High Impact Profiling: Estimating Heterogeneous Treatment Effects Amongst Kenyan Aid Programs

Keaton Corder-Swanson, Nick Hughes,
Che Lopez, Tiadora Miric,
Max Sohl, and Phoebe Vavoulis

June 10, 2022

Abstract

A sophisticated understanding of the impact of humanitarian aid programs on beneficiaries can help drive enhanced program design. Focusing our inquiry on heterogeneous treatment effects, we analyze Kenyan refugee household survey data from 2017 to 2021 to ascertain the impact of aid programs implemented by the United Nations High Commissioner for Refugees. By examining the data with both logistic (logit) and causal forest models, we determine the impact of this aid on refugees' ability to increase income and savings. We find that the selected aid program is minimally correlated with refugee monetary gain. Further analysis identifies relationships between individual traits and the level of impact an individual is likely to experience. Some of these relationships offer actionable insight, while others are more nuanced. We conclude by identifying key characteristics of 'high impact' beneficiaries and make suggestions for future research.

Keywords: Africa, Kenya, Refugee, Humanitarian Aid, Causal Forest, Logit, Regression, Econometric Analysis

1. Introduction

After being forced to flee one's home country due to conflict or to achieve a higher standard of living, it is their livelihood that may take the longest time to replenish and improve. With refugees being one of the most vulnerable groups in the labor market, and often over 50% of refugees worldwide being unemployed, refugee host communities have an opportunity to vastly improve their solution to this problem (UNHCR Livelihoods Information System, 2022). One of the most effective ways they can achieve this is by supporting refugees in rebuilding their lives with dignity and in peace through the opportunity to work and earn a living (UNHCR Livelihoods Information System, 2022).

In the following paper we will explore the effects of the UNHCR's implementation of a self-employment and business-related intervention program on the livelihoods of refugees in Kenya. Specifically, we will study the effects of the program on one's perceived increase in savings and income prior to and following implementation to answer the questions: does the implementation of self-employment and business-related intervention programs cause an increase in one's ability to save and earn money? And if so, which specific characteristics contribute the most to the probability that an individual benefits from the intervention program?

To study this question, we accessed the Integrated Refugee and Forcibly Displaced Livelihoods Information system shared publicly by the UNHCR, which monitors the impact of implemented livelihood programs through a set of global indicators. Some of these indicators include "increased access to livelihood assets," "increased self-employment," and "increased access to markets." These indicators are further broken down into sub-indicators such as "social assets provided," "access to financial services facilitated", "Income from wage employment facilitated", and "savings from wage employment facilitated". As mentioned, we will specifically focus on refugees displaced in Kenya to evaluate the effectiveness of the program on sub-indicators related to change in income and change in saving. It is the intent that the results drawn from the following study will be extended to future policy implementations relating to employment efforts made for refugees.

We retrieved data from the UNHCR information system in the form of Beneficiary survey responses. To measure the impact of the program, a survey was conducted once before program implementation and once after implementation to the same beneficiaries. The observations pre-program observations are considered our "control group" and the post-program observations are considered our "treatment group" for the purposes of our analysis. The selection of refugees for data collection was semi-random. That is, the selection for those who received treatment may not have been random, but the selection of beneficiaries who were chosen for data collection was random. This has important implications since this is not a true randomized control trial.

For our economic analysis, we will first present estimates for changes in savings and income which rely on a traditional logit model. We will then compare these to estimates computed by studying the heterogeneous effects of the program using a causal forest, a more innovative machine learning approach. The idea here is that our logit model will develop an estimate for an overall treatment effect on income and savings changes without considering heterogeneous effects, and we will use a causal forest approach as a robustness check, as well as to study how program implementation outcomes vary across different demographics in Kenya. We expected to find that the program has a positive correlation with an increase in savings and income, as well as significant heterogeneous effects which would help refugee aid agencies allocate funds in the most efficient way.

The initial result produced by our analyses is that the refugee aid program indeed has a positive effect on the probability that one increases their income and savings. In addition, both models agreed that the program had a greater effect on income than it did on savings. While the effects were quite small, a similar result was drawn from both models, indicating that the findings are robust. This tells us that the implementation of a refugee program which focuses on developing skills in self-employment and business development could have a positive impact on refugee camps all over the world. It is important to note, however, first that the effects are small and second that our results could be biased due to data collection methods as well as certain model intricacies that are discussed later in the paper.

The main result generated by the causal forest analysis provided a multitude of interesting and important heterogeneous effects. Most notably, the treatment effect was highest in individuals with low dependency scores. That is, individuals who benefit most from the program come from households where working members provide for fewer people within the household. In addition, individuals who are not married tend to benefit more from the program. A few additional characteristics were associated with higher treatment effects, including beneficiaries who are older, have a larger family size, and lower education levels. This result has important implications for the future of aid programs, as supporting agencies will be able to use this causal forest analysis to maximize the benefits received from the dollars spent on refugee program implementation. That is, agencies like the UNHCR can take these heterogeneous effects into account when selecting who receives aid and at what rate. This would ensure that decisions about investments into refugee communities are informed and optimal.

In the following pages, we will take you through the literature on logit and causal forest models for a better understanding of how these models are successful in similar types of analyses. Following that, we will go into specifics about the data that we used as well as empirical methodology. We then give a more detailed overview of logit and causal forest models followed by the results they produced. Lastly, we provide a brief discussion on the robustness check provided by the two models.

2. Causal Forest Analysis and Logit Models of Variable Impact

Many studies involving refugees examine the process that asylum-seekers must go through to assimilate to a new culture. Studies about the English proficiency of refugees are paramount with logit models. Elements of life before and after immigration factor into a refugee's ability to successfully learn and speak English, particularly in the United States (Espenshade and Fu, 1997). However, age and education were both major factors in this study that indicated the probability that a given non-native English speaker would pick up the language. Fennelly and Palasz (2004) reinforced this idea when they found that asylum-seekers with college diplomas, "were more than 29 times more likely than non-high school graduates to speak English well, and more than 20 times as likely to read well." Like our logit model, age and education are also contributing variables that play a role in predicting a refugee's ability to generate and retain income.

While logit models have been used for decades, the use of causal forest analysis in economic studies is relatively new. Athey and Imbens (2016) pioneered this space with their study on the effect of intervention on students' academic achievements. We first sort our data into bins and then examine the average characteristics of the given individuals within each bin. We model our methodology done by the work of Athey and Imbens to uncover the conditional treatment effect by looking at high and low impacted groups of individuals. Because causal forest analysis aims to uncover the heterogeneous treatment effects, many of these studies look at sociological or health-related impacts. Elek and Bíróa (2021) explored the relationship amongst regional differences of diabetes in Europe and how they related to socio-economic, health and lifestyle variables. Their causal forest model uncovered strong heterogenous regional differences and determined that there is a much higher chance of having diabetes in Eastern and Southern Europe than there is in the West. Causal forest analysis was used to test multiple datasets to determine the most significant dependency relationships among clinical variables for a thyroid disease diagnosis (Wang et al., 2017). Like our project, the authors of this study chose to compare their model to other types of regressions, finding that their causal forest model was more accurate than a variety of Bayesian network classifiers. However, our project expands on this by instead using the same set of data between multiple models to check the robustness of each regression.

Our study presents a novel approach to combining logit and causal forest by examining the effect of humanitarian aid on Kenyan refugees' abilities to generate income and increase their savings. We contribute to literature through our unique use of causal forest analysis. While causal forests are great for their analysis of heterogeneous treatment effects, our study is unique because to the best of our knowledge, we are one of the first to apply this analysis to refugee data. In addition, many causal forest studies analyze the effect of discrete treatment variables, but our study is unique in that we also analyze a discrete outcome.

3. Data

Our analysis utilizes survey data from the United Nations High Commissioner for Refugees (UNHCR) "Integrated Refugee and Forcibly Displaced Livelihoods Information System". The survey is designed to monitor the impact of aid programs through a set of global indicators. Data is available for several countries for cross-country comparison: however, this paper studies the effects of an aid program specifically on asylum seekers, refugees, internally displaced persons (IDPs) and citizens of Kenya from 2017-2021.

We explore the effects of the program on two dependent variables, which include (1) whether the beneficiary saw a change in income relative to the previous year and (2) whether they saw a change in savings relative to the previous year. Our control group consists of the data observations assigned to the 'Baseline' group, which are beneficiaries who have been selected for aid programs but have not yet participated in an intervention at the time of the survey. The treatment group consists of the data observations assigned to the 'Endline' group, which are the beneficiaries who have completed participation in an aid program at the time of the survey. The sampling methodology for survey respondents follows a criterion laid out by the UNHCR's

Livelihood Monitoring Framework. For each aid program, if the number of beneficiaries was less than 100, then all beneficiaries were surveyed, otherwise, 10% of the total beneficiaries of the program were used as the sample size (whichever number was greater). The sample beneficiaries were selected as randomly as possible (i.e., if a beneficiary list was available, every household was pre-selected, and if a list was not available, every X^{th} household receiving the intervention was selected). Unfortunately, we lack the initial selection criteria for beneficiaries. This means that we can reasonably assume that our data is the product of a random sample of beneficiaries, but the beneficiary pool is not a random sample of the refugee and host community population at large.

Covariates included in our analyses are gender, refugee status, household size, household dependency ratio, country of origin, marital status, education level, and the year the survey was conducted. Household dependency ratio is defined as the number of family members ages 0-14 or over 64 divided by the number of family members ages 15-64 that are contributing to household income. Education level was recorded from the respondents as categorical data, ranging from no schooling up to post-university degrees. To perform our analysis, we transformed education into the average number of years of schooling based on reported education level, as shown in Table 1. These values represent our best estimation of years of education and do not attempt to mimic true population statistics. Rather, they proxy the categorical responses from the respondents to allow for more robust interpretation of results. The inclusion of the year that the survey was conducted assists to capture macroeconomic factors in the country that affect the ability to earn and save, like varying unemployment rates and national GDP.

The aid programs covered in our dataset are divided into three main intervention categories: Agricultural Production, Self-Employment/Business, and Wage Employment. After aggregating all the observational data available between 2017 and 2021 we found that 454 were missing intervention assignment. This is due to participants dropping out of the program after initial baseline surveys, which causes the concern of attrition bias. The remaining 11,042 respondents were assigned to the Self Employment/Business category, designated as the “O2” intervention type in the dataset. Our analysis focuses on survey respondents who belong to the Self Employment/Business category, as all the other programs have sparse data and far too few observations for any quantitative analysis to be feasible. An interesting analysis could be performed on all three parent categories but availability and access to complete data is a limiting factor that must be acknowledged.

4. Empirical Methods

The focus of this paper is to explore the effect of the UNHCR's Kenya refugee aid programs on survey respondents' ability to earn income and produce savings. Refugee aid programs address the needs of their beneficiaries using a wide variety of approaches and levers for change. Within our selected intervention type, Self-Employment/Business, the UNHCR facilitated various training programs, disbursed cash grants as well as productive assets for income-earning activities. With the objective to build capacity for self-reliance, which is indicated by the beneficiary's ability to increase their income or savings relative to before the intervention was rendered. Should a respondent that took part in the program report an ability to earn more income or to save more money, this indicates that the program was effective. Conversely, should a respondent that took part in the program report no change or a decreased capacity for earning or saving, this would indicate that the program was ineffective. Following this logic, we assign our outcome variable as a 0 if the respondent reported a diminished or unchanged capacity and a 1 if the respondent reported an increased capacity.

$$Y_i = \begin{cases} 1, & Y_i > 0 \\ 0, & Y_i \leq 0 \end{cases} \tag{1}$$

In this paper we analyze two outcomes relative to this indicator variable. The change in an individual's income and savings, denoted by $Y_{i,I}$ and $Y_{i,S}$.

4.1 Logit Model

A logit model is well suited for our analysis and has the advantage of being easy to understand and reproduce. An ordinary least squares (OLS) regression would result in a biased and inaccurate model as there is multicollinearity within our data. Logistic regressions solve for classification and can predict the outcome of a categorical dependent variable. The estimated average marginal effects from a logit model can be compared to estimates produced in the causal forest model. This allows us to conduct a deeper analysis and produce stronger claims from the indicators and their correlations. We estimate the logit model as,

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_1 X_1 + \cdots + \beta_n X_n \tag{2}$$

where $\frac{\pi}{1-\pi}$ represents the probability of a change in an individual's savings or income. To find π, we used the equation:

$$\pi = \frac{\exp(Y)}{\exp(Y) + 1} \tag{3}$$

Each β estimate does not represent the estimated effect of the covariate. It can only tell us if the effect is positive or negative. Instead, we must look at the average marginal errors (AMEs) to understand the estimated effect of the covariate X_i on the outcome. Once the AMEs are calculated, we can use those estimates to compare against the results of the causal forest regression. Under the unconfoundedness assumption, we interpret these estimates as causal effects. The logit AMEs come from the analytical derivative and computing the average effect of each X_i on π.

We aimed to measure the effect of humanitarian aid on individual monetary gain and retention. So, we decided to run two logit models. One looked at the change in an individual's savings while the other looked at the change in an individual's income. As with the causal forest model, we began preparing the data for the logit model by creating dummy variables for all data that was non-numerical. So, instead of an individual with 'Syria' as their 'Country of Origin' value, a new indicator titled 'Syria' was created and the same individual received a value of '1'. Additionally, instead of 'Man' or 'Woman', the 'Gender' variable became '0' for men or '1' for women. This resulted in a total of 138 covariates that could possibly inform our model. Our data included 9075 observations, but after breaking out many of these factors into columns of dummy variables, many of the columns would include less than 30 observations that differed from the others. For example, out of 9075 observations, there was only one observation that indicated that the refugee had arrived at camp in 1986, and only two people who arrived from Angola. This created a very sparse data frame.

Logit models are not robust to sparse data. The low variance of many of the columns in our data frames would often cause the model to confuse the dichotomous variable for a constant, which would halt the operation of the model. To eliminate this problem, we implemented code that created a variance threshold for the covariates that could be included in the model. Setting the variance to .002 reduced the number of covariates to 84. We chose our variance threshold through iterative testing and found that any threshold that was set to .001 or smaller would result in covariates with data that was too sparse to successfully run our logit model. Extremely low variance indicated that a given covariate would likely not produce a large effect on the outcome variable, which made it safe to exclude from the model. A full list of variables that were included and excluded in the income and savings logit models can be found in Appendix 1.

4.2 Causal Forest Model

To gain a richer understanding of the impact aid programs have on refugees, we focus our secondary analysis on estimating the heterogeneous effects across varying demographics of refugee respondents. Understanding the distribution of conditional treatment effects across demographics provides value in two keyways. First, by identifying 'high-impact' beneficiaries, aid organizations can tailor the selection process employed in future programs to maximize impact per dollar spent. Second, identifying 'low-impact' demographics allows for targeted program reform to benefit the lives of underserved beneficiaries more effectively. We label respondents 'high impact' if they demonstrate an estimated conditional treatment effect relatively higher than other demographics. Conversely, demographics that align with lower conditional treatment effect estimates are deemed 'low-impact' beneficiaries.

We perform our analysis utilizing an ensemble method which combines the causal forest methodology developed by Athey and Imbens (2016) with double machine learning (DML) residualization from the work of Chernozhukov et al (2017). Causal forests and DML offer an opportunity to employ cutting edge methods from recent literature that merges econometric analysis with machine learning. Firstly, causal forests are ultimately an extension of generalized random forests (Athey et al., 2019), which are known to be highly accurate, generalized prediction methods well-suited to models with large numbers of features (Biau, 2012). Another feature of random forests is that they can produce consistent estimates even in the presence of sparse data, although these estimates' rate of convergence does hinge on the presence of features with demonstrable predictive power (Biau, 2012). This is favorable for our data which contains many indicator variables and displays sparsity. Finally, DML is a simple, robust method for estimating approximately unbiased and normally distributed treatment effects. It is a generally simple framework which makes it approachable and easy to implement in tandem with many other Machine Learning techniques (Chernozhukov, 2017).

Applied to our refugee dataset, we can estimate the conditional treatment effect for each individual. We then sort all the observations in our sample into 5 bins based on the varying estimated heterogeneous treatment effects. This allows us to inspect the average characteristics of the individuals within each bin and explore the differences observed between 'high impact' and 'low impact' bins. We calculate Shapley values from our model, adding another dimension to our analysis of which beneficiary features are identifiers of 'high impact' and 'low impact' individuals. Shapley values offer a popular framework for explaining the contribution of features to model predictions in a manner that is comprehensible even to readers that are not well versed in machine learning (Fryer et al., 2021). Finally, we use the Logit and Causal Forest models to estimate the constant treatment effect across all individuals, comparing these estimates for a robustness check of our methods.

5 Results

5.1 Logit Results

The results of our income and savings logit models reveal that UNCHR aid programs may have little positive correlation with increase refugees' income and savings. After running the logit model and estimating the AMEs, we found that all the coefficients on the effects for both models were estimated to be relatively small. Specifically, the treatment variable in both models was estimated to be smaller than 0.1. This would indicate that the aid provided to refugees in Kenya likely had a very minimal effect on their ability to generate income and retain savings. Summaries of the regression results for both income and savings change are shown in Table 2 and Table 3, respectively. For simplicity, only the treatment variable and the variables with the largest estimated effects are shown in Table 2 and Table 3.

The income and savings logit models produced conditional treatment effect estimates for each of the individuals in our dataset that ranged from -0.3387 to 0.2415 for the savings model and from -0.3789 to 0.2073 for the income model. In addition, the largest 95% confidence interval in the logit income model produced by an individual is between -0.567 and 0.346. The largest 95% confidence interval in the logit savings model produced by an individual is between -0.372 and

0.624. In addition, all standard errors across both models were no larger than 0.254 for each estimated conditional treatment effect. These small widths of 0.913 and 0.996, along with the low standard errors indicate that our model likely produced very precise estimates of the variable effects.

Having an arrival year of 1994 was estimated to be one of the largest indicators of having both savings and income decrease or stay the same. This could be for several reasons. Potentially, people who arrived in 1994 would be assumed to be more established than people who arrived more recently. In this case, being more established would give access to resources that might allow for an individual to increase savings or income more readily, given the help from the aid program. For the income model, having taken the survey in 2021 was also an indicator correlated with decreasing or stagnant income. In reporting these results, it is important to acknowledge that the low-variance variables that were left out of both of our logit models likely create biased estimates.

5.2 Causal Forest Results

We found the intra-bin average of feature values across all observations. creating approximate profiles for individuals in each bin. Most features varied across bins in a manner that rendered little insight. A select few differed in an interpretable manner between 'high impact' and 'low impact' beneficiaries. We designate these as our 'Features of Interest', which include age, gender, dependency score, marital status, and our indicator variables for the residency status of individuals, for which we included the categories of refugee, asylum seeker and host community member. In Figure 1 and Figure 2, we graph the distribution of Features of Interest across bins for a visual representation of the relationships encountered in the data.

Across the bins, certain features have average values at the extremes (in Bins 1 and 5) which are markedly different from the central three bins (Bin 2, 3 and 4). We refer to features with this characteristic as 'Divisive Features'. The level of education, measured in years of schooling, is a divisive feature. Bins at the extremes contain observations with generally low education levels, while the center three bins all contain beneficiaries with generally high levels of education. Within the center three bins, there is a clear positive relationship between the years of education that an individual has and the benefits rendered by the aid program. Gender is another divisive feature, with a clear divide between the outer and inner bins. The majority of individuals in the highest and lowest impact bins are women, while the center three bins have equal representation between the genders. While these Divisive Features do not offer actionable insight, per se, they do serve as a starting point for future inquiry.

Bins 4 and 5 contained individuals who averaged the two lowest average dependency score values relative to the other three bins. This makes intuitive sense: individuals who benefit most from the program come from households where working members provide for fewer people within the household. Although there is not a drastic difference between bins in terms of marital status, there is a slight negative trend. Greater benefits are generally realized by individuals who are not married. These two Features of Interest indicate that a high impact beneficiary can be described as unmarried and belonging to household with a low dependency score. As a note, due to arrival year being captured by many indicator variables, it was omitted from the Features of Interest list because of computational challenges in reporting average values for each bin due to the sheer diversity in years of arrival. The opposite is true for legal status, which is also represented by indicator variables, but instead, all of beneficiaries belong to just three groups: refugees, asylum seekers and members of the host communities.

The composition of beneficiaries' legal status varied distinctly across bins. As we move from low impact bins to high impact bins, we see that refugees make up a decreasing percentage of the population within each bin. The opposite is true for asylum seekers. In Bins 1 and 2 (our low impact bins), the number of asylum seekers was negligible. Their representation jumps to around 10% of the population in Bins 3 and 4 and jumps again to approximately 40% of the population for our high impact bin, Bin 5. Finally, host community members have representation in each of the bins in a manner that offers little intuitive insight. It could be characterized as a mildly divisive feature, where most host community members fell into the center three bins. This would indicate that host community members benefit from the program at a level that is closer to the average treatment effect. Overall, we can conclude that regarding legal status, high impact beneficiaries are asylum seekers.

Figure 3 demonstrates clear relationships between a handful of key features and the predicted individual effects on the ability of beneficiaries to increase their income. Some of this information is hard to leverage in aid program design. The year the survey was conducted, for instance, has a clear relationship with program effectiveness, but cannot feasibly be integrated into improved program design. Figure 3 establishes that the age, family size and education level of beneficiaries are relatively stable predictors of program effect, all of which can aid in effective program design. Age and education level have more consistent effects, where older beneficiaries with higher levels of education tend to see a larger impact. With family size, we see high values (in red) on both tails of the cluster, indicating that these values swayed the estimated effects of the program in both directions. Values falling near the center depict a clearer picture, with larger families generally seeing a higher impact from the programs.

Figure 4 demonstrates clear relationships between a handful of key features and the predicted individual effects on the ability of beneficiaries to increase their savings. As with income, particular survey years were predictive of impact levels, but aid program design cannot feasibly incorporate the macroeconomic factors driving these relationships as they fluctuate year to year. We focus again on age, family size, and education levels. The age of the beneficiary contributed to estimated treatment effect in no clear manner, with all ages being represented across the spectrum of treatment effects. Family size has a clear positive relationship with the estimated treatment effects, where larger families generally saw higher estimated treatment effects and vice versa. Education level had a mixed effect near the mean, but a clear effect at the tails, where high education levels were associated with low treatment effects and vice versa.

Average treatment effects for the aid program were estimated in both the logit and causal forest model and are reported in Table 5 along with their standard errors. Comparing the values serves as a robustness check for our two models. All average treatment effects calculated were positive, with both models indicating that the aid program had a higher effect on income than it did on savings. The parametric logit model produced lower standard errors than the non-parametric causal forest, which is to be expected. This brief but effective comparison reinforces confidence in both the integrity of our models, as well as the notion that the UNHCR's aid programs in Kenya had a positive, albeit small, average effect on the income and savings of the beneficiaries they reached.

6. Conclusion

In this paper, we establish a richer understanding of the impact that the UNHCR's refugee aid program in Kenya has on increasing beneficiary income and savings. Our Logit model serves as a traditional econometric approach, estimating that the program increased the probability of greater income by 9.01%. In addition, the aid was estimated to increase the probability of having greater savings by 7.13%. These treatment effect estimates are consistent with the estimated treatment effect in the causal forest model, which served as a complementary, machine learning technique. Under the causal forest model, we estimate that the aid program increased the chances of an individual increasing their income and savings by 12% and 9%, respectively.

While the logit identified the beneficiary's year of arrival as the best predictor of the conditional treatment effect, the causal forest model focused on a batch of Features of Interest across binned conditional treatment effects. These Features of Interest showcased the clearest relationships between treatment effect and the legal status of the beneficiary, their education level, gender, dependency score and marital status. The relationships are nuanced, with high impact individuals most easily being characterized as unmarried asylum seekers with low household dependency scores. Finally, Shapley values calculated add another dimension, indicating that older beneficiaries with larger households and lower levels of education would have higher conditional treatment effects for earning income. The same is true for income regarding family size and education levels, but age is ineffective as a predictor of conditional treatment effects.

The disparity between Shapley values and binned analysis from the causal forest model draws intrigue and is a stepping off point for future inquiry. In aid and development work at large, an opportunity lies in exploring the nuances of the characteristics we identified as Divisive Features in the binning analysis. Identifying the factors that differentiate the demographics consistently found at both extremes of the range of conditional treatment effects could prove to be pivotal in developing robust, targeted interventions. Finally, analyses performed in the development space suffer from data scarcity and analytically naïve selection of metrics that limit the researchers' capacity to produce insight of value. We recommend and encourage aid and development organizations to integrate quantitative specialists in the design of their surveys and evaluation methods to create richer and more robust datasets that are suited for rigorous analysis from the start.

References

Athey, Susan, and Guido Imbens. "Recursive Partitioning for Heterogeneous Causal Effects." *Proceedings of the National Academy of Sciences* 113, no. 27 (July 5, 2016): 7353–60. https://doi.org/10.1073/pnas.1510489113.

Athey, Susan, Julie Tibshirani, and Stefan Wager. "Generalized Random Forests." *The Annals of Statistics* 47, no. 2 (2019). https://doi.org/10.1214/18-aos1709.

Biau, Gerard. "Analysis of a Random Forests Model." *Journal of Machine Learning Research* 13 (April 2012): 1063–95.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. "Double/Debiased Machine Learning for Treatment and Structural Parameters." *The Econometrics Journal* 21, no. 1 (January 16, 2017): C1–C68. https://doi.org/10.3386/w23564.

Elek, Péter, and Anikó Bíró. "Regional Differences in Diabetes across Europe – Regression and Causal Forest Analyses." *Economics & Human Biology* 40 (January 2021): 100948. https://doi.org/10.1016/j.ehb.2020.100948.

Espenshade, Thomas J., and Haishan Fu. "An Analysis of English-Language Proficiency among U.S. Immigrants." *American Sociological Review* 62, no. 2 (April 1997): 288–305. https://doi.org/10.2307/2657305.

Fennelly, Katherine, and Nicole Palasz. "English Language Proficiency of Immigrants and Refugees in the Twin Cities Metropolitan Area." *International Migration* 41, no. 5 (January 23, 2004): 93–125. https://doi.org/10.1111/j.0020-7985.2003.00262.x.

Lis.unhcr.org. 2022. *UNHCR Livelihoods Information System*. [online] Available at: <https://lis.unhcr.org/home>.

Wang, LiMin, FangYuan Cao, ShuangCheng Wang, MingHui Sun, and LiYan Dong. "Using K-Dependence Causal Forest to Mine the Most Significant Dependency Relationships among Clinical Variables for Thyroid Disease Diagnosis." *PLOS ONE* 12, no. 8 (August 17, 2017). https://doi.org/10.1371/journal.pone.0182070.

Appendix 1

Variables excluded from the income and savings logit models:

(Note: variables that are simply a country name indicate a refugee's country of origin)

> 'Algeria', 'Angola', 'ArrivalYear_1976', 'ArrivalYear_1986', 'ArrivalYear_1989', 'ArrivalYear_1990', 'ArrivalYear_1990.1', 'ArrivalYear_1993', 'ArrivalYear_1993.1', 'ArrivalYear_1994.1', 'ArrivalYear_1995', 'ArrivalYear_1995.1', 'ArrivalYear_1996.1', 'ArrivalYear_1997', 'ArrivalYear_1999', 'ArrivalYear_2001', 'ArrivalYear_2002', 'ArrivalYear_2020.1', 'BE', 'Bahamas', 'Bermuda', 'Brunei_Darussalam', 'Cameroon', 'Canada', 'Central_African_Republic', 'Cook_Islands', 'CountryOrigin', 'Dominican_Republic', 'Educ_PostUniversity_No_Diploma', 'Equatorial_Guinea', 'Eritrea', 'Estonia', 'Gender', 'Germany', 'Guinea', 'Host_Community', 'IDP', 'Income_Change', 'Kenya', 'LegalStatus', 'MaritalStatus', 'Mauritania', 'Monaco', 'Naturalised', 'Oman', 'Panama', 'Returnee', 'Saving_Change', 'South_Africa', 'St._Pierre_and_Miquelon', 'Stateless', 'Tanzania', 'Virgin_Islands_(U.S.)'

Variables included in the income and savings logit models:

(Note: 'Year' variables indicate the year that the survey was conducted)

> 'Age', 'FamilySize', 'Dependency', 'treatment', 'gender_indicator', 'Married_ind', 'Year_2017.0', 'Year_2018.0', 'Year_2019.0', 'Year_2020.0', 'Year_2021.0', 'Educ_Middle', 'Educ_Middle_No_Diploma', 'Educ_No_Educaiton', 'Educ_PostUniversity', 'Educ_Primary', 'Educ_Primary_No_Diploma', 'Educ_Secondary', 'Educ_Secondary_No_Diploma', 'Educ_TVET', 'Educ_TVET_No_Diploma', 'Educ_University', 'Educ_University_No_Diploma', 'ArrivalYear_1991', 'ArrivalYear_1992', 'ArrivalYear_1994', 'ArrivalYear_1996', 'ArrivalYear_1998', 'ArrivalYear_2000', 'ArrivalYear_2003', 'ArrivalYear_2004', 'ArrivalYear_2005', 'ArrivalYear_2006', 'ArrivalYear_2007', 'ArrivalYear_2008', 'ArrivalYear_2009', 'ArrivalYear_2010', 'ArrivalYear_2011', 'ArrivalYear_2012', 'ArrivalYear_2013', 'ArrivalYear_2014', 'ArrivalYear_2015', 'ArrivalYear_2016', 'ArrivalYear_2017', 'ArrivalYear_2018', 'ArrivalYear_2019', 'ArrivalYear_2020', 'ArrivalYear_2021', 'ArrivalYear_1991.1', 'ArrivalYear_1992.1', 'ArrivalYear_1997.1', 'ArrivalYear_1998.1', 'ArrivalYear_1999.1', 'ArrivalYear_2000.1', 'ArrivalYear_2001.1', 'ArrivalYear_2002.1', 'ArrivalYear_2003.1', 'ArrivalYear_2004.1', 'ArrivalYear_2005.1', 'ArrivalYear_2006.1', 'ArrivalYear_2007.1', 'ArrivalYear_2008.1', 'ArrivalYear_2009.1', 'ArrivalYear_2010.1', 'ArrivalYear_2011.1', 'ArrivalYear_2012.1', 'ArrivalYear_2013.1', 'ArrivalYear_2014.1', 'ArrivalYear_2015.1', 'ArrivalYear_2016.1', 'ArrivalYear_2017.1', 'ArrivalYear_2018.1', 'ArrivalYear_2019.1', 'Burundi', 'Congo', 'Congo_DR', 'Ethiopia', 'Rwanda', 'Somalia', 'South_Sudan', 'Sudan', 'Uganda', 'Asylum_Seeker', 'Refugee'

Table 1. Transformation of years of education

Reported Education	Data Input
None/Never attended	0
Primary School, no diploma/certificate	3
Primary School with diploma/certificate	5
Middle School, no diploma/certificate	6.5
Middle School with diploma/certificate	8
Secondary School, no diploma/ certificate	10
Secondary with diploma	12
Technical and Vocational Education and Training (TVET) School, no diploma	12.5
Technical and Vocational Education and Training (TVET) School with diploma	13
University, no diploma	14
University with diploma	16
Post University, no diploma	17
Post University, with diploma	18

Table 2. The Effect of Aid on Income (Logit)

	Estimate	Standard Error	95% Confidence Interval
Arrival Year: 1994	-0.3789	0.174	[-0.720, -0.038]
Survey Year: 2021	-0.3161	0.232	[-0.770, 0.138]
Treatment	0.0907	0.011	[0.070, 0.112]

N = 9075

Table 3. The Effect of Aid on Savings (Logit)

	Estimate	Standard Error	95% Confidence Interval
Arrival Year: 1994	-0.3387	0.177	[-0.686, 0.009]
Education: Middle School	-0.3059	0.205	[-0.707, 0.095]
Treatment	0.0713	0.011	[0.050, 0.093]

N = 9075

Table 4. Causal Forest Bins

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
Average (Income)	-0.615	-0.186	0.242	0.67	1.10
Average τ (Savings)	-0.841	-0.339	-0.034	0.313	0.805

Table 5. Comparing models

	Income		Savings	
Average Treatment Effect	Estimate	Standard Error	Estimate	Standard Error
Logit	0.0907	(0.011)	0.0713	(0.011)
Causal Forest	0.1288	(0.131)	0.0983	(0.135)

Figure 1. Legal Status of Beneficiaries Across Bins

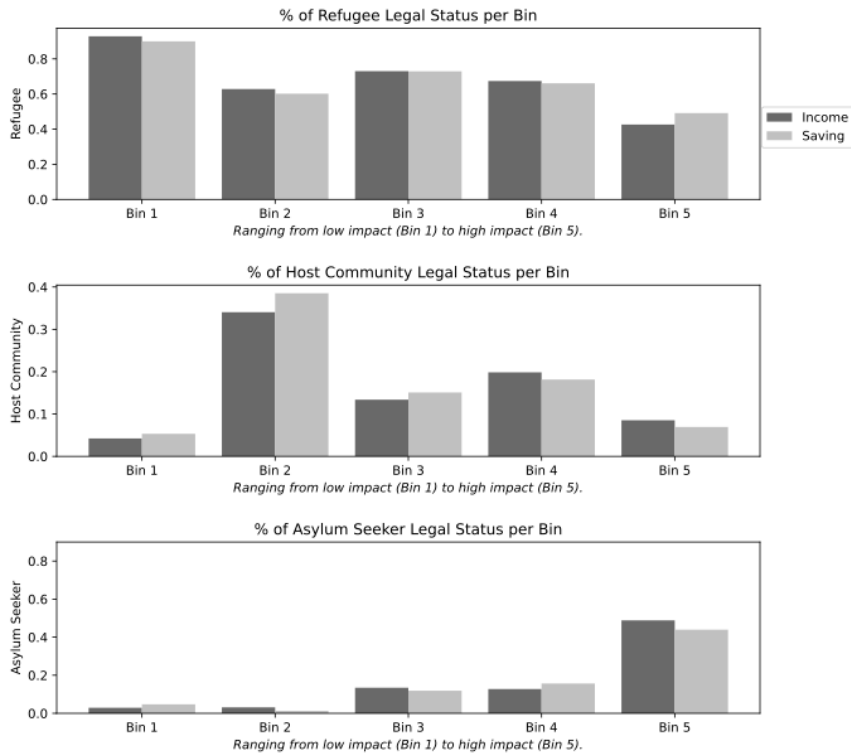

Figure 2. Features of Interest Across Bins

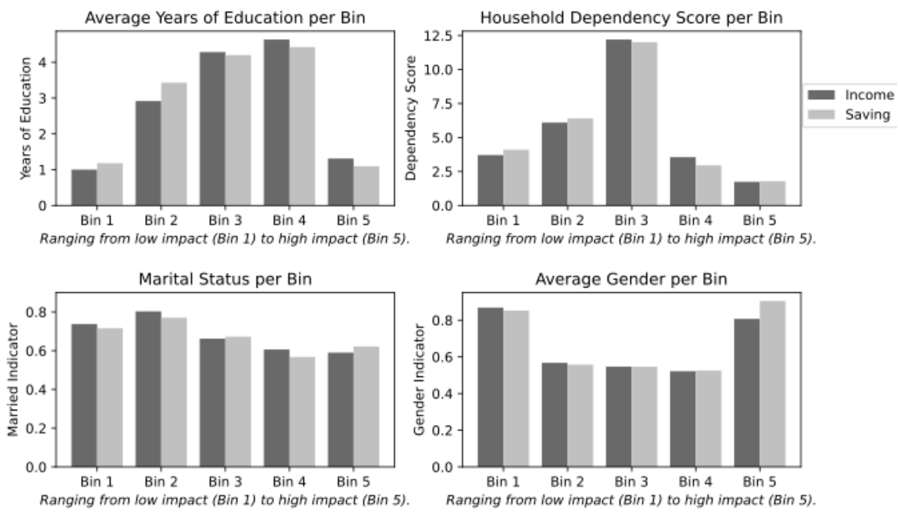

Figure 3. Predictors of program effectiveness

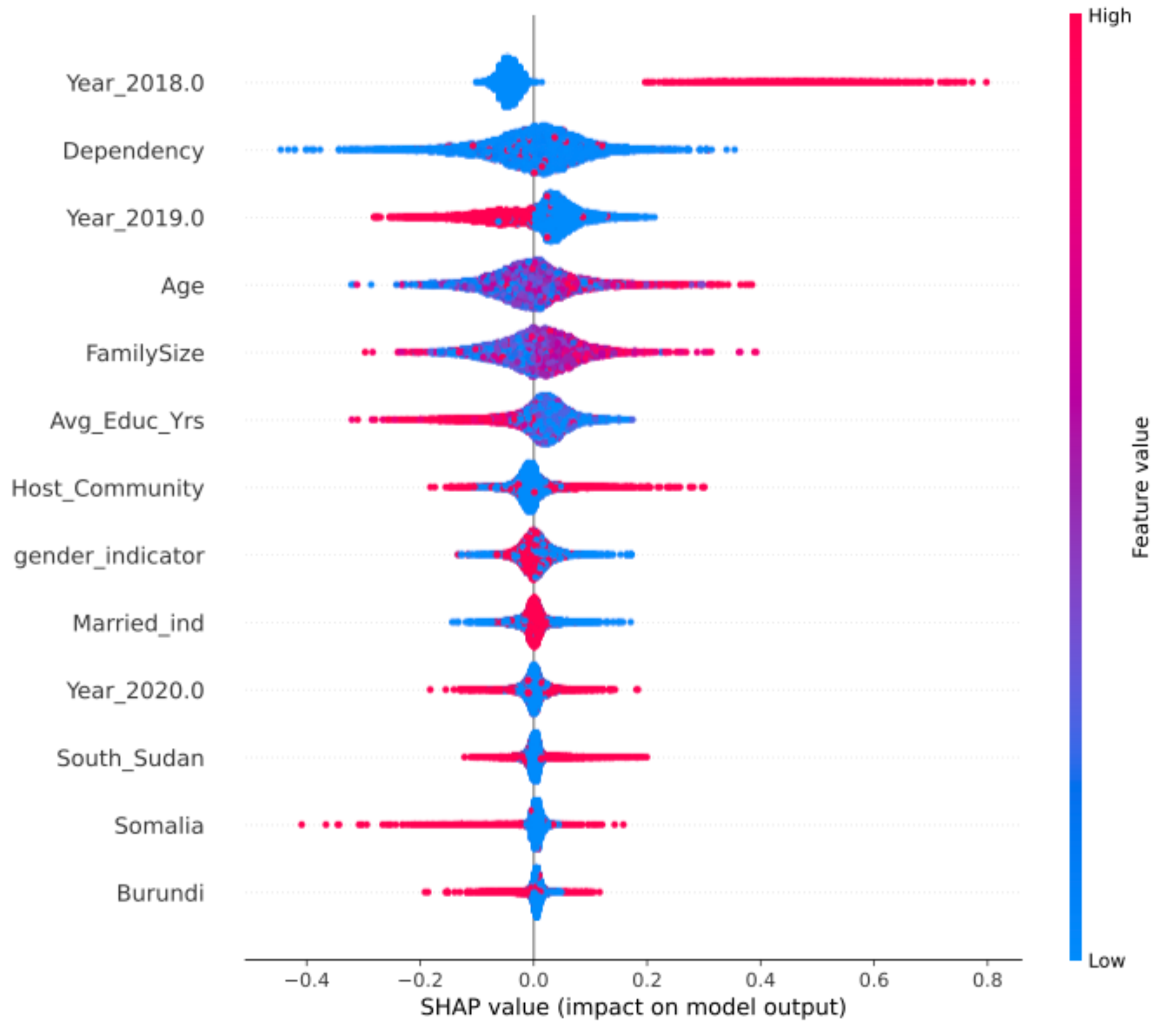

Figure 4. Key features and effects