Placement Optimization in Refugee Resettlement

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

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
Revised: July 2020

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 tens of 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 decisions. We integrate machine learning and integer optimization into an innovative software tool, *Annie* Moore, that assists a US resettlement agency with matching refugees to their initial placements. Our software suggests optimal placements while giving substantial autonomy to the resettlement staff to fine-tune recommended matches, thereby streamlining their resettlement operations. Initial backtesting indicates that *Annie* can improve short-run employment outcomes by 22%–38%. We conclude by discussing several directions for future work.

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

1. Introduction

In 2018 there were 20.4 million refugees—the highest number ever recorded—under the mandate of the United Nations High Commission for Refugees (UNHCR) (UNHCR 2019b). Of those, the UNHCR considers 1.44 million refugees to be in need of *resettlement*—permanent relocation from their asylum country to a third country (UNHCR 2019a). The number of cases submitted by the UNHCR for resettlement in 2018, however, was just over 81,000, with fewer than 56,000 refugees departing for resettlement (UNHCR 2019a). 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 2019a, Annex 3). Currently, most refugees departing for resettlement are Syrians who seek asylum in Turkey, Lebanon, and Jordan, but there are also thousands of resettled refugees from the Democratic Republic of the Congo, Iraq, Somalia, and Myanmar.

Dozens of countries, including the United States (US), Canada, the United Kingdom (UK), Australia, France, Norway, and Sweden, resettle refugees (for refugee allocation mechanisms *across* countries, see Moraga and Rapoport (2014) and Jones and Teytelboym (2017b)). 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 best initial match between the refugee family and the community is crucial for social, economic, and humanitarian perspectives. Even so, 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 into a software package that we call *Annie* Moore (Matching and Outcome Optimization for Refugee Empowerment), named after Annie Moore, the first immigrant on record at Ellis Island, New York in 1892. *Annie* is, to the best of our knowledge, the first software designed for resettlement agency pre-arrival staff to recommend data-driven, optimized matches between refugees and local affiliates while respecting refugee needs and affiliate capacities. *Annie* was developed in close collaboration with representatives from all levels of the US resettlement agency HIAS (founded as the Hebrew Immigrant Aid Society), where the first version was deployed in May 2018. It was presented in August 2018 to the US Department of State and all staff at HIAS, with new features being regularly added.

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 end-user autonomy by the end-user. This flexibility empowers staff to focus more of their resources on refugee families that might be more challenging to match (for example due to complex medical conditions). Backtesting indicates that *Annie* would have been able to increase employment outcomes among refugees resettled by HIAS in 2017 by between 22% and 38%, 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, estimation procedures and diagnostics, and a variety of experiments using different objective functions and testing the sensitivity of our modeling.

2. Background Context and Previous Work

Annie helps HIAS to resettle refugees in the US, so we briefly describe the US resettlement program.

2.1 Refugee Resettlement in the United States

The US has historically been, by a wide margin, the world's largest destination of resettled refugees, with 22,491 arriving in 2018, down from 53,716 in 2017, and 84,994 in 2016 (in this manuscript, all references to years of refugee data are presented in terms of the *fiscal* year, that is from October 1 through September 30) (U.S. Department of State 2020). In terms of refugees resettled per capita (calendar year), the US trails a number of countries including Sweden, Norway, Canada, New Zealand, Iceland, and Australia. The US Refugee Admissions Program (USRAP) resettles US refugees and is managed by the Bureau of Population, Refugees and Migration (PRM) of the US Department of State, with the assistance of 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, previously known as *voluntary agencies*. In addition to HIAS, these agencies include Church World Service (CWS), Ethiopian Community Development Council (ECDC), Episcopal Migration Ministries (EMM), International Rescue Committee (IRC), Lutheran Immigration and Refugee Service (LIRS), US Conference of Catholic Bishops (USCCB), US Committee for Refugees and Immigrants (USCRI), and World Relief (WR). HIAS handles around 5% of all refugees in the US, resettling 2,038 refugees in 2017 and 3,844 refugees in 2016. The resettlement agencies are responsible for developing their own networks of *affiliates* in local communities that welcome refugees and help them integrate into a new life in the US. Affiliates offer resettlement capacity voluntarily, although affiliate capacity is monitored and approved by the US government. There are currently around 360 affiliates in approximately 200 local communities across the US, and HIAS operates 20 of them at the time of this writing.

Resettlement 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 of an immediate *family* of one or more members (we use *case* and *family* interchangeably). A significant portion (roughly one third) of these are free cases, that is, they have no relatives in the US. Such cases are especially vulnerable, as the absence of family support exacerbates the challenges of lacking language skills and independent financial means. Thus, the responsible agency must carefully leverage its affiliate network to inform their case selection. After each agency selects their set of weekly cases, staff manually assess—on a one-by-one basis—the feasibility and fit of cases to locations in their network. 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 available placement capacity, housing availability, slots for English language instruction, and employment prospects.

This manual process creates multiple inefficiencies that motivated the development of *Annie*. First, it is organizationally demanding for HIAS staff to keep in mind various support attributes such as languages, nationalities, family composition, and medical needs for all affiliates. This information overload at times 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. Established indicators include employment and economic sufficiency, developed social networks, and civic engagement activities like voting (see, e.g., Ager and Strang 2008, Lichtenstein et al. 2016). Hence, refugees are often not placed to the best available affiliate even according to well-defined outcome metrics. Third, inefficiencies arise from processing refugees case-by-case, in sequential fashion, rather than matching all arriving refugees to affiliates simultaneously. We show that *Annie* resolves or mitigates each of these inefficiencies.

2.2 Related Literature

Our work builds on a number of contemporary studies in humanitarian matching systems. One recent example is a tool to match children in state custody 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 (see also Delacrétaz et al. (2019), Trapp et al. (2018), and Nguyen et al. (2019)). 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 2017a, Andersson and Ehlers 2017, Aziz et al. 2018, Roth 2018, Delacrétaz et al. 2019, Nguyen et al. 2019).

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, similar to other humanitarian operations problems, placement optimization also diverges from the traditional stance of optimizing a single financial metric, and may consider alternative objectives such as those based on equity (see, e.g., Besiou and Van Wassenhove 2020, Cao et al. 2016, Çelik et al. 2012). 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.

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.

3.1 Formal Problem Setup

We use i, j, k, and ℓ as indices for family (case), member, service and affiliate, respectively. For any placement period, let $\mathcal{F} = \left\{F^1, F^2, \ldots, F^i, \ldots, F^{|\mathcal{F}|}\right\}$ be the set of refugee families to be placed. Each family F^i is a set of refugees consisting of one or more members, $F^i = \left\{f^{i,1}, f^{i,2}, \ldots, f^{i,j}, \ldots f^{i,|F^i|}\right\}$. For clarity of exposition, we refer to member j of family F^i as f^{ij}. Denote as N_w^i the set of working-age refugees in family F^i, where $N_w^i \subseteq F^i$. Denote the set of all refugees as $\mathcal{R} = \bigcup_{i \in \{1,2,\ldots,|\mathcal{F}|\}} \bigcup_{j \in \{1,2,\ldots,|F^i|\}} \left\{f^{ij}\right\}$. Moreover, let the set of affiliates (localities) to which families are resettled be $\mathcal{L} = \left\{L^1, L^2, \ldots, L^\ell, \ldots L^{|\mathcal{L}|}\right\}$.

A family F^i requires various *capacitated services* from a set $\mathcal{S} = \left\{S^1, S^2, \ldots, S^k, \ldots S^{|\mathcal{S}|}\right\}$. The needs of family F^i are summarized by vector s^i, where s_k^i denotes the required units of service k. The set $\mathcal{S}$ may include services such as 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 of local affiliate L^ℓ, at most $\bar{s}_k^\ell$ units may be filled by families placed in affiliate L^ℓ. There may also be a requirement of at least $\underline{s}_k^\ell$ units of the service S^k to be filled by the families placed in affiliate L^ℓ (we assume $\underline{s}_k^\ell \leq \bar{s}_k^\ell$); in practice, nonzero lower bounds exist for certain services, such as ensuring regular, positive refugee placement in affiliates.

For every family F^i and local affiliate L^ℓ, let binary variable z_ℓ^i equal 1 if family F^i is matched to affiliate L^ℓ, and 0 otherwise. Let $a_\ell^i \in \{0, 1\}$ indicate whether family F^i can be feasibly placed in affiliate L^ℓ. The value of a_ℓ^i is a priori 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 denote these community support services as *binary services*.

We attribute to each refugee-affiliate match a single number called the *quality score*. The function $q : \mathcal{R} \times \mathcal{L} \to \mathbb{R}_{\geq 0}$ defines quality score q_ℓ^{ij} for any $f^{ij} \in \mathcal{R}$ and any $L^\ell \in \mathcal{L}$. We are 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. In Section 7 we discuss the sole use of employment data to generate these estimates (indeed, no other data related to integration outcomes is systematically available). We aggregate the refugee level quality scores q_ℓ^{ij} of each family F^i and affiliate ℓ into a case-level *value* (or weight) v_ℓ^i. The primary means of aggregation that we consider is the *sum* of individual scores over each family $v_\ell^i = \sum_{j=1}^{|F^i|} q_\ell^{ij}$ (SUM). Discussions on alternative interpretations of case-level quality scores can be found in Appendix D.

3.2 Placement Optimization

We now present integer optimization problem RefMatch, represented by (1a)–(1e):

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

Objective function (1a) maximizes the total value over all matched families to affiliates. Constraint set (1b) ensures that families are placed in at most one affiliate. Constraint set (1c) ensures that lower and upper bounds are respected for all capacitated services and affiliates. Constraint set (1d) 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 (1e). Finally, let $z^\star$ be the optimized match outcome, that is, the optimal solution representing the assignment of families to affiliates that optimizes objective function (1a).

While integer optimization problem RefMatch bears similarity to a variety of knapsack-like problem classes, we are unaware of another application of this 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 $\underline{s}_k^\ell = 0 \ \forall\, \ell$, it is the NP-hard *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).

- When $|\mathcal{L}| = 1$ and $\underline{s}_k^\ell = 0 \; \forall \; k$, it is the NP-hard *multidimensional 0–1 knapsack problem* which features knapsack items that consume integer resources along multiple dimensions (Fréville 2004).

- When $\underline{s}_k^\ell = 0 \; \forall \; \ell, k$, it is the NP-hard *multiple multidimensional knapsack problem* and combines features of both, that is, multiple knapsacks along multiple dimensions (Song et al. 2008). If in addition, $\sum_{\ell=1}^{|\mathcal{L}|} z_\ell^i = 1 \; \forall \; i$, it is the NP-hard *multiple-choice multidimensional knapsack problem* (Hifi et al. 2004); in our setting, there is no requirement (in theory) for every family to be placed in an affiliate.

Integer optimization problem RefMatch generalizes the *multiple multidimensional knapsack problem* of Song et al. (2008), as it allows for positive lower bounds $\underline{s}_k^\ell$ for any services and affiliates. The existence of such lower bounds differentiates it from the multiple multidimensional knapsack problem, as it may lead to infeasibility. The formulation is valid over any operational period. Due to its generality, our model can be customized to specific refugee resettlement settings. Section 5 shows the results of testing the sensitivity of our model under three different scenarios. First, we test the effect of relaxing upper bounds (1c) for the number of total resettled refugees. Second, we test the effects of lower bounds (1c) expressed as distributional requirements (such as minimum average case sizes across affiliates) and as lower bounds on the total number of resettled refugees. We also consider the effects of relaxing the binary service constraints (1d). We discuss alternative models and objective functions and conduct sensitivity checks in Appendices D, E, and F.

4. Estimation of Counterfactual Employment Probabilities

We use the estimated probability of employment of refugee f^{ij} in each affiliate L^ℓ as a measure of quality score, or:

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

where y_{ij} is (binary) outcome data indicating employment status of refugee f^{ij} within 90 days of arrival in the United States, and $\mathbf{X}_{ij}$ is 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 A.

Using expected potential outcomes rather than stated preferences for our counterfactual analysis creates two challenges. First, y_{ij} is unobserved for incoming refugees. Second, even for past refugees we only observe $y_{ij}|x^{ij}_{\ell^\star}$, that is, employment status of refugee f^{ij} in $\ell^\star$, the affiliate to which they were actually assigned in the data. We do not observe the corresponding potential outcome distribution $y_{ij} \mid x^{ij}_{\ell} \; \forall \; \ell \neq \ell^\star$. 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 to compute $\hat{q}^{ij}_{\ell}$, the estimated probability of employment of refugee f^{ij} in affiliate L^{ℓ}. 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, which are those refugees (individuals or families) that the resettlement agency can assign to any of the affiliates. We therefore exclude refugees with pre-existing family ties, which are almost always pre-assigned to the affiliate where their pre-existing connection resides. This choice restricts the samples we use to train and test the models to 2,486 and 498 refugees , respectively.

While it may be tempting to increase the number of available observations for model estimation by including all refugees resettled by HIAS, the additional refugees will likely differ from the free cases to which *Annie* will be applied, and including them in the estimation might introduce bias and 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. By restricting our sample to free cases, we align the sample used for estimation with the sample on which *Annie* will be applied.

We estimate effects on employment for the seven (out of twenty) affiliates receiving at least 200 refugees up to 2016, and aggregate the remaining affiliates into a single partition ℓ_0. In a parametric approach, it is possible to 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 use 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. The single LASSO hyper-parameter disciplines both main and interaction terms with the same weight, biasing them towards zero (and thus biasing predictions towards the mean). 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 (we internally calibrate constraint strength for LASSO, and the learning rate and pre-pruning level for GBRT). We choose hyper-parameter values for each model by maximizing the area under a Receiver Operating Characteristic (ROC) curve.

We benchmark both models against the performance of a naïve constant estimator (see, e.g., Bansak et al. 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 effects. 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. 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. The area under the ROC is highest for LASSO, but overall both models exhibit similar predictive power.

Insert Table 1 Here

Insert Figure 1 Here

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 of refugee f^{ij} to $\ell^{\star}$. 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 of refugee f^{ij} to $\ell^{\star}$ is 157.93 for LASSO, it is only 142.96 for GBRT (calibration plots appear in Appendix B). 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 (and in Appendix C, we replicate employment gains under the predictions of GBRT).

5. Counterfactual Optimization Outcomes

We now describe the counterfactual impact of using our integer optimization problem RefMatch. 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 157.93 employed refugees versus 159 who were actually employed in the testing data—our optimization is a meaningful counterfactual exercise.

Our objective function (1a) maximizes the total expected number of employed refugees. Our binary service constraints (1d) are: language, nationality, single-parent, and large-family support. We set the capacity constraints (1c) 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 (1c). 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, a common aim of the agencies is 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 cases. However, binary service constraints, especially language constraints, can be important to ensure successful refugee integration. We therefore test two values: {binary service constraints are activated (ON), binary service constraints are not activated (OFF) }.

Minimum average case size in each affiliate. A placement that maximizes the total expected number of employed refugees could potentially pack 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. Therefore, to capture such equity considerations, we experiment with the implementation of a minimum average case size in each affiliate. The average case size in our 2017 test dataset across all affiliates is 2.55. We therefore test five values: {no minimum average case size, observed minimum average case size at each affiliate, 2, 2.5, 3}.

In total, we have $3 \times 2 \times 5 = 30$ counterfactual test scenarios. Akin to Bansak et al. (2018), we conduct our experiments using capacity levels for the period of one year. All experiments were run on a laptop computer with an Intel(R) Core(TM)i5-8365U 1.60GHz processor and 16GB RAM running 64-bit Microsoft Windows 10 Enterprise. The Gurobi Optimizer v9.0.0 (Gurobi 2020) and Python 3.7.4 was used for all counterfactual optimization testing in Section 5, and the optimality gap tolerance parameter MIPGap was set to 0. We summarize our results in Table 2.

Insert Table 2 Here

Insert Figure 2 Here

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 (almost) first-order stochastically dominates the pre-optimized estimated distribution. Therefore, the estimated probabilities of employment increase for all refugees 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.

Insert Figure 3 Here

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 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. It should be noted that these are precisely the instances for which many cases and refugees are unplaced, thus causing reduced optimal objective function values and corresponding gains.

Insert Figure 4 Here

We report the runtime as the time (in seconds) to both build the optimization model and solve it to global optimality using Gurobi (Gurobi 2020). It can be immediately observed that for the entire FY17 dataset (839 refugees / 329 cases / 498 working-age refugees / 20 affiliates), the combined build and solve times in Gurobi finish in well under one minute (actually, under 30 seconds), with a median combined runtime of less than four seconds.

Insert Figure 5 Here

Overall, our optimization produces a substantial gain in employment, ensures that refugee binary services are better satisfied, and important distributional considerations can be respected. Moreover, the resettlement agency may impose any subset of the binary service constraints, or introduce constraints on the number of refugees with certain regional origins (although regional constraints were formerly officially considered in US placements, they are no longer specified).

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 agency could select a different employment objective function, such as maximizing the sum of *maximum* employment probabilities within every matched case. In Appendix D we provide further experiments that optimize over several reasonable (including equity-based) measures based on derived from the individual refugee-level quality scores q_ℓ^{ij}; we find that all perform fairly well with respect to gains in employment. Further details on these experiments can be found in Appendix D.

We also recognize that there is inherent uncertainty in the modeling environment with respect to estimating the quality score $\hat{q}_\ell^{ij}$ for each f^{ij} in affiliate L^ℓ. In Appendix E, we investigate how the objective function changes under our optimized placement outcome $z^\star$ when we resample $\hat{q}_\ell^{ij}$ from the estimated distribution. In particular, we observe that the refugee allocation determined by our approach produces stable employment gains, and that these employment gains are not artificially inflated by uncertainty in the estimation of employment probabilities. The average expected employment given uncertainty in our predicted probabilities is within 2% of that obtained in our original backtesting for almost all of the considered scenarios.

Finally, we note that these outcomes were obtained by optimizing placement of all refugees in FY17 without splitting into multiple periods, that is, over the entire year ($n = 1$), on par with experiments reported in Bansak et al. (2018). While desirable, experiments with $n > 1$ placement periods in a given year introduced some additional nuances that required equally detailed implementation strategies. Even so, we present such experiments in Appendix F. The key takeaways include that increasing the number of periods to $n \in \{4, 12, 52\}$ (that is, quarterly, monthly, and weekly) for placing refugees, and thereby allowing for the innate arrival stochasticity present in FY17 data, reveal encouraging results. While gains are indeed largest for $n = 1$, our methods perform very well for $n = 4$ and $n = 12$, and respectably even for $n = 52$. We refer the reader to Appendix F for additional details.

6. Operationalizing Placement Software at US Resettlement Agency

Integer optimization and machine learning hold great promise of solving the operational challenge of improving placement outcomes in refugee resettlement. While these methods offer 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 and financial resources, lack of exposure to technology, and data scarcity. Humanitarian and non-profit organizations must be responsive to crisis events, immediate needs, and changes in political and donor climates. These realities can make it fairly prohibitive for resettlement agencies to be proactive in pursuing, and implementing, advanced technological innovations.

Successful integration of operations research methods 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 problem RefMatch is modeled entirely within the PuLP Python modeling environment (Mitchell et al. 2020) and solved using the CBC (COIN-OR 2020) solver. The machine learning models described in Section 4 are based on 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 2020). The front-end is a combination of HTML, CSS, and JavaScript. We made this choice of technology because it is modern, stable, accessible, and easy to build upon. 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.

Insert Figure 6 Here

6.2 Interactive Optimization

Representing overall match quality in objective function (1a) is by no means trivial. Employment probabilities for refugees will always be estimated with error margins (see Appendix E for experiments and related discussions around uncertainty in employment probability estimation, and corresponding sensitivity of match outcomes). Even if the employment probabilities could be perfectly estimated, any algorithmic solution should be carefully evaluated before actual implementation, as the overall livelihoods of refugees are at stake. 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 (see, e.g., Meignan et al. 2015). The user of *Annie* is intimately involved in the matching process and can fine-tune the optimization results. 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.

6.3 Features of *Annie*

Operationally, *Annie* optimizes for the expected number of employed *refugees* throughout the network of affiliates at HIAS. Alternative objective functions, such as those discussed in Appendix D, can be easily implemented.

Insert Figure 7 Here

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 seen in the *View Results* view depicted in the front right of Figure 6, where the total expected number of employed refugees is prominently displayed near the top.

Insert Figure 8 Here

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, along with the ability to quickly adjust capacity. 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 the next. Moreover, the total expected number of employed refugees is also dynamically updated at the top of the *View Results* view. Hence, at a glance, the effect of moving cases to alternative affiliates is easily and clearly visualized.

Insert Figure 9 Here

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 facilitated through a lock icon on the case tile that resettlement staff can click, which locks desired case-affiliate matches. Figure 9 depicts this capability.

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, with affiliate capacities adjusted down from any locked cases, via a color-coded gray reoptimize button that indicates the non-optimized state (see Figure 9). Thus, any "final" matches can be locked, and all remaining cases can be rematched using the remaining available capacity.

Insert Figure 10 Here

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 (note that these are cases with US ties, as previously described in Section 4). 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. For any two cases i, i' that are cross-referenced, *Annie* sets $z_\ell^i = z_\ell^{i'}$ for all local communities L^ℓ; and if i, i' are cross-referenced to a particular local community ℓ', *Annie* sets $z_{\ell'}^i = z_{\ell'}^{i'}$ only for local community $L^{\ell'}$. Figure 10 depicts an example where two cases are cross-referenced not only to one another (e.g., adult siblings), but also to an affiliate.

Insert Figure 11 Here

If a case tile is moved into an affiliate but there is a lack of compatibility between this case and the new affiliate in terms of binary community support services, the background color of the case becomes red as an indication, and an exclamation mark icon appears in the bottom left of the case tile (see Figure 11). Hovering over this exclamation mark icon displays a new list that shows the unsupported needs for that particular case-community match.

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, integer optimization, and interactive visualization—while arming human decision-makers with all available information to facilitate the decision-making process.

7. Conclusion

Refugee resettlement is a complex humanitarian challenge 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 US. We introduce the innovative software tool, *Annie* MOORE, that assists the US resettlement agency HIAS 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. Because *Annie* matches on refugee employment outcomes, we expect refugees to more quickly integrate economically into each affiliate, as well as make more productive economic and societal contributions such as creating new jobs and generating tax revenues, benefitting local communities.

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, enabling resettlement staff to place greater emphasis on cases that need greater attention, such as those with severe medical conditions.

Technological solutions, including machine learning and integer optimization, have enormous potential to help tackle humanitarian operations problems, such as placement optimization in refugee resettlement. 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 a 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 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 annual approved arrival capacities 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 over 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, the social objective considered in this paper is to maximize employment. Even if it can be argued that "there is no single, generally accepted definition, theory or model of immigrant and refugee integration" (Castles et al. 2002, p.114), it is also clear that there are key aspects of integration beyond employment. Ager and Strang (2008), for example, argue that there are ten established integration indicators, including health, housing, and education. This additional information—such as housing information, social networks, or new job opportunities—likely exists to at least some degree at the local community level, and could prove very useful in supplementing the decision process. Moreover, regular and sustained engagement of local communities and associated stakeholders can also produce valuable insights that augment decision outcomes (see, e.g., Johnson et al. 2018). So, while 90-day binary employment outcomes are at present the only data available to estimate future integration outcomes, additional integration factors may be possible to integrate in the future, and we thus leave the analysis of such models for future research.

Fifth, recent theoretical work on refugee matching (Aziz et al. 2018, Delacrétaz et al. 2019, Jones and Teytelboym 2017a, Olberg and Seuken 2019) suggests that preferences of refugees should be explicitly taken into account, because refugee families themselves know best where they would like to settle and where they are most likely to thrive. Refugee preferences could ideally 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 versus rural, crime, amenities, and quality of schools). Unfortunately, refugee preferences are not elicited either by UNHCR or by the US Department of State. In any case, the consideration of refugee preferences should be handled with care. Including preferences while optimizing for a particular observable outcome can in itself be problematic (Biró and Gudmundsson 2020), and it is also unclear how preferences should be elicited based on the reported information. Allowing refugees to report complete preferences may also be overly challenging. Hence, although it is clear that the approach we adopt has room for improvement, we believe it to be a reasonable approach in line with the growing evidence that the initial placement of refugee families greatly affects lifetime employment which, in turn, profoundly alters lifetime welfare (Åslund and Rooth 2007, Damm 2014, Ferwerda and Gest 2017).

A final challenge with eliciting refugees' preferences—and a main theme in the book by Roth (2016)—is that agents often find it "unsafe" to report true information. Rather than strategically misrepresenting information to "game" the system, agents may be reluctant to report complete information simply because a lack of knowledge on how the information will used, how it will be spread, or trust that the reported information will be used in their best interest. This is surprisingly often the case even in applications where the outcome has a large impact on future welfare and life quality, such as in school choice (Abdulkadiroğlu and Sönmez 2003) and kidney exchange (Roth et al. 2004). It is in general difficult to design systems where all agents find it "safe" to report true and complete preferences (see, e.g., Alcalde and Barberà 1994, Barberà and Jackson 1995, Roth 1982, Sönmez 1999). Recent work on refugee matching with preferences also indicates that it can be difficult to design matching systems in which refugees have no incentive to misreport their preferences (e.g. Andersson and Ehlers 2017, Delacrétaz et al. 2019).

Annie has primarily been developed to assist HIAS in their initial refugee placements, and there is significant potential to expand to additional resettlement contexts, both within the US, as well as beyond—the most direct being with other US resettlement agencies which face analogous placement decision challenges. *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 (43,745 such transfers were made in 2018). 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 carefully designed optimization and matching systems should be adopted (indeed, *Annie* could be adapted for the Swedish context; the authors of this paper have already presented the first version of *Annie* at the Swedish Ministry of Finance for potential adoption). Finally *Annie* may, for example, be adapted for distributing asylum seekers who are currently at reception centers in host countries (such as Germany, or the southern border of the US).

8. Acknowledgments

We are incredibly grateful to many for helping make this research successful. 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. Trapp and Ahani are grateful for the support of the National Science Foundation (Operations Engineering) grant CMMI-1825348, as well as that provided by the Foisie Business School at Worcester Polytechnic Institute. Mike Mitchell and Karen Monken and other members of HIAS staff have patiently worked with us since early 2017 to build and implement *Annie*. Emily Aiken provided invaluable software development assistance for an early prototype. Comments and suggestions from Barbara Day, David Delacrétaz, Pat Flaherty, Daniel Freund, Carina Hausladen, Scott Duke Kominers, Will Jones, Amin Khademi, Kayse Lee Maass, Karen Monken, Ariel Procaccia, Hillel Rapoport, Alvin Roth, Andrew Schaefer, Vince Slaugh, Tayfun Sönmez, and M. Utku Ünver significantly improved this paper.

Appendices

A 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 2010 through 2017, inclusive.
- 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.

- Age Upon Arrival

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 4 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; note that 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). 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). Note that we add non-adjusted unemployment rates to capture seasonality in employment probabilities; whether we adjust employment ratios or unemployment rates for seasonality does not matter for our predictions. 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.

B 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 at least 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 model 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.

Insert Figure 12 Here

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 0.5 (Bansak et al. 2018). LASSO is well calibrated 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.

Insert Figure 13 Here

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, LASSO constraints penalize coefficients different than zero. Given two strongly correlated features contributing similarly to the outcome predictions, a strong enough LASSO constraint will force one of the two associated coefficients to be equal to zero, and rely solely on the other feature for prediction. While this selection often improves the predictive performance of the model by reducing model complexity, it does not imply that the feature whose coefficient was pushed to zero has no predictive power at all. This selection simply implies that the information carried in that feature could be expressed as a function of other features without a strong loss in the in-sample explanatory power.

Insert Table 3 Here

Insert Table 4 Here

C Counterfactual Optimization Outcomes for GBRT

The outcomes from optimizing the SUM objective: $v_\ell^i = \sum_{j=1}^{|F^i|} q_\ell^{ij}$ using estimates $\hat{q}$ from the GBRT model are detailed in Table 5, which is organized in the same manner as Table 2. The baseline employment levels when considering the actual (manual) placements of cases to affiliates in the data is 142.96 when using the GBRT model.

Insert Table 5 Here

D Exploration of Alternative Objective Functions

Optimizing the placement of cases to affiliates allows for multiple interpretations for translating the individual refugee-level quality scores q_ℓ^{ij} into a case-level value (or weight), v_ℓ^i for each family F^i and affiliate ℓ. While we prioritize $v_\ell^i = \sum_{j=1}^{|F^i|} q_\ell^{ij}$ (SUM) for the case-level quality score v_ℓ^i appearing in (1a), as mentioned in Section 5 other reasonable interpretations exist. These include $v_\ell^i = \max_j q_\ell^{ij}$ (MAX); $v_\ell^i = \min_j q_\ell^{ij}$ (MIN); and $v_\ell^i = 1/|N_w^i| \sum_{j \in N_w^i} q_\ell^{ij}$ (MEAN), which because they position $v_\ell^i \in [0, 1]$, are more appropriately interpreted as information about the case. We now conduct a formal study of these alternatives and explore the associated tradeoffs.

Defining $v_\ell^i = \max_j q_\ell^{ij}$ (MAX) assigns to v_ℓ^i the highest employment likelihood of the members in family F^i for affiliate L^ℓ. Correspondingly, maximizing $\sum_{i=1}^{|\mathcal{F}|} \sum_{\ell=1}^{|\mathcal{L}|} v_\ell^i z_\ell^i$ is interpreted as emphasizing placements of families into communities according to their most employable member. Alternatively, taking $v_\ell^i = \min_j q_\ell^{ij}$ (MIN) assigns to v_ℓ^i the lowest employment likelihood of the members in family F^i for affiliate ℓ. Then, maximizing $\sum_{i=1}^{|\mathcal{F}|} \sum_{\ell=1}^{|\mathcal{L}|} v_\ell^i z_\ell^i$ produces outcomes whereby each member of every placed family can do at least as well as the lowest employment likelihood of the family. We note that both MAX and MIN have equity connotations: while the former attempts to maximize the number of families in which at least one refugee is likely to gain employment, the latter seeks to maximize the number of families in which all of the adults have some chance of getting employment. Finally, defining $v_\ell^i = 1/|N_w^i| \sum_{j \in N_w^i} q_\ell^{ij}$ (MEAN) assigns to v_ℓ^i the average employment likelihood of the (working-age) members in family F^i for affiliate ℓ. Maximizing $\sum_{i=1}^{|\mathcal{F}|} \sum_{\ell=1}^{|\mathcal{L}|} v_\ell^i z_\ell^i$ in this context can be interpreted as placing families into communities so as to ensure that the sum of mean likelihoods is as high as possible.

Insert Table 6 Here

Insert Table 7 Here

Insert Table 8 Here

The outcomes from optimizing using the MAX, MIN, and MEAN objectives are presented in Tables 6, 7, and 8, which are organized in the same manner as Table 2. Using the actual FY17 data, the count of cases for which at least one refugee was employed within 90 days was 151, whereas the baseline employment levels when evaluating the actual (manual) placements of cases to affiliates using these alternative objectives are 137.61 for MAX, 97.57 for MIN, and 117.49 for MEAN.

Overall, they exhibit similar performance to the SUM objective that is detailed in Table 2. With the exception of one test scenario for the MIN objective (at 62.74 seconds), all combined build and solve runtimes complete in well under one minute. We observe that the same test scenarios seem to perform comparatively well across all of the objectives. That said, for the MEAN objective some test scenarios approach and even exceed 40% gains. Unsurprisingly, the gains for the MIN objective are comparatively lower, and for a few of the scenarios with very restrictive constraints (e.g. minimum average case size of 3), negative gains are apparent in some of the objectives. Importantly, these low gain scenarios are precisely the instances for which many cases and refugees remained unplaced, thus excluding their contribution in the objective.

Insert Table 9 Here

Table 9 reports how the optimal solutions obtained by optimizing using the SUM, MAX, MIN, or MEAN objectives perform with respect to being evaluated in the alternative objectives, for each of the 30 test scenarios. In particular, we organize Table 9 so that for each of these objectives, one can readily compare the objective function values for the various optimal solutions. We also report the percent loss in objective function value from using an optimal solution to an alternative objective function.

As can be observed in Table 9, in general the optimal solutions to alternative objective functions perform quite well across all experiments—typically experiencing a loss of only a few percent when comparing SUM, MAX, and MEAN. The exception is the MIN objective, where more variation is observed—either in the high single digits or low double digits, up to a max of 16.9% under the test scenario with observed capacity, minimum average case size of three, de-activated binary service constraints, and evaluating the optimal solution to the MAX objective in the MIN objective. With respect to the MIN objective, unsurprisingly the evaluation in the MIN objective of optimal solutions obtained from optimizing the MEAN objective seem to perform best, with all percent losses in the single digits.

E Counterfactual Optimization Experiments with Uncertainty around Predictions

In this section we allow for uncertainty in the modeling environment as it pertains to estimating the quality score for each refugee and affiliate. We evaluate the performance of our optimized placement outcome $z^\star$ with respect to alternative objective functions sampled from the same distribution of estimated probabilities. In other words, we hold the optimal refugee allocation fixed, and evaluate the change in employment gains as we perturb and allow for errors in the estimated employment probabilities.

The bootstrapping approach described in Section 4 generated 1,000 sets of models that predict alternative employment probabilities for each refugee-affiliate pair. This exercise generates 1,000 objective functions, all of which have slightly different quality scores than that of the objective function obtained from the original estimated $\hat{q}_\ell^{ij}$ quality scores. To assess the performance of our optimization with respect to the uncertainty in predicted employment likelihoods, we then evaluate the optimized placement outcome $z^\star$ using each of the 1,000 bootstrapped instances for each of the thirty test scenarios.

We obtain a distribution of employment gains, which we would expect to produce lower employment gains on average than if our employment probabilities were predicted exactly, as in this latter case the optimal allocation is obtained using optimization. Nonetheless, this exercise shows that our approach produces remarkably stable employment gains even under uncertainty.

Insert Figure 14 Here

Figure 14 depicts box plots for 10 of the 30 scenarios for which there were feasible solutions. These box plots show the distribution of employment gains given uncertainty in our predicted employment probabilities, holding the optimal allocation fixed. The figure shows that the performance of $z^\star$ clusters tightly around the median values, and is well within the first and third quartiles. The performance of $z^\star$ exceeds the median since it was obtained using optimization. However, the average employment gains under uncertainty are very near (within 2% for most scenarios) to those obtained assuming that the predicted probabilities were at their means. We can therefore conclude that optimal allocations computed by our approach produce stable employment gains.

F Counterfactual Optimization Experiments with Multiple Placement Periods

To evaluate the performance of our approach with respect to inherent uncertainty in refugee arrivals, we consider counterfactual optimization of the SUM objective with $n > 1$ placement periods in FY17 data. Specifically, we consider $n = \{4, 12, 52\}$ placement periods over the 30 test scenarios, and compare with $n = 1$, that is, the results detailed in Table 2.

Given the limited size of the data, such experiments introduced some nuanced challenges that required creative handling. In particular, larger values of n may cause insufficient per-period capacities at certain affiliates to accommodate certain families having many members. This situation is further compounded by single-person families that have relatively high employment likelihoods, that are simply able to fill (pack) the available capacity at affiliates that offer superior employment prospects. Note that this was rarely the case under manual placements: affiliate capacity was regularly exceeded for any given period, so long as later periods use respectively less capacity.

Therefore, in the following experiments we incorporated some guidelines to accommodate the above challenges. All refugee families that are unplaced in a given period are included in the next placement period. For each period, the total period capacity over all affliates is set to the number of arriving refugees for that placement period, in addition to any unplaced refugees from earlier periods. The per-period capacity for each affiliate is then obtained by simply scaling the total period capacity by the affiliate's annual observed proportion (and rounding to the nearest integer). For any placement period, if any affiliate has already reached its annual capacity, its capacity for the present and future placement periods is set to zero, and the proportional share of per-period capacity is evenly distributed among other affiliates that have not yet used their annual capacity.

Insert Table 10 Here

Table 10 details the results of our experiments and is organized in the same manner as Table 2, with each row reporting one of the 30 test scenarios. There are four sets of columns, each of which reports the total expected employed refugees, percent gains over the predicted baseline of 157.93, and number of unplaced cases and refugees. The first set of columns repeats these values first reported in Table 2, whereas the last three report these values for $n = 4$, $n = 12$, and $n = 52$, respectively.

As can be seen in Table 10, the percent gains with respect to predicted employed refugees (157.93) remain high for $n = 4$ and $n = 12$. In particular, for the routine operational scenario that the second row embodies of using observed capacity, no minimum average case size constraints, and activated binary service constraints, we see that the gains drop only slightly from 31.86% for $n = 1$, to 30.52% for $n = 4$, and then again to a respectable 27.67% for $n = 12$. The earlier described effect of insufficient per-period capacities begins to be noticed for $n = 52$, which is why a percent gain of only 16.10% is achieved, and also why there are many more unplaced cases and refugees. In general, greater numbers of unplaced refugees and cases naturally lead to lower gains, which may even be negative in extreme scenarios. With greater numbers of refugees to be placed, we naturally would expect these gains to be higher.

References

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

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

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

Tommy Andersson and Lars Ehlers. Assigning refugees to landlords in Sweden: Efficient, stable, and maximum matchings. *The Scandinavian Journal of Economics*, 122(3):937–965, 2017.

Tommy Andersson, Lars Ehlers, and Alessandro Martinello. Dynamic refugee matching. Technical report, Lund University, 2018.

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

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

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

Olof Åslund, Per-Anders Edin, Peter Fredriksson, and Hans Grönqvist. Peers, neighborhoods, and immigrant student achievement: Evidence from a placement policy. *American Economic Journal: Applied Economics*, 3(2):67–95, 2011.

Haris Aziz, Jiayin Chen, Serge Gaspers, and Zhaohong Sun. Stability and Pareto optimality in refugee allocation matchings. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, pages 964–972. International Foundation for Autonomous Agents and Multiagent Systems, 2018.

Kirk Bansak, Jeremy Ferwerda, Jens Hainmueller, Andrea Dillon, Dominik Hangartner, Duncan Lawrence, and Jeremy Weinstein. Improving refugee integration through data-driven algorithmic assignment. *Science*, 359(6373):325–329, 2018.

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

Maria Besiou and Luk N Van Wassenhove. Humanitarian operations: A world of opportunity for relevant and impactful research. *Manufacturing & Service Operations Management*, 22(1):135–145, 2020.

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

Péter Biró and Jens Gudmundsson. Complexity of finding Pareto-efficient allocations of highest welfare. *European Journal of Operational Research*, 2020.

David Bolt. An inspection of the vulnerable persons resettlement scheme. Technical report, Independent Chief Inspector of Borders and Immigration, 2018.

Wenwei Cao, Melih Çelik, Özlem Ergun, Julie Swann, and Nadia Viljoen. Challenges in service network expansion: An application in donated breastmilk banking in South Africa. *Socio-Economic Planning Sciences*, 53:33–48, 2016.

Stephen Castles, Maja Korac, Ellie Vasta, and Steven Vertovec. Integration: Mapping the field. Technical report, Centre for Migration and Policy Research and Refugee Studies Centre, University of Oxford, 2002.

Melih Çelik, Özlem Ergun, Ben Johnson, Pınar Keskinocak, Álvaro Lorca, Pelin Pekgün, and Julie Swann. Humanitarian logistics. In *New Directions in Informatics, Optimization, Logistics, and Production*, pages 18–49. INFORMS, 2012.

COIN-OR. CBC user guide, 2020. URL `http://www.coin-or.org/cbc`.

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

David Delacrétaz, Scott Duke Kominers, and Alexander Teytelboym. Matching mechanisms for refugee resettlement. Technical report, University of Oxford, 2019.

Per-Anders Edin, Peter Fredriksson, and Olof Åslund. Ethnic enclaves and the economic success of immigrants—evidence from a natural experiment. *Quarterly Journal of Economics*, 118(1):329–357, February 2003.

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

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

Gurobi. *Gurobi Optimizer 9.0.0 Reference Manual*. Gurobi Optimization, Inc., 2020.

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

Michael P. Johnson, Gerald Midgley, and George Chichirau. Emerging trends and new frontiers in community operational research. *European Journal of Operational Research*, 268:1178–1191, 2018.

Will Jones and Alexander Teytelboym. The local refugee match: Aligning refugees' preferences with the capacities and priorities of localities. *Journal of Refugee Studies*, 31:152–178, 2017a.

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

Gary Lichtenstein, Jini Puma, Amy Engelman, and Maggie Miller. The refugee integration survey evaluation project (RISE) year five: Final report, 2016.

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

David Meignan, Sigrid Knust, Jean-Marc Frayret, Gilles Pesant, and Nicolas Gaud. A review and taxonomy of interactive optimization methods in operations research. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 5(3):1–43, 2015.

Stuart Mitchell, Michael O'Sullivan, and Iain Dunning. PuLP: A linear programming toolkit for Python, 2020. URL `https://github.com/coin-or/pulp`.

Jesús Fernández-Huertas Moraga and Hillel Rapoport. Tradable immigration quotas. *Journal of Public Economics*, 115:94–108, 2014.

Thành Nguyen, Hai Nguyen, and Alexander Teytelboym. Stability in matching markets with complex constraints. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 61–61. ACM, 2019.

Nils Olberg and Sven Seuken. Enabling trade-offs in machine learning-based matching for refugee resettlement. Technical report, University of Zurich, 2019.

Eleonora Patacchini and Yves Zenou. Ethnic networks and employment outcomes. *Regional Science and Urban Economics*, 42(6):938–949, November 2012.

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

Armin Ronacher. Flask (python microframework), 2020. URL `https://palletsprojects.com/p/flask/`.

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

Alvin E. Roth. *Who Gets What - And Why: Understand the Choices You Have, Improve the Choices You Make*. William Collins, 2016.

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

Alvin E. Roth, Tayfun Sönmez, and M. Utku Ünver. Kidney exchange. *Quarterly Journal of Economics*, 119:457–488, 2004.

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

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

Tayfun Sönmez. 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.

Andrew C. Trapp, Alexander Teytelboym, Narges Ahani, and Tommy Andersson. 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 2020. Technical report, United Nations High Commissioner for Refugees, June 2019a.

UNHCR. Global Trends: Forced Displacement in 2018. Technical report, United Nations High Commissioner for Refugees, 2019b.

U.S. Department of State. Refugee Arrivals by State and Nationality as of Mar. 6, 2020. `http://www.wrapsnet.org/admissions-and-arrivals`, 2020. Accessed: 2020-03-18.

	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). All of these measures refer to a binary classification with a threshold set at the standard value of 0.5. Because our measure of quality scores uses predicted probabilities of employment, this specific threshold does not affect optimal allocations. *AUC-ROC* measures the area under the Receiver Operating Characteristic Curve for each model (ROC curves appear in Appendix B).

Table 1: Model performance.

Capacity Adjustment	Min Avg Case Size	Binary Service Constraints	Total Expected Employed Refugees	Gains wrt to Predicted Employed Refugees (157.93)	St Dev in Avg Case Size Across Affiliates	# of Unplaced Cases / Refugees	#of Affiliates Violating 90% Capacity	# and % of Cases/Refugees Violating Constraints	Build and Run Time (s)
Observed	None	Off	213.02	34.89%	1.29	0/0	0	70/209 (21.28%/24.91%)	0.48
Observed	None	On	208.25	31.86%	1.35	3/10	1	0/0 (0.00%/0.00%)	0.36
Observed	2	Off	206.28	30.61%	1.04	1/1	0	81/220 (24.62%/26.22%)	0.92
Observed	2	On	202.03	27.92%	1.17	2/9	1	0/0 (0.00%/0.00%)	0.93
Observed	2.5	Off	196.76	24.59%	0.33	1/1	0	97/265 (29.48%/31.59%)	5.43
Observed	2.5	On	192.95	22.17%	0.14	3/7	0	0/0 (0.00%/0.00%)	3.87
Observed	3	Off	172.83	9.43%	0.65	78/86	6	71/217 (21.58%/25.86%)	7.78
Observed	3	On	169.64	7.42%	0.65	79/89	6	0/0 (0.00%/0.00%)	4.97
Observed	Observed	Off	199.34	26.22%	0.84	2/2	0	81/232 (24.62%/27.65%)	5.40
Observed	Observed	On	195.65	23.89%	1.09	4/8	1	0/0 (0.00%/0.00%)	2.80
$\leq 110\%$	None	Off	218.06	38.07%	1.40	0/0	1	71/199 (21.58%/23.72%)	0.78
$\leq 110\%$	None	On	212.96	34.84%	1.42	2/9	2	0/0 (0.00%/0.00%)	0.55
$\leq 110\%$	2	Off	212.39	34.48%	1.16	0/0	2	75/226 (22.80%/26.94%)	1.09
$\leq 110\%$	2	On	207.72	31.53%	0.95	2/9	3	0/0 (0.00%/0.00%)	0.85
$\leq 110\%$	2.5	Off	202.75	28.38%	0.38	0/0	1	87/222 (26.44%/26.46%)	5.19
$\leq 110\%$	2.5	On	198.84	25.90%	0.66	3/7	3	0/0 (0.00%/0.00%)	3.23
$\leq 110\%$	3	Off	177.51	12.40%	0.00	78/86	5	65/191 (19.76%/22.77%)	5.66
$\leq 110\%$	3	On	174.27	10.34%	0.90	79/89	6	0/0 (0.00%/0.00%)	3.58
$\leq 110\%$	Observed	Off	204.27	29.34%	0.83	0/0	3	81/207 (24.62%/24.67%)	6.35
$\leq 110\%$	Observed	On	200.49	26.95%	1.07	3/7	4	0/0 (0.00%/0.00%)	8.03
$[90\%, 110\%]$	None	Off	218.06	38.07%	1.36	0/0	0	68/189 (20.67%/22.53%)	1.23
$[90\%, 110\%]$	None	On	212.91	34.82%	1.14	1/2	0	0/0 (0.00%/0.00%)	1.01
$[90\%, 110\%]$	2	Off	212.39	34.48%	0.95	0/0	0	72/194 (21.88%/23.12%)	1.40
$[90\%, 110\%]$	2	On	207.58	31.44%	1.05	2/6	0	0/0 (0.00%/0.00%)	1.67
$[90\%, 110\%]$	2.5	Off	202.75	28.38%	0.32	0/0	0	79/198 (24.01%/23.60%)	6.57
$[90\%, 110\%]$	2.5	On	198.81	25.89%	0.61	2/3	0	0/0 (0.00%/0.00%)	5.95
$[90\%, 110\%]$	3	Off	Infeasible instance						
$[90\%, 110\%]$	3	On	Infeasible instance						
$[90\%, 110\%]$	Observed	Off	204.26	29.34%	0.86	5/5	0	74/200 (22.49%/23.84%)	24.71
$[90\%, 110\%]$	Observed	On	200.36	26.86%	1.10	6/7	0	0/0 (0.00%/0.00%)	12.60

Table 2: Results of counterfactual employment optimization under various scenarios using the SUM objective and LASSO model.

	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.

	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 at least 200 refugees before 2017.

Table 4: Estimated coefficients in the LASSO model.

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	Build and Run Time (s)
Observed	None	Off	175.62	22.85%	1.47	0/0	0	59/198 (17.93%/23.60%)	0.45
Observed	None	On	174.08	21.77%	1.60	2/3	0	0/0 (0.00%/0.00%)	0.41
Observed	2	Off	167.59	17.23%	0.77	1/1	0	75/195 (22.80%/23.24%)	0.94
Observed	2	On	166.21	16.26%	0.83	1/2	0	0/0 (0.00%/0.00%)	1.03
Observed	2.5	Off	159.32	11.45%	0.33	2/2	0	81/185 (24.62%/22.05%)	6.97
Observed	2.5	On	157.93	10.47%	0.14	3/4	0	0/0 (0.00%/0.00%)	7.03
Observed	3	Off	142.18	-0.55%	0.00	80/92	6	53/137 (16.11%/16.33%)	10.57
Observed	3	On	141.23	-1.21%	1.07	81/95	9	0/0 (0.00%/0.00%)	9.84
Observed	Observed	Off	161.68	13.09%	0.85	2/2	0	70/163 (21.28%/19.43%)	7.34
Observed	Observed	On	160.33	12.15%	0.84	2/3	0	0/0 (0.00%/0.00%)	6.81
$\leq 110\%$	None	Off	178.71	25.01%	1.76	0/0	3	57/179 (17.33%/21.33%)	0.85
$\leq 110\%$	None	On	177.12	23.90%	1.34	1/2	1	0/0 (0.00%/0.00%)	0.73
$\leq 110\%$	2	Off	171.96	20.28%	1.61	0/0	4	57/155 (17.33%/18.47%)	1.41
$\leq 110\%$	2	On	170.57	19.31%	1.04	1/2	1	0/0 (0.00%/0.00%)	1.25
$\leq 110\%$	2.5	Off	163.33	14.25%	0.12	0/0	1	84/197 (25.53%/23.48%)	6.25
$\leq 110\%$	2.5	On	161.93	13.27%	0.36	2/3	2	0/0 (0.00%/0.00%)	6.32
$\leq 110\%$	3	Off	145.20	1.57%	0.65	80/92	5	66/177 (20.06%/21.10%)	8.14
$\leq 110\%$	3	On	144.19	0.86%	1.07	81/95	6	0/0 (0.00%/0.00%)	6.53
$\leq 110\%$	Observed	Off	164.79	15.27%	0.99	0/0	3	80/191 (24.32%/22.77%)	11.41
$\leq 110\%$	Observed	On	163.44	14.33%	1.03	2/3	4	0/0 (0.00%/0.00%)	7.01
$[90\%, 110\%]$	None	Off	178.71	25.01%	1.27	0/0	0	53/154 (16.11%/18.36%)	0.81
$[90\%, 110\%]$	None	On	177.12	23.90%	1.22	1/2	0	0/0 (0.00%/0.00%)	1.14
$[90\%, 110\%]$	2	Off	171.96	20.28%	1.12	0/0	0	62/183 (18.84%/21.81%)	1.71
$[90\%, 110\%]$	2	On	170.57	19.31%	0.96	1/2	0	0/0 (0.00%/0.00%)	1.19
$[90\%, 110\%]$	2.5	Off	163.33	14.25%	0.32	0/0	0	72/189 (21.88%/22.53%)	4.09
$[90\%, 110\%]$	2.5	On	161.93	13.27%	0.33	2/3	0	0/0 (0.00%/0.00%)	8.83
$[90\%, 110\%]$	3	Off			Infeasible instance				
$[90\%, 110\%]$	3	On			Infeasible instance				
$[90\%, 110\%]$	Observed	Off	164.79	15.27%	0.83	4/4	0	70/181 (21.28%/21.57%)	19.75
$[90\%, 110\%]$	Observed	On	163.44	14.32%	0.86	5/6	0	0/0 (0.00%/0.00%)	13.59

Table 5: Results of counterfactual employment optimization under various scenarios using the SUM objective and GBRT model.

Capacity Adjustment	Min Avg Case Size	Binary Service Constraints	Total Expected Employed Refugees	Gains wrt to Predicted Employed Refugees (137.61)	St Dev in Avg Case Size Across Affiliates	# of Unplaced Cases / Refugees	#of Affiliates Violating 90% Capacity	# and % of Cases/Refugees Violating Constraints	Build and Run Time (s)
Observed	None	Off	184.68	34.21%	1.33	0/0	0	55/160 (16.72%/19.07%)	0.48
Observed	None	On	180.98	31.52%	1.28	3/10	0	0/0 (0.00%/0.00%)	0.36
Observed	2	Off	173.98	26.43%	0.81	1/1	0	76/205 (23.10%/24.43%)	0.92
Observed	2	On	171.22	24.43%	1.16	3/10	1	0/0 (0.00%/0.00%)	0.93
Observed	2.5	Off	163.18	18.58%	0.36	4/4	0	90/222 (27.36%/26.46%)	5.43
Observed	2.5	On	160.98	16.98%	0.20	3/7	0	0/0 (0.00%/0.00%)	3.87
Observed	3	Off	138.74	0.82%	0.65	79/89	8	54/153 (16.41%/18.24%)	7.78
Observed	3	On	137.27	-0.25%	0.00	80/92	6	0/0 (0.00%/0.00%)	4.97
Observed	Observed	Off	166.41	20.93%	0.84	2/2	0	83/222 (25.23%/26.46%)	5.40
Observed	Observed	On	164.30	19.39%	1.08	5/6	1	0/0 (0.00%/0.00%)	2.80
$\leq 110\%$	None	Off	188.36	36.88%	1.29	0/0	2	68/214 (20.67%/25.51%)	0.78
$\leq 110\%$	None	On	184.55	34.11%	1.31	2/9	3	0/0 (0.00%/0.00%)	0.55
$\leq 110\%$	2	Off	179.45	30.41%	1.11	0/0	3	70/188 (21.28%/22.41%)	1.09
$\leq 110\%$	2	On	176.23	28.06%	1.11	2/9	4	0/0 (0.00%/0.00%)	0.85
$\leq 110\%$	2.5	Off	168.34	22.33%	0.34	0/0	3	85/199 (25.84%/23.72%)	5.19
$\leq 110\%$	2.5	On	166.02	20.64%	0.59	2/3	6	0/0 (0.00%/0.00%)	3.23
$\leq 110\%$	3	Off	142.64	3.66%	0.00	79/89	5	66/205 (20.06%/24.43%)	5.66
$\leq 110\%$	3	On	140.99	2.45%	0.65	80/92	5	0/0 (0.00%/0.00%)	3.58
$\leq 110\%$	Observed	Off	170.47	23.88%	1.05	0/0	4	83/210 (25.23%/25.03%)	6.35
$\leq 110\%$	Observed	On	168.28	22.28%	1.09	2/3	4	0/0 (0.00%/0.00%)	8.03
$[90\%, 110\%]$	None	Off	188.36	36.88%	1.32	0/0	0	60/176 (18.24%/20.98%)	1.23
$[90\%, 110\%]$	None	On	184.49	34.07%	1.28	1/2	0	0/0 (0.00%/0.00%)	1.01
$[90\%, 110\%]$	2	Off	179.45	30.41%	0.97	0/0	0	73/183 (22.19%/21.81%)	1.40
$[90\%, 110\%]$	2	On	176.14	28.00%	1.05	1/2	0	0/0 (0.00%/0.00%)	1.67
$[90\%, 110\%]$	2.5	Off	168.34	22.33%	0.34	0/0	0	78/198 (23.71%/23.60%)	6.57
$[90\%, 110\%]$	2.5	On	166.02	20.64%	0.33	2/3	0	0/0 (0.00%/0.00%)	5.95
$[90\%, 110\%]$	3	Off			Infeasible instance				
$[90\%, 110\%]$	3	On			Infeasible instance				
$[90\%, 110\%]$	Observed	Off	170.47	23.88%	0.89	5/5	0	73/204 (22.19%/24.31%)	24.71
$[90\%, 110\%]$	Observed	On	168.24	22.26%	1.10	6/7	0	0/0 (0.00%/0.00%)	12.60

Table 6: Results of counterfactual employment optimization under various scenarios using the MAX objective and LASSO model.

Capacity Adjustment	Min Avg Case Size	Binary Service Constraints	Total Expected Employed Refugees	Gains wrt to Predicted Employed Refugees (97.57)	St Dev in Avg Case Size Across Affiliates	# of Unplaced Cases / Refugees	#of Affiliates Violating 90% Capacity	# and % of Cases/Refugees Violating Constraints	Build and Run Time (s)
Observed	None	Off	120.60	23.60%	1.23	0/0	0	58/160 (17.63%/19.07%)	0.46
Observed	None	On	118.17	21.12%	1.22	1/2	0	0/0 (0.00%/0.00%)	0.43
Observed	2	Off	109.42	12.15%	0.86	1/5	0	75/210 (22.80%/25.03%)	10.75
Observed	2	On	107.38	10.06%	0.78	4/11	1	0/0 (0.00%/0.00%)	3.27
Observed	2.5	Off	98.27	0.72%	0.33	1/5	0	84/242 (25.53%/28.84%)	30.56
Observed	2.5	On	97.12	-0.46%	0.19	7/19	1	0/0 (0.00%/0.00%)	11.66
Observed	3	Off	76.14	-21.96%	0.00	88/116	10	55/176 (16.72%/20.98%)	14.07
Observed	3	On	75.38	-22.75%	0.65	88/116	7	0/0 (0.00%/0.00%)	9.00
Observed	Observed	Off	101.60	4.13%	0.85	2/3	0	83/238 (25.23%/28.37%)	19.68
Observed	Observed	On	100.45	2.95%	0.84	7/19	1	0/0 (0.00%/0.00%)	8.46
$\leq 110\%$	None	Off	120.99	24.00%	1.32	0/0	1	66/189 (20.06%/22.53%)	0.33
$\leq 110\%$	None	On	118.57	21.52%	1.22	1/2	3	0/0 (0.00%/0.00%)	0.34
$\leq 110\%$	2	Off	111.57	14.35%	0.82	0/0	1	71/206 (21.58%/24.55%)	5.75
$\leq 110\%$	2	On	109.52	12.25%	0.86	1/2	4	0/0 (0.00%/0.00%)	1.09
$\leq 110\%$	2.5	Off	101.15	3.67%	0.32	0/0	1	83/223 (25.23%/26.58%)	14.18
$\leq 110\%$	2.5	On	99.84	2.33%	0.29	6/15	3	0/0 (0.00%/0.00%)	6.83
$\leq 110\%$	3	Off	77.61	-20.46%	0.00	89/119	4	58/178 (17.63%/21.22%)	6.95
$\leq 110\%$	3	On	76.98	-21.11%	0.65	89/119	5	0/0 (0.00%/0.00%)	5.45
$\leq 110\%$	Observed	Off	103.67	6.26%	0.90	0/0	2	78/223 (23.71%/26.58%)	19.99
$\leq 110\%$	Observed	On	102.52	5.07%	0.91	7/19	3	0/0 (0.00%/0.00%)	8.32
$[90\%, 110\%]$	None	Off	120.99	24.00%	1.15	0/0	0	68/195 (20.67%/23.24%)	0.44
$[90\%, 110\%]$	None	On	118.57	21.52%	1.16	1/2	0	0/0 (0.00%/0.00%)	0.54
$[90\%, 110\%]$	2	Off	111.57	14.35%	0.79	0/0	0	73/215 (22.19%/25.63%)	8.31
$[90\%, 110\%]$	2	On	109.52	12.25%	0.89	1/2	0	0/0 (0.00%/0.00%)	1.36
$[90\%, 110\%]$	2.5	Off	101.15	3.67%	0.32	0/0	0	93/247 (28.27%/29.44%)	14.36
$[90\%, 110\%]$	2.5	On	99.70	2.18%	0.33	3/5	0	0/0 (0.00%/0.00%)	40.22
$[90\%, 110\%]$	3	Off			Infeasible instance				
$[90\%, 110\%]$	3	On			Infeasible instance				
$[90\%, 110\%]$	Observed	Off	103.67	6.26%	0.83	1/1	0	84/229 (25.53%/27.29%)	62.74
$[90\%, 110\%]$	Observed	On	102.44	5.00%	0.86	5/9	0	0/0 (0.00%/0.00%)	37.80

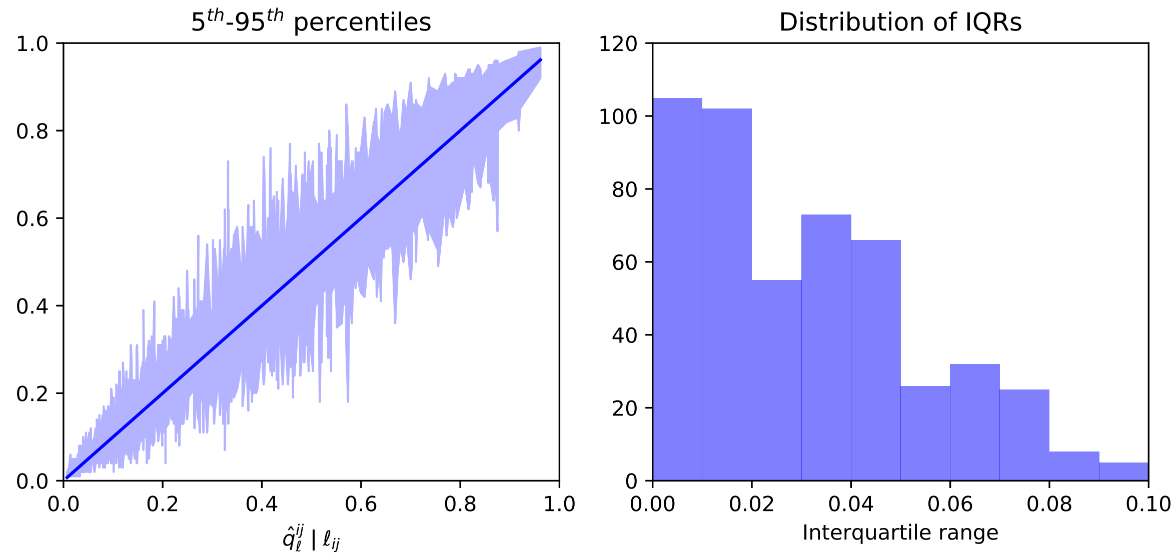
Table 7: Results of counterfactual employment optimization under various scenarios using the MIN objective and LASSO model.

Capacity Adjustment	Min Avg Case Size	Binary Service Constraints	Total Expected Employed Refugees	Gains wrt to Predicted Employed Refugees (117.49)	St Dev in Avg Case Size Across Affiliates	# of Unplaced Cases / Refugees	#of Affiliates Violating 90% Capacity	# and % of Cases/Refugees Violating Constraints	Build and Run Time (s)
Observed	None	Off	163.81	39.42%	1.39	0/0	0	65/200 (19.76%/23.84%)	1.62
Observed	None	On	160.67	36.75%	1.30	2/9	0	0/0 (0.00%/0.00%)	0.49
Observed	2	Off	151.73	29.14%	0.79	3/3	0	77/219 (23.40%/26.10%)	2.43
Observed	2	On	149.16	26.96%	1.06	3/10	1	0/0 (0.00%/0.00%)	1.46
Observed	2.5	Off	140.19	19.32%	0.45	4/4	0	91/245 (27.66%/29.20%)	12.60
Observed	2.5	On	138.07	17.51%	0.46	5/6	0	0/0 (0.00%/0.00%)	8.57
Observed	3	Off	115.36	-1.81%	0.00	80/92	7	60/186 (18.24%/22.17%)	5.92
Observed	3	On	113.87	-3.08%	0.65	81/95	5	0/0 (0.00%/0.00%)	3.66
Observed	Observed	Off	143.51	22.15%	0.85	2/2	0	79/220 (24.01%/26.22%)	7.52
Observed	Observed	On	141.50	20.44%	0.84	4/10	2	0/0 (0.00%/0.00%)	4.46
$\leq 110\%$	None	Off	166.39	41.62%	1.40	0/0	1	63/191 (19.15%/22.77%)	0.66
$\leq 110\%$	None	On	163.17	38.88%	1.47	2/9	2	0/0 (0.00%/0.00%)	1.16
$\leq 110\%$	2	Off	156.33	33.06%	1.48	0/0	4	72/210 (21.88%/25.03%)	1.42
$\leq 110\%$	2	On	153.41	30.57%	1.40	2/9	2	0/0 (0.00%/0.00%)	1.38
$\leq 110\%$	2.5	Off	144.58	23.06%	0.32	0/0	1	88/228 (26.75%/27.18%)	13.24
$\leq 110\%$	2.5	On	142.31	21.12%	0.34	2/3	2	0/0 (0.00%/0.00%)	5.34
$\leq 110\%$	3	Off	118.41	0.79%	0.00	80/92	4	64/193 (19.45%/23.00%)	7.34
$\leq 110\%$	3	On	116.85	-0.55%	0.90	81/95	9	0/0 (0.00%/0.00%)	6.30
$\leq 110\%$	Observed	Off	146.87	25.00%	0.90	0/0	4	69/198 (20.97%/23.60%)	10.13
$\leq 110\%$	Observed	On	144.83	23.27%	0.99	2/3	5	0/0 (0.00%/0.00%)	7.52
$[90\%, 110\%]$	None	Off	166.39	41.62%	1.38	0/0	0	57/162 (17.33%/19.31%)	1.32
$[90\%, 110\%]$	None	On	163.16	38.87%	1.26	1/2	0	0/0 (0.00%/0.00%)	1.42
$[90\%, 110\%]$	2	Off	156.33	33.06%	0.94	0/0	0	79/215 (24.01%/25.63%)	1.53
$[90\%, 110\%]$	2	On	153.39	30.56%	0.99	1/2	0	0/0 (0.00%/0.00%)	1.76
$[90\%, 110\%]$	2.5	Off	144.58	23.06%	0.34	0/0	0	88/242 (26.75%/28.84%)	14.74
$[90\%, 110\%]$	2.5	On	142.31	21.12%	0.33	2/3	0	0/0 (0.00%/0.00%)	5.20
$[90\%, 110\%]$	3	Off			Infeasible instance				
$[90\%, 110\%]$	3	On			Infeasible instance				
$[90\%, 110\%]$	Observed	Off	146.87	25.00%	0.83	1/1	0	88/240 (26.75%/28.61%)	8.88
$[90\%, 110\%]$	Observed	On	144.76	23.21%	0.86	2/3	0	0/0 (0.00%/0.00%)	12.13

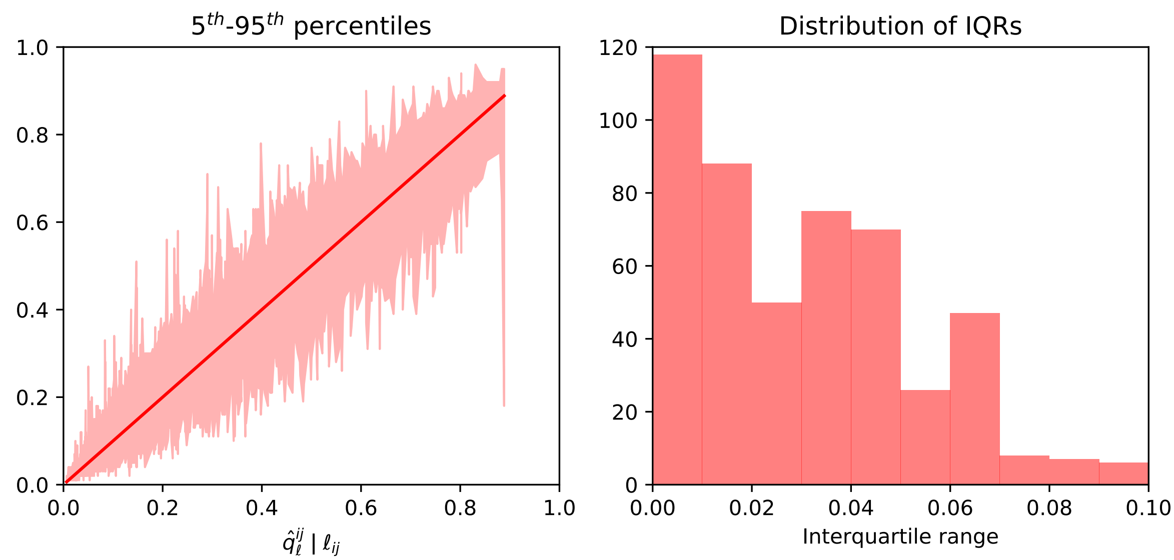
Table 8: Results of counterfactual employment optimization under various scenarios using the MEAN objective and LASSO model.

Capacity Adjustment	Min Avg Case Size	Binary Service Constraints	Relative to SUM Objective				Relative to MAX Objective				Relative to MIN Objective				Relative to MEAN Objective			
			SUM	MAX (% loss)	MIN (% loss)	MEAN (% loss)	MAX	SUM (% loss)	MIN (% loss)	MEAN (% loss)	MIN	SUM (% loss)	MAX (% loss)	MEAN (% loss)	MEAN	SUM (% loss)	MAX (% loss)	MIN (% loss)
Observed	None	Off	**213.02**	207.64 (2.53%)	193.82 (9.02%)	209.80 (1.51%)	**184.68**	180.63 (2.19%)	169.90 (8.01%)	183.10 (0.86%)	**120.60**	116.14 (3.70%)	117.85 (2.28%)	118.91 (1.40%)	**163.81**	161.00 (1.71%)	162.47 (0.82%)	152.73 (6.76%)
Observed	None	On	**208.25**	203.85 (2.11%)	192.89 (7.37%)	205.40 (1.37%)	**180.98**	177.16 (2.11%)	168.67 (6.80%)	179.03 (1.08%)	**118.17**	114.53 (3.08%)	115.40 (2.35%)	116.63 (1.30%)	**160.67**	157.93 (1.70%)	159.18 (0.92%)	151.23 (5.88%)
Observed	2	Off	**206.28**	201.65 (2.24%)	181.75 (11.89%)	201.73 (2.20%)	**173.98**	170.38 (2.07%)	158.94 (8.64%)	172.04 (1.11%)	**109.42**	102.17 (6.63%)	101.14 (7.57%)	102.86 (6.00%)	**151.73**	149.70 (1.34%)	150.18 (1.02%)	141.15 (6.97%)
Observed	2	On	**202.03**	197.48 (2.25%)	181.13 (10.34%)	197.70 (2.14%)	**171.22**	168.16 (1.79%)	158.25 (7.58%)	169.35 (1.10%)	**107.38**	100.48 (6.43%)	99.93 (6.94%)	100.98 (5.96%)	**149.16**	147.60 (1.05%)	147.79 (0.92%)	139.62 (6.40%)
Observed	2.5	Off	**196.76**	191.02 (2.92%)	173.53 (11.81%)	192.09 (2.37%)	**163.18**	159.43 (2.30%)	150.87 (7.54%)	161.25 (1.18%)	**98.27**	88.12 (10.32%)	88.63 (9.81%)	90.22 (8.19%)	**140.19**	137.47 (1.94%)	138.05 (1.53%)	131.79 (5.99%)
Observed	2.5	On	**192.95**	187.22 (2.97%)	170.86 (11.45%)	187.03 (3.07%)	**160.98**	157.47 (2.18%)	147.91 (8.12%)	158.99 (1.24%)	**97.12**	87.53 (9.87%)	87.94 (9.45%)	89.89 (7.44%)	**138.07**	135.70 (1.71%)	136.14 (1.39%)	129.42 (6.27%)
Observed	3	Off	**172.83**	167.26 (3.22%)	147.03 (14.93%)	167.45 (3.11%)	**138.74**	134.78 (2.85%)	124.41 (10.33%)	136.52 (1.60%)	**76.14**	63.52 (16.57%)	63.28 (16.90%)	66.02 (13.29%)	**115.36**	112.37 (2.59%)	113.10 (1.96%)	106.07 (8.05%)
Observed	3	On	**169.64**	163.72 (3.49%)	146.33 (13.74%)	164.45 (3.06%)	**137.27**	134.28 (2.18%)	124.00 (9.66%)	134.58 (1.96%)	**75.38**	63.19 (16.17%)	63.67 (15.54%)	65.53 (13.07%)	**113.87**	111.49 (2.09%)	111.85 (1.78%)	105.28 (7.55%)
Observed	Observed	Off	**199.33**	193.90 (2.73%)	177.11 (11.15%)	193.66 (2.85%)	**166.41**	162.79 (2.18%)	153.71 (7.63%)	164.50 (1.15%)	**101.60**	92.63 (8.82%)	92.41 (9.05%)	94.47 (7.02%)	**143.51**	141.11 (1.67%)	141.63 (1.31%)	134.62 (6.20%)
Observed	Observed	On	**195.65**	190.42 (2.67%)	175.11 (10.50%)	190.19 (2.79%)	**164.30**	161.05 (1.98%)	152.96 (6.90%)	162.38 (1.17%)	**100.45**	92.03 (8.38%)	91.64 (8.77%)	93.09 (7.33%)	**141.50**	139.48 (1.43%)	139.72 (1.26%)	133.56 (5.61%)
$\leq 110\%$	None	Off	**218.05**	212.38 (2.60%)	197.89 (9.25%)	215.37 (1.23%)	**188.36**	184.24 (2.18%)	172.78 (8.27%)	186.45 (1.01%)	**120.99**	117.57 (2.83%)	118.21 (2.30%)	119.21 (1.47%)	**166.39**	163.98 (1.45%)	165.04 (0.81%)	154.64 (7.06%)
$\leq 110\%$	None	On	**212.96**	208.12 (2.27%)	195.13 (8.37%)	210.51 (1.15%)	**184.55**	181.21 (1.81%)	170.71 (7.50%)	182.71 (1.00%)	**118.57**	115.34 (2.72%)	115.98 (2.18%)	117.02 (1.31%)	**163.17**	160.95 (1.36%)	161.73 (0.88%)	152.80 (6.35%)
$\leq 110\%$	2	Off	**212.39**	207.34 (2.38%)	185.02 (12.89%)	208.25 (1.95%)	**179.45**	175.71 (2.09%)	161.95 (9.76%)	177.61 (1.03%)	**111.57**	105.13 (5.77%)	103.96 (6.82%)	105.73 (5.23%)	**156.33**	154.47 (1.19%)	154.53 (1.15%)	143.69 (8.08%)
$\leq 110\%$	2	On	**207.72**	202.57 (2.48%)	184.66 (11.10%)	203.97 (1.80%)	**176.23**	173.18 (1.73%)	162.19 (7.96%)	174.35 (1.07%)	**109.52**	103.46 (5.53%)	101.99 (6.87%)	103.55 (5.45%)	**153.41**	151.82 (1.03%)	151.84 (1.02%)	142.83 (6.90%)
$\leq 110\%$	2.5	Off	**202.75**	196.42 (3.12%)	176.22 (13.08%)	198.17 (2.26%)	**168.34**	164.77 (2.12%)	153.79 (8.64%)	166.24 (1.25%)	**101.15**	91.11 (9.93%)	90.97 (10.07%)	93.67 (7.40%)	**144.58**	142.02 (1.77%)	142.28 (1.59%)	134.75 (6.80%)
$\leq 110\%$	2.5	On	**198.84**	192.86 (3.00%)	176.14 (11.42%)	194.09 (2.39%)	**166.02**	162.92 (1.87%)	152.72 (8.01%)	163.82 (1.32%)	**99.84**	90.21 (9.65%)	90.14 (9.71%)	92.37 (7.49%)	**142.30**	140.22 (1.47%)	140.41 (1.33%)	133.57 (6.14%)
$\leq 110\%$	3	Off	**177.51**	172.20 (2.99%)	149.22 (15.94%)	172.23 (2.97%)	**142.64**	139.04 (2.52%)	126.76 (11.13%)	140.06 (1.81%)	**77.60**	64.83 (16.46%)	64.74 (16.58%)	67.43 (13.11%)	**118.41**	115.34 (2.59%)	116.19 (1.88%)	108.34 (8.51%)
$\leq 110\%$	3	On	**174.27**	168.96 (3.05%)	147.58 (15.32%)	169.80 (2.56%)	**140.99**	137.76 (2.29%)	126.53 (10.25%)	138.35 (1.87%)	**76.98**	64.89 (15.70%)	65.06 (15.48%)	67.85 (11.86%)	**116.85**	114.24 (2.23%)	114.88 (1.68%)	107.65 (7.87%)
$\leq 110\%$	Observed	Off	**204.27**	198.15 (2.99%)	181.01 (11.39%)	198.62 (2.77%)	**170.47**	166.62 (2.26%)	157.26 (7.75%)	168.20 (1.33%)	**103.67**	94.51 (8.84%)	93.92 (9.41%)	96.70 (6.73%)	**146.87**	144.47 (1.63%)	144.66 (1.50%)	137.91 (6.10%)
$\leq 110\%$	Observed	On	**200.49**	194.91 (2.79%)	178.71 (10.86%)	196.01 (2.23%)	**168.28**	164.94 (1.98%)	154.97 (7.91%)	166.08 (1.31%)	**102.52**	93.84 (8.46%)	92.99 (9.29%)	95.17 (7.17%)	**144.83**	142.59 (1.55%)	143.07 (1.22%)	136.02 (6.08%)
$[90\%, 110\%]$	None	Off	**218.05**	212.38 (2.60%)	194.95 (10.60%)	215.37 (1.23%)	**188.36**	184.24 (2.18%)	171.00 (9.22%)	186.45 (1.01%)	**120.99**	117.57 (2.83%)	118.21 (2.30%)	119.21 (1.47%)	**166.39**	163.98 (1.45%)	165.04 (0.81%)	153.11 (7.98%)
$[90\%, 110\%]$	None	On	**212.91**	208.03 (2.29%)	193.31 (9.21%)	210.61 (1.08%)	**184.49**	181.12 (1.82%)	168.79 (8.51%)	182.53 (1.06%)	**118.57**	115.33 (2.73%)	115.98 (2.18%)	116.89 (1.41%)	**163.16**	160.92 (1.37%)	161.69 (0.90%)	151.82 (6.95%)
$[90\%, 110\%]$	2	Off	**212.39**	207.34 (2.38%)	187.11 (11.90%)	208.25 (1.95%)	**179.45**	175.71 (2.09%)	163.60 (8.83%)	177.61 (1.03%)	**111.57**	105.13 (5.77%)	103.96 (6.82%)	105.73 (5.23%)	**156.33**	154.47 (1.19%)	154.53 (1.15%)	144.63 (7.49%)
$[90\%, 110\%]$	2	On	**207.58**	202.61 (2.39%)	184.64 (11.05%)	203.90 (1.77%)	**176.14**	173.08 (1.74%)	161.48 (8.32%)	174.25 (1.07%)	**109.52**	103.66 (5.35%)	102.31 (6.58%)	103.60 (5.41%)	**153.39**	151.92 (0.96%)	152.12 (0.83%)	142.64 (7.01%)
$[90\%, 110\%]$	2.5	Off	**202.75**	196.42 (3.12%)	178.99 (11.72%)	198.17 (2.26%)	**168.34**	164.77 (2.12%)	155.58 (7.58%)	166.24 (1.25%)	**101.15**	91.11 (9.93%)	90.97 (10.07%)	93.67 (7.40%)	**144.58**	142.02 (1.77%)	142.28 (1.59%)	135.95 (5.97%)
$[90\%, 110\%]$	2.5	On	**198.81**	192.86 (2.99%)	176.69 (11.13%)	194.09 (2.38%)	**166.02**	162.75 (1.97%)	152.98 (7.85%)	163.82 (1.32%)	**99.70**	90.40 (9.32%)	90.14 (9.58%)	92.37 (7.35%)	**142.30**	140.33 (1.39%)	140.41 (1.33%)	133.65 (6.08%)
$[90\%, 110\%]$	3	Off	Infeasible instance															
$[90\%, 110\%]$	3	On	Infeasible instance															
$[90\%, 110\%]$	Observed	Off	**204.26**	198.15 (2.99%)	179.80 (11.98%)	198.62 (2.76%)	**170.47**	166.80 (2.16%)	156.74 (8.06%)	168.20 (1.33%)	**103.67**	94.77 (8.59%)	93.92 (9.41%)	96.70 (6.73%)	**146.87**	144.69 (1.48%)	144.66 (1.50%)	137.34 (6.49%)
$[90\%, 110\%]$	Observed	On	**200.36**	194.91 (2.72%)	179.09 (10.62%)	196.31 (2.02%)	**168.24**	165.07 (1.89%)	156.38 (7.05%)	166.10 (1.27%)	**102.44**	94.16 (8.09%)	92.99 (9.23%)	95.06 (7.21%)	**144.76**	142.75 (1.39%)	143.01 (1.21%)	136.39 (5.78%)

Table 9: Comparison of performance of alternative SUM, MAX, MIN, and MEAN objectives of interest. Analysis is with respect to the upper headings and four subcolumns. Under each of these headings, entries in **bold** are the maximized value according to that objective, and remaining columns report the performance of the optimal solution of each alternative objective, when evaluated according to the objective of the upper heading, along with associated percent loss.

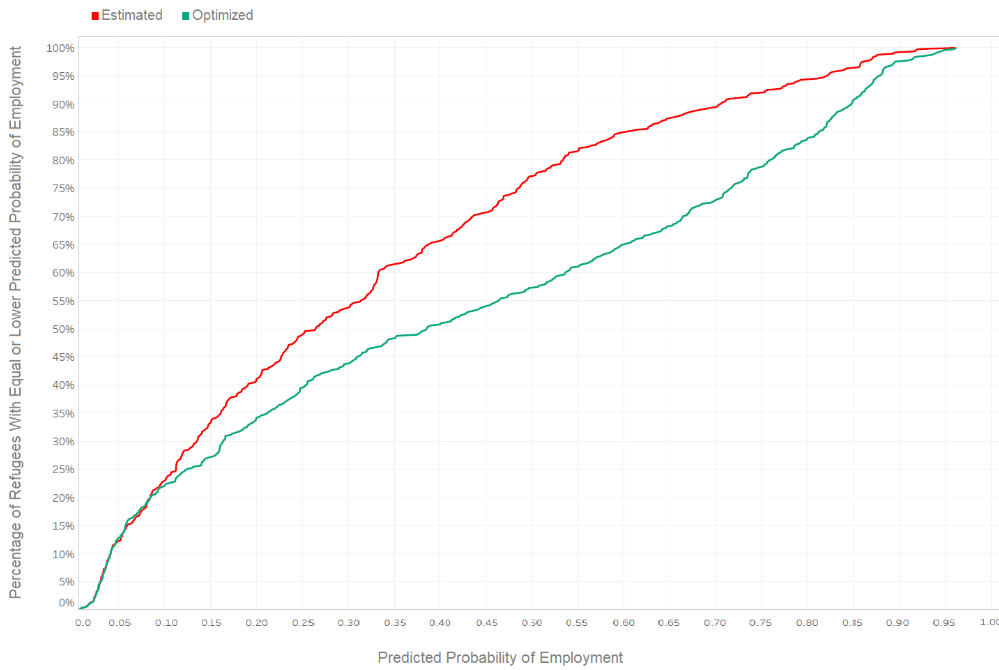


(a) LASSO

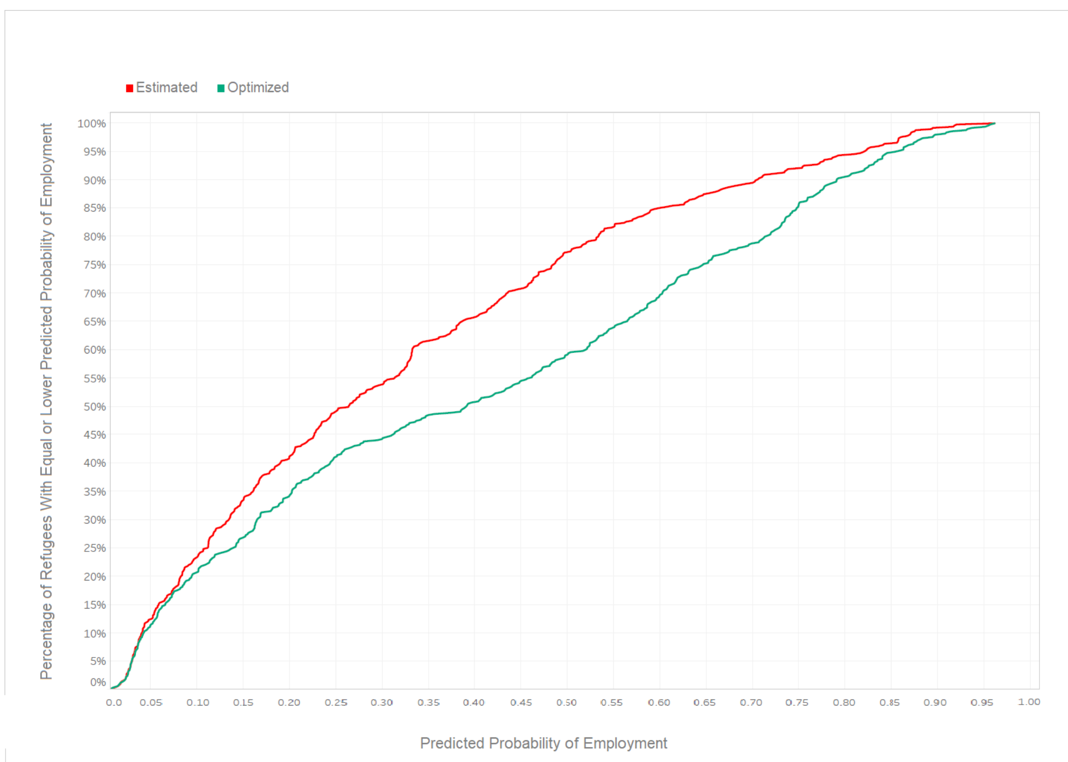


(b) Gradient Boosted Regression Tree (GBRT)

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.



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



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

Figure 2: Employment gains from optimizing refugee placement.

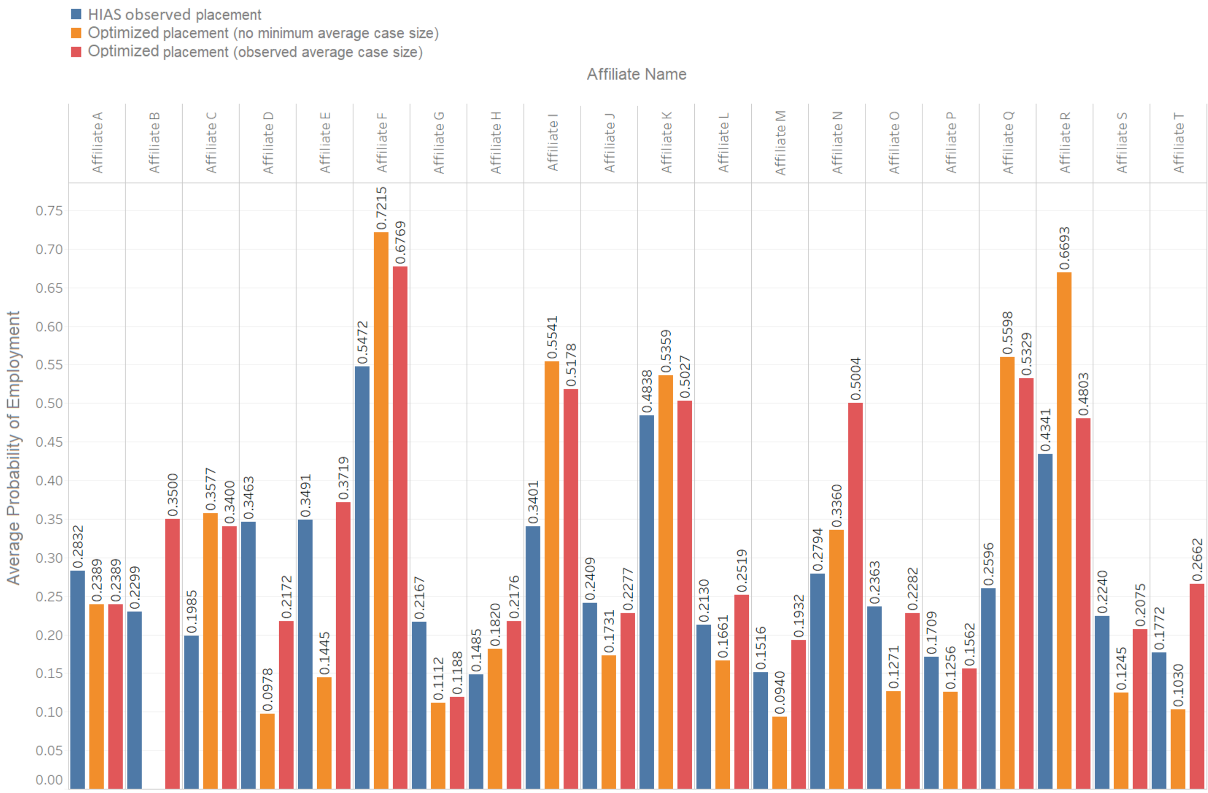


Figure 3: Average probability of employment at each affiliate. Blue bar: estimated probabilities under HIAS placement. Orange bar: average probability of employment for observed capacity, activated binary service constraints, no minimum average case size scenario. Red bar: average probability of employment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

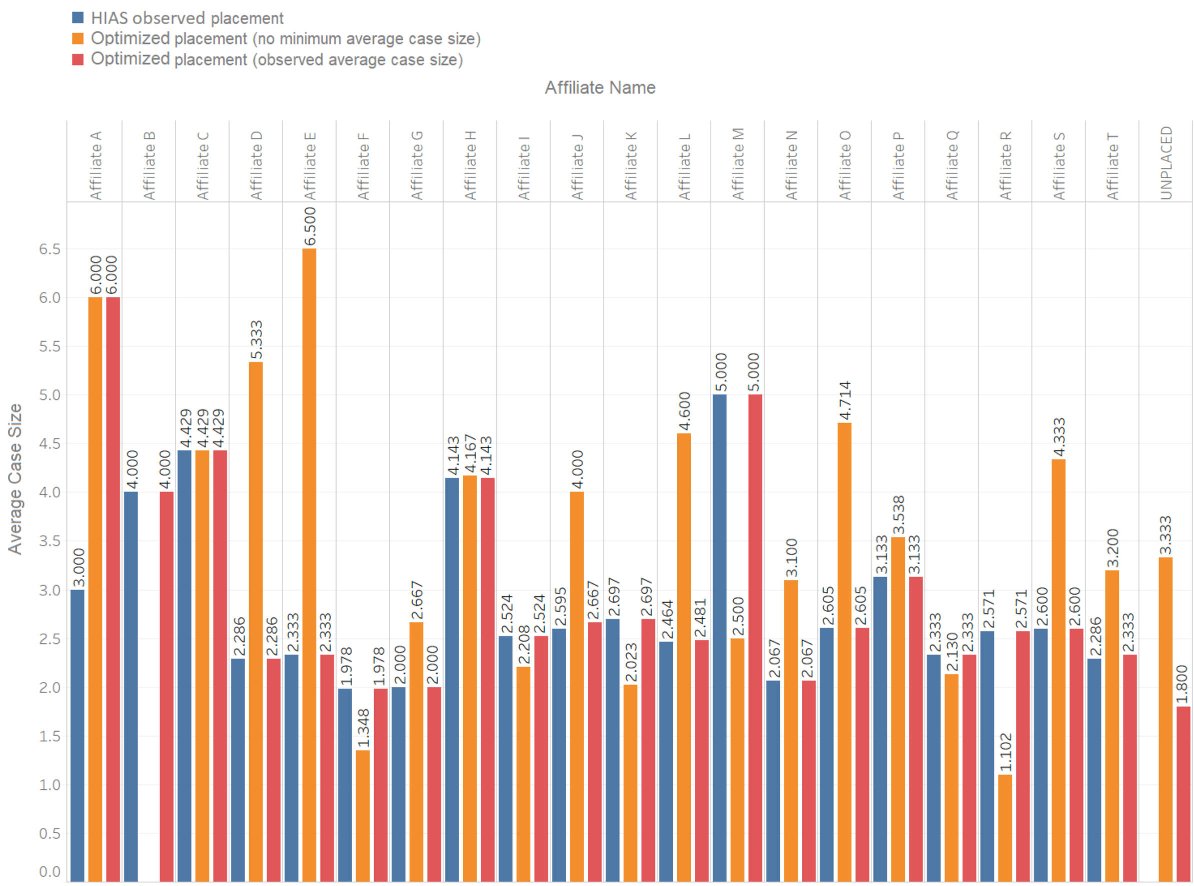
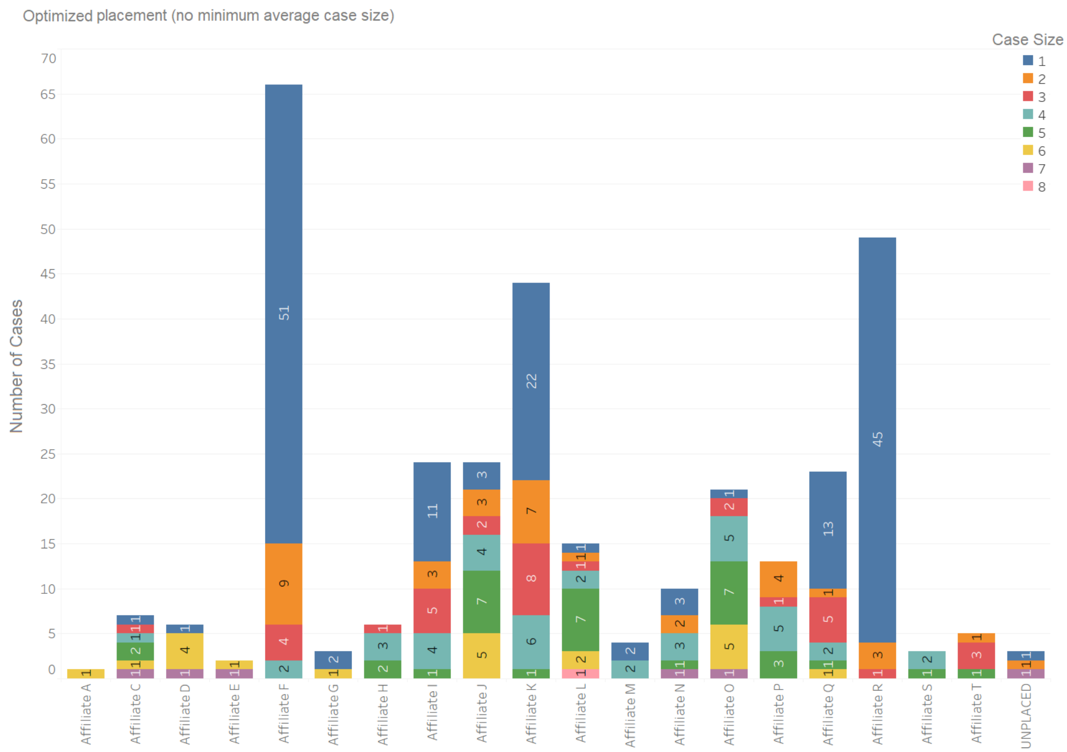
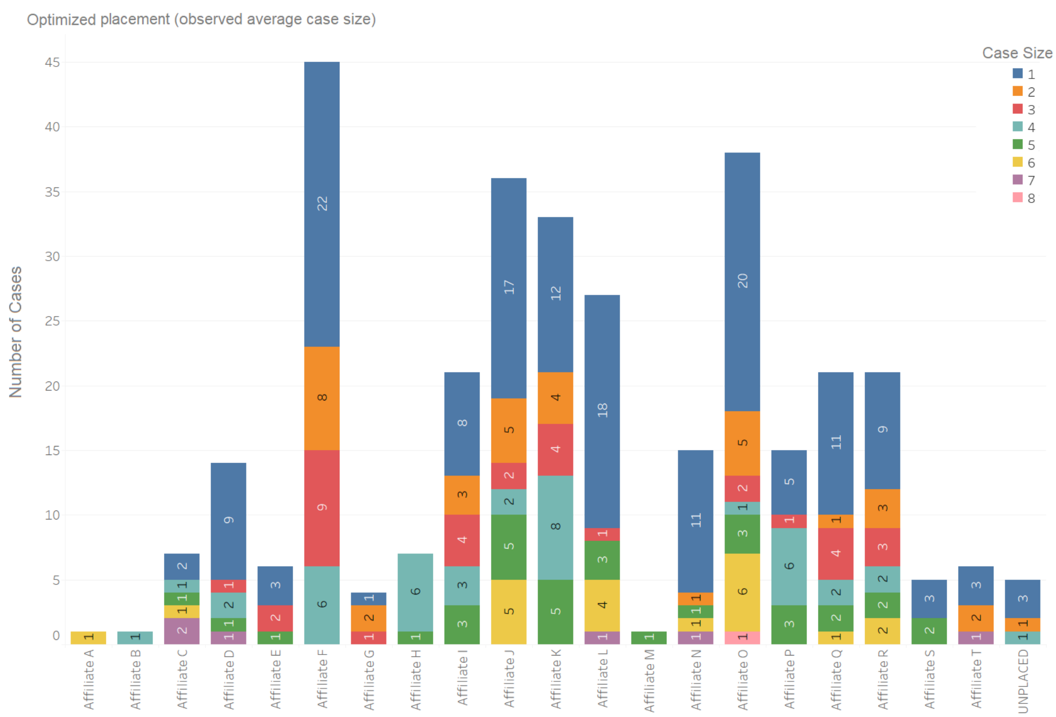


Figure 4: Average case size at each affiliate. Blue bar: observed average case size under HIAS placement. Orange bar: average case size for {observed capacity, activated binary service constraints, no minimum average case size} scenario. Red bar: average case size for {observed capacity, activated binary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



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



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

Figure 5: Distribution of case sizes at each affiliate.

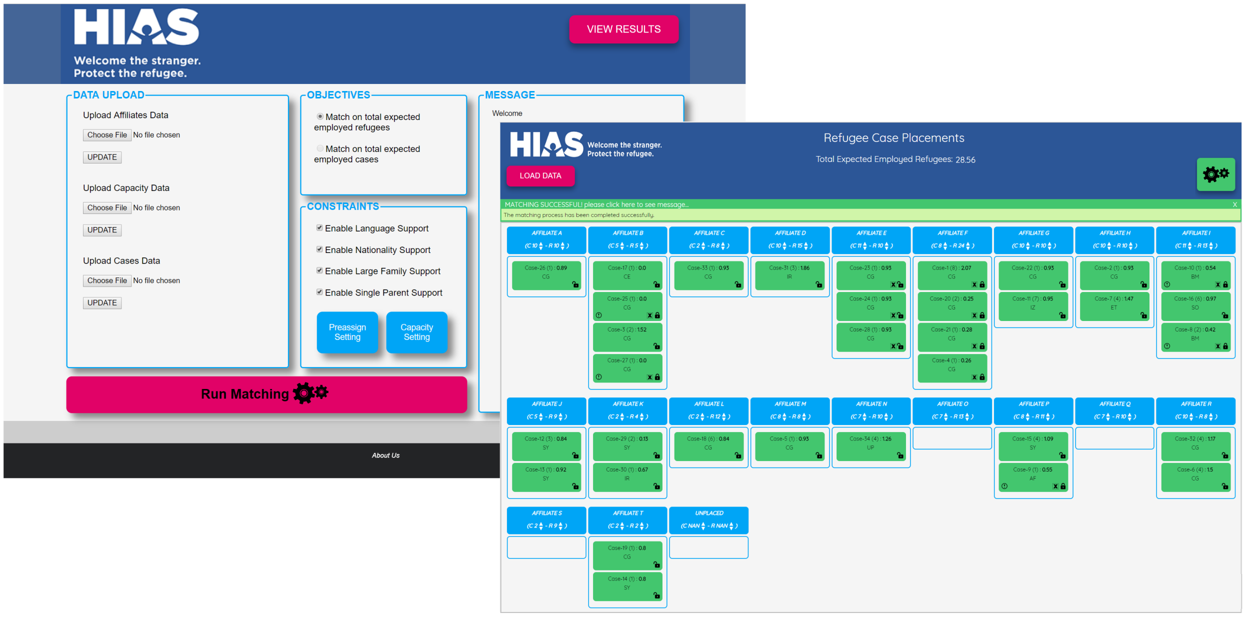


Figure 6: *Annie* Interface.

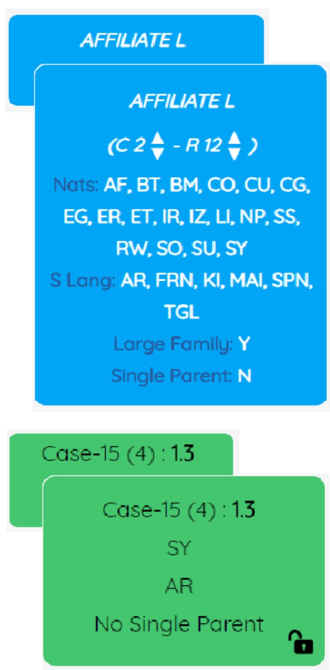
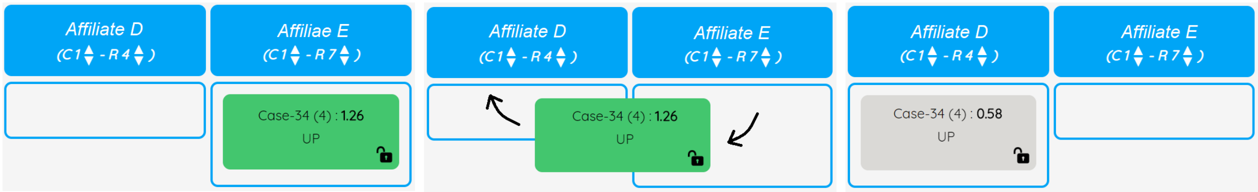


Figure 7: Expanding tiles: refugee and affiliate data.



(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.

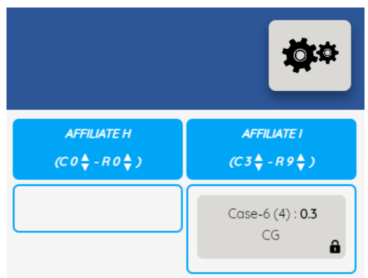


Figure 9: Locking case tiles and reoptimizing.

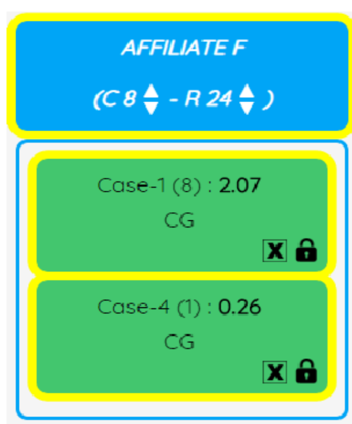


Figure 10: Cross-referencing cases to Affiliate F.

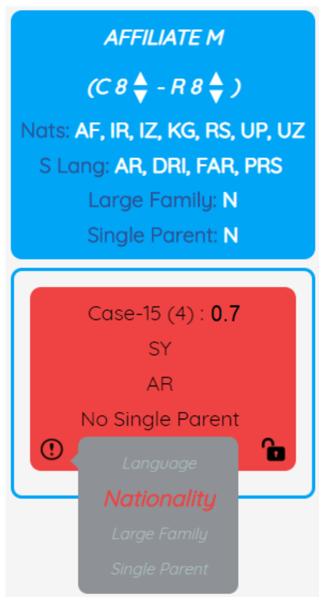
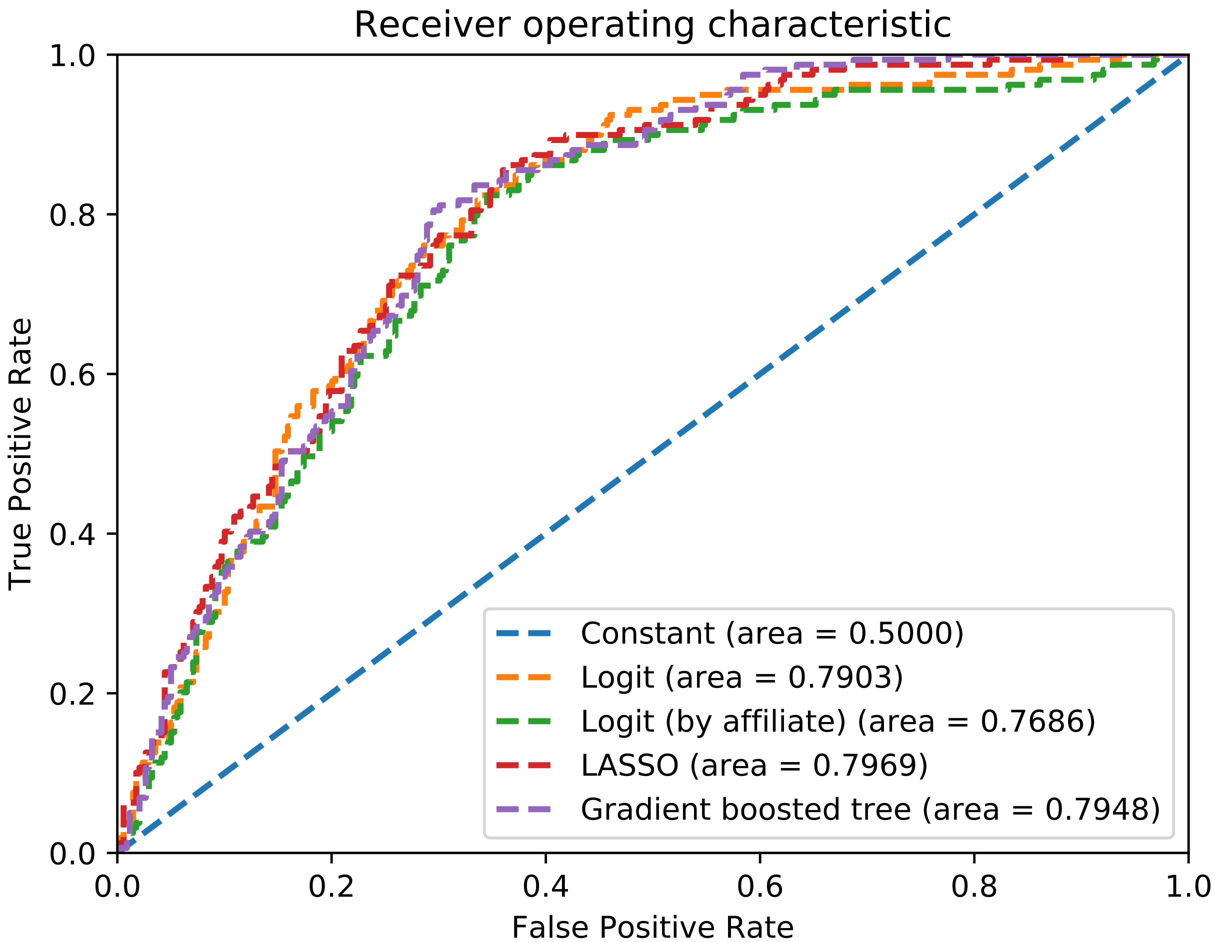
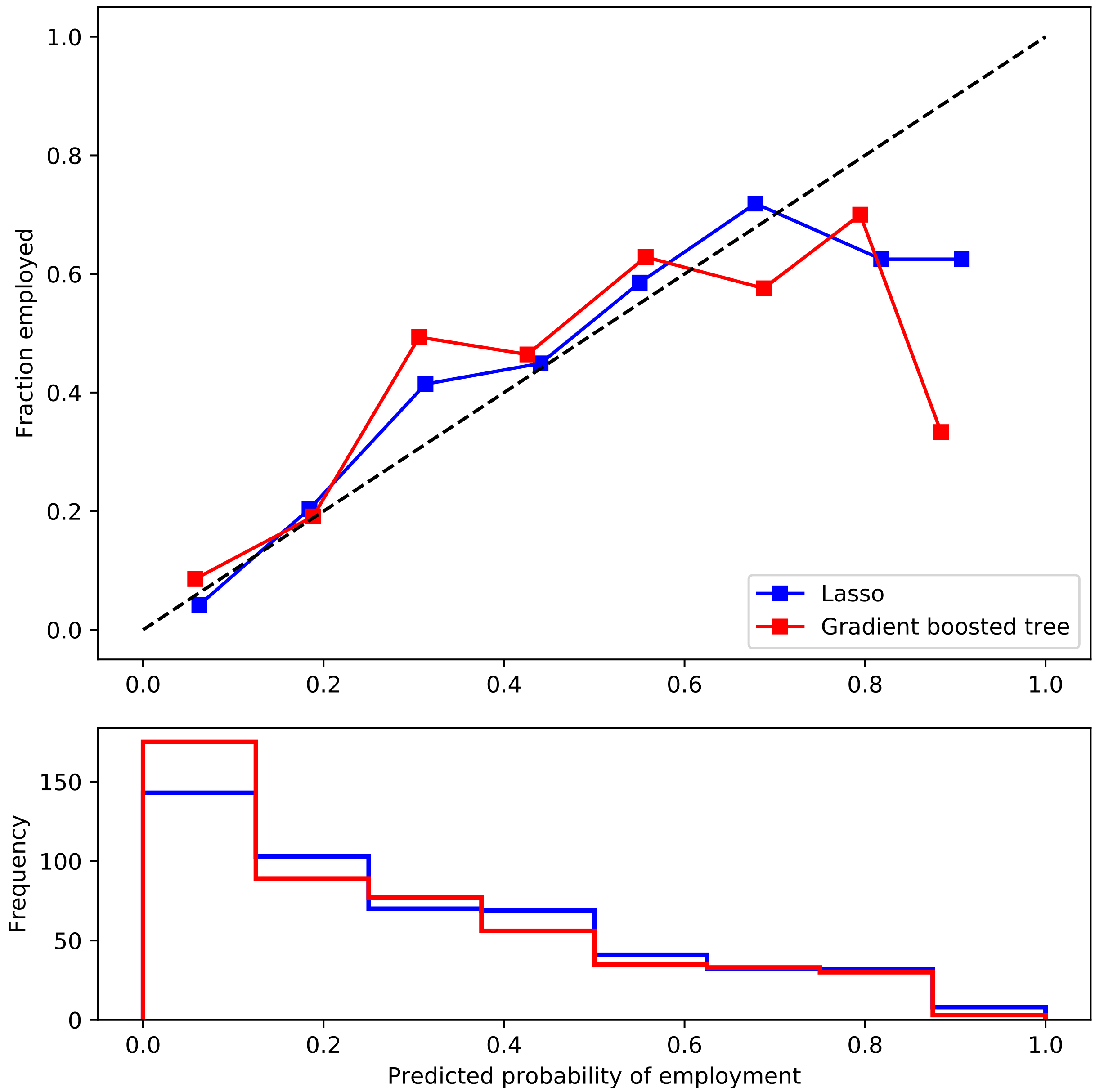


Figure 11: Case tile changes color when placed into affiliate that violates binary service constraints. Hovering over exclamation point reveals additional details.



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.



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 (FY17 data).

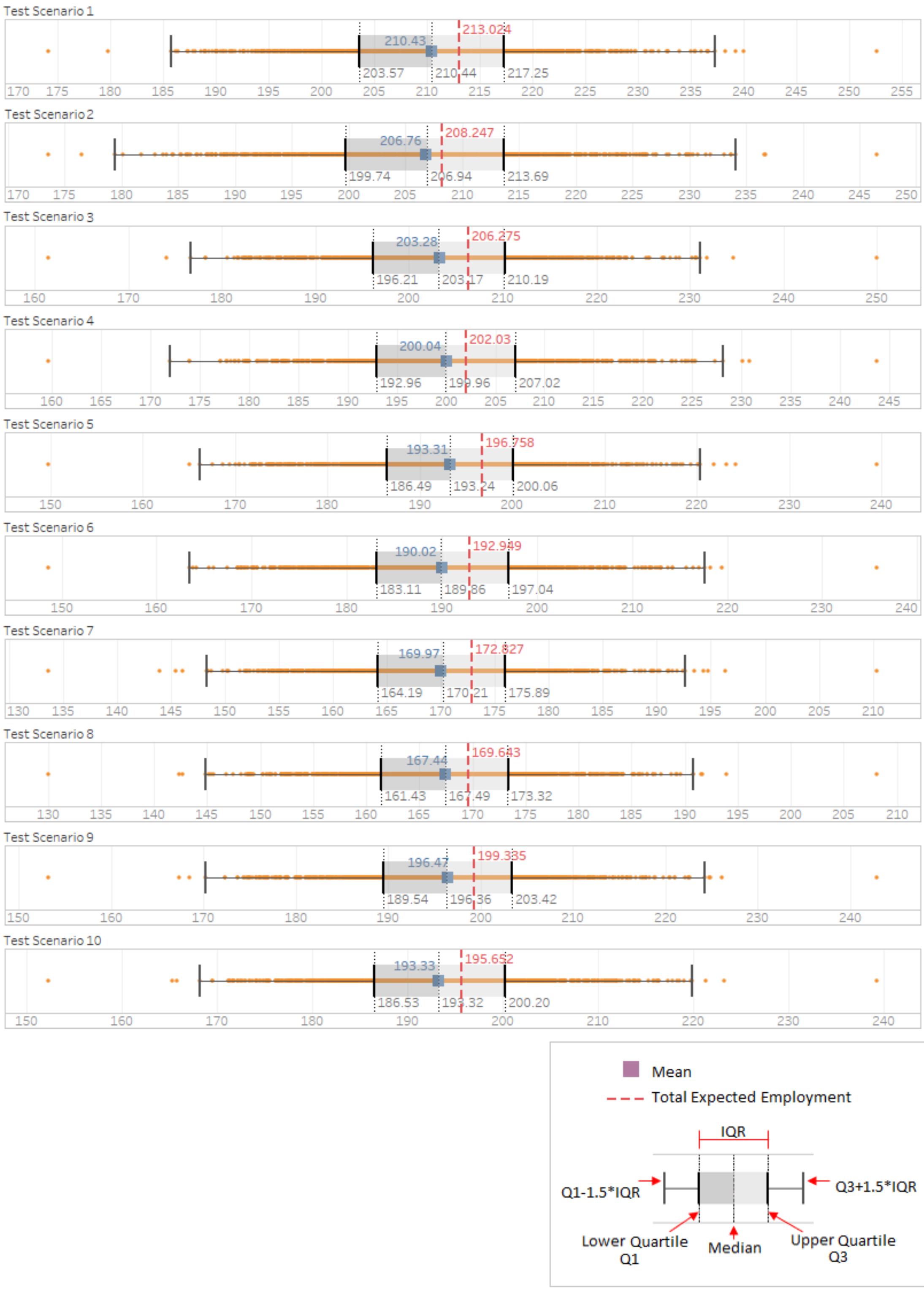


Figure 14: Uncertainty in the objective function illustrated across the first ten scenarios, namely those that use "Observed" capacity levels. Box plots depicting the distribution of the evaluation of the optimized $z^\star$ solution in 1,000 bootstrapped objective functions.

			Yearly ($n = 1$, Single Optimization)			Quarterly ($n = 4$ Optimizations)			Monthly ($n = 12$ Optimizations)			Weekly ($n = 52$ Optimizations)		
Capacity Adjustment	Min Avg Case Size	Binary Service Constraints	Total Expected Employed Refugees	Gains wrt to Predicted Employed Refugees (157.93)	# of Unplaced Cases / Refugees	Total Expected Employed Refugees	Gains wrt to Predicted Employed Refugees (157.93)	# of Unplaced Cases / Refugees	Total Expected Employed Refugees	Gains wrt to Predicted Employed Refugees (157.93)	# of Unplaced Cases / Refugees	Total Expected Employed Refugees	Gains wrt to Predicted Employed Refugees (157.93)	# of Unplaced Cases / Refugees
Observed	None	Off	213.024	34.89%	0/0	209.875	32.89%	4/23	207.985	31.69%	3/18	194.658	23.26%	3/19
Observed	None	On	208.247	31.86%	3/10	206.127	30.52%	7/29	201.633	27.67%	8/38	183.349	16.10%	21/59
Observed	2	Off	206.275	30.61%	1/1	198.688	25.81%	7/27	193.738	22.67%	10/10	182.168	15.35%	14/14
Observed	2	On	202.03	27.92%	2/9	195.252	23.63%	11/38	187.07	18.45%	18/44	169.204	7.14%	36/68
Observed	2.5	Off	196.758	24.59%	1/1	186.658	18.19%	24/37	179.93	13.93%	35/37	164.228	3.99%	58/58
Observed	2.5	On	192.949	22.17%	3/7	182.574	15.60%	28/43	171.824	8.80%	42/58	153.915	-2.54%	73/103
Observed	3	Off	172.827	9.43%	78/86	167.598	6.12%	78/86	161.751	2.42%	78/82	146.847	-7.02%	97/101
Observed	3	On	169.643	7.42%	79/89	165.656	4.89%	78/85	153.858	-2.58%	90/107	136.769	-13.40%	109/139
Observed	Observed	Off	199.335	26.22%	2/2	189.913	20.25%	24/27	178.564	13.07%	41/43	164.037	3.87%	62/67
Observed	Observed	On	195.652	23.89%	4/8	183.585	16.24%	31/51	171.916	8.86%	49/67	156.85	-0.68%	73/107
$\leq 110\%$	None	Off	218.055	38.07%	0/0	215.816	36.65%	2/12	210.658	33.39%	4/24	198.996	26.00%	6/37
$\leq 110\%$	None	On	212.959	34.84%	2/9	210.652	33.38%	4/18	205.737	30.27%	7/38	190.165	20.41%	17/43
$\leq 110\%$	2	Off	212.389	34.48%	0/0	206.649	30.85%	4/14	193.656	22.62%	11/39	181.962	15.22%	30/50
$\leq 110\%$	2	On	207.72	31.53%	2/9	203.967	29.15%	4/22	192.196	21.70%	12/40	177.023	12.09%	36/56
$\leq 110\%$	2.5	Off	202.75	28.38%	0/0	194.902	23.41%	14/20	178.237	12.86%	38/49	168.257	6.54%	56/57
$\leq 110\%$	2.5	On	198.837	25.90%	3/7	189.295	19.86%	20/33	177.391	12.32%	36/49	160.813	1.83%	65/89
$\leq 110\%$	3	Off	177.511	12.40%	78/86	172.399	9.16%	78/86	166.175	5.22%	78/84	151.179	-4.28%	93/93
$\leq 110\%$	3	On	174.266	10.34%	79/89	169.213	7.15%	81/91	161.88	2.50%	84/93	142.874	-9.53%	107/136
$\leq 110\%$	Observed	Off	204.268	29.34%	0/0	196.716	24.56%	16/22	180.308	14.17%	40/45	168.245	6.53%	55/56
$\leq 110\%$	Observed	On	200.49	26.95%	3/7	191.986	21.56%	19/37	176.133	11.53%	42/61	163.035	3.23%	67/95
$[90\%, 110\%]$	None	Off	218.055	38.07%	0/0		Infeasible instance			Infeasible instance			Infeasible instance	
$[90\%, 110\%]$	None	On	212.914	34.82%	1/2		Infeasible instance			Infeasible instance			Infeasible instance	
$[90\%, 110\%]$	2	Off	212.389	34.48%	0/0		Infeasible instance			Infeasible instance			Infeasible instance	
$[90\%, 110\%]$	2	On	207.575	31.44%	2/6		Infeasible instance			Infeasible instance			Infeasible instance	
$[90\%, 110\%]$	2.5	Off	202.75	28.38%	0/0		Infeasible instance			Infeasible instance			Infeasible instance	
$[90\%, 110\%]$	2.5	On	198.81	25.89%	2/3		Infeasible instance			Infeasible instance			Infeasible instance	
$[90\%, 110\%]$	3	Off		Infeasible instance			Infeasible instance			Infeasible instance			Infeasible instance	
$[90\%, 110\%]$	3	On		Infeasible instance			Infeasible instance			Infeasible instance			Infeasible instance	
$[90\%, 110\%]$	Observed	Off	204.259	29.34%	5/5		Infeasible instance			Infeasible instance			Infeasible instance	
$[90\%, 110\%]$	Observed	On	200.357	26.86%	6/7		Infeasible instance			Infeasible instance			Infeasible instance	

Table 10: Results of counterfactual employment optimization over the entire year, over 4 quarters, over 12 months, and over 52 weeks using the SUM objective and LASSO model.