coverage criteria. Indeed, if programs were to be administered based on purely legal and entitlements criteria, these organizations should cover the totality of refugees and coverage would be universal. Targeting is therefore a second best option determined by budget constraints.

As an example of targeting based on budget criteria, Fig. 5 plots cut points (or probability thresholds) corresponding to different mixes of coverage and leakage rates on the *y-axis* versus different budget scenarios on the *x-axis*. For simplicity, the budget on the *x-axis* varies between 0 and 100, where 100 corresponds to the budget needed for universal coverage. For example, if we assume that the WFP has a budget sufficient to cover only 80 percent of the refugee population (the vertical line in the figure), the probability threshold that would achieve that outcome is 0.42 (the horizontal line in the figure).

Alternatively, one can focus on optimizing coverage and leakage. This can serve organizations that do not have a sufficient budget to cover all potential beneficiaries but can choose between different levels of expenditure based on efficiency criteria. In this case, one can plot coverage and leakage rates against different budget scenarios as shown in Fig. 6. By drawing a vertical line that corresponds to the maximum distance between coverage and leakage rates, we can derive the optimal budget. In the example proposed, this budget is equivalent to 58.7% of the budget necessary for universal coverage. This is the budget that simultaneously maximizes coverage and minimizes leakage.

On the contrary, a decision maker may prefer to fix a priori the percentage of coverage. This is shown in Table 3 where, for a given coverage rate, the corresponding percentage of the universal budget, as well as the leakage rate and the cut point are given. For example, in order to reach 80% of coverage, one needs to use 62.25% of the universal budget, which corresponds to 22% of leakage and to a probability threshold equal to 66%.

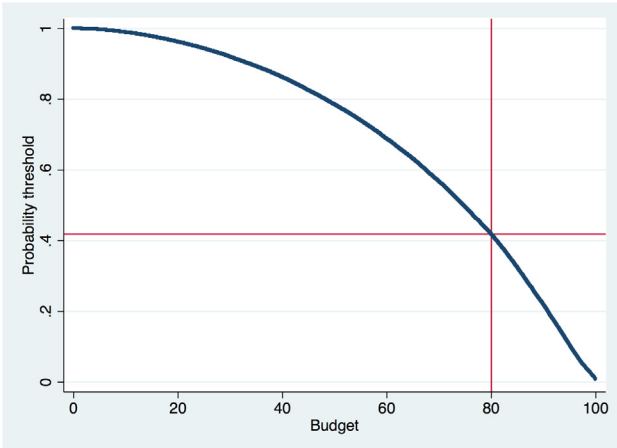

Fig. 5. Probability threshold for different budget scenarios. Source: Authors' estimations based on Jordan UNHCR data.

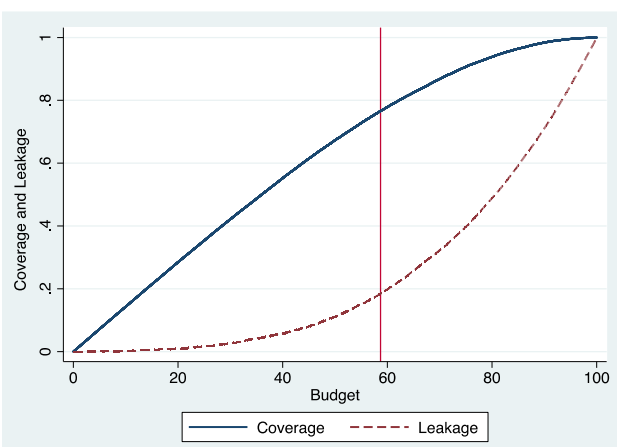

Fig. 6. Coverage rate and leakage rate for different budget scenarios. Source: Authors' estimations based on Jordan UNHCR data.

Table 3
Budget, leakage, probability threshold for different % of the universal coverage.

% Coverage	Budget (as % of universal budget)	Leakage	Probability threshold
100%	100.00	1.00	0.00
80%	62.25	0.22	0.66
60%	43.81	0.07	0.83
50%	3.47	0.00	1.00

Source: Authors' estimations based on Jordan UNHCR data.

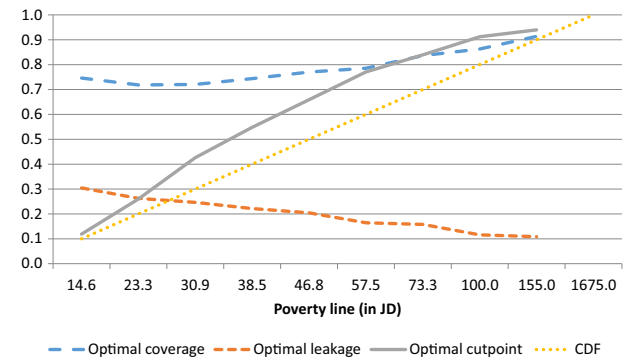

Fig. 7. Optimal coverage rate, leakage rate and cut point (%). Source: Authors' estimations based on Jordan UNHCR data.

Summarizing alternative options, Table 4 shows the change in coverage, leakage, probability threshold and poverty rates in correspondence of different budgets (represented as different percentages of the universal budget). The first line shows that, in case of universal budget, every case receives assistance (coverage and leakage are equal to one and cut point is equal to zero) and the poverty rate corresponds to 32.3%. Moreover, 69.2% of the overall budget is used for coverage whereas leakage costs 30.8%. When, instead, only 80% of the universal budget is used, the poverty rate increases to 38.4%, and the budget is split between 81.2% for coverage and 18.8% for leakage. Reducing again the percentage of the universal budget to 50%, the poverty rate rises to 50.1%, coverage absorbs 93.1% of the budget and leakage 6.9%. Therefore, this approach provides a costing of leakage in addition to coverage and the corresponding outcomes in terms of poverty rates. A table like Table 4 is a useful tool for donors and decision makers called to make difficult budget choices.

Adjusting the poverty line may be another device to optimize coverage and leakage. One can repeat the whole exercise of estimating the poverty model for a set of different poverty lines, thus obtaining the corresponding predicted probabilities and the cut point that optimizes coverage and leakage, and see whether marginally adjusting the poverty line can result in significant gains in terms of coverage and/or leakage. This is illustrated in Fig. 7. As an example, we use ten alternative poverty lines in correspondence of each decile of expenditure per capita and estimate the optimal probability threshold and the corresponding coverage and leakage rate for each poverty line. We then connect these ten points to show the resulting curves. A policy maker would normally seek the maximum distance between the coverage and leakage rate and, as one would expect, this distance grows as we increase the poverty line.

However, the increase in this distance is not linear. For example, using the second decile's mean value as poverty line (23.3 JD), would result in a coverage rate of 72 percent and a leakage rate of 26 percent. If the poverty line is increased to the third decile's mean value (39.9 JD), this would not change the coverage rate and would decrease the leakage rate by only 1 percent. In other words, there is

Table 4
Coverage, leakage, probability threshold and poverty rate for different% of the universal budget.

% Universal budget	Coverage rate	Leakage rate	Probability threshold	Poverty rate (%)	% Budget for coverage	% Budget for leakage
100%	1.00	1.00	0.00	32.30	69.2%	30.8%
80%	0.94	0.49	0.42	38.39	81.2%	18.8%
60%	0.78	0.20	0.69	45.65	89.8%	10.2%
50%	0.67	0.11	0.78	50.10	93.1%	6.9%
0%	0.00	0.00	1.00	69.18		

Source: Authors' estimations based on Jordan UNHCR data.

Please cite this article in press as: Verme, P., & Gigliarano, C. Optimal targeting under budget constraints in a humanitarian context. *World Development* (2018), https://doi.org/10.1016/j.worlddev.2017.12.012

almost no gain in increasing the poverty line from 23.3 to 30.9 JD, although some poor households would be excluded by the program. Instead, shifting the poverty line from the sixth to the seventh decile's mean value would increase the coverage rate by 2 percentage points while decreasing leakage by 4 percentage points, a significant gain. Poverty lines established for targeting are generally of an absolute nature and based on basic needs assessments. However, targeting or budget considerations may justify adjusting the poverty line to optimize the use of resources. For this purpose, Fig. 7 can be a useful instrument for policy makers.

Fig. 7 also shows the Cumulative Distribution Function (CDF) built on deciles of expenditure per capita (a straight line by construction). The CDF is useful in that it shows the poverty rate on the (*y-axis*) for each possible poverty line (*x-axis*). For example, with a poverty line of approximately 42 JD, about 50 percent of the population would be under the poverty line. Therefore, when one adjusts the poverty line to optimize coverage and budget, it is also possible to monitor what the effect of this change would be on the poverty rate and, consequently, on the number of people under the poverty line. Again, this is a relatively simple device that supplies a set of critical information for policy makers aiming to fine-tune targeting decisions.

6. Conclusions

Humanitarian organizations such as the WFP and the UNHCR rely entirely on donors' contributions to administer programs for refugees and IDPs. These contributions are increasingly hard to come by and, when crises are multi-dimensional, contributions are typically below the needs generated by the sum of these dimensions. This is the case, for example, of the recent crises in Nigeria, South Sudan, Somalia and the Republic of Yemen. The conflicts that affected these countries generated the displacement of millions of people who became IDPs or refugees. The more recent droughts that affected all these countries are contributing to further misery and the combination of conflict and droughts led to food shortages that are resulting in famine.

The needs of the affected populations, whether refugees and IDPs or host populations, grow disproportionally when these calamities occur simultaneously, donors find themselves stretched between different calls for funds and humanitarian organizations remain under-funded. In these cases, targeting is an obliged choice and optimal and efficient targeting can make the difference between life and death for some of the displaced populations. This justifies refining targeting techniques making the most of the available technology.

In addition to being technically sound, targeting requires tools that make choice and implementation relatively simple for policy makers and field staff. One cannot expect all humanitarian staff working on the field to be knowledgeable in econometrics or statistics techniques. Hence the need for relatively simple visual devices that can be used for making normative targeting choices based on positive criteria without necessarily be cognizant of the econometrics and statistics that lie behind these devices.

This paper has proposed the use of ROC curves and related indices to refine targeting when budgets are constrained and has developed relatively simple graphs that can be used by policy makers to make decisions on coverage, budgets and poverty lines when targeting is based on welfare criteria. We show existing trade-offs between optimal coverage and leakage rates, the optimal use of existing budgets and small adjustments that can result in large gains in terms of poverty reduction. The methods proposed can be applied to food voucher programs as illustrated in this paper or to any other cash program. We also showed that such methods can be applied using existing data collected by the UNHCR in the framework of existing programs such as the WFP food voucher program. In other words, these are tools that can be readily applied and do not necessarily require the collection of new data or the administration of special programs.

It is also clear that the method proposed cannot be applied in all contexts. The Jordan example provided in this paper relies on a set of data (proGres and home visits data) that is atypical in humanitarian contexts. While proGres data are available in most countries where the UNHCR operates, the quality of these data is very variable (the Jordan data are known to be among the best quality data available). Also, home visits data that contain information on income or expenditure are the exception rather than the rule in humanitarian contexts. Therefore, at present, the replicability of the method proposed elsewhere is limited to selected areas and countries. However, all humanitarian operations with budget limitations that use cash or food vouchers as a form of social protection face the same targeting challenges described in this paper. In the absence of quality data on income or consumption, these operations have to rely on alternative and less accurate targeting criteria the outcomes of which (in terms of poverty, coverage and leakage) are non-measurable. This paper has implicitly shown that the collection of income or consumption data can lead to improvements in the measurement of outcomes and in the effectiveness of targeting.

This paper has also ignored aspects of political economy, administration and other outcomes that may be equally important for beneficiaries. Governments have often a bias for geographical targeting covering areas relevant for selected constituencies rather than focusing on the poor. Administrative constraints such as complex logistics or lack of administrative budgets can make reaching the poor impossible in some cases making geographical or other forms of targeting more appealing. Outcomes different from poverty reduction such as nutrition levels may be as important or more important than poverty reduction in a humanitarian and emergency context. These are all elements to factor in when considering the method proposed in this paper.

Conflict of interest

We hereby declare that neither authors (Paolo Verme or Chiara Gigliarano) have any conflict of interest in relation to the paper submitted: "Optimal Targeting under Budget Constraints: The Case of a Refugee Food Voucher Program".

Acknowledgement

This work is part of the program "Building the Evidence on Protracted Forced Displacement: A Multi-Stakeholder Partnership". The program is funded by UK aid from the United Kingdom's Department for International Development (DFID), it is managed by the World Bank Group (WBG) and was established in partnership with the United Nations High Commissioner for Refugees" (UNHCR). The scope of the program is to expand the global knowledge on forced displacement by funding quality research and disseminating results for the use of practitioners and policy makers. This work does not necessarily reflect the views of DFID, the WBG or UNHCR. The authors would like to thank Lidia Ceriani, Xavier Devictor, Ugo Gentilini and Phillippe Leite for very useful comments.

Appendix A.

Table A
Poverty models.

		MODEL 1		MODEL 2		MODEL 3	
		Coef.	*z*	*Coef.*	*z*	*Coef.*	*z*
Case size		0.581	49.000	0.639	53.190	0.629	51.580
Case size squared		−0.009	−10.450	−0.014	−16.890	−0.014	−15.800
Case size * Age of PA		−0.003	−14.300	−0.003	−14.150	−0.003	−13.970
Age of PA		0.007	9.150	0.009	10.370	0.008	9.170
Proportion of children < 18 years		0.362	15.470	0.343	13.950	0.295	11.800
Highest education of PA (in years)		−0.079	−20.310	−0.051	−12.420	−0.046	−11.250
Employment of PA (Ref. None)	*Low Skilled*	−0.001	−0.030	−0.043	−2.200	−0.039	−1.990
	Skilled	−0.053	−3.280	−0.045	−2.660	−0.047	−2.720
	High Skilled	−0.149	−9.170	−0.119	−7.000	−0.120	−7.020
	Professional	−0.239	−13.350	−0.213	−11.370	−0.213	−11.320
Marital status of PA (Ref. Married or engaged)	*Divorced or separated*	0.243	7.100	0.215	5.950	0.189	5.200
	Single	0.111	6.000	0.059	3.000	0.049	2.500
	Widowed	0.132	7.400	0.094	4.970	0.080	4.230
Origin (Ref. Damascus)	*Al-hasakeh*	0.032	0.620	−0.175	−3.060	−0.122	−2.100
	Aleppo	0.085	4.650	0.087	4.570	0.105	5.460
	Ar-raqqa	−0.168	−5.100	−0.231	−6.690	−0.188	−5.370
	Dar'a	0.054	3.610	0.076	4.920	0.064	4.080
	Hama	0.609	28.130	0.083	3.330	0.082	3.270
	Homs	0.140	9.230	0.192	12.100	0.158	9.830
	Idleb	0.613	16.690	0.170	4.090	0.158	3.770
	Rural Damascus	0.083	5.180	0.074	4.430	0.064	3.770
	Tartous	0.288	1.770	0.263	1.570	0.248	1.470
	As-sweida	0.225	2.010	0.056	0.460	0.024	0.200
	Deir-ez-zor	0.072	1.370	−0.017	−0.310	−0.012	−0.210
	Lattakia	−0.100	−1.360	−0.058	−0.770	−0.059	−0.780
	Quneitra	−0.121	−2.040	−0.088	−1.450	−0.079	−1.290
Formal arrival		−0.254	−23.720	−0.199	−17.770	−0.193	−17.070
Destination (Ref. Amman)	*Ajloun*	0.472	14.530	0.549	16.260	0.553	16.250
	Aqabah	−0.139	−2.400	−0.161	−2.610	−0.136	−2.170
	Balqa	0.121	6.400	0.063	3.190	0.081	4.060
	Irbid	0.116	10.430	0.125	10.690	0.123	10.330
	Jarash	0.434	15.160	0.486	16.370	0.502	16.840
	Karak	0.422	15.910	0.477	17.130	0.448	15.980
	Maan	0.303	8.250	0.376	9.560	0.340	8.580
	Madaba	0.390	14.400	0.366	12.700	0.391	13.510
	Mafraq	0.288	22.410	0.093	6.620	0.031	2.120
	Tafilah	0.725	12.590	0.883	14.350	0.913	14.770
	Zarqa	0.485	36.840	0.442	31.870	0.422	29.810
Border crossing point (Ref. Airport)	*Ruwaished-Hadallat*	0.109	5.710	−0.019	−0.940	−0.026	−1.280
	Tal Shihab	0.150	8.380	0.117	6.300	0.094	5.010
	Nasib-official or unofficial	0.177	13.500	0.143	10.510	0.113	8.190
	Other or no data	0.098	6.530	0.063	4.050	0.044	2.790
House: electricity				−0.120	−8.920	−0.121	−8.930
House: sanitary				−0.221	−19.850	−0.205	−18.270
House: concrete house				−0.277	−12.410	−0.294	−13.040
House: for rent or owned				−1.025	−42.580	−1.033	−42.620
House area: square meters per capita (Ref. < 10 sq meters)	*10–15 sq meters*			−0.099	−9.840	−0.097	−9.570
	>15 sq meters			−0.285	−29.890	−0.280	−29.150
Wash: water through piped and piped sewerage				−0.086	−8.700	−0.081	−8.180
Wash: having adequate and functioning latrine				−0.109	−9.490	−0.114	−9.840
Receiving any kind of NFIs						0.093	9.710
Coping strategy: humanitarian assistance						0.297	25.970
Coping strategy: host community						0.091	10.510
Coping strategy: dropping children from school						0.095	7.180
UNHCR financial assistance		0.648	52.190	0.725	56.100	0.627	46.180
	Constant	−2.137	−46.740	−0.599	−11.390	−0.572	−10.790
	N	174,125		174,125		166,616	
	Pseudo R-squared	0.314		0.314		0.348	
	AUC	0.858		0.873		0.875	

Source: Authors' estimations based on Jordan UNHCR data. Note: PA = Principal Applicant.

Please cite this article in press as: Verme, P., & Gigliarano, C. Optimal targeting under budget constraints in a humanitarian context. *World Development* (2018), https://doi.org/10.1016/j.worlddev.2017.12.012

References

Arias, M. A., Ibáñez, A. M., & Zambrano, A. (2017). Agricultural production amid conflict: Separating the effects of conflict into shocks and uncertainty. *World Development*, forthcoming.

Brück, T., & d'Errico, M. (2017). Food security and conflict: Introduction to the special issue. *World Development*, forthcoming.

Brück, T., d'Errico, M., & Pietrelli, R. (2017). The effects of violent conflict on household resilience and food security: Evidence from the 2014 Gaza conflict. *World Development*. forthcoming.

Coady, D., Grosh, M., & Hoddinott, J. (2004). *Targeting of transfers in developing countries: Review of lessons and experience*. Washington DC: World Bank and International Food Policy Research Institute (IFPRI).

Coll-Black, S., Gilligan, D. O., Hoddinott, J. F., Kumar, N., Taffesse, A. S., & Wiseman, W. (2012). In: *Food and agriculture in Ethiopia: Progress and policy challenges*, chapter 10 International Food Policy Research Institute (IFPRI).

Gigliarano, C., Figini, S., & Muliere, P. (2014). Making classifier performance comparisons when ROC curves intersect. *Computational Statistics and Data Analysis, 77*, 300–312.

Hand, D. J. (2009). Measuring classifier performance: a coherent alternative to the area under the ROC curve. *Machine Learning, 77*, 103–123.

Hand, D. J. (2010). Evaluating diagnostic tests: The area under the ROC curve and the balance of errors. *Statistics in Medicine, 29*, 1502–1510.

Hand, D. J. (2012). Assessing the performance of classification methods. *International Statistical Review, 80*, 400–414.

Hand, D. J., & Anagnostopoulos, C. (2013). When is the area under the receiver characteristic curve an appropriate measure of classifier performance? *Pattern Recognition, 80*, 400–414.

Krzanowski, W. J., & Hand, D. J. (2009). *ROC curves for continuous data*. New York: CRC Press—Chapman and Hall.

Lusted, L. B. (1971). Signal detectability and medical decision-making. *Science, 171*, 1217–1219.

Martin-Shields, C., & Stojetz, W. (2017). Framing the relationship between food security and violent conflict data: A clustering and typology approach. *World Development* (forthcoming).

Maxwell, D., Young, H., Jaspars, S., Burns, J., & Frize, J. (2011). Targeting and distribution in complex emergencies: Participatory management of humanitarian food assistance. *Food Policy, 36*(4), 535–543.

Mercier, M., Ngenzebukez, R. L., & Verwimpx, P. (2017). Violence exposure and deprivation: Evidence from the Burundi civil war. *World Development* (forthcoming).

Morris, S., Levin, C., Armar-Klemesu, M., Maxwell, D., & Ruel, M. T. (1999). Does geographic targeting of nutrition interventions make sense in cities? Evidence from Abidjan and Accra. *World Development, 27*(11), 2011–2019.

Shemyakina, O. N. (2017). Political violence and child health: Results from Zimbabwe. *World Development* (forthcoming).

Thomas, L. C. (2009). *Consumer credit models: Pricing, profit, and portfolios*. Oxford University Press.

Tranchant, J.-P., Bliznashka, A. G. L., Diallo, A. S., Sacko, M., Assima, A., Siegel, E. H., et al. (2017). The impact of food assistance on food insecure populations during conflict: Evidence from a quasi-experiment in Mali. *World Development*, forthcoming.

van Weezel, S. (2017). Food security and armed conflict: a cross-country analysis, School of Economics, University College Dublin, mimeo. Available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2934177>.

Verme, P., Gigliarano, C., Wieser, C., Hedlund, K., Petzoldt, M., & Santacroce, M. (2016). *The welfare of Syrian refugees: Evidence from Jordan and Lebanon*. Washington, DC: Sage. https://openknowledge.worldbank.org/handle/10986/23228 License: CC BY 3.0 IGO.

World Food Programme (WFP) (2014). Verification exercise, preliminary results, March 2014. WFP, Rome. Available at: <http://foodsecuritycluster.net/document/wfp-verification-exercise-preliminary-results-march-2014>.

Wodon, Q. (1997). Targeting the poor using ROC curves. *World Development, 25*(12), 2083–2092.

Please cite this article in press as: Verme, P., & Gigliarano, C. Optimal targeting under budget constraints in a humanitarian context. *World Development* (2018), https://doi.org/10.1016/j.worlddev.2017.12.012