Working Paper 2018:23

Department of Economics
School of Economics and Management

Placement Optimization in Refugee Resettlement

Andrew C. Trapp
Alexander Teytelboym
Alessandro Martinello
Tommy Andersson
Narges Ahani

October 2018

Lund University

Placement Optimization in Refugee Resettlement

Andrew C. Trapp
Worcester Polytechnic Institute
Worcester, MA USA
atrapp@wpi.edu

Alex Teytelboym
University of Oxford
Oxford, UK
alexander.teytelboym@economics.ox.ac.uk

Alessandro Martinello
Lund University
Lund, Sweden
alessandro.martinello@nek.lu.se

Tommy Andersson
Lund University
Lund, Sweden
tommy.andersson@nek.lu.se

Narges Ahani
Worcester Polytechnic Institute
Worcester, MA USA
nahani@wpi.edu

Abstract: Every year thousands of refugees are resettled to dozens of host countries. While there is growing evidence that the initial placement of refugee families profoundly affects their lifetime outcomes, there have been few attempts to optimize resettlement destinations. We integrate machine learning and integer optimization technologies into an innovative software tool that assists a resettlement agency in the United States with matching refugees to their initial placements. Our software suggests optimal placements while giving substantial autonomy for the resettlement staff to fine-tune recommended matches. Initial back-testing indicates that *Annie* can improve short-run employment outcomes by 22%–37%. We discuss several directions for future work such as incorporating multiple objectives from additional integration outcomes, dealing with equity concerns, evaluating potential new locations for resettlement, managing quota in a dynamic fashion, and eliciting refugee preferences.

Keywords: Refugee Resettlement; Matching; Integer Optimization; Machine Learning; Humanitarian Operations

1. Introduction

In 2017, there were 18.5 million refugees—the highest number ever recorded—under the mandate of the United Nations High Commission for Refugees (UNHCR) (UNHCR 2018). Of those, the UNHCR considers 1.2 million refugees to be in need of *resettlement*—permanent relocation from their asylum country to a third country (UNHCR 2017). Refugees in need of resettlement are particularly vulnerable: a quarter are survivors of torture and a third face persecution in their country of origin (UNHCR 2017, Annex 3). Currently, most refugees departing for resettlement are Syrians who seek asylum in Jordan and Lebanon, but there are also thousands of resettled refugees from the Democratic Republic of the Congo, Iraq, Somalia, and Myanmar. In 2016, the number of resettlement submissions reached 165,000 (a twenty-year high) and 125,800 people departed for resettlement (UNHCR 2017).

Dozens of countries, including the United States (US), Canada, the United Kingdom (UK), Australia, France, Norway, and Sweden, resettle refugees.[1] There is ample empirical evidence that the initial placement of refugees within the host countries determines their lifetime employment, education, and welfare outcomes (Åslund and Rooth 2007, Åslund and Fredriksson 2009, Åslund et al. 2010, 2011, Damm 2014, Ferwerda and Gest 2017). Therefore, ensuring the optimality of the initial match between the refugee family and the community is crucial for social, economic, and humanitarian perspectives. However, resettlement capacity offered by communities is rarely being used to maximize either the welfare of refugees or of the host population.

This paper integrates machine learning and integer optimization technologies into the software *Annie* MOORE (Matching and Outcome Optimization for Refugee Empowerment), named after Annie Moore, the first immigrant on record at Ellis Island, circa 1892. *Annie* is, to the best of our knowledge, the first software designed for resettlement agencies pre-arrival staff to recommend data-driven, optimized matches between refugees and local affiliates while respecting refugee capacities. *Annie* was developed in close collaboration with representatives from all levels of Hebrew Immigrant Aid Society (HIAS), where a first version was deployed in May 2018. New features were regularly added until August 2018 when it was presented to the US State Department and all staff at HIAS.

We combined techniques from operations research, machine learning, econometrics, and interactive visualization to create *Annie*. The software is distinctive in that it blends rigorous analysis with careful attention to the detail of the day-to-day resettlement process for resettlement staff. As such, *Annie* integrates the generation of data-informed recommendations with substantial autonomy by the end-user. This flexibility empowers staff to focus their resources on difficult cases (for example due to complex medical conditions). Back-testing indicates that *Annie* would have been able to increase employment outcomes among refugees resettled by HIAS in 2017 by between 22% and 37%,

[1]For refugee allocation mechanisms *across* countries, see Moraga and Rapoport (2014) and Jones and Teytelboym (2017).

depending on the constraints activated by the agency staff. *Annie* also alleviates inefficiencies in the manual matching process, and holds much promise for future impact in refugee resettlement—both domestically and abroad—as well as for new applications, such as asylum matching.

The paper proceeds as follows. Section 2 describes the specific context of refugee resettlement in the US, and places our work in the greater context of humanitarian operations problems. Section 3 sets up the integer optimization model that guides the matching recommendations. In Section 4, we explain how we estimate counterfactual employment probabilities from data. Section 5 discusses the backtesting we conducted to validate our approach. Section 6 describes the implementation and features of our software, while Section 7 concludes and points to many directions for further work. Appendices include detailed data descriptions, as well as estimation procedures and diagnostics.

2. Background Context and Previous Work

HIAS primarily resettles refugees in the United States. Because *Annie* is presently used only in the United States, we briefly describe the US resettlement program.

2.1 Refugee Resettlement in the United States

The United States has historically been, by a wide margin, the world's largest destination of resettled refugees, with 78,340 admitted in 2016.[2] The refugee resettlement program is managed by the United States Refugee Admissions Program (USRAP) comprising the Bureau of Population, Refugees and Migration (PRM) of the US Department of State, the US Citizenship and Immigration Services (USCIS) of the US Department of Homeland Security, and the Office of Refugee Resettlement (ORR) of the US Department of Health and Human Services (HHS). Alongside the UNHCR and the International Organization for Migration, these agencies coordinate identifying refugees, conducting security checks, and arranging for travel funding from the refugees' destinations.

The actual matching of refugees to their initial placements is delegated to nine resettlement agencies, known as *voluntary agencies*.[3] HIAS is one of these agencies, and resettles around 5% of all refugees in the US.[4] The voluntary agencies are responsible for developing their own networks

[2]In terms of per capita refugee resettlement, the US is behind Canada, Norway, and Australia.

[3]The other agencies are: Church World Service (CWS), Ethiopian Community Development Council (ECDC), Episcopal Migration Ministries (EMM), International Rescue Committee (IRC), Lutheran Immigration and Refugee Services (LIRS), US Conference of Catholic Bishops (USCCB), US Committee for Refugees and Immigrants (USCRI), and World Relief Corporation (WR).

[4]HIAS resettled 3,702 refugees in 2016, and 2,038 refugees in 2017.

of *affiliates*—local communities that welcome refugees and help them integrate into a new life in the United States. Affiliates offer resettlement capacity voluntarily although affiliate capacity is monitored and approved by the US government. There are currently around 360 affiliates across the United States and HIAS operates 20 of them.

Voluntary agencies match refugees to affiliates during the resettlement process largely by hand. Resettlement staff from each agency meet weekly to select, in round-robin fashion, from a pool of "cleared for arrival" refugee *cases*. Each case consists in an immediate *family* of one or more members (we use *case* and *family* interchangeably). Roughly half of these cases have no relatives in the United States, making these cases especially vulnerable, as they typically have lack language skills, family support, and independent financial means. Therefore, the responsible agency must carefully leverage its affiliate network to inform their case selection. After each agency selects their set of weekly cases (families), staff manually assess—on a one-by-one basis—the feasibility and fit of families to locations in their network of affiliates. In addition to integration factors such as language and nationality feasibility, the fit between the affiliate and the family depends on various community capacities, such as refugee processing, housing availability, slots for school children, and English language instruction.

This manual process creates multiple inefficiencies that motivated the development of *Annie*. First, keeping in mind support attributes such as languages, nationalities, family composition, and medical needs for all affiliates is mentally taxing for the staff. This information overload often results in not meeting the needs of refugees and in stretching the provision capacity of the affiliates. Second, while established indicators exist to assess the degree to which a refugee has successfully integrated into their new surroundings, estimating and optimizing these welfare outcomes manually is prohibitive.[5] Hence, refugees are often not placed to the best available affiliate even according to well-defined outcome metrics. Third, inefficiencies arise from processing refugees sequentially throughout the year rather than assigning all arriving refugees to affiliates simultaneously. We show that *Annie* solves or mitigates each of these inefficiencies.

2.2 Related Literature

Our work builds on a number of recent contributions in humanitarian matching systems. One recent example of such a humanitarian matching system is a tool to match children in state custody

[5] Established indicators include employment and economic sufficiency, developed social networks, and civic engagement activities like voting, see, for example, Ager and Strang (2008), Lichtenstein et al. (2016).

to families for adoption used by the Pennsylvania Adoption Exchange (Slaugh et al. 2016). Bansak et al. (2018) first proposed to use machine learning and linear programming for refugee resettlement based on employment data from the US and Switzerland. Using a similar dataset to theirs, we expand on their estimation techniques, while extending their optimization methods. Our integer optimization model extends the multiple multidimensional knapsack model for refugee matching described in Delacrétaz et al. (2016) and Trapp et al. (2018). However, as we focus on outcome optimization, our work differs substantially from papers that suggested preference-based matching systems for refugee resettlement (Moraga and Rapoport 2014, Jones and Teytelboym 2016, Andersson and Ehlers 2016, Delacrétaz et al. 2016, Roth 2018).

Placement optimization in refugee resettlement shares many common features with other problems in humanitarian operations (Pedraza-Martinez and Van Wassenhove 2016, Besiou et al. 2018). Typical challenges in this sector include severe lack of resources—financial, labor, time, and data—as well as complex decision environments. The refugee resettlement decision environment includes refugees as well as local communities, non-profit organizations, donors, and federal, state and local governments. Hence, like other humanitarian operations problems, placement optimization similarly diverges from the traditional stance of optimizing a financial metric. Refugee resettlement is perhaps most differentiated by its particular exposure and sensitivity to shifting political climates and attitudes, both domestic and abroad. This volatility generates significant uncertainty with respect to the operating and planning environments of resettlement agencies.

Because of these factors, only solutions satisfying a number of specific requirements can succeed in placement optimization in refugee resettlement. The design of the solution needs to be attractive, lightweight, and intuitive to use, so as to engage resettlement staff. The design cycle ought to be transparent and attentive to the practical, operational details that resettlement staff face. It should be data-driven, responsive to the dynamic resettlement environment, requiring careful attention to the data and machine learning techniques to derive accurate estimates of refugee integration. Proper optimization modeling is needed to account for the welfare-maximizing matching problem at hand in light of varying capacities. Finally, due to the severe lack of resources, the technologies comprising the solution ought to be carefully united via an open-source implementation that allows for extensive end-user interaction.

3. Integer Optimization for Refugee Resettlement

We formulate the operational challenge of matching refugee families to local communities, or affiliates, presently solved manually by resettlement agencies, using mathematical optimization. This formulation extends core ideas from Delacrétaz et al. (2016).

3.1 Formal Problem Setup

We use i, j, k, and ℓ as indices for family (case), member, service and affiliate, respectively. Let $\mathcal{F} = \{F^1, F^2, \ldots, F^i, \ldots\}$ be a finite set of refugee families, or cases. Family F^i consists of members $\{f^{i,1}, f^{i,2}, \ldots, f^{i,j}, \ldots\}$ and has size $|F^i|$. For clarity of exposition, we refer to j of family F^i as f^{ij}. We denote the set of all refugees as $\mathcal{R}$, that is, $\mathcal{R} = \bigcup_{i \in \{1,2,\ldots,|\mathcal{F}|\}} \bigcup_{j \in \{1,2,\ldots,|F^i|\}} f^{ij}$. Moreover, there exists a finite set of affiliates (localities) $\mathcal{L} = \{L^1, L^2, \ldots, L^\ell, \ldots\}$ to which families are resettled.

A family F^i requires various ***capacitated*** services from a set $\mathcal{S} = \{S^1, S^2, \ldots, S^k, \ldots\}$. The needs of family F^i are summarized by a vector s^i, with a typical element denoted by s^i_k. Services may include raw weekly refugee processing capacity at affiliates, slots in foreign language instruction (such as ESL), school seats for children in the family, and housing availability. For every service S^k provided by local affiliate L^ℓ, at most $\bar{s}^\ell_k$ units may be filled by families placed in affiliate L^ℓ. There may also be a requirement of at least $\underline{s}^\ell_k$ units of the service S^k to be filled by the families placed in affiliate L^ℓ (we assume $\underline{s}^\ell_k \leq \bar{s}^\ell_k$); in practice, nonzero lower bounds exist for certain services, such as ensuring regular, positive refugee placement in affiliates.

For every refugee f^{ij} and affiliate L^ℓ, let the binary variable x^{ij}_ℓ equal 1 if refugee f^{ij} is matched to local affiliate L^ℓ, and 0 otherwise. Similarly, for every family F^i and local affiliate L^ℓ, let binary variable z^i_ℓ equal 1 if family F^i is matched to affiliate L^ℓ, and 0 otherwise. As it is customary to resettle all refugees from a family unit to the same affiliate, we establish constraints to ensure this outcome. We define a feasibility indicator a^i_ℓ if family F^i can be feasibly placed in affiliate L^ℓ. The value of a^i_ℓ is determined by evaluating the compatibility of family F^i with various binary community support services at affiliate L^ℓ, such as language and nationality, as well as large family and single parent support conditions (should these be present in the family). We will denote these community support services as ***binary*** services.

The value of each refugee-affiliate match is summarized with a single number called the *quality score*. The function $q : \mathcal{R} \times \mathcal{L} \to \mathbb{R}_{\geq 0}$ defines quality score q^{ij}_ℓ for any $f^{ij} \in \mathcal{R}$ and any $L^\ell \in \mathcal{L}$. We will be interested in the scenario where q represents the employment outcome of refugee f^{ij} in affiliate L^ℓ and can be estimated from data using observable affiliate and family characteristics.

We assume a nondecreasing objective function $\mathbf{q}(x)$ that represents overall match quality. While $\mathbf{q}(x)$ can take many forms, we consider maximizing the linear function:

$$\mathbf{q}(x) = \sum_{i=1}^{|\mathcal{F}|} \sum_{j=1}^{|F^i|} \sum_{\ell=1}^{|\mathcal{L}|} q_\ell^{ij} x_\ell^{ij}. \tag{1}$$

3.2 Placement Optimization

With this notation we formulate the following integer optimization problem that maximizes a welfare function over all matched refugees:

$$\begin{aligned}
\text{maximize} \quad & \mathbf{q}(x) && \text{(2a)}\\
\text{subject to} \quad & \sum_{\ell=1}^{|\mathcal{L}|} z_\ell^i \le 1, \ \forall\, i, && \text{(2b)}\\
& \underline{s}_k^\ell \le \sum_{i=1}^{|\mathcal{F}|} s_k^i z_\ell^i \le \bar{s}_k^\ell, \ \forall\, \ell, \ \forall\, k, && \text{(2c)}\\
& \sum_{j=1}^{|F^i|} x_\ell^{ij} = |F^i| z_\ell^i, \ \forall\, i, \ \forall\, \ell, && \text{(2d)}\\
& z_\ell^i \le a_\ell^i, \ \forall\, i, \ \forall\, \ell, && \text{(2e)}\\
& x_\ell^{ij} \in \{0,1\}, \ \forall\, i, j, \ \forall\, \ell; \ z_\ell^i \in \{0,1\}, \ \forall\, i, \ \forall\, \ell. && \text{(2f)}
\end{aligned}$$

Constraint set (2b) ensures that families are placed in at most one affiliate. Constraint set (2c) ensures that lower and upper bounds are respected for all capacitated services and affiliates. Constraint set (2d) links the refugee and family variables by ensuring that whenever families are placed in an affiliate, the constituent family members are also placed there, and conversely, no refugees from a family may be placed in an affiliate, unless the family is placed there. Constraint set (2e) ensures that family-affiliate matches can only occur when the affiliate can support the needs of the family, that is, the necessary binary services exist. Variable domains are specified in (2f).

While formulation (2a)–(2f) bears similarity to a variety of knapsack-like problem classes, we are unaware of another with its particular form. When $|\mathcal{S}| = 1$, $\underline{s}_k^\ell = 0 \ \forall\, \ell$, and $s_k^i = 1 \ \forall\, i$, the optimization problem can be solved via linear programming (Bansak et al. 2018). When $|\mathcal{S}| = 1$ and $\underline{s}_k^\ell = 0 \ \forall\, \ell, k$, our problem becomes the *multiple 0–1 knapsack problem* which features multiple knapsacks and items that consume integer resources for the knapsack in which they are placed (Martello and Toth 1980). It is NP-hard. When $|\mathcal{L}| = 1$ and $\underline{s}_k^\ell = 0 \ \forall \ \ell, k$, we have a *multidimensional 0–1 knapsack problem* which features knapsack items that consume integer resources along multiple dimensions (Fréville 2004). It is also NP-hard. When $\underline{s}_k^\ell = 0 \ \forall \ \ell, k$, our problem is called the *multiple multidimensional knapsack problem* combines features of both, that is, multiple knapsacks along multiple dimensions (Song et al. 2008, Delacrétaz et al. 2016). Our problem generalizes the *multiple multidimensional knapsack problem* of Song et al. (2008), as beyond the integer s_k^i values representative of family size and number of children needing slots in schools, we also allow for positive lower bounds $\underline{s}_k^\ell$ for any services and affiliates. Due to potential lower bounds, our problem, unlike the multiple multidimensional knapsack problem, may have no feasible solution. Our model is also distinct from the *multichoice multidimensional knapsack problem* (Hifi et al. 2004) because we do not require (in theory) that every family is placed in some affiliate.

Formulation (2a)–(2f) is valid over any operational period (weekly placements, annual counterfactual outcomes). While general, our problem can be customized to specific refugee resettlement settings. In this paper, we will test the sensitivity of our objective under three different scenarios. First, we will test the effect of relaxing upper bounds (2c) for the number of total resettled refugees. Second, we will test the effects of lower bounds (2c) expressed as distributional requirements (such as minimum average case sizes across affiliates) and as lower bounds on the total number of resettled refugees. Finally, we will look at the effects of relaxing one or more of the binary service constraints (2e).

4. Estimation and Empirics

Let ℓ_{ij} denote the affiliate that refugee f^{ij} was assigned to in the data. We use the expected probability of employment of refugee f^{ij} in each affiliate ℓ as a measure of quality score, or:

$$q_\ell^{ij} = E\left[y_{ij} \mid \mathbf{X}_{ij}, \ell\right], \tag{3}$$

where y_{ij} is employment status and $\mathbf{X}_{ij}$ a set of observable refugee characteristics and quarterly macroeconomic variables. We use national employment ratio and unemployment rate as macroeconomic variables, which are common to all refugees arriving in a given quarter. Further details on the available data appear in Appendix 9.1.

Using expected potential outcomes instead of stated preferences creates two challenges. First, y_{ij} is unobserved for incoming refugees. Second, even for past refugees we only observe $y_{ij}|x^{ij}_{\ell_{ij}}$, that is employment status of refugee f^{ij} in the affiliate they were actually assigned to. We do not observe the corresponding potential outcome distribution $y_{ij} \mid x^{ij}_{\ell} \quad \forall\, \ell \neq \ell_{ij}$. Moreover, the functional form connecting y_{ij}, $\mathbf{X}_{ij}$, and ℓ is unknown. Specific synergies may exist between refugee characteristics and affiliates that affect refugee integration. Following Bansak et al. (2018), we thus exploit machine learning approaches in the estimation of $\hat{q}^{ij}_{\ell}$. Using data on refugees arriving between 2010 and 2016, we estimate both semi- and non-parametric functions $\hat{f}_{\ell} : \mathcal{R} \rightarrow \mathbb{R}_{\geq 0}$ such that $\hat{q}^{ij}_{\ell} = \hat{f}_{\ell}(\mathbf{X}_{ij})$. We then test the performance of these models on refugees arriving in 2017.

In the estimation process we only use free cases, that is, those refugees that the resettlement agency could in principle assign to any of the affiliates. We therefore exclude refugees with family ties, which are almost always assigned to the affiliate where their pre-existing connection resides. This choice, while restricting the samples we use to train and test the models to 2,486 and 498 refugees, has two key advantages. First, we focus on the relevant refugee-affiliate synergies, those of refugees that can actually be assigned to multiple affiliates. Second, including endogenously assigned refugees would likely overestimate existing synergies for free cases. For example, because of pre-existing networks, family reunifications enjoy particular advantages (Edin et al. 2003, Patacchini and Zenou 2012) that would bias our estimates.

We estimate synergies for the seven (out of twenty) affiliates receiving more than 200 refugees up to 2016, and aggregate the remaining affiliates in a single partition ℓ_0. In a parametric approach, one could estimate a fully saturated logit model for employment where flexible transformations of refugee characteristics $\mathbf{X}_{ij}$ are interacted with $\ell - 1$ affiliate dummies. Such an approach would, however, estimate an overly complex model, with poorly identified coefficients, and therefore yield poor predictive properties.

We thus estimate two alternative machine learning models. First, we introduce a Least Absolute Shrinkage and Selection Operator (LASSO) constraint to the interacted logit model to reduce model complexity. Second, we follow Bansak et al. (2018) and estimate a Gradient Boosted Regression Tree (GBRT), an iterative ensemble of classification trees. We set the hyper-parameters of these models via 5-fold cross-validation on our training sample.[6] We choose hyper-parameter values by maximizing the area under a model's Receiver Operating Characteristic (ROC) curve.

We benchmark both models against the performance of a naïve constant estimator (Bansak et al.

[6] We internally calibrate constraint strength for LASSO, as well as the learning rate and pre-pruning level for GBRT.

	Training data	Test data			
	Misc. error	Misc. error	Recall (1)	Precision (1)	AUC-ROC
Constant	0.259	0.319	0.000	0.000	0.500
Logit	0.240	0.263	0.491	0.609	0.790
Logit (by affiliate)	0.177	0.281	0.547	0.561	0.769
LASSO	0.159	0.201	0.453	0.637	0.799
Gradient boosted tree	0.129	0.209	0.396	0.624	0.791

NOTE: *Misclassification error* is the proportion of observations incorrectly classified. *Recall* measures the proportion of correctly predicted employed refugees among refugees actually employed (true positives over true positives plus false negatives). *Precision* measures the proportion of correctly predicted employment cases among all predicted employment cases (true positives over true positives plus false positives). *AUC-ROC* measures the area under the Receiver Operating Characteristic Curve for each model (ROC curves appear in Appendix 9.2).

Table 1: Model performance.

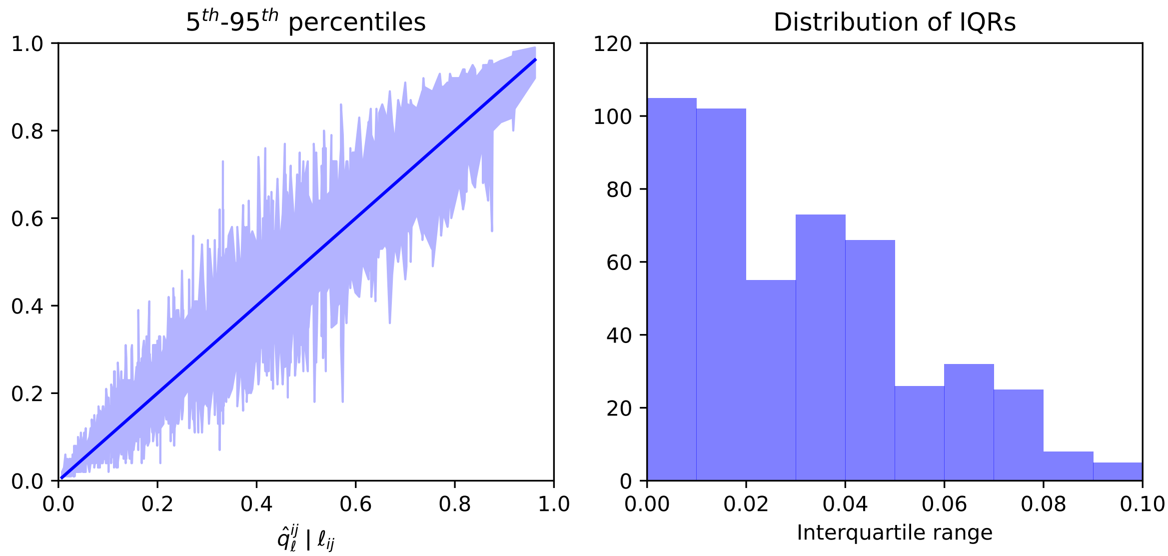
2018), as well as two second-best standards. The first benchmark model is a standard logit model that includes all variables in $\mathbf{X}_{ij}$, but does not attempt to estimate affiliate-specific synergies. The second benchmark model is a logit model with no LASSO constraint, where $\mathbf{X}_{ij}$ interacts with all ℓ affiliates. Table 1 shows that both LASSO and GBRT outperform the second-best benchmarks by over 20% in terms of misclassification error when applied to 2017 refugees.[7] The area under the ROC is highest for LASSO, but overall both models exhibit similar predictive power.

LASSO, however, produces slightly more stable and well-calibrated predictions, particularly for observations with high predicted employment probabilities. We obtain these results by bootstrapping the distribution of predictions for each data point in the test set given assignment to ℓ_{ij}. In each of a thousand iterations, we re-sample with replacement the training dataset, re-estimate each model and compute a new predicted probability of employment. The right panels of Figure 1 show the 5^{th} to 95^{th} percentiles of the prediction distributions for each data point in the test sample. The left panels show the distribution of bootstrapped interquartile ranges for each data point.

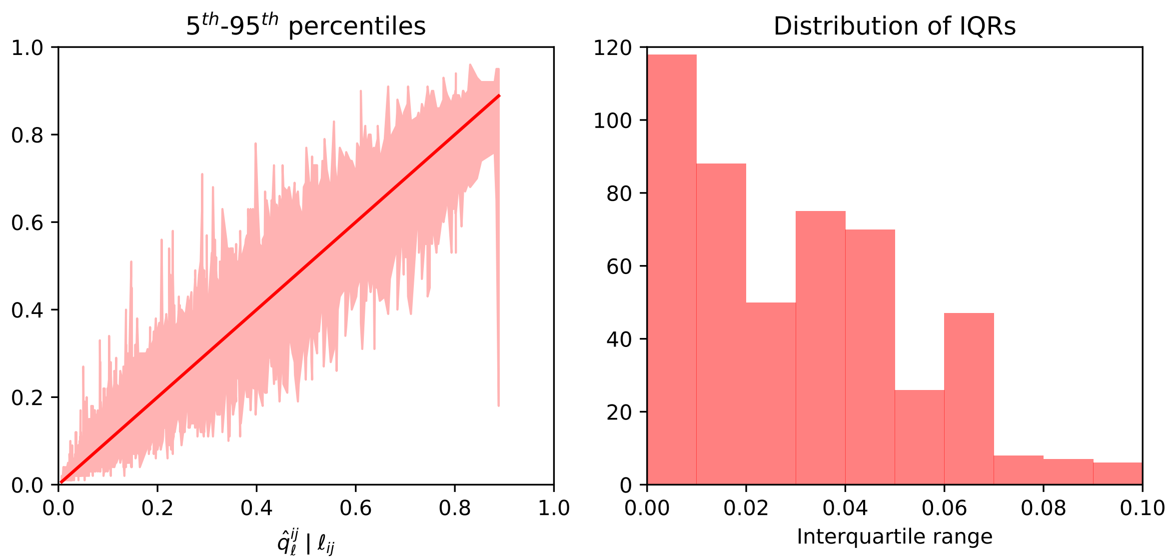
LASSO tends to produce more narrow predictions for refugees with high baseline probability of employment, which are highly relevant for the quantification of employment gains. LASSO is also better calibrated than GBRT—with 159 employed refugees in our test set, whereas the sum of predicted employment probabilities given assignment to ℓ_{ij} is 157.9 for LASSO, it is only 142.9 for GBRT.[8] Thus, while using either model has very similar consequences for optimal refugee assignment, in the remainder of the paper we quantify employment gains given the quality scores predicted by LASSO. We replicate these results given the predictions of GBRT in Appendix 9.3.

[7] With respect to the constant-logit benchmark used by Bansak et al. (2018) we obtain a 37% and 34% improvement using LASSO and GBRT respectively, which is comparable to the 28% they obtain in their US data.

[8] Calibration plots appear in Appendix 9.2.



(a) LASSO



(b) Gradient Boosted Regression Tree (GBRT)

Figure 1: Bootstrapped uncertainty of predicted employment probabilities in 2017 for LASSO and GBRT model. Left panels: prediction distributions (5th-95th percentile) for each data point in test sample. Right panels: distribution of interquartile ranges for each data point in test sample.

5. Counterfactual Optimization Outcomes

We now describe the counterfactual impact of using our placement optimization formulation (2a)–(2f). We create test scenarios that result from varying three constraint sets. To quantify the impact of optimally reassigning refugees to affiliates, we use the employment probabilities for each affiliate estimated in Section 4. We compute the counterfactual gain in employment relative to our prediction from the LASSO model for 2017. Since our prediction is very close to the actual employment values—the LASSO model predicts 158 employed refugees versus 159 who were actually employed in the testing data—our optimization is a meaningful counterfactual exercise.

All experiments were run on a laptop computer with an Intel(R) Core(TM)i5-4300U 2.50GHz processor and 8GB RAM running 64-bit Windows 10 Enterprise. The Gurobi optimizer (Gurobi 2018) and Python 2.7 was used for all counterfactual optimization testing in Section 5. Our objective function (2a) is the total expected number of employed refugees. Our binary service constraints (2e) are: language, nationality, single-parent, and large-family support. We set the capacity constraints (2c) for each affiliate relative to the observed capacity in 2017. Moreover, we specify minimum average case sizes to enforce distributional constraints via the lower bounds in (2c). We vary the following three factors to create our test scenarios.

Affiliate capacity. Affiliate capacity is federally approved, but can be exceeded by up to 10% without further pre-approval. Moreover, agencies aim to fill at least 90% of the approved capacity at each affiliate. In 2017, somewhat unusually, approved capacity was much higher than the observed number of arriving refugees. We therefore use the observed placements at each affiliate to set sensible counterfactual capacities. We test three values: {observed capacity with no lower bound; 110% of the observed capacity with no lower bound; and 110% of observed capacity with a lower bound of 90% of observed capacity}.

Binary service constraints. In the observed 2017 placements, binary service constraints were violated 38 times (26 language constraints, 1 nationality constraint, 8 single-parent constraints, and 3 large-family constraints), representing approximately 12% of resettled refugees. However, binary service constraints, especially language constraints, can be important to ensure successful refugee integration. We therefore test two values: {binary service constraints are imposed, binary service constraints are not imposed}.

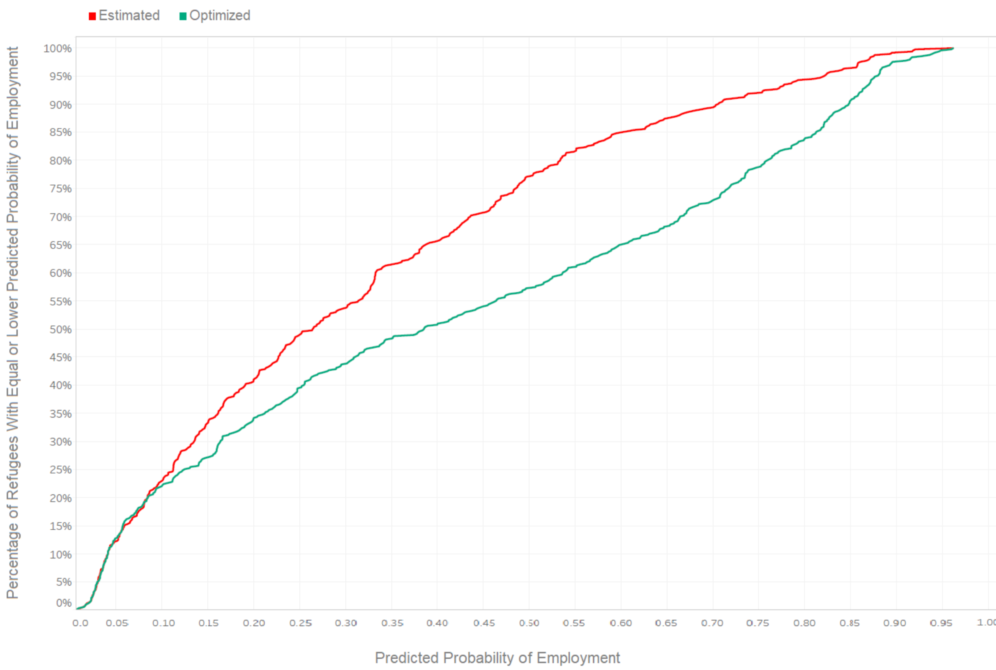
Minimum average case size in each affiliate. A placement that maximizes the number of employed refugees could potentially place many single-refugee cases or large-family cases into the same affiliate. This could be seen as unfair by the agencies, reduce support for resettlement, and stymie refugee integration. The average case size in our test dataset is 2.55 (FY 2017). We therefore test five values: {no minimum average case size, observed average case size (2.55), 2, 2.5, 3}.

Capacity Adjustment	Min Avg Case Size	Binary Service Constraints	Total Expected Employed Refugees	Gains wrt to Predicted Employed Refugees (158)	St Dev in Avg Case Size Across Affiliates	# of Unplaced Cases / Refugees	#of Affiliates Violating 90% Capacity	# and % of Cases/Refugees Violating Constraints
Observed	None	Off	213.094	34.87%	1.311	0/0	0	72/194(21.88%/23.12%)
Observed	None	On	208.301	31.84%	1.637	3/10	2	0/0
Observed	2	Off	206.279	30.56%	1.088	1/1	0	77/215(23.40%/25.63%)
Observed	2	On	202.06	27.89%	1.086	2/9	1	0/0
Observed	2.5	Off	196.775	24.54%	0.325	5/5	0	94/234(28.57%/27.89%)
Observed	2.5	On	192.949	22.12%	0.8	4/8	0	0/0
Observed	3	Off	172.827	9.38%	0	78/86	9	62/191(18.84%/22.77%)
Observed	3	On	169.635	7.36%	0.9	79/89	9	0/0
Observed	Observed	Off	199.58	26.32%	0.828	2/2	0	80/216(24.32%/25.74%)
Observed	Observed	On	195.888	23.98%	1.079	5/9	2	0/0
$\leq$ 110%	None	Off	217.713	37.79%	1.587	0/0	6	67/180(20.36%/21.45%)
$\leq$ 110%	None	On	212.618	34.57%	1.62	2/9	5	0/0
$\leq$ 110%	2	Off	211.778	34.04%	1.265	0/0	5	72/183(21.88%/21.81%)
$\leq$ 110%	2	On	207.167	31.12%	1.014	2/9	5	0/0
$\leq$ 110%	2.5	Off	201.985	27.84%	0.588	0/0	6	93/234(28.27%/27.89%)
$\leq$ 110%	2.5	On	198.084	25.37%	0.817	3/7	5	0/0
$\leq$ 110%	3	Off	177.506	12.35%	0	78/86	11	66/204(20.06%/24.31%)
$\leq$ 110%	3	On	174.266	10.29%	1.071	79/89	8	0/0
$\leq$ 110%	Observed	Off	203.871	29.03%	0.909	0/0	4	87/221(26.44%/26.34%)
$\leq$ 110%	Observed	On	200.084	26.64%	1.162	3/7	4	0/0
[90%, 110%]	None	Off	217.695	37.78%	1.416	0/0	0	69/194(20.97%/23.12%)
[90%, 110%]	None	On	212.535	34.52%	1.539	1/2	0	0/0
[90%, 110%]	2	Off	211.789	34.04%	1.125	0/0	0	76/211(23.10%/25.15%)
[90%, 110%]	2	On	206.947	30.98%	1.2	1/2	0	0/0
[90%, 110%]	2.5	Off	201.969	27.83%	0.329	0/0	0	97/239(29.48%/28.49%)
[90%, 110%]	2.5	On	198.052	25.35%	0.334	2/3	0	0/0
[90%, 110%]	3	Off			Model is infeasible			
[90%, 110%]	3	On			Model is infeasible			
[90%, 110%]	Observed	Off	203.714	28.93%	0.861	5/5	0	86/226(26.14%/26.94%)
[90%, 110%]	Observed	On	199.752	26.42%	0.862	5/6	0	0/0

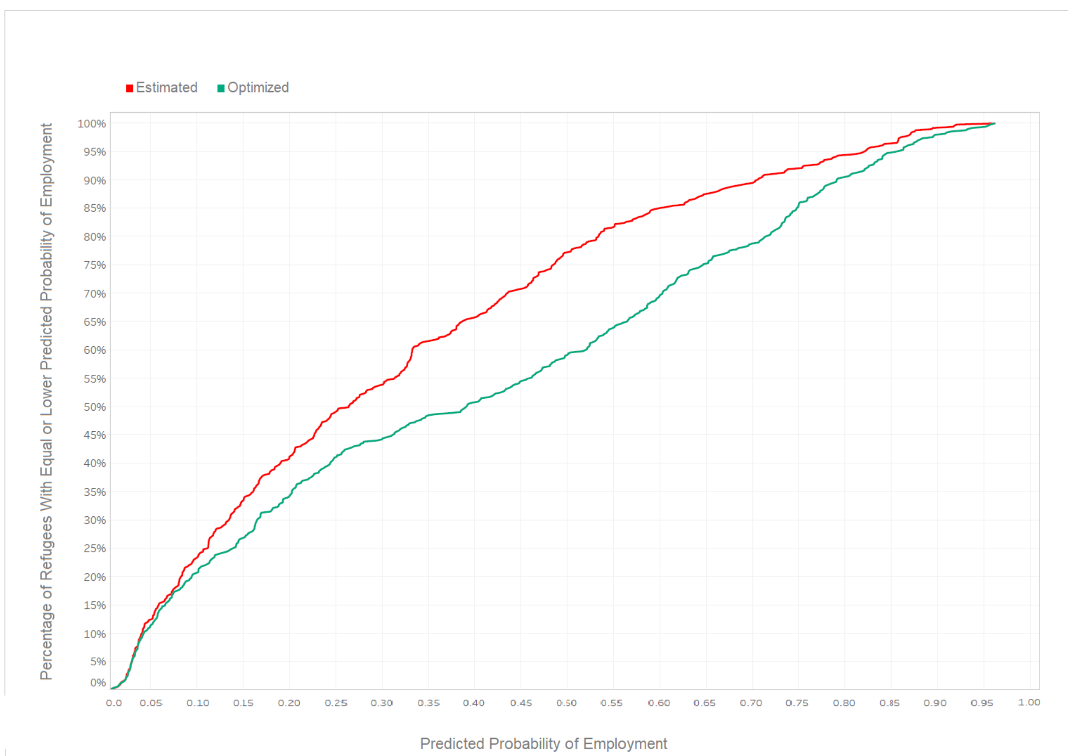
Table 2: Results of counterfactual employment optimization under various scenarios (using the LASSO model).

In total, we have $3 \times 2 \times 5 = 30$ counterfactual test scenarios. The results are summarized in Table 2. First, note that without minimum average case size constraints, the gain in employment from optimization is over 30% in all scenarios. As Figures 2(a) and 2(b) show, the employment probability distribution after optimization first-order stochastically dominates the pre-optimized estimated distribution. Therefore, the probabilities of employment increase across the distribution after optimization. Moreover, Figure 3 shows that employment rates rise in nearly two-thirds of the affiliates after optimization. Table 2 further indicates that, if we do not impose binary service constraints, they are violated for around a quarter of the refugees—a rate much higher than in the test data (approximately 12%). However, the presence of binary service constraints and of increasing capacity has a fairly small impact on employment gains. Indeed, because in some cases our model leaves some refugees unplaced (meaning that they would need to be placed manually by agency staff), our employment gain estimates should be even higher.

However, in these scenarios the optimization suggests rather unequal placement. Figure 4 compares the distribution of average case sizes in each affiliate to the distribution under our second



(a) Cumulative distribution of employment probabilities. Red: estimated probabilities under HIAS placement. Green: optimized probabilities for {observed capacity, service constraints on, no mininum average case size} scenario.



(b) Cumulative distribution of employment probabilities. Red: estimated probabilities under HIAS placement. Green: optimized probabilities for {observed capacity, service constraints on, at least observed average case size} scenario.

Figure 2: Employment gains from optimizing refugee placement.

counterfactual optimization which produces the largest variance in average case sizes. Figure 5(a) shows that without distributional constraints, many single-person cases are placed in just three affiliates that offer a high probability of obtaining employment to many types of refugees. Other affiliates get much larger cases on average. This allocation may not be acceptable to a resettlement agency. Thus, we evaluated the placement optimization by enforcing minimum average case size constraints. At low values (up to 2.5) and at observed 2017 average case size values, the optimization is still able to realize employment gains of well over 20% (see also Figure 5(b)). This is extremely encouraging because it shows that our optimization performs well even under tight distributional constraints. However, at high average case sizes, the constraints bind harder and either reduce the performance of the model substantially (by not placing many refugees), or simply cause infeasibility.

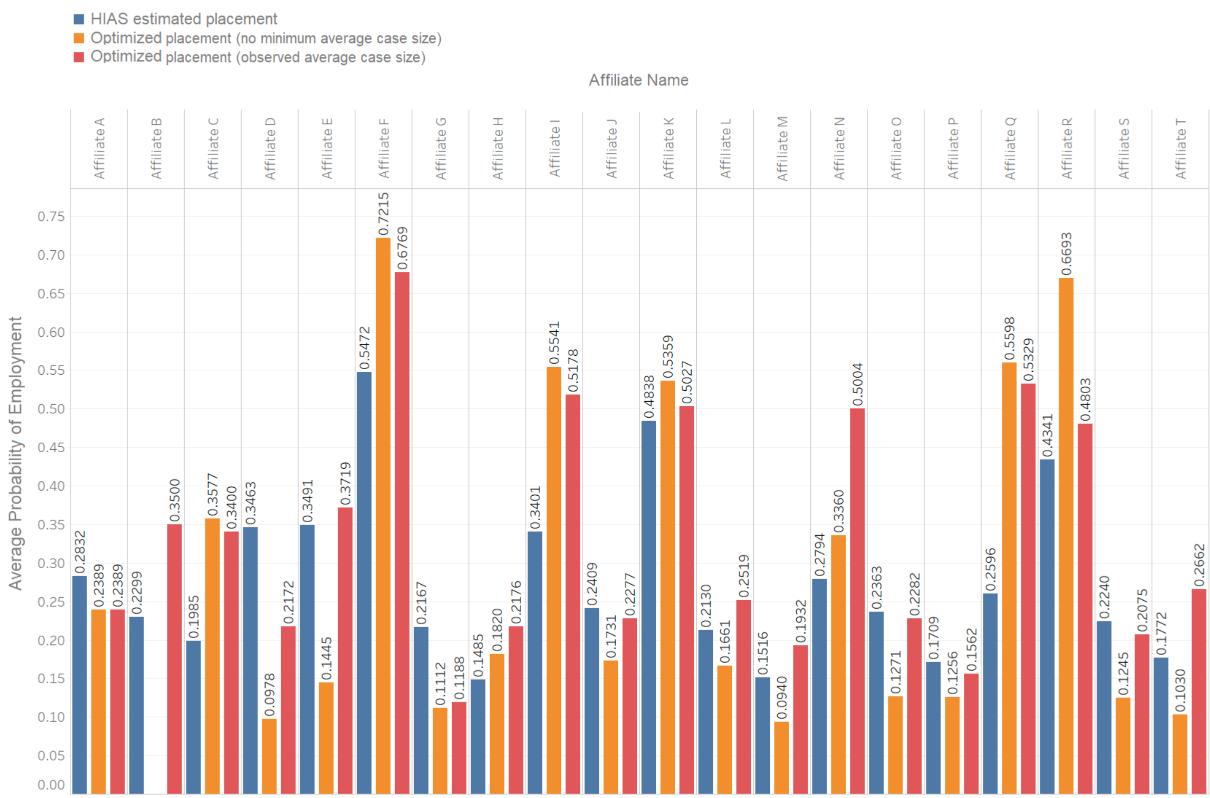


Figure 3: Average probability of employment at each affiliate. Blue bar: estimated probabilities under HIAS placement. Orange bar: average probability of employment for observed capacity, service constraints on, no minimum average case size scenario. Red bar: average probability of employment for {observed capacity, service constraints on, at least observed average case size} scenario.

It is worth emphasizing that the space of objective functions and constraints that the resettlement agency can impose within our model is much richer than what we have presented here. For example, the resettlement agency could impose any subset of the binary service constraints or introduce constraints on number of refugees with certain regional origins.[9] Alternatively, the agency could select a different employment objective function, for example maximizing the sum of minimum employment probabilities within every case.

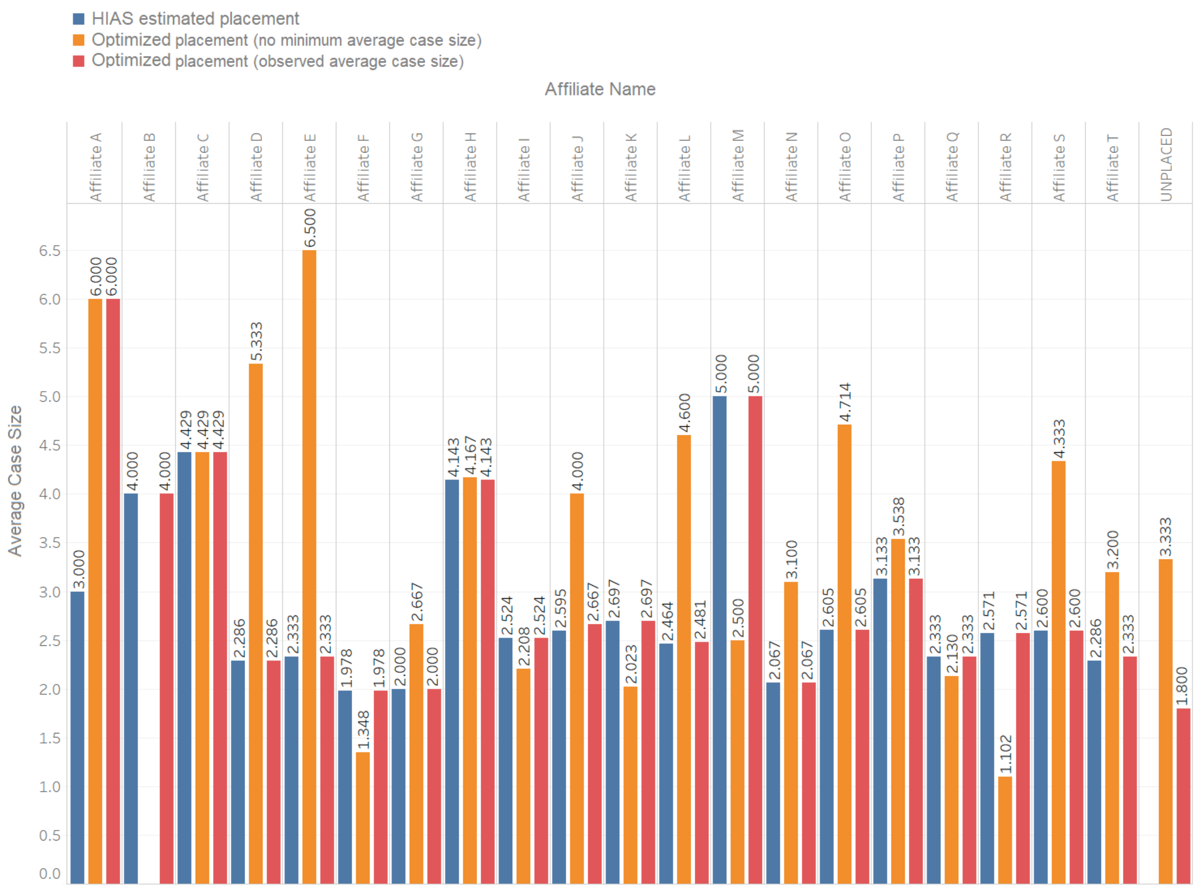
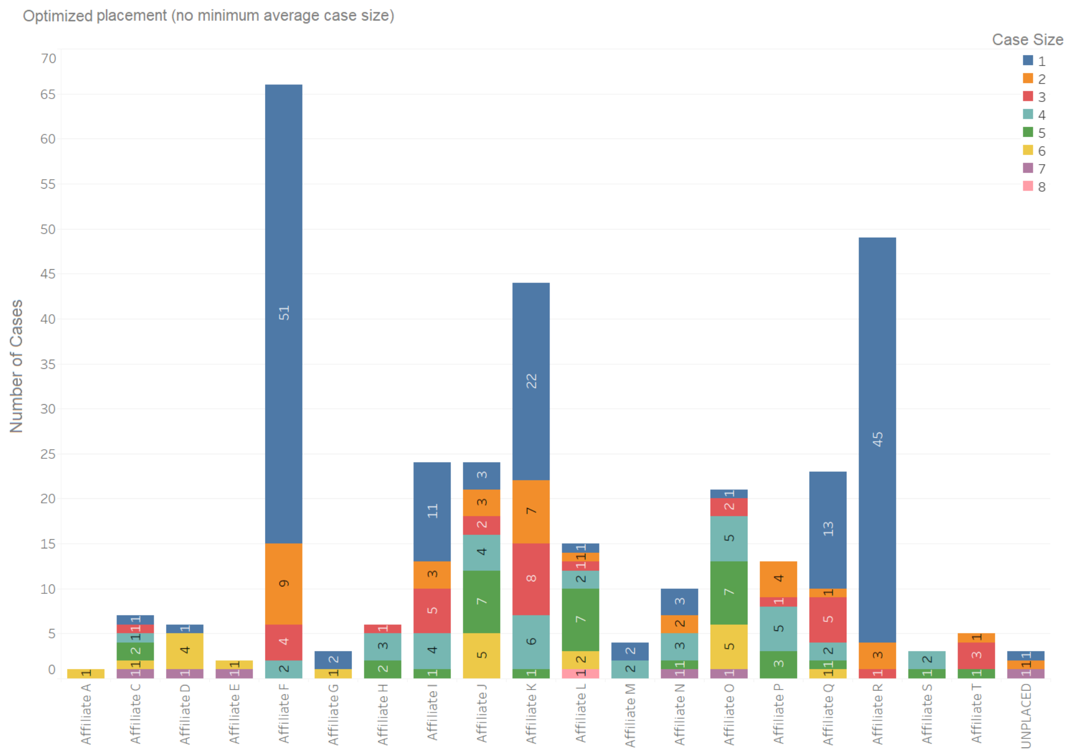


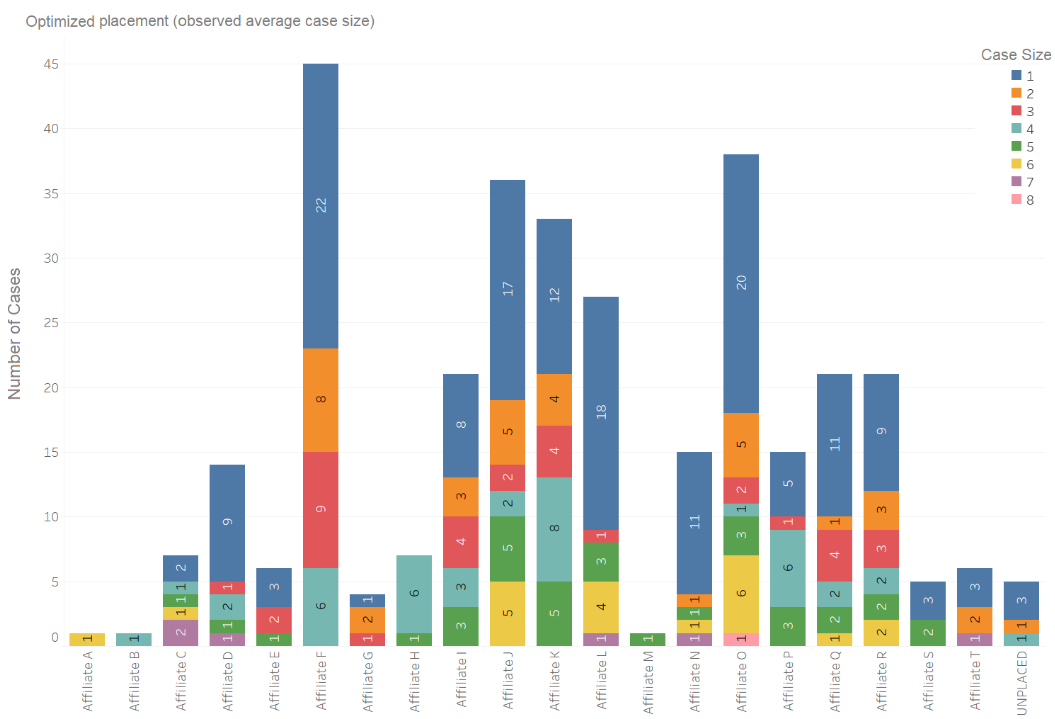
Figure 4: Average case size at each affiliate. Blue bar: observed average case size under HIAS placement. Orange bar: average case size for {observed capacity, service constraints on, no minimum average case size} scenario. Red bar: average case size for {observed capacity, service constraints on, at least observed average case size} scenario.

Overall, our optimization produces a substantial gain in employment, ensures that refugee binary services are better satisfied, and important distributional considerations can be respected. We must stress that we were able to optimize placement of all refugees within a given year simultaneously rather than considering weekly decisions under arrival uncertainty that the resettlement agency faced. Therefore, the level of our employment gains might be hard to replicate in practice. Dynamic quota management is an interesting area for further work.

[9]Although regional constraints used to be officially considered in US placements, they are no longer specified or tracked.



(a) Distribution of case sizes for {observed capacity, service constraints on, no minimum average case size} scenario.



(b) Distribution of case sizes for {observed capacity, service constraints on, at least observed average case size} scenario.

Figure 5: Distribution of case sizes at each affiliate.

6. Operationalizing Placement Software at Resettlement Agency

Integer optimization and machine learning techniques offer great promise of solving the operational challenge of improvement placement outcomes in refugee resettlement. While these technologies provide significant value, expertise is needed for successful implementation. In the private sector, this expertise is readily available. On the other hand, operations research in humanitarian environments, including refugee resettlement, typically feature significant challenges, such as lack of human or financial resources, lack of exposure to technology, and data scarcity. Humanitarian organizations must be responsive to crisis events and immediate needs, and reactive to changes in the political climate. These realities can make it fairly prohibitive to be proactive in pursuing, and implementing, advancing technological innovations.

We maintain that successful integration of operations research technologies in a humanitarian environment requires cultivating and sustaining partnerships with stakeholders that include both management, as well as practitioners that will use the technology. The authors of this paper worked closely with many dedicated members of staff at HIAS for many months to develop *Annie* into an innovative, interactive optimization environment for refugee resettlement. Our close working relationship built a level of rapport that allowed us to understand and remedy, real operational challenges faced by resettlement staff. We believe these are key elements for creating a successful software solution for improving humanitarian operations.

6.1 Technologies Involved in the Creation of *Annie*

Annie represents the confluence of several open-source technologies, critical for this resource-constrained environment. In particular, the integer optimization formulation (2a)–(2f) is modeled entirely within the PuLP Python modeling environment (Mitchell et al. 2018) and solved using the CBC (COIN-OR 2018) solver. The machine learning models described in Section 4 were developed entirely using the Python scikit-learn package. We chose to develop the interactive environment of *Annie* as a web application. The back-end is implemented in Python 3 using the Flask framework, with Jinja2 as the templating engine (Ronacher 2018). The front-end is a combination of HTML, CSS, and JavaScript. We made this choice of technology because it is modern and stable, accessible, and easy to build on. The only installation that is needed is (the free) Python 3 and some freely available packages and libraries. Moreover, it is a light technology: The front-end operates entirely within a browser rather than as a downloadable, executable file. By combining core open-source integer optimization and machine learning technology within a flexible, modern interface, we were able to achieve a completely free, lightweight software solution for HIAS.

6.2 Interactive Optimization

Representing overall match quality in objective function (2a) is by no means trivial. The best efforts toward estimating refugee and case employment outcomes, including substantial efforts to leverage as much of the inherent information available in the data, still leave approximate match scores. Even with perfect knowledge of how to represent match quality, vulnerable refugee lives are at stake, and any algorithmic solution should be carefully evaluated before actual implementation. Therefore there is a need for an interactive optimization environment, where resettlement staff can interact with various facets of the problem context. Without compromising on the insights afforded by the theory and data, *Annie* was designed to accommodate the real needs of the practitioner. The purpose of developing *Annie* as an interactive optimization tool is to translate advanced analytical methods into effective decision tools (Meignan et al. 2015). The user of *Annie* is intimately involved in the matching process and can fine-tune the result of the optimization. We believe that *Annie* strikes the right balance. Our close interactions with HIAS allow us to iteratively develop and test multiple versions of the software via remote updating. Moreover, our predictive models can be refined as more data on 90-day employment outcomes arrive over time.

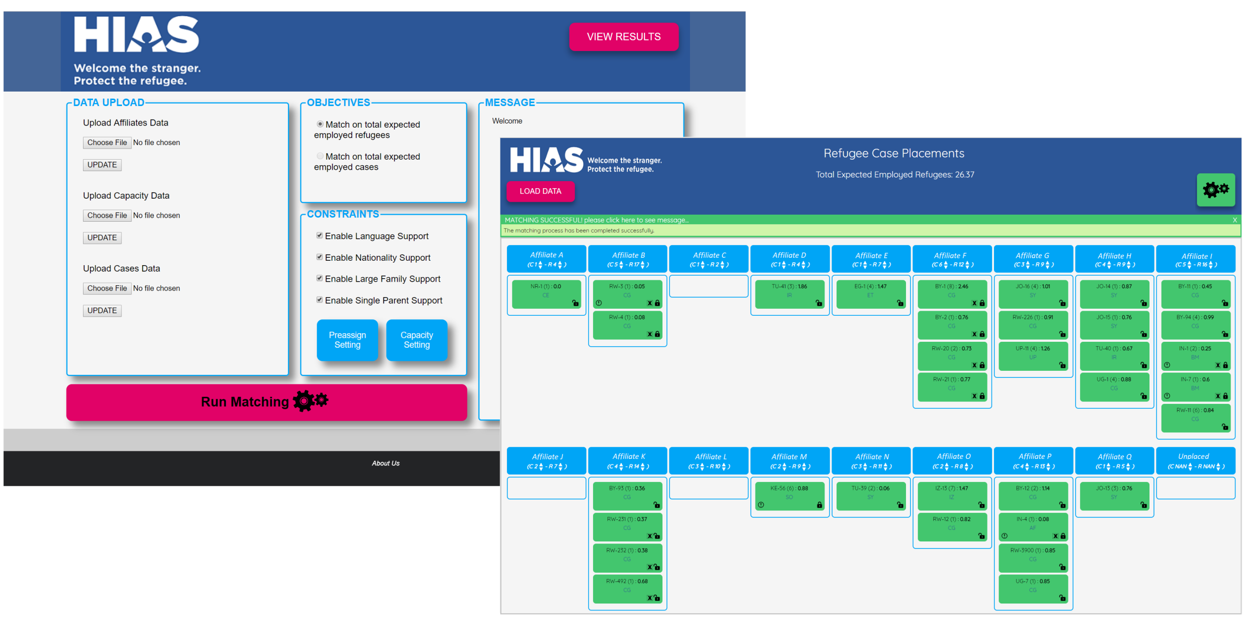


Figure 6: *Annie* Interface.

6.3 Features of *Annie*

The first version of *Annie* was delivered in early May 2018. We regularly added new features to *Annie* until August 2018 when it was presented to the US State Department and all staff at HIAS. Currently, *Annie* presently has two options for optimization. In addition to optimizing matches for the total employed *refugees*, *Annie* can optimize for total expected number of employed *cases* across the network of HIAS affiliates. We believe the former option to be preferable as it factors multiple refugees from a given case into the objective function match scores, which are individually estimated according to the predictive modeling of Section 4.

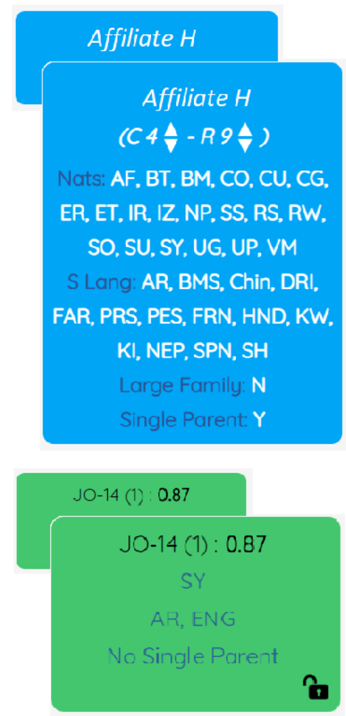
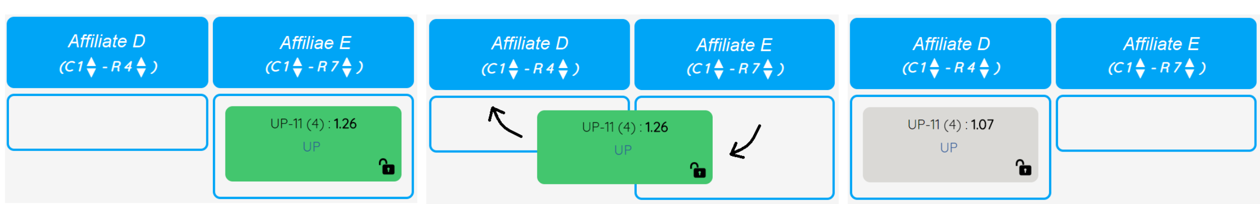


Figure 7: Expanding tiles: refugee and affiliate data.

The *Load Data* view is depicted in the rear left of Figure 6, where the optimization environment can be configured for the matching process, including the activation of binary support services. The matching results can be observed at the *View Results* view depicted in the front right of Figure 6, where the total number of expected employed cases is prominently displayed near the top.

The output of the matching engine results in cases being optimally assigned to affiliates, depicted with user-friendly *tiles*. Figure 7 displays both case and affiliate tiles. Case tiles show language, nationality, and other attributes unique to the family, whereas affiliate tiles show support features offered by affiliates. Clicking on the tiles expands their size to reveal detailed information at a quick glance. Case tiles can be moved to other affiliates as desired. Figure 8 illustrates the ability to dynamically view changes in the match scores as refugee case tiles are moved from one affiliate to



(a) Case assigned to Affiliate E. (b) Moving case tile to Affiliate D. (c) Case tile moved to Affiliate D.

Figure 8: Case tiles can be moved by dragging to an alternate affiliate tile. Upon moving, the match scores dynamically update. The background of the case tile changes to gray to indicate a non-optimized state.

the next. Moreover, the total expected number of employed refugees is also dynamically updated. Hence, at a glance, the effect of moving cases to alternative affiliates is easily and clearly visualized.

Perhaps the most important feature of *Annie* is its ability for interactive optimization. Resettlement staff may interact with intermediate solver output in a manner that progresses toward eventual convergence of a finalized assignment of refugee cases to affiliates. This is enabled through a lock icon on the case tile that resettlement staff can click, which locks desired case-affiliate matches. Figure 9 depicts this capability.

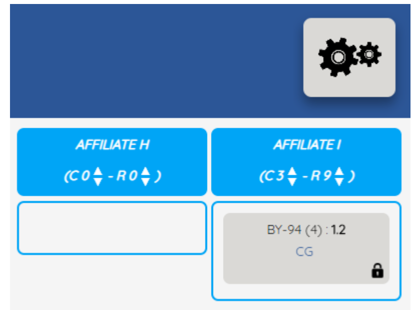


Figure 9: Locking case tiles and reoptimizing.

When locked, that case is temporarily "assigned" to that affiliate, and is literally unable to be moved elsewhere, until unlocked. After locking certain case-affiliate matches (this essentially assigns $z_\ell^i = 1$ for family F^i and location L^ℓ), any remaining unlocked cases may be rematched, adjusting down affiliate capacities from any locked cases, via a color-coded reoptimize button that indicates the non-optimized state (see Figure 9). Hence, any "final" matches can be locked, and all remaining cases can be rematched using the remaining available capacity.

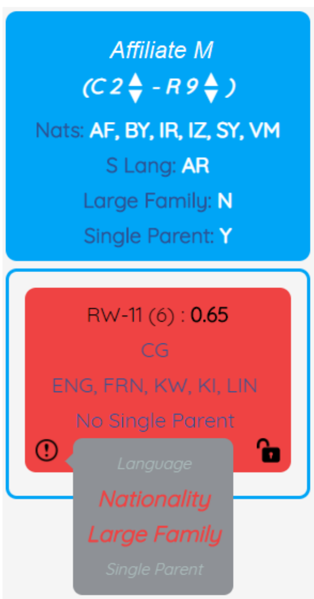


Figure 10: Case tile changes color when placed into affiliate that violates binary service constraints. Hovering over exclamation point reveals additional details.

If a case tile is moved into an affiliate but there is a mismatch between this case and the new affiliate in terms of binary community support services, the color of the case becomes red as an indication and an exclamation icon appears in the bottom left of the case tile (see Figure 10). Hovering over this exclamation icon displays up a new list that shows the unsupported needs for that particular case-community match.

We also enable cross-referencing. Cross-referencing occurs when refugee cases are linked to other cases that a) have previously been resettled to a specific local affiliate, or b) are among the pool of cases that are presently to be resettled to the same affiliate. In either case, *Annie* visually depicts cases that are associated with a) an affiliate or b) other cases via unique yellow borders upon hovering over a large, boxed X icon, for associated case tiles. Figure 11 depicts an example where two cases are cross-referenced not only to one another, but also to an affiliate.

Throughout the development process, we have firmly maintained that *Annie* is a tool that augments the perspective of resettlement staff at HIAS. That is, matches generated by *Annie* are suggestive in nature. HIAS has complete discretion to match and rematch cases according to their expert judgment. In this way, we allow for the best of both worlds: leveraging the strengths of modern computational technology—machine learning and integer optimization—while arming human decision-makers with all available information to facilitate the decision-making process.

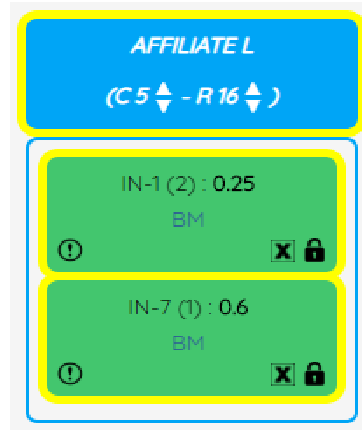


Figure 11: Cross-referencing cases to Affiliate L.

7. Conclusion

Refugee resettlement is a complex humanitarian problem that requires insights from a number of disciplines, including operations research, statistics, economics, political science, and sociology. Much work is urgently needed to improve the livelihoods of resettled refugees and the communities into which they integrate. In this paper, we show how combining tools from machine learning, integer optimization, and interactive visualization can improve refugee outcomes within the United States. We expect that local communities will benefit more by welcoming refugees that more closely match their needs, available resources, and opportunities. Moreover, because our matching is based on refugee employment outcomes, refugees will more quickly integrate economically into each affiliate, as well as make more productive economic and societal contributions such as paying taxes.

Annie has analytically enhanced the placement decision-making process at HIAS, having largely eliminated the inefficiencies of the former manual placement process. The operational process of placing refugees has improved considerably, allowing resettlement staff to effectively automate the placement of easier cases (such as those without major accommodations), and instead focus their time on those cases that need greater attention, such as those with several medical conditions.

Technological solutions for humanitarian operations problems, such as placement optimization in refugee resettlement, have the potential for profound societal impact. In particular, the mature technologies of machine learning and integer optimization offer incredible potential. While the humanitarian sector offers many opportunities for impact, any solution must properly account for the severe lack of resources—including financial, labor, time, and data. These factors must be carefully considered in designing solutions, to afford the best opportunity of effecting change.

Particular solution design features that we advocate include being lightweight, open-source, and designed with the end-user in mind by incorporating important aspects of their regular operations.

There are several directions for further work. First, as is often the case in the humanitarian context, data has been difficult to obtain due to the severely resource-constrained environment. Indeed, data collection appears to be under-prioritized across the resettlement agencies. We used the only existing outcome data from previous US placements, namely a refugee-specific binary indicator for employment measured 90 days after arrival. While we went through great efforts to make the most we got out of the available data, the relative lack thereof necessarily hampered our prediction ability. Further work could apply our techniques to data on other outcomes, such as longer-term employment, physical and mental health, education, and household earnings. Unfortunately, at the time of writing, no data on these objectives for resettled refugees arriving in the US appears to be systematically available. However, we anticipate to be able to better process other constraints like free-form text fields to discern whether refugees require medical accommodations such as wheelchair access.

Second, while agreed upon annual quotas exist for affiliates, refugees arrive stochastically over the course of a year. Therefore, it is important to schedule the arrival of refugees given the partial information about future arrival in the course of the whole year. Andersson et al. (2018) tackle this problem in the Swedish context.

Third, it is interesting to consider which features of local areas offer the best potential to host refugees. For example, we could analyze to what extent local unemployment or community demographics affect refugee outcomes. This could help refugee agencies target areas for new affiliates.

Fourth, we could explicitly include preferences of refugees and priorities of affiliates (Delacrétaz et al. 2016, Jones and Teytelboym 2016, Aziz et al. 2017). Preferences could be collected during the refugee pre-arrival orientation using a questionnaire that elicits how refugees might trade off features of areas (such as climate, urban / rural, crime, amenities, and quality of schools). While desirable, including preferences is not unproblematic. For example, including preferences while optimizing for a particular observable outcome can in itself be a challenging problem (Biró and Gudmundsson 2018). It is also unclear how preferences should be elicited based on the reported information. Allowing refugees to report complete preferences—as often is the case in school choice problems (Abdulkadiroğlu and Sönmez 2003)—may be too challenging. On the other hand, limited preference information—like the dichotomous preference environment commonly used in kidney exchange problems (Roth et al. 2004)— may not be very informative for this particular application. Furthermore, it is well documented that whenever agents are allowed to report preferences, they often take advantage of the situation and misrepresent information to manipulate the outcome of the allocation mechanism in their advantage. In general, it is difficult to design mechanisms that are non-manipulable by all agents (see, e.g., Alcalde and Barberà 1994, Barberà and Jackson 1995, Roth 1982, Sönmez 1999) and these strategic concerns are exacerbated in the complex matching systems required for refugee resettlement (Andersson and Ehlers 2016, Delacrétaz et al. 2016).

While *Annie* has primarily been developed to assist HIAS in their initial refugee placements, there are potential uses for the software beyond refugee resettlement in the US. *Annie* could be used to help improve placement in the (Syrian) Vulnerable Persons Resettlement Scheme operated by the British government between 2015 and 2020. A recent report by the UK Independent Chief Inspector of Borders and Immigration recommended that the Home Office "improve the geographical matching process" of refugees in this resettlement scheme (Bolt 2018, p.12). In Sweden, asylum seekers who enter are temporarily placed at Migration Board accommodation facilities in anticipation of either a deportation order or a residence permit. If a residence permit is granted, the legal responsibility for asylum seekers (such as finding housing and schooling) is transferred from the Migration Board facility to one of the 290 municipalities in Sweden (around 50,000 such transfers were made in 2017). This system is, in a sense, a version of refugee resettlement in which asylum seekers are resettled within Sweden. While the current Swedish system is not based on sophisticated matching techniques, a recent report by the Swedish Government (SOU 2018, p.280) recommends that cleverly designed optimization and matching techniques should be adopted.[10] Finally *Annie* may, for example, be adapted to resettle asylum seekers who have crossed the southern border of the United States or the southern border of Germany and require shelter prior to judicial processing.

8. Acknowledgments

We are incredibly grateful to many for helping make this research successful. Emily Aiken provided invaluable software development assistance for an early prototype of *Annie*. We are also thankful for institutional and financial support. Andersson gratefully acknowledges the financial support of the Jan Wallander and Tom Hedelius Foundation (Research Grant P2016-0126:1) and the Ragnar Söderberg Foundation (E8/13). Teytelboym gratefully acknowledges the Skoll Centre for Social Entrepreneurship Research Accelerator Grant ("Implementing the Local Refugee Match"). This work was supported by the Economic and Social Research Council grant number ES/R007470/1.

[10] *Annie* could be adapted for the Swedish context and the authors of this paper have already presented the first version of *Annie* at the Swedish Ministry of Finance for potential adoption.

Trapp and Ahani are grateful for the support of the National Science Foundation (grant CMMI-1825348), as well as that provided by the Foisie Business School at Worcester Polytechnic Institute. Mike Mitchell and Karen Monken from HIAS worked patiently with us for over a year to build and implement this software. David Delacrétaz, Scott Duke Kominers, Will Jones, Ariel Procaccia, Hillel Rapoport, Alvin Roth, Vince Slaugh, Tayfun Sönmez, and M. Utku Ünver gave invaluable comments on the first draft.

9. Appendices

9.1 Data Appendix

We obtain anonymized data on all individual refugees relocated by HIAS between 2010 and 2017. We focus on free cases, that is, refugees that can be freely allocated across affiliates as they have no pre-existing family ties. As stated in the main text, we use refugees arriving until 2016 to train our models, and those arriving in 2017 as a test sample. Note that the quota-relevant year starts on October 1. Therefore, 2017 refugees are those arriving from October 1 2016 to September 30 2017. After the split, we observe 2,486 refugees in the training sample and 498 refugees in the test sample.

What follows is a list of data features and definitions.

- Arrival Date: The years span FY 2010 through FY 2017.
- Case Number: This is an anonymized, unique identifier for each family; in total, there are 1,896 families and 5,326 refugees.
- Relationship Code: The relationship to the principal applicant for each individual in a family; these include Principal Applicant (PA), Husband (HU), Wife (WI), Daughter (DA), Son (SO), Stepdaughter (SD), Stepson (SN).
- Gender Code: Genders include Male and Female.
- Nationality: There are 33 nationalities represented.
- Language: There are 133 languages represented, with proficiency levels for reading, speaking, and writing.
- Education Level: Levels include kindergarten, primary, intermediate, secondary, technical school, pre-university, university, professional, and graduate school.

- Medical Condition: There are at least 31 types of medical conditions.
- Treatment Urgency: There are several levels indicating the degree of treatment urgency, including Ongoing, Immediate, Urgent.
- Urgency Code: This is how fast the case must be assured by the resettlement agency. Values include both normal and expedited (such as medical, protection, etc.).
- Affiliate: This is the local community to which family is resettled.
- Employed: This is a binary value indicating whether the refugee was employed 90 days after arrival.
- Date Of Birth

Summary statistics for the above features include:

- Average case size: The average size differs among nationalities, affiliates, and year of arrival. Across all cases, the average size is approximately 2.809.
- Average age: The average age is approximately 23 years; 42.81% of refugees are under the age of 18, 55.97% are between 18 to 64, and 1.22% are beyond 64 years of age.
- Total number of nationalities: The refugees originate from 33 different nationalities; 96% of which derive from 13 countries.
- Total number of languages: There are 133 different languages among all refugees.
- Fraction with tertiary education: 6.04% of all refugees (10.57% of adult refugees) have a tertiary education.

To estimate counterfactual employment probabilities (Section 3 of the paper), we recode and transform some of the observed features. From Relationship Code we create an indicator of being a single parents, and a counter (censored at 5) of the number of children in the household. From Language we obtain an indicator for English speaking and a counter of the number of languages spoken. From Medical Condition we create an indicator for whether the refugee suffers from any medical condition, and a counter (censored at 5) of the total number of medical conditions reported. We recode Education Level into four groups (less than secondary schooling, secondary schooling, advanced—but not college—degrees, and university and college level degrees). Finally, we use the primary NATIONALITY to group refugees in their area of origin (Africa, Middle East, Asia, or Other).[11] For estimating LASSO, we also manually construct interactions between these variables and add a second order polynomial in age. The full list of features used in the LASSO and GBRT models appears in Tables 3 and 4.

To correctly account for changes in the average level of employment over time, we add to the data quarter-specific macro-economic variables, that is, average US employment level (adjusted for seasonality) and average unemployment rate (not adjusted by seasonality).[12] In the interacted logit and LASSO models these macro variables do not interact with affiliates, as their purpose is simply to adjust the varying average employment level of refugees over time.

9.2 Machine Learning Models: Procedure and Diagnostics

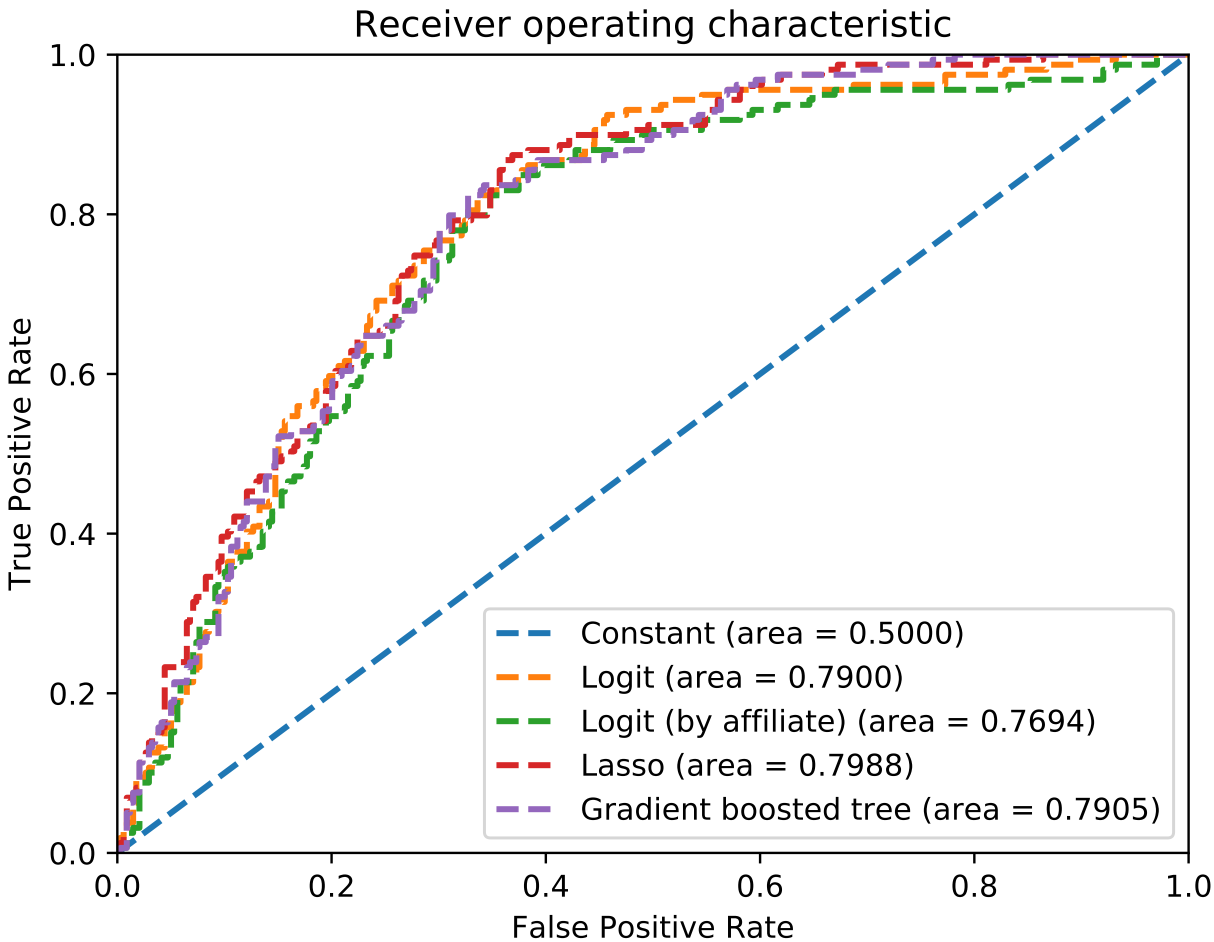
As stated in the main text, we restrict our data to refugees arriving between 2010 and 2016 for training our models, and test them on data for refugees arriving in 2017. For LASSO, we build a series of feature interactions, and then again fully interact this data matrix for each of the seven affiliates receiving more than 200 refugees until 2016. We standardize each feature such that it ranges from 0 to 1 in the training data (we use maxima and minima of the training set to standardize the test set). We use 5-fold cross-validation targeting the in-sample area under the Receiver Operating Characteristic (ROC) curve to tune the models' hyper-parameters.

Figure 12 shows Receiver Operating Characteristic (ROC) curves for LASSO, GBRT, and all benchmark models in the test data. ROC curves plot the achievable fraction of true positives as a function of the admissible false positives. The higher the fraction of true positives achievable for a given fraction of false positive is, the better is the performance of the model. Thus, curves to the northwest of the graph dominate the others. The graph shows that both LASSO and GBRT produce higher AUC-ROC than the benchmark models.

For both the GBRT and LASSO models, Figure 13 also shows calibration plots, depicting the average number of employed refugees in the test set for given predicted probabilities of employment. It is apparent that the predicted probabilities of employment after 90 days can be high for refugees and range from zero to approximately 0.8. This range of predicted probabilities for the US is in stark contrast with that observable in Europe, where predicted probabilities of employment rarely exceed

[11] We classify Oman, Lebanon, Iraq, Yemen, Iran, Bahrain, Syria, Qatar, Jordan, Kuwait, Israel, U.A.E. and Saudi Arabia as Middle East rather than Asia to better differentiate refugees from the Arabian peninsula and those from East Asia

[12] We add not-adjusted unemployment rates to capture seasonality in employment probabilities. Whether we adjust employment ration or unemployment rates for seasonality does not matter for our predictions.



Note: The figure plots the fraction of achievable true positives as a function of the fraction of false positives for each estimated model. The constant model is the benchmark used by Bansak et al. (2018). The logit model uses the same features used in LASSO for predicting employment, but without a LASSO constraint and affiliate-specific interactions. The logit by affiliate model uses the same features used in the LASSO model (including affiliate-specific interactions), but without a LASSO constraint. We compute all functions on refugees arriving in 2017 (test sample).

Figure 12: Receiver Operating Characteristics (ROC) curves.

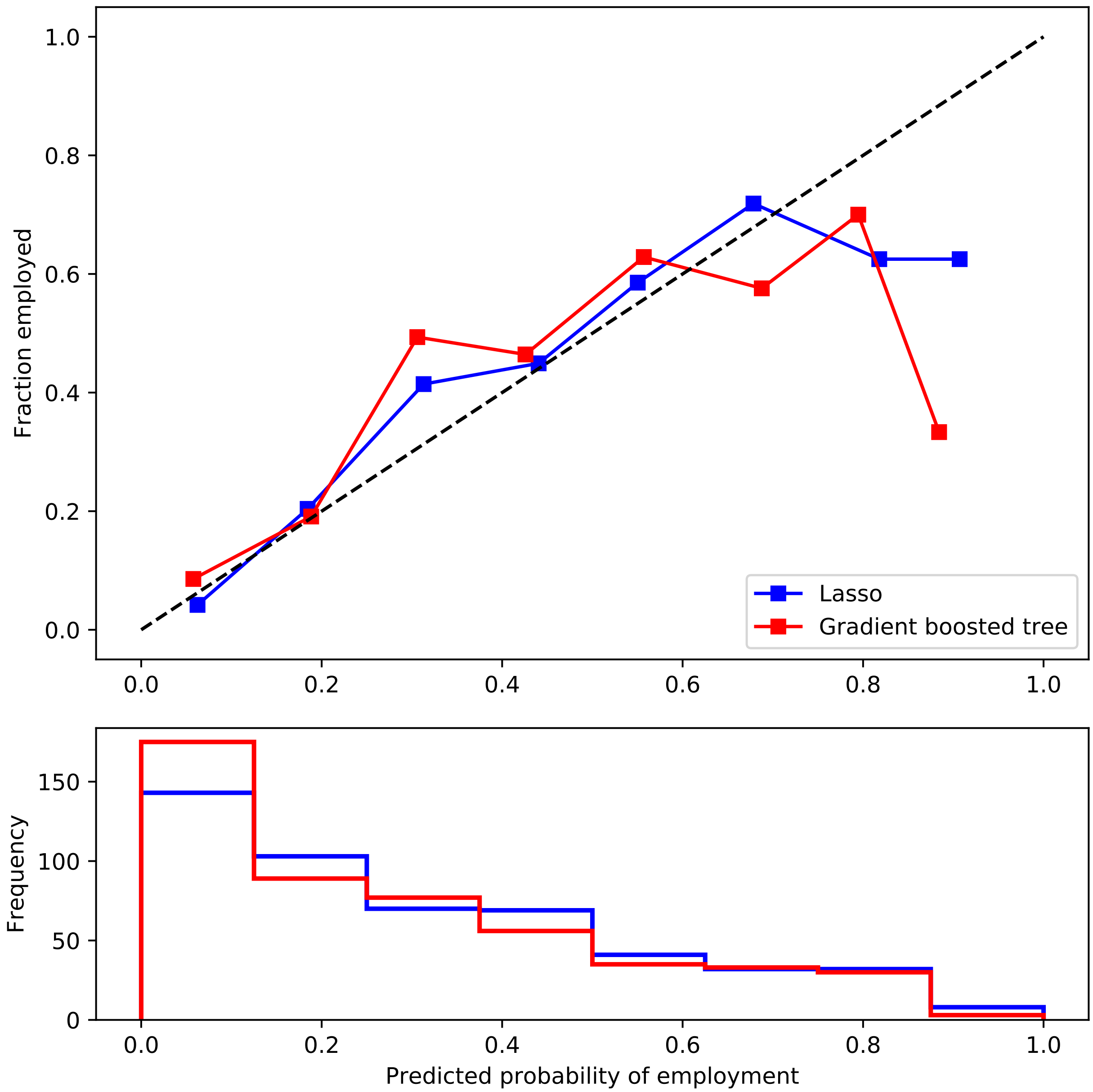
	Gini Importance
age	0.243
male	0.057
education level	0.065
case size	0.042
number of children	0.039
continent	0.069
affiliate	0.176
number of conditions	0.054
number of languages	0.024
English speaking	0.020
urgency code	0.010
primary applicant	0.022
unemployment rate (unadjusted)	0.131
employment ratio	0.049

NOTE: The table shows the normalized importance measure for each feature in the Gradient Boosted Regression Tree model. The coefficients sum to one. These measures are calculated as the average across all trees of mean decrease impurity scores for each node in which a given feature serves to split the data.

Table 3: Feature importance in the Gradient Boosted Regression Tree (GBRT) model.

0.5 (Bansak et al. 2018). LASSO is well calibrated until up to very high predicted probabilities, for which in the test data we observe a lower rate of employment than predicted. This behavior is primarily due to our out-of-sample extrapolation using macro-economic data for the affiliates for which we have little data. Without the inclusion of macro data as model features both LASSO and GBRT models are better calibrated, but tend to under-predict average employment levels.

The remainder of this Appendix reports normalized feature (Gini) importance scores for GBRT and model coefficients for LASSO. Note that these scores and coefficients, while broadly indicative of the amount of explanatory power contained in each feature, should not be taken as direct measures of feature relevance, especially as most features in our data are strongly correlated with one another. This point is particularly relevant for LASSO coefficients. While we standardize all model features such that they range from zero to one in the training sample, their standard deviation varies considerably. Moreover, as LASSO constraints penalize coefficients different than zero, whether the model selects and estimates a coefficient for one of two correlated variables or the other is irrelevant for the performance of the model.



NOTE: Both panels in the figure plot predicted employment probabilities by either LASSO or GBRT in the x-axis for the test data (refugees arriving in 2017). The top panel of the figure plots for each predicted employment probability the average number of effectively employed refugees in 2017. The bottom panel shows the histogram of the predicted employment probabilities in the test sample.

Figure 13: Calibration plots of LASSO and GBRT models (2017 data).

	Baseline	Affiliate C	Affiliate F	Affiliate I	Affiliate K	Affiliate N	Affiliate Q	Affiliate R
age	0.277				0.923			
male	1.289			-0.055	0.288			
medical condition			-0.647		-0.505	0.037	-0.207	
case size					0.922	-0.056		
number of children	-1.966		-0.243			-0.295		
single parent	0.438				0.030	0.083	-0.832	
number of conditions	-0.417			-0.797			-0.416	
number of languages	0.153				0.007			
English speaking	0.225	0.183	0.108	0.114	0.208		0.130	-0.032
urgency code		-0.334		0.715	-0.276			
age2	-2.050					-0.103		
primary applicant	0.136	-0.147		-0.021	0.637	0.040		0.642
education level_1-less than secondary				0.134		0.125		
education level_2-secondary	0.051	-0.318		-0.016				0.435
education level_3-advanced	0.719							
education level_4-university			1.003			0.680		
continent_Asia	-0.140							
continent_Middle east	-0.639							
continent_other	-0.400	1.236			0.278		1.424	0.613
1.education level#1.male	0.004	0.390						0.568
2.education level#1.male	0.120		0.203			0.022		
3.education level#1.male					-0.195			
4.education level#1.male	0.017							-0.001
c.number of children#1.male	1.134			-0.218		-0.122		
c.age#1.male								
1.primary applicant#1.male	-0.076			-0.367				
1.single parent#1.male	-0.655	1.105						
c.number of conditions#1.male		0.096		-0.220		0.213		
unemployment rate (unadjusted)								
employment ratio	0.939							
constant	-0.974	-0.289	1.805	0.713	0.208			0.373

NOTE: The table shows the estimated nonzero coefficients in the LASSO model. The first column shows the baseline coefficients of the model, while the other columns show the estimated interactions with each of the seven affiliates for which we observe more than 200 refugees before 2017.

Table 4: Estimated coefficients in the LASSO model.

9.3 Counterfactual Optimization Outcomes for GRBT

Capacity Adjustment	Min Avg Case Size	Binary Service Constraints	Total Expected Employed Refugees	Gains wrt to Predicted Employed Refugees (143)	St Dev in Avg Case Size Across Affiliates	# of Unplaced Cases / Refugees	#of Affiliates Violating 90% Capacity	# and % of Cases/Refugees Violating Constraints
Observed	None	Off	175.621	23%	1.257	0/0	0	64/200(19.45%/23.84%)
Observed	None	On	174.084	22%	1.465	2/3	0	0/0
Observed	2	Off	167.593	17%	0.916	3/3	0	72/186(21.88%/22.17%)
Observed	2	On	166.206	16%	0.922	1/2	0	0/0
Observed	2.5	Off	159.319	11%	0.326	4/4	0	80/186(24.32%/22.17%)
Observed	2.5	On	157.915	10%	0.326	4/7	0	0/0
Observed	3	Off	142.172	-1%	0	80/92	9	53/144(16.11%/17.16%)
Observed	3	On	141.222	-1%	0.654	81/95	10	0/0
Observed	Observed	Off	161.749	13%	1.079	4/4	1	80/202(24.32%/24.08%)
Observed	Observed	On	160.409	12%	1.001	3/4	1	0/0
$\leq$ 110%	None	Off	178.488	25%	1.314	0/0	3	59/180(17.93%/21.45%)
$\leq$ 110%	None	On	176.899	24%	1.451	1/2	6	0/0
$\leq$ 110%	2	Off	171.457	20%	0.862	0/0	7	61/154(18.54%/18.36%)
$\leq$ 110%	2	On	170.072	19%	1.245	1/2	7	0/0
$\leq$ 110%	2.5	Off	162.735	14%	0.786	0/0	6	68/168(20.67%/20.02%)
$\leq$ 110%	2.5	On	161.347	13%	0.735	2/3	5	0/0
$\leq$ 110%	3	Off	145.202	2%	0.654	80/92	8	52/133(15.81%/15.85%)
$\leq$ 110%	3	On	144.183	1%	0.654	81/95	10	0/0
$\leq$ 110%	Observed	Off	164.461	15%	1.17	0/0	4	69/166(20.97%/19.79%)
$\leq$ 110%	Observed	On	163.116	14%	1.08	2/3	6	0/0
[90%, 110%]	None	Off	178.479	25%	1.44	0/0	0	57/177(17.33%/21.10%)
[90%, 110%]	None	On	176.897	24%	1.242	1/2	0	0/0
[90%, 110%]	2	Off	171.454	20%	0.843	0/0	0	62/169(18.84%/20.14%)
[90%, 110%]	2	On	170.072	19%	1.175	1/2	0	0/0
[90%, 110%]	2.5	Off	162.74	14%	0.338	0/0	0	80/184(24.32%/21.93%)
[90%, 110%]	2.5	On	161.339	13%	0.334	2/3	0	0/0
[90%, 110%]	3	Off			Model is infeasible			
[90%, 110%]	3	On			Model is infeasible			
[90%, 110%]	Observed	Off	164.425	15%	0.86	4/4	0	79/196(24.01%/23.36%)
[90%, 110%]	Observed	On	163.058	14%	0.881	5/6	0	0/0

Table 5: Results of counterfactual employment optimization under various scenarios (using the GBRT model).

References

Abdulkadiroğlu, A. and Sönmez, T. School choice: A mechanism design approach. *American Economic Review*, 93:729–747, 2003.

Ager, A. and Strang, A. Understanding integration: A conceptual framework. *Journal of Refugee Studies*, 21(2):166–191, 2008.

Alcalde, J. and Barberà, S. Top dominance and the possibility of strategy-proof stable solutions to the marriage problem. *Economic Theory*, 4:417–435, 1994.

Andersson, T. and Ehlers, L. Assigning refugees to landlords in Sweden: Efficient stable maximum matchings. Technical report, 2016.

Andersson, T., Ehlers, L., and Martinello, A. Dynamic refugee matching. Technical report, 2018.

Åslund, O., Edin, P.-A., Fredriksson, P., and Grönqvist, H. Peers, neighborhoods, and immigrant student achievement: Evidence from a placement policy. *American Economic Journal: Applied Economics*, 3 (2):67–95, 2011.

Åslund, O. and Fredriksson, P. Peer effects in welfare dependence quasi-experimental evidence. *Journal of Human Resources*, 44(3):798–825, 2009.

Åslund, O., Östh, J., and Zenou, Y. How important is access to jobs? Old question, improved answer. *Journal of Economic Geography*, 10(3):389–422, 2010.

Åslund, O. and Rooth, D.-O. Do when and where matter? Initial labour market conditions and immigrant earnings. *The Economic Journal*, 117(518):422–448, 2007.

Aziz, H., Chen, J., Gaspers, S., and Sun, Z. Stability and Pareto optimality in refugee allocation matchings. Technical report, Mimeo, 2017.

Bansak, K., Ferwerda, J., Hainmueller, J., Dillon, A., Hangartner, D., Lawrence, D., and Weinstein, J. Improving refugee integration through data-driven algorithmic assignment. *Science*, 359(6373):325–329, 2018.

Barberà, S. and Jackson, M. Strategy-proof exchange. *Econometrica*, 63:51–87, 1995.

Besiou, M., Pedraza-Martinez, A. J., and Van Wassenhove, L. N. OR applied to humanitarian operations. *European Journal of Operational Research*, 269:397–405, 2018.

Biró, P. and Gudmundsson, J. Complexity of Finding Pareto-Efficient Allocations of Highest Welfare. Technical report, March 2018.

Bolt, D. An inspection of the vulnerable persons resettlement scheme, 2018.

COIN-OR. CBC User Guide, 2018. URL `http://www.coin-or.org/cbc`.

Damm, A. P. Neighborhood quality and labor market outcomes: Evidence from quasi-random neighborhood assignment of immigrants. *Journal of Urban Economics*, 79:139–166, 2014.

Delacrétaz, D., Kominers, S. D., and Teytelboym, A. Refugee resettlement. Technical report, University of Oxford, 2016.

Edin, P.-A., Fredriksson, P., and Åslund, O. Ethnic Enclaves and the Economic Success of Immigrants—Evidence from a Natural Experiment. *Q J Econ*, 118(1):329–357, February 2003. ISSN 0033-5533. URL `https://academic.oup.com/qje/article/118/1/329/1917042`.

Ferwerda, J. and Gest, J. Location, location: Refugee resettlement and employment outcomes in the United States. Technical report, Dartmouth College, 2017.

Fréville, A. The multidimensional 0–1 knapsack problem: An overview. *European Journal of Operational Research*, 155(1):1–21, 2004.

Gurobi. *Gurobi Optimizer 8.0.1 Reference Manual*. Gurobi Optimization, Inc., 2018.

Hifi, M., Michrafy, M., and Sbihi, A. Heuristic algorithms for the multiple-choice multidimensional knapsack problem. *Journal of the Operational Research Society*, 55(12):1323–1332, 2004.

Jones, W. and Teytelboym, A. The local refugee match: Aligning refugees' preferences with the capacities and priorities of localities. *Journal of Refugee Studies*, 2016.

Jones, W. and Teytelboym, A. The international refugee match: A system that respects refugees' preferences and the priorities of states. *Refugee Survey Quarterly*, 36(2):84–109, 2017.

Lichtenstein, G., Puma, J., Engelman, A., and Miller, M. The refugee integration survey evaluation project (RISE) year five: Final report, 2016.

Martello, S. and Toth, P. Solution of the zero-one multiple knapsack problem. *European Journal of Operational Research*, 4(4):276–283, 1980.

Meignan, D., Knust, S., Frayret, J.-M., Pesant, G., and Gaud, N. A review and taxonomy of interactive optimization methods in operations research. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 5(3):17, 2015.

Mitchell, S., O'Sullivan, M., and Dunning, I. PuLP: A linear programming toolkit for Python, 2018. URL `https://code.google.com/p/pulp-or/`.

Moraga, J. F.-H. and Rapoport, H. Tradable immigration quotas. *Journal of Public Economics*, 115:94–108, 2014.

Patacchini, E. and Zenou, Y. Ethnic networks and employment outcomes. *Regional Science and Urban Economics*, 42(6):938–949, November 2012. ISSN 0166-0462. URL `http://www.sciencedirect.com/science/article/pii/S0166046212000075`.

Pedraza-Martinez, A. J. and Van Wassenhove, L. N. Empirically grounded research in humanitarian operations management: The way forward. *Journal of Operations Management*, 45:1–10, 2016.

Ronacher, A. Flask (python microframework), 2018. URL `http://flask.pocoo.org/`.

Roth, A. E. The economics of matching: Stability and incentives. *Mathematics of Operations Research*, 7: 617–628, 1982.

Roth, A. E. Marketplaces, markets, and market design. *American Economic Review*, 108(7):1609–58, 2018.

Roth, A. E., Sönmez, T., and Ünver, M. U. Kidney exchange. *Quarterly Journal of Economics*, 119:457–488, 2004.

Slaugh, V. W., Akan, M., Kesten, O., and Ünver, M. U. The Pennsylvania Adoption Exchange improves its matching process. *Interfaces*, 46(2):133–153, 2016.

Song, Y., Zhang, C., and Fang, Y. Multiple multidimensional knapsack problem and its applications in cognitive radio networks. In *MILCOM 2008-2008 IEEE Military Communications Conference*, pages 1–7. IEEE, 2008.

Sönmez, T. Strategy-proofness and essentially singled-valued cores. *Econometrica*, 67:677–689, 1999.

SOU. Ett ordnat mottagande – gemensamt ansvar för etablering eller återvändande. Statens Offentliga Utredningar, SOU 2018:22,, Stockholm, 2018.

Trapp, A. C., Teytelboym, A., Ahani, N., and Andersson, T. Refugee resettlement via machine learning and integer optimization. In *OR60 Annual Conference – Keynote Papers and Extended Abstracts*. The OR Society, 2018.

UNHCR. Global Resettlement Needs 2018. Technical report, United Nations High Commissioner for Refugees, June 2017.

UNHCR. Mid-year trends, June 2017. Technical report, UHNCR, March 2018.