

UC San Diego

UC San Diego Previously Published Works

Title

Combining Outcome-Based and Preference-Based Matching: A Constrained Priority Mechanism

Permalink

<https://escholarship.org/uc/item/86d8c02n>

Journal

Political Analysis, 30(1)

ISSN

1047-1987

Authors

Acharya, Avidit
Bansak, Kirk
Hainmueller, Jens

Publication Date

2022

DOI

10.1017/pan.2020.48

Peer reviewed

IPL Working Paper Series

**Matching Refugees to Host Country Locations
Based on Preferences and Outcomes**

Avidit Acharya, Kirk Bansak, and Jens Hainmueller

Working Paper No. 19-03

February 2019

arXiv:1902.07355v1 [econ.GN] 20 Feb 2019

IPL working papers are circulated for discussion and comment purposes. They have not been formally peer reviewed.

© 2019 by Avidit Acharya, Kirk Bansak, and Jens Hainmueller. All rights reserved.

Matching Refugees to Host Country Locations Based on Preferences and Outcomes*

Avidit Acharya[†] Kirk Bansak[‡] Jens Hainmueller[§]

February 21, 2019

Abstract

Facilitating the integration of refugees has become a major policy challenge in many host countries in the context of the global displacement crisis. One of the first policy decisions host countries make in the resettlement process is the assignment of refugees to locations within the country. We develop a mechanism to match refugees to locations in a way that takes into account their expected integration outcomes and their preferences over where to be settled. Our proposal is based on a priority mechanism that allows the government first to specify a threshold g for the minimum level of expected integration success that should be achieved. Refugees are then matched to locations based on their preferences subject to meeting the government's specified threshold. The mechanism is both strategy-proof and constrained efficient in that it always generates a matching that is not Pareto dominated by any other matching that respects the government's threshold. We demonstrate our approach using simulations and a real-world application to refugee data from the United States.

1 Introduction

The global refugee crisis is one of the most pressing social problems of our time. The United Nations reports that there are currently 68.5 million displaced persons globally, including 25.4 million refugees (United Nations, 2018). This crisis has led to tremendous

*We acknowledge funding from the Rockefeller Foundation, Schmidt Futures, the 2018 HAI seed grant program from the Stanford AI Lab, Stanford School of Medicine, and Stanford Graduate School of Business. The funders had no role in the data collection, analysis, decision to publish, or preparation of the manuscript. The U.S. refugee data were provided to us under a collaboration research agreement with the Lutheran Immigration and Refugee Service (LIRS). This agreement requires that we do not transfer or disclose the data. We thank LIRS for access to data and guidance. We are grateful to Fuhito Kojima and Shunya Noda for help and guidance.

[†]Stanford Political Science Department; avidit@stanford.edu

[‡]Stanford Political Science Department and Immigration Policy Lab; kbansak@stanford.edu

[§]Stanford Political Science Department and Immigration Policy Lab; jhain@stanford.edu.

suffering among the displaced population (Lindert et al., 2016). It has also resulted in major policymaking challenges in host countries that are struggling to facilitate the successful integration of refugees into the local economy and society (Commission, 2016).

There have been several policy proposals to improve the integration of refugees (Mousa, 2018). One prominent idea is to assign refugees to resettlement locations through a matching process (Moraga and Rapoport, 2014, Fernández-Huertas Moraga and Rapoport, 2015, Delacrétaz et al., 2016, Andersson and Ehlers, 2016, Bansak et al., 2018, Roth, 2018). When refugees are admitted to a host country, government officials typically decide the location to which a refugee is assigned within the country. Although the processes vary across countries, this assignment usually is determined by capacity constraints or proportional distribution keys. The idea of refugee matching is to select locations that are likely to be a good fit for a given refugee to thrive. Extant research has shown that the place of initial settlement has a profound impact on the long-term integration success of refugees (Åslund and Rooth, 2007, Damm, 2014, Bansak et al., 2018).

Two main approaches to refugee matching have emerged: preference-based and outcome-based matching. Preference-based matching uses market design algorithms, like those used in school choice problems (Abdulkadiroğlu and Sönmez, 2003, Abdulkadiroğlu et al., 2009), to assign refugees to locations based on the preferences of the refugees or the preferences of the locations (Moraga and Rapoport, 2014, Fernández-Huertas Moraga and Rapoport, 2015, Delacrétaz et al., 2016, Andersson and Ehlers, 2016). This approach is appealing because it allows refugees to select locations they think would be a good match. In addition, preference-based matching may facilitate successful integration if refugees have accurate private information about which location is best for them. However, to our knowledge such schemes have not been implemented in the refugee context, and there are several practical limitations. Governments want to ensure that refugees become self-sufficient and are typically reluctant to let them freely choose where to settle due to concerns that this could result in a highly uneven regional distribution and the creation of ethnic enclaves. In addition, there currently exists no systematic data on refugee preferences, and some refugees might have limited information with which to choose their best location.

The second approach is outcome-based matching (Bansak et al., 2018, Gözl and Procaccia, 2018). Here the assignment seeks to maximize refugees' predicted integration success as measured by, for example, employment or earnings. Data-driven algorithms train supervised learners on historical data to discover synergies between places and types of refugees. The learned models are then used for newly arriving refugees to predict their expected integration success and optimally match them to locations where they have the highest probability of success subject to capacity and other constraints. Outcome-based matching is appealing because it harnesses historical data to maximize expected integration success and does not require collecting data on refugee preferences. The Swiss government has recently implemented a randomized test to examine the performance of data-driven algorithms for outcome-based assignment. However, a pure outcome-based

approach does not take preferences into account and does not utilize private information that refugees may possess regarding which location would work best for them.

While preference-based and outcome-based matching are often discussed as contrasting approaches, they are not mutually exclusive. We propose a method that draws on the strengths of both approaches and incorporates them into a unified framework. Our assignment mechanism allows governments to harness the power of data-driven assignment to ensure some minimum level of expected integration success while taking into account the location preferences of refugees. Our mechanism integrates the data-driven matching algorithm of Bansak et al. (2018) into a priority mechanism (Satterthwaite and Sonnenschein, 1981) for preference-based matching. The government first proposes a metric of integration success (e.g. refugee employment, earnings, health outcomes, etc.), and a minimum level of expected integration success that should be achieved. Refugees express preferences over locations. The algorithm then maps preferences to a feasible matching by serially assigning refugees to locations in a way that accommodates their preferences subject to being able to maintain the minimum average level of expected integration success. We illustrate our mechanism using simulations and refugee data from the United States.

Our mechanism has several desirable properties. First, it strikes a compromise between the need of governments to ensure a minimum level of integration success and the appeal of incorporating refugee preferences. In this sense our approach improves policy through a marriage of machine-learning-based predictive analytics and preference-based matching from theories of market design (Milgrom and Tadelis, 2018). Second, despite the added complexity of accounting for the government’s constraint, our mechanism inherits the desirable properties of priority mechanisms: it is constrained Pareto-efficient (subject to the government’s constraint), immune to strategic manipulation through false reporting of preferences, and computationally feasible. It also allows refugees to rely purely on the algorithmic assignment or to express preferences without the requirement that they strictly rank all locations. This flexibility is important since there may be a large degree of heterogeneity as to whether refugees have distinct preferences over locations. Third, our algorithm can be implemented by governments with only minor additions to their existing assignment processes. It only requires the additional step of eliciting refugees’ top choices. Some governments, such as the Netherlands, already collect such information as part of their interviews with refugees.

2 g -Constrained Priority Mechanism

2.1 Preliminaries

There are n refugee families labeled $1, \dots, n$, each of which has to be assigned to a location in the host country. Let L denote the finite set of locations. Each location $l \in L$ has a capacity $q_l \geq 1$ as to how many families it can accommodate. We assume that $n \leq \sum_l q_l$ so that it is feasible to assign all families. For each family i , let $g_i(l)$ be

a measure of being successfully integrated in location l when assigned to that location. Integration success may be related to the family's preferences, but is a key consideration for the host government. For example, it could represent the probability that the head of family i 's household will be employed in the assigned location. We refer to $g_i(l)$ as the government's outcome score.

Each family i has a complete and transitive preference ordering \succsim_i over the set of locations.¹ Indifference and strict preference relations are denoted \sim_i and \succ_i , respectively, and $\succsim = (\succsim_1, \dots, \succsim_n)$ denotes the vector of preferences.

We make the assumption on families' preferences that the only indifferences are over the worst-ranked locations. That is, apart from possibly having ties among a set of locations that a family deems to be the worst, each family has a strict preference over all of the other locations. Formally, for all families i , if $l \sim_i l'$ for some $l' \neq l$, there is no l'' such that $l \succ_i l''$. This still allows for a family to be indifferent over all locations. This assumption is suited to our application: refugees often do not have full information on all possible locations, but they may have (strict) preferences over a limited set of top choices. In addition, in a practical application governments would likely limit preference elicitation to a set of top choices that refugees can express in an application form.

Define the set $S_i = L \setminus \{l \in L : \exists l' \sim_i l\}$ which are all of the locations except any that family i is indifferent over. Family i has a strict preference across all locations in S_i and if any location is left out of S_i then it must have been ranked worst.

A matching μ maps the set of individuals to locations. A matching μ is

1. **feasible** if it satisfies the capacity constraints:

$$|\mu^{-1}(l)| \leq q_l, \forall l$$

2. **g -acceptable** if the average outcome score is not lower than g :

$$\frac{1}{n} \sum_i g_i(\mu(i)) \geq g.$$

g -acceptability reflects the idea that the government wants the average outcome score not to fall below a specified threshold g . It wants to ensure that the allocation is such that refugee families have some minimum level of expected outcomes (e.g. a minimum expected employment rate).

Note that not all values of g can produce a feasible matching. Let \bar{g} denote the highest possible average outcome score that can be generated by a feasible matching:

$$\bar{g} := \max_{\mu} \frac{1}{n} \sum_i g_i(\mu(i)) \text{ subject to } |\mu^{-1}(l)| \leq q_l, \forall l \quad (1)$$

Feasible g -acceptable matchings exist only for $g \leq \bar{g}$.

¹We assume that all families prefer to be assigned to a location rather than not assigned, so we can omit non-assignment from the set of possible outcomes for each family.

2.2 The Mechanism

Given a value of $g \leq \bar{g}$, the algorithm starts with family 1 and works down to family n in a sequence of n steps before completing in either the n th or an additional $(n+1)$ th step. At Step $i \leq n$, family i is either assigned to a location or put on hold by being added to a set of temporarily unassigned families that will all get assigned simultaneously at Step $n+1$. At each Step i , let N_i denote the set of families $j < i$ that have been put on hold. $N_1 = \emptyset$ since at the start of the algorithm no family is on hold.

If family $j < i$ was assigned a location prior to Step i , then let $\alpha_i(j)$ denote the location and $(j, \alpha_i(j))$ the assignment, viewing α_i as a function. Refer to this function as the completed assignment at Step i . Note that $\alpha_1 = \emptyset$, so the completed assignment at Step 1 is trivial. A remaining assignment β_i at Step i is a mapping of the unassigned families $\{i, \dots, n\} \cup N_i$ to locations such that

$$\mu_{(\alpha_i, \beta_i)}(j) := \begin{cases} \alpha_i(j) & \text{if } j < i \\ \beta_i(j) & \text{if } j \in \{i, \dots, n\} \cup N_i \end{cases}$$

is a matching. We refer to $\mu_{(\alpha_i, \beta_i)}$ as the matching associated with the pair of completed and remaining assignments (α_i, β_i) . The existence of these matchings will be guaranteed recursively by the algorithm.

At each Step $i \leq n$, given α_i define the set

$$L_i^g(\alpha_i) = \{l \in L : \exists \beta_i \text{ s.t. } l = \beta_i(i) \text{ and } \mu_{(\alpha_i, \beta_i)} \text{ is a feasible } g\text{-acceptable matching}\}$$

This is the set of locations that are not at full capacity and for which there is a way to finish assigning all unassigned families so as to create a feasible g -acceptable matching.

Let q_l^i be the remaining capacity of location l after any individuals ahead of i (i.e., $j < i$) have been assigned in the previous $i-1$ steps. At the start we have $q_l^1 = q_l$ for all l . It will also be convenient to define the following problems: for all Steps $i = 1, \dots, n+1$, and given a vector $q^i := (q_l^i)_{l \in L}$,

$$G_i(q^i) := \max_{\beta_i} \sum_{j \in \{i, \dots, n\} \cup N_i} g_j(\beta_i(j))$$

subject to $|\beta_i^{-1}(l)| \leq q_l^i, \forall l$ (2)

with the convention that $\{i, \dots, n\} := \emptyset$ if $i = n+1$. At each Step i , the problem in (2) finds the remaining assignment that maximizes the total outcome score subject to the updated capacity constraints at Step i . The solution to this problem at each step determines whether the associated matching is g -acceptable. In fact, to verify whether or not a location l belongs in $L_i^g(\alpha_i)$ we must first check whether the highest possible value of the average outcome score that can be achieved under the remaining assignment is at least g ; i.e., whether

$$\bar{g}_i(l) := \frac{1}{n} \left(G_{i+1}(q^{i+1}) + g_i(l) + \sum_{j < i \text{ s.t. } j \notin N_i} g_j(\alpha_i(j)) \right) \geq g$$

where $q_l^{i+1} = q_l^i$ for all $l' \neq l$ and $q_l^{i+1} = q_l^i - 1$. If indeed $\bar{g}_i(l) \geq g$ and $q_l^i > 0$, then l belongs to $L_i^g(\alpha_i)$; otherwise it does not. Constructing $L_i^g(\alpha_i)$ at each Step $i = 1, \dots, n+1$ therefore requires solving the problems given in (2). In addition, to verify whether $g < \bar{g}$ also requires solving one of these problems since the problem in (1) equals $G_1(q^1)/n$.

The steps of the algorithm are as follows.

Step 0. Verify that $g \leq \bar{g}$ and proceed only if it holds.

Step $i \leq n$. If $S_i \cap L_i^g(\alpha_i)$ is empty (meaning that there is no location that family i ranked strictly to which it could be assigned, and we can find a remaining assignment that generates a feasible g -acceptable matching), then place family i on hold. In this case, set

$$N_{i+1} = N_i \cup \{i\}, \alpha_{i+1} = \alpha_i, q_l^{i+1} = q_l^i \forall l$$

and move on to Step $i + 1$. Otherwise, if $S_i \cap L_i^g(\alpha_i)$ is nonempty, then it contains a unique best location from the perspective of family i – i.e., a location l_i^* such that $l_i^* \succ_i l$ for all $l \in S_i \cap L_i^g(\alpha_i)$. This follows from the fact that i ranks the elements of S_i strictly. Assign family i to l_i^* , and set

$$N_{i+1} = N_i, \alpha_{i+1} = \alpha_i \cup \{(i, l_i^*)\}, \\ q_{l_i^*}^{i+1} = q_{l_i^*}^i - 1, \text{ and } q_l^{i+1} = q_l^i \forall l \neq l_i^*$$

If $i < n$, then move to Step $i + 1$. If $i = n$, then move to Step $n + 1$ only if a family was ever put on hold (i.e., $N_{n+1} \neq \emptyset$); otherwise, stop.

Step $n+1$. At this stage the only unassigned families are those that were put on hold in N_{n+1} . Here, choose any remaining assignment that maximizes the average outcome score given the completed assignment and the capacity constraints; that is, solve (2) for $i = n + 1$ and stop.

For any preference vector \succsim satisfying our assumptions, our algorithm produces a matching, namely $\mu_{(\alpha_s, \beta_s)}$, where $s \in \{n, n + 1\}$ was the step at which the algorithm stopped. The algorithm then defines a “mechanism” φ , which, given the other parameters of the model, is a mapping from preference vectors to feasible matchings. We refer to the mechanism as g -constrained priority, since it is a modification of the usual priority mechanism (Satterthwaite and Sonnenschein, 1981) for our application.

2.3 Properties of the Mechanism

Let $\varphi(\succsim)$ denote the matching produced by the g -constrained priority mechanism for any preference vector \succsim that satisfies our assumptions, and $\varphi(\succsim)(i)$ the location assignment of family i under this matching. By construction, the matching produced by this mechanism is feasible and g -acceptable. In addition, the mechanism satisfies two key properties. It is:

1. **constrained efficient** in the sense that for all preference vectors \succsim that satisfy our assumptions, $\varphi(\succsim)$ is not Pareto dominated by another feasible g -acceptable

matching μ . That is, it is not the case that $\mu(i) \succeq_i \varphi(\zeta)(i)$ for all families i , and $\mu(i) \succ_i \varphi(\zeta)(i)$ for some family i .

2. **strategy-proof** in the sense that truthful reporting is a dominant strategy of the induced preference reporting game. That is, for every preference vector ζ satisfying our assumptions, every family i , and every alternative preference ζ'_i that i could report that also satisfies our assumptions, $\varphi(\zeta)(i) \succeq_i \varphi(\zeta'_i, \zeta_{-i})(i)$.

The proof that the mechanism is constrained efficient and strategy proof is straightforward, but for completeness we include it in the Supplementary Information (SI).

3 Applications

To illustrate the mechanism, we apply it both to simulated data and real-world data from refugees in the United States.

Our mechanism requires governments to select a value for g , and this choice implies a tradeoff between an outcome-based and preference-based matching. It is desirable to achieve the highest possible value of g to ensure that refugees' integration outcomes are optimized. However, setting a higher value of g comes at the cost of assigning refugees to locations that are, in expectation, lower in their preference rankings. That is, while the mechanism simultaneously attempts to optimize for both outcomes and preferences, there is a tradeoff between the two, where the balance of that tradeoff changes as g increases.

The precise nature of the tradeoff also depends upon the joint distribution of refugees' preference rankings and their outcome scores. Two measures, in particular, play an important role: the correlation between outcome scores and preference rankings within families (i.e. the degree to which a family's preferred locations align with the locations where that family would achieve their best outcomes) and the correlation between preference rankings across families (i.e. the degree to which families have similar preference rankings). We apply the mechanism to simulated data to show these properties.

In addition, to illustrate how the mechanism could perform in a real-world scenario, we apply the mechanism to data from refugees in the United States. Early employment is a core goal of the U.S. resettlement program, which strives to quickly transition refugees into self-sufficiency after arrival. This application illustrates how our mechanism could hypothetically be employed in the United States to achieve a desired level of early employment while geographically assigning refugees based on their location preferences.

3.1 Simulation Data

For simplicity, our simulations involve assigning 100 families to 100 locations with one slot each. For each family, we randomly generate a preference rank vector (with 1 indicating the most desired location and 100 the worst) and an outcome score vector (with values in $[0, 1]$). The simulations vary both the correlation between preference and

outcome vectors (-0.5 , 0 , and 0.5) and the correlation between preference vectors across families (0 , 0.5 , and 0.8).² This yields nine different scenarios, and in each we apply our mechanism to make the assignment for various values of g . See the Supplemental Information (SI) for details.

3.2 Refugee Data

Our real-world refugee data includes de-identified information on working-age refugees (ages 18 to 64; $N = 33,782$) who have been resettled to the United States during the 2011-2016 period by one of the largest U.S. refugee resettlement agencies. Over this time period, the agencies' placement officers centrally assigned refugees to one of approximately 40 resettlement locations in the agency's network. The data contain details on the refugee characteristics such as age, gender, origin, and education. It also includes the assigned resettlement location, whether the refugee was employed at 90 days after arrival, and whether the refugee migrated from the initial location within 90 days.

We applied our mechanism to data on the refugee families who arrived in the third quarter (Q3) of 2016, specifically focusing on refugees who were free to be assigned to different resettlement locations (561 families), in contrast to refugees who were predestined to specific locations on the basis of existing family or other ties. To generate each family's outcome score vector across each of the locations, we employed the same methodology in Bansak et al. (2018), using the data for the refugees who arrived from 2011 up to (but not including) 2016 Q3 to generate models that predict the expected employment success of a family (i.e. the mean probability of finding employment among working-age members of the family) at any of the locations, as a function of their background characteristics. These models were then applied to the families who arrived in 2016 Q3 to generate their predicted employment success at each location, which comprise their outcome score vectors. See the SI and Bansak et al. (2018) for details.

Our mechanism also requires data on location preferences of refugees. To the best of our knowledge, such data do not currently exist in the United States, where refugees are assigned to locations by the resettlement agencies. We therefore infer revealed location preferences from secondary migration behavior. Specifically, we use the same modeling procedures used in the outcome score estimation, simply swapping in out-migration in place of employment as the response variable. This allows us to predict for each refugee family that arrived in 2016 Q3 the probability of out-migration at each location as a function of their background characteristics. For each family, we then rank locations such that the location with the lowest (highest) probability of out-migration is ranked first (last).

²The correlation between preference and outcome vectors treats higher preferences (i.e. closer to 1) as more positive values, such that a positive correlation between preferences and outcomes indicates more highly preferred locations are those that also result in higher outcome scores.

4 Results

4.1 Simulations

Figure 1 depicts the results for nine different simulation scenarios that vary the correlation between preferences and outcome scores within families and the correlation between preferences across families. In addition, to model a real-world scenario, in which families can indicate only a limited number of top locations in an application form, the preference vectors are truncated such that only the top 10 ranks are retained and indifference is established among the remaining locations. The top panel shows the proportion of families who were assigned to one of their top three locations given various levels of g , the government’s threshold for the minimum average outcome score. The bottom panel shows the mean outcome score for families in their assigned locations for the same levels of g . The curves end once \bar{g} has been reached and hence no feasible assignment is possible.

There is a clear tradeoff between realized preference ranks and outcome scores in all simulations. As g is increased, the realized mean outcome score eventually increases. This is a mechanical result of increasing g and hence enforcing the requirement for a higher mean outcome value. Simultaneously, as soon as the mean outcome score is impacted, the proportion of families assigned to one of their preferred locations also begins to decrease. This occurs because enforcing the requirement for a higher value of g requires the mechanism to deviate from the preference-based optimization.

Figure 1 also shows how the immediacy and severity of the tradeoff can vary substantially depending upon the joint distribution of preferences and outcome scores.³ First, focusing on the top panel, we see that the higher the correlation between families’ preferences, the worse is the achievable baseline proportion of families that can be assigned to one of their top locations at the lowest values of g . This result, which holds regardless of the correlation between preferences and outcome scores, is intuitive: the more similar are different families’ preferences, the more rivalrous is the matching procedure, and hence the more difficult it is to match families to one of their top-ranked locations given limited capacity in each location.

Second, the more positive the correlation between preferences and outcome scores, the less severe is the tradeoff in the sense that the tradeoff does not kick in until higher levels of g are enforced. The intuition for this result is that if preferences and outcomes are positively correlated, then matching based on preferences should indirectly also lead to outcome-based matching, and hence deviation from the preference-based solution will not occur until a higher level of g is reached. This is a useful finding from the standpoint of a real-world implementation of the mechanism. If, in advance of their preference reporting, refugees were given information on their predicted outcomes in each location, they could incorporate such information into their preference determination. If this

³It can also depend on the number of slots available in each location and the extent to which each location contributes to the correlations.

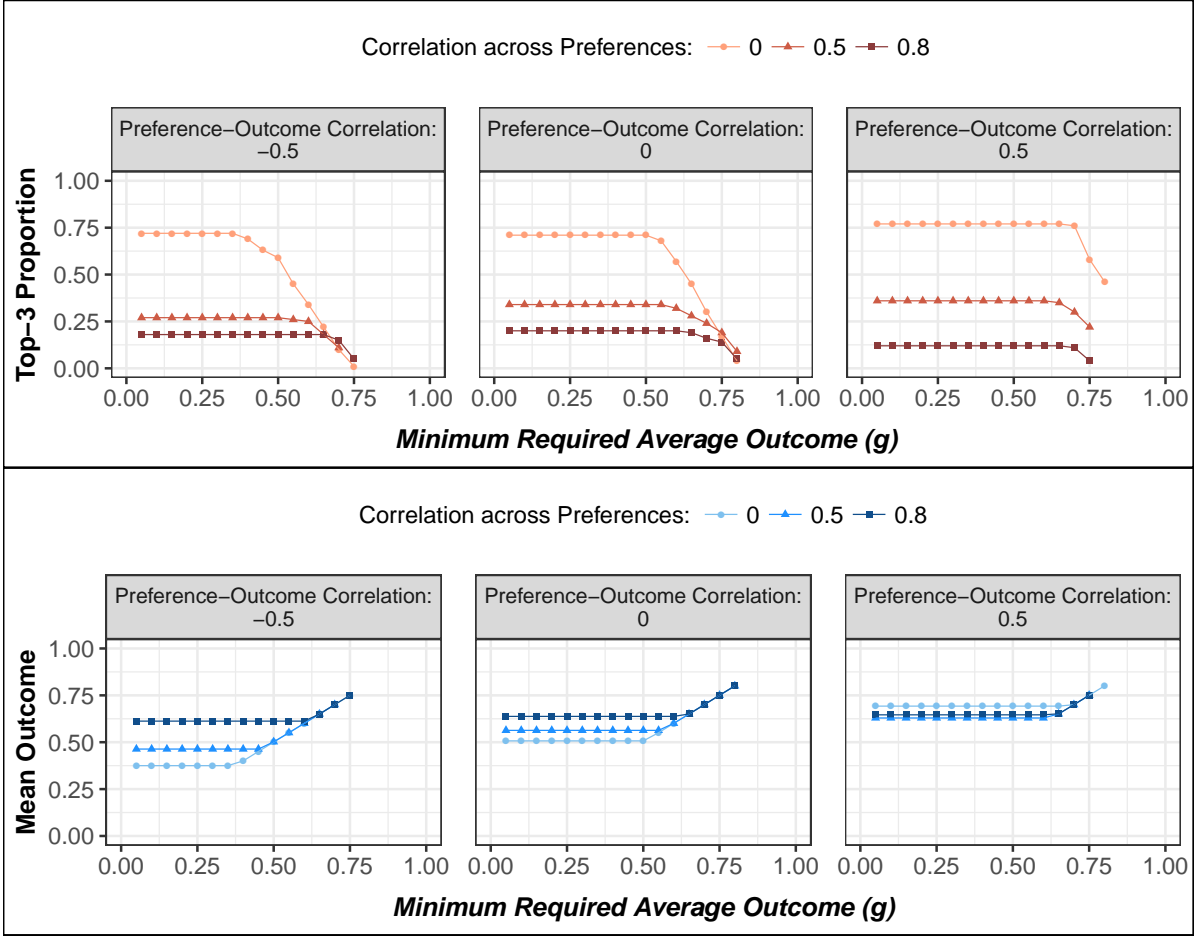


Figure 1: Results from applying our assignment mechanism to simulated data that varies the correlations between location preference and integration outcome vectors and the correlations between preference vectors across families. Upper panel shows the average probability that a family was assigned to one of its top three locations. Lower panel shows the realized average integration outcomes, i.e. the average projected probability of employment. $N=100$.

results in a closer alignment of preferences and outcomes, that would help alleviate the tradeoff in the mechanism.

Third, turning to the bottom panel in Figure 1, we see that once the tradeoff kicks in, the realized mean outcome curves trace closely along the identity line; that is, upon enforcing a level of g that deviates from the preference-based assignment, the mechanism will find an alternative assignment that optimizes for preferences subject to just barely satisfying the g constraint. The realized mean outcome results also mirror the trends on realized preference ranks: The more positive is the correlation between preference and outcome vectors, the later the tradeoff kicks in.

Fourth, we see that given a negative correlation between preferences and outcome scores, the correlation across preference vectors has a significant impact on how the tradeoff affects the realized mean outcome score, with the tradeoff being more severe with a low correlation across preference vectors. This result can be explained as follows. A negative correlation between preference and outcome vectors implies that preference-based assignment is counter to the goal of optimizing for realized outcome scores. However, if there is also a positive correlation across families' preferences, that means that different kinds of families generally prefer the same locations, and hence also that the locations that result in low outcome scores are also similar across people, thus limiting the degree to which matching based on preferences will actually hurt realized outcome scores on average. If, in contrast, there is no correlation across preferences, then there is greater latitude for the mechanism to assign families to their higher-ranked locations, which also happen to be the locations that are the worst for their outcome scores. As the correlation between preference and outcome vectors becomes more positive, this dynamic begins to disappear. However, the reason it does not reverse in the bottom-right panel of Figure 1 is due to the existence of trailing indifferences in the preference rank vectors, which means the families who could not be matched to one of their strictly ranked locations are assigned using outcome-based optimization, thereby limiting the effect of the phenomenon described above.⁴

4.2 Application Using U.S. Data

Figure 2 shows features of the joint distribution of the refugee families' outcome score and preference rank vectors. The top panel pertains to the correlation between the families' outcome and preference vectors. For each family, a correlation is computed between its two vectors, and the panel displays the distribution of those correlations. The distribution is roughly centered around zero (the mean correlation is 0.03). This suggests, perhaps surprisingly, a relatively limited relationship between the locations refugees prefer and those where they would actually achieve better employment outcomes. This is an interesting finding and also has a key policy implication. Providing refugees with information on which locations are beneficial for their employment outcomes would allow them to formulate more informed preferences. If this results in a closer correlation between preference and outcome vectors, this would help strengthen our mechanism since a more positive correlation alleviates the tradeoff between outcome- and preference-based matching.

The middle panel in Figure 2 shows the distribution of pairwise correlations between families' preference vectors. The correlations are mostly highly positive, with a mean correlation of 0.67. This shows that preference vectors are relatively similar across the families; many refugees would more or less prefer to be placed in similar locations. Given

⁴The SI includes the results of the same simulations without truncating the preference rank vectors. In that case, we do see the expected reversal across the lower three panels.

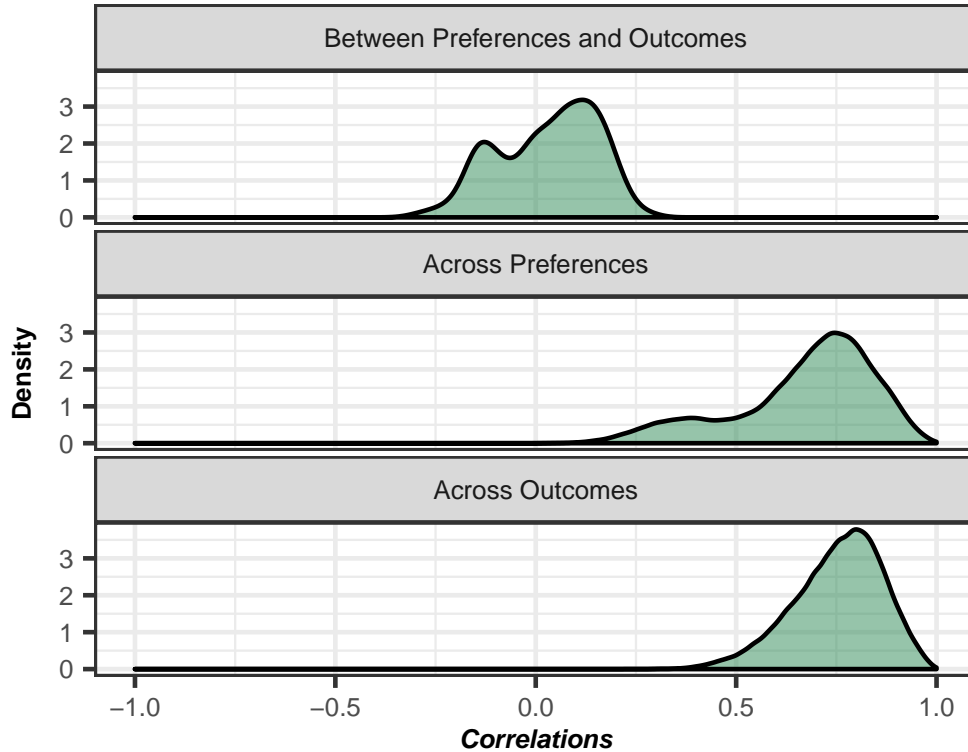


Figure 2: Shows the distribution of pairwise correlations between refugee family location preferences, integration outcomes (i.e. employment), and preferences and outcomes. $N=561$ refugee families who arrived in the United States in Q3 of 2016.

the existence of location capacity constraints, this is an inconvenient finding from the standpoint of preference-based assignment.

The bottom panel in Figure 3 shows the distribution of all pairwise correlations between families' outcome vectors. As can be seen, the correlations are overwhelmingly positive (with a mean correlation of 0.75), highlighting the fact already shown elsewhere (Bansak et al., 2018) that certain locations are generally better than other locations for helping refugees to achieve positive employment outcomes. However, the fact that there still is meaningful variation across different families' outcome score vectors indicates that certain locations do indeed make a better match for different refugee families, depending on their personal characteristics, which is the foundation for the outcome-optimization matching procedure introduced by Bansak et al. (2018).

In applying our mechanism to the 2016 Q3 refugee data, we impose real-world assignment constraints, giving each location capacity for the same number of families as were sent to those locations in actuality. We also truncate each family's preference vectors such that only the first 10 ranks are retained and indifference is established among the remaining locations.

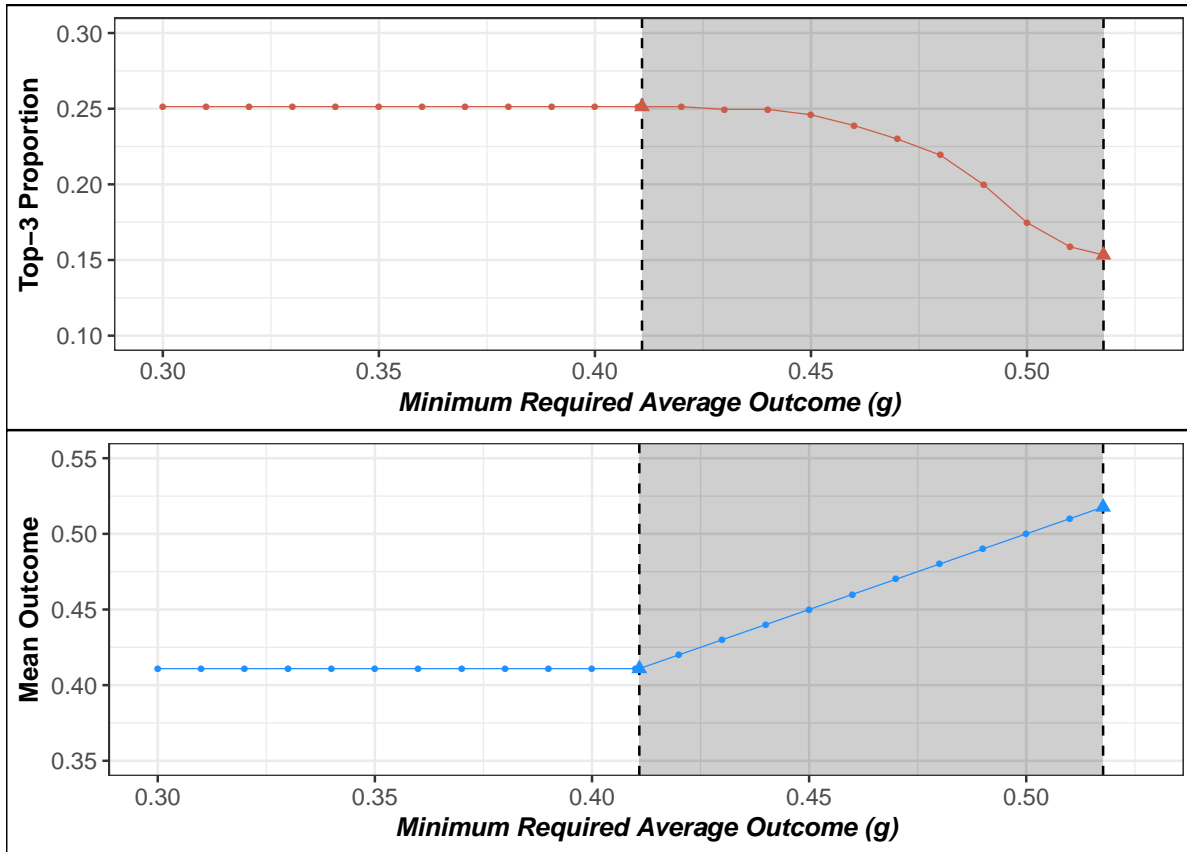


Figure 3: Results of applying the assignment mechanism to refugee families in the United States for various specified thresholds for the expected minimum level of average integration outcomes (g). Upper panel shows the average probability that a refugee got assigned to one of their top three locations. Lower panel shows the realized average integration outcomes, i.e. the average projected probability of employment. $N=561$ families who arrived in Q3 of 2016.

Figure 3 displays the results of applying the mechanism to these refugee families (in a randomly drawn order). As before, the mechanism is applied at various levels of g , which is denoted by the x -axis. The y -axis of the top panel denotes the proportion of cases assigned to one of their top three locations, while the y -axis in the bottom panel denotes the mean realized outcome score, i.e. the average predicted probability of employment, based on the assignment. The two dashed vertical lines highlight the tradeoff interval, where altering the value of g impacts both preferences and outcomes, and the interval ends when g is raised above \bar{g} .

Given a predominantly preference-based assignment (i.e. setting g to any value below the value at which the tradeoff interval begins), a mean outcome score of 0.41 is achieved,

meaning the predicted average employment rate is 41%.⁵ On the opposite end of the spectrum, a purely outcome-driven optimization would yield the highest feasible g (\bar{g}), which is just below 0.52. The fact that it is not possible to raise g even further is, of course, the result of the full distribution of the refugee families’ outcome vectors, namely the fact that they feature a large positive correlation with one another.

As before, within the tradeoff interval, the mean outcome score curve in the bottom panel traces closely along the identity line. However, the preference curve in the top panel features a gradient that more gradually steepens, with the tradeoff becoming increasingly more severe as g is increased.

5 Conclusion

Refugee matching has become a prominent policy innovation proposed to help facilitate the successful integration of refugees into the host country’s economy and society. However, there is disagreement over whether integration is best served by matching on refugee preferences or expected integration outcomes.

We have developed a mechanism that incorporates the strengths of both approaches into a unified framework to assign refugees based on optimizing both refugee preferences and expected outcomes. Our mechanism strikes a compromise in that it allows governments to ensure a minimum level of expected integration success (g) while at the same time respecting refugee preferences to the extent possible. It is also strategy proof, does not require refugees to rank all locations, and could be incorporated into existing assignment mechanisms by eliciting refugee preferences for their top locations. In a real-world implementation, governments could either fix a feasible value of g in advance or review the projected results along a sequence of g values, as in Figure 3, and choose the final preferred assignment according to their own criteria.

Our mechanism contributes to the literature on refugee matching and also more generally to the study of market design. For refugee matching in particular, our mechanism provides governments with an actionable and cost-efficient tool to improve the welfare of refugees and the communities in which they reside. More generally, our mechanism provides an example of how predictive analytics from machine learning can be fruitfully combined with the preference-based allocation schemes common in market design. The marriage of these two approaches can provide a powerful tool to improve allocations in a way that incorporates information from preferences about what people want while harnessing the statistical learnings from what the historical data suggest would be the best options. Given the heterogeneity in information levels and the richness of historical data on outcomes, we envision that such a combined approach could lead to better

⁵Setting g to a value below the tradeoff interval does not result in a purely preference-based assignment given the trailing indifferences in the preference rank vectors. We also applied the mechanism to the same data without truncating the preference vectors. The result is a purely preference-based assignment at the lowest values of g , which yields a mean outcome score of 0.37. See SI.

allocations in a variety of settings compared to schemes that rely only on preferences or only on expected outcomes.

Supplemental Information (SI)

A Proof of the Mechanism's Properties

The following presents the proof that the g -constrained priority mechanism is constrained efficient and strategy-proof.

Constrained efficient Suppose that φ is not g -constrained efficient, so that for some preference profile \succsim , $\varphi(\succsim)$ is Pareto-dominated by a feasible g -acceptable matching μ .

For all families i , let $M_i = \{j < i : j \notin N_i\}$ be the families ahead of i that were already assigned a location under $\varphi(\succsim)$, and let $\underline{i} = \min\{i : \mu(i) \succ_i \varphi(\succsim)(i)\}$ be the first family to which μ assigns it a location that it strictly prefers to the one it gets under $\varphi(\succsim)$. (Such a family must exist if μ Pareto-dominates $\varphi(\succsim)$.) By construction $\mu(i) = \varphi(\succsim)(i)$ for all $i \in M_{\underline{i}}$. So for μ to be feasible and g -acceptable, it must be that $\mu(\underline{i}) \in S_{\underline{i}} \cap L_{\underline{i}}^g(\alpha_{\underline{i}})$, where $\alpha_{\underline{i}}$ is the completed assignment under $\varphi(\succsim)$ at Step \underline{i} . This means that $S_{\underline{i}} \cap L_{\underline{i}}^g(\alpha_{\underline{i}}) \neq \emptyset$ so $\varphi(\succsim)$ must have assigned the best location $l_{\underline{i}}^*$ in this set to family \underline{i} . But since $\mu(\underline{i}) \succ_i \varphi(\succsim)(\underline{i}) = l_{\underline{i}}^*$, this contradicts the assumption that $l_{\underline{i}}^*$ is the best location for \underline{i} in $S_{\underline{i}} \cap L_{\underline{i}}^g(\alpha_{\underline{i}})$.

Strategy-proof Suppose that there is some i for whom reporting a different preference \succsim'_i produces a strictly better location assignment: $\varphi(\succsim'_i, \succsim_{-i}) \succ_i \varphi(\succsim)(i)$.

Let $l'_i = \varphi(\succsim'_i, \succsim_{-i})$ and note that $S_j \cap L_j^g(\alpha_j)$ is independent of i 's reported preference for all $j < i$. Therefore, $N_i = N'_i$ where N_i is the set of families on hold at Step i under the truthfully reported profile \succsim and N'_i are those on hold at Step i under the profile $(\succsim'_i, \succsim_{-i})$. In addition, $\varphi(\succsim'_i, \succsim_{-i})(j) = \varphi(\succsim)(j)$ for all $j \in N_i$. This implies that $\alpha'_i = \alpha_i$, where α'_i is the completed assignment at Step i under preference profile $(\succsim'_i, \succsim_{-i})$ and α_i is the completed assignment at Step i under preference profile \succsim . Therefore, $L_i^g(\alpha_i) = L_i^g(\alpha'_i) =: L'_i$.

Let S'_i be the locations that i ranks strictly under \succsim'_i and S_i the locations that i ranks strictly under \succsim_i . If $S_i \cap L'_i = \emptyset$, then all of the locations in L'_i are ones that i ranks worst, and i is guaranteed to be assigned one of these locations regardless of which location i reports. Therefore it cannot be that $\varphi(\succsim'_i, \succsim_{-i}) \succ_i \varphi(\succsim)(i)$.

On the other hand, if $S_i \cap L'_i \neq \emptyset$ then $\varphi(\succsim'_i, \succsim_{-i}) \succ_i \varphi(\succsim)(i)$ and $L_i^g(\alpha_i) = L'_i(\alpha'_i)$ implies that $l'_i \in S_i \cap L_i(\alpha_i)$. But then $l'_i \succ_i \varphi(\succsim)(i) = l_i^*$ contradicts the fact that l_i^* is the unique best location in $S_i \cap L_i(\alpha_i)$ under preference \succsim_i .

B Verifying g -Acceptability

As described in the main text, implementing the g -constrained priority mechanism involves iteratively verifying that the next assignment of a family to a particular location can be performed without compromising the possibility of a g -acceptable final matching. This process requires solving the maximization problem in Equation 2 of the main text:

$$G_i(q^i) := \max_{\beta_i} \sum_{j \in \{i, \dots, n\} \cup N_i} g_j(\beta_i(j))$$

$$\text{subject to } |\beta_i^{-1}(l)| \leq q_l^i, \forall l \quad (2)$$

This involves computing the maximum possible total outcome score for any remaining set of units and the remaining location capacities.

In implementing the mechanism, Equation 2 can be solved by employing a standard linear sum assignment problem (LSAP) (Burkard et al., revised reprint, 2012). Specifically, the LSAP formulation is applied to an augmented cost matrix, whereby the rows correspond to the remaining units and the columns correspond to location capacity slots (i.e. each column is replicated according to the number of capacity slots belonging to the associated location). Each element $[i, v]$ of the cost matrix corresponds to the complement of the outcome score for the i th unit when assigned to the location to which the v th column pertains.

Various algorithms have been developed for solving the LSAP, beginning with the introduction of the Hungarian algorithm in the 1950s (Kuhn, 1955, Munkres, 1957). We employ the RELAX-IV cost flow solver developed by Bertsekas and Tseng (Bertsekas and Tseng, 1994) and implemented in R by the `optmatch` package (Hansen and Klopfer, 2006).

C Simulation Application: Additional Details

The follow describes the data-generating process employed in the simulations.

First a number N is chosen, denoting the number of families. For simplicity, the same number of locations is also used, each with capacity for one family. In addition, ρ_p and ρ_{op} are both chosen, denoting the pre-specified correlation between preferences across families and the correlation between preferences and outcome scores within families.

Next, N different N -dimensional latent variable vectors are generated, and these vectors are column-bound into an $N \times N$ matrix, which we denote by \mathbf{P} , representing a simulated preference matrix. Specifically, each vector is a multivariate normal random vector, using a mean vector of 0, and a covariance matrix with 1 for all the diagonal elements and ρ_p for all the off-diagonal elements. Let \vec{z}_l denote the l th N -dimensional latent variable vector, which pertains to the l th location and comprises the l th column of \mathbf{P} . For any given vector, the i th element pertains to the i th family.

By generating the $N \times N$ matrix \mathbf{P} in this way, each row represents a family and each column represents a location. Thus, the i th row, $\mathbf{P}[i,]$, denotes a latent preference vector

for family i , with higher (more positive) values corresponding to a higher preference and vice versa. By construction, for any two families (rows), the pairwise correlation between the two vectors will be ρ_p in expectation, imposing a correlation of ρ_p across families' location preferences.

Let \vec{s}_i denote the i th family's outcome score vector. The outcome score vectors are constructed such that $\vec{s}_i = \text{sign}(\rho_{op}) \cdot (\mathbf{P}[i, \cdot] + \vec{\epsilon})$, where the elements of $\vec{\epsilon}$ are independently distributed normal with mean 0 and variance σ_ϵ^2 . The value of σ_ϵ^2 is determined such that it, in combination with the $\text{sign}(\rho_{op})$ operator, produces an expected pairwise correlation of ρ_{op} between \vec{s}_i and $P[i, \cdot]$, thereby inducing the correlation of ρ_{op} between a family's preferences and outcome scores. The outcome score vectors are then row-bound to create an $N \times N$ outcome score matrix \mathbf{S} , where each row represents a family and each column represents a location.

In applying our mechanism to the simulated data, the \mathbf{S} matrix is first normalized such that its elements are in the interval $[0, 1]$, and the \mathbf{P} matrix is mapped to preference ranks (i.e. each row $\mathbf{P}[i, \cdot]$ is transformed into ranks such that the most positive value becomes 1 and the most negative value becomes N).

For simplicity, the simulations presented in the study employ $N = 100$ (i.e. 100 families assigned to 100 locations each with one slot). In addition, to mimic reality, in which families are likely to be able to report only a limited number of location preferences, the preference vectors for each family are truncated such that only the top 10 ranks are retained and indifference is established among the remaining locations. The simulations vary both the correlation between preference and outcome vectors (three values of ρ_{op} : -0.5, 0, and 0.5) and the correlation between preference vectors across families (three values of ρ_p : 0, 0.5, and 0.8). This yields nine different scenarios, and in each we apply our mechanism to make the assignment for various values of g . Figure 1 in the main text displays the results.

In addition, Figure S1 in this SI shows the results of the same simulations when the preference rank vectors are not truncated.

D U.S. Refugee Application

D.1 Background Information on U.S. Resettlement

Resettled refugees in the United States are assigned to locations based on collaboration between the Department of State and nine voluntary resettlement agencies. During a regular draft, refugees are first allocated to one of the nine agencies according to specific quotas. Agencies are then responsible for assigning refugees to locations within their networks. Typically refugees are assigned as cases, where a case is a family. The assignment varies based on whether the refugee has family ties in the United States. Refugees with ties are placed at the location most proximate to the tie. Refugees without such ties, so-called "free cases," are assigned on a case-by-case basis and can be assigned to any location in the network. Placement officers consider special characteristics of

the case (nationality, case structure, medical needs) and consult with the local offices on whether they can accommodate a case (e.g. some offices may lack interpreters for particular languages). Among the offices that can accommodate a case, the case is then typically assigned to offices with the smallest proportion of their yearly capacity currently filled. Note that a different process applies to refugees with Special Immigrant Visas (SIVs).

Once a refugee case has been assigned, the local office then provides placement and reception services for 90 days beginning after arrival as mandated by the U.S. Resettlement Program. The duration is 180 days for refugees assigned to the matching grant program. Agencies are mandated to report employment outcomes to the Department of State after the conclusion of the placement and reception period. If a refugee leaves the area before the placement and reception period ends, they may no longer receive the benefits associated with the placement and reception service.

D.2 Registry Data

Our data includes all refugees that were resettled by one of the largest resettlement agencies and arrived between quarter 1, 2011 and quarter 3, 2016. The same data is used in Bansak et al. (2018). We restrict the sample to those aged between 18 and 64 years at the time of arrival (i.e. working age). We also remove a small number of duplicates and locations that have had less than 200 refugees assigned to them over the entire period. In the final data there are 33,782 refugees from 22,144 cases. Of those, 9,506 refugees are from free cases.

Table S1 shows the descriptive statistics for our sample. Below is a list of variables and measures used:

- *Male*: Binary variable coded as 1 for males and 0 for females.
- *Speaks English*: Binary variable coded as 1 for refugees who speak English at the time of arrival and 0 otherwise.
- *Age at arrival*: Age at arrival measured in years.
- *Education*: Highest level of educational attainment at arrival. Categories include: None/Unknown, Less than Secondary, Secondary, Advanced, and University.
- *Country of origin*: Country of origin or nationality.
- *Employed*: Binary variable coded as 1 for refugees who are employed at 90 days after arrival, and 0 otherwise.
- *Year of arrival*: Year of arrival (continuous).
- *Month of arrival*: Month of arrival (continuous).
- *Free case*: Binary variable coded as 1 for refugees who are free cases with no U.S. ties, and 0 otherwise.

D.3 Applying the Mechanism

We applied our mechanism to the data on the refugee families who arrived in the third quarter (Q3) of 2016, specifically focusing on refugees who were free to be assigned to different resettlement locations (561 families, 919 working-age individuals). To generate each family’s outcome score vector across each of the locations, we employed the same methodology in Bansak et al. (2018), using the data for the refugees who arrived from 2011 up to (but not including) 2016 Q3 to train gradient boosted tree models that predict the expected employment success of a family (i.e. the mean probability of finding employment among working-age members of the family) at any of the locations, as a function of their background characteristics. These models were then applied to the families who arrived in 2016 Q3 to generate their predicted employment success at each location, which comprise their outcome score vectors.

To generate preference rank vectors, we infer revealed location preferences from secondary migration behavior. Specifically, we use the same modeling procedures used in the outcome score estimation, simply swapping in out-migration in place of employment as the response variable. This allows us to predict for each refugee family that arrived in 2016 Q3 the probability of out-migration at each location as a function of their background characteristics. For each family, we then rank locations such that the location with the lowest (highest) probability of out-migration is ranked first (last).

In applying our mechanism to the 2016 Q3 refugee data, we impose real-world assignment constraints, giving each location capacity for the same number of families as were sent to those locations in actuality. We also truncate each family’s preference rank vector such that only the first 10 ranks are retained and indifference is established among the remaining locations. Figure 3 in the main text displays the results. In addition, Figure S2 in this SI shows the results of the same simulations when the preference rank vectors are not truncated.

More details on the procedures used to generate the outcome score and preference rank vectors can be found below.

D.4 Generating Outcome Scores and Preference Ranks

The methods used for estimating the predicted probabilities of employment and out-migration in this study are the same as those employed in ?. The following material describes the procedures and is modified directly from the Supplementary Materials document of ?.

D.5 Training vs. Prediction Data Designation

Let \mathbf{T} (training data) be the matrix of refugee data, in which the unit of observation is a single refugee, that will be used for model training. The \mathbf{T} matrix contains the data for all working age refugees in our data who arrived starting in 2011 and up to (but not including) the third quarter of 2016. For each refugee we observe her assigned location,

response variables of interest (employment for the outcome score and out-migration for the preference rank), and her full set of covariates.

Let \mathbf{R} (prediction data) be the matrix of data for the working age, free case refugees who arrived during the third quarter of 2016. This comprises the set of refugees to whom we applied our mechanism in this application. In a real-world application, these \mathbf{R} matrix data would correspond to new refugee arrivals and must include the same set of covariates as in the model training data. In contrast to the model training data, however, these prediction data need not include refugees’ response variables. In fact, in a real-world prospective implementation of the mechanism, refugees belonging to these prediction data will not have yet been assigned to a resettlement location.

Note that when applying our mechanism both the model training and prediction data should be subsetting to the group of refugees for whom the outcomes of interest are relevant. In our application the integration outcome is employment and therefore the population of interest is working-age refugees. In addition, the prediction data should be subsetting only to those refugees who are free to be assigned to different resettlement locations—in contrast to refugees with predetermined geographic destinations due to family ties and other special circumstances—as this is the subset for whom the mechanism is designed to help with the assignment process. That said, the model training data need not be restricted to only free cases. Free-case and non-free-case refugees might be sufficiently dissimilar that forecasting free-case refugees’ outcomes with models built using non-free-case data may seem problematic. This issue is addressed, however, by including case type as a predictor variable in the model building process (see below).

D.6 Modeling

The training data is used to build a bundle of learners that predict refugees’ probabilities of the response variables (employment and out-migration), and those learned models are then applied to the prediction data to generate their predicted probabilities.

The modeling is implemented on a location-by-location basis. For each resettlement location, the training data are first subsetting to those refugees who were assigned to that location, and a statistical model is then fit that uses those refugees’ characteristics to predict the response. That fitted model is then applied to the prediction data (2016 Q3 refugees) to predict the probability of the response for these refugee arrivals if they were hypothetically sent to the location in question. This process is performed separately for each individual location, which yields for each refugee in the prediction data a vector of predicted probabilities, one for each location. Collectively for all refugees in the prediction data, the final result is then a matrix of predicted probabilities (\mathbf{M} matrix) with rows representing individual refugees and columns representing resettlement locations. Note that there are two \mathbf{M} matrices: one for probabilities of employment and one for probabilities of out-migration.

More formally, for each refugee $r = 1, \dots, n_T$, let the response of interest (e.g. employment) be denoted by $y_r \in \{0, 1\}$ and the location assignment denoted by $w_r \in \{1, \dots, k\}$, for a total of k possible resettlement locations. Let \vec{x}_r denote a p -dimensional feature

vector comprised of the characteristics of refugee r , and x_{rm} denote the m th feature in \vec{x}_r , where $m = 1, \dots, p$. The goal of the modeling process is to learn the function $\theta_l(\vec{x}_r) = P(y_r = 1 | \vec{x}_r, w_r = l)$. The following describes the steps in the modeling stage.

1. Designate the historical model training data and denote it by the matrix \mathbf{T} :

$$\mathbf{T} = \begin{bmatrix} y_1 & w_1 & x_{11} & \cdots & x_{1m} & \cdots & x_{1p} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ y_r & w_r & x_{r1} & \cdots & x_{rm} & \cdots & x_{rp} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ y_{n_T} & w_{n_T} & x_{n_T1} & \cdots & x_{n_Tm} & \cdots & x_{n_Tp} \end{bmatrix}$$

2. Train a set of k models, $\Theta = \{\hat{\theta}_1(\vec{x}_r), \dots, \hat{\theta}_l(\vec{x}_r), \dots, \hat{\theta}_k(\vec{x}_r)\}$ as follows.
For $l = 1, \dots, k$:

- (a) Subset \mathbf{T} to refugees for whom $w_r = l$ (i.e. refugees assigned to l -th location), and call this \mathbf{T}_l :

$$\mathbf{T}_l = \begin{bmatrix} y_1 & x_{11} & \cdots & x_{1m} & \cdots & x_{1p} \\ \vdots & \vdots & & \vdots & & \vdots \\ y_r & x_{r1} & \cdots & x_{rm} & \cdots & x_{rp} \\ \vdots & \vdots & & \vdots & & \vdots \\ y_{n_l} & x_{n_l1} & \cdots & x_{n_lm} & \cdots & x_{n_lp} \end{bmatrix}_{w=l} = \begin{bmatrix} y_1 & \vec{x}_1 \\ \vdots & \vdots \\ y_r & \vec{x}_r \\ \vdots & \vdots \\ y_{n_l} & \vec{x}_{n_l} \end{bmatrix}_{w=l}$$

where n_l denotes the number of refugees for whom $w_r = l$.

- (b) Using the data in \mathbf{T}_l (the outcome y_r and feature vector \vec{x}_r for all n_l refugees in \mathbf{T}_l), model and estimate the function $\hat{\theta}_l(\vec{x}_r)$.

3. Designate the data on new refugee arrivals and denote them by the matrix \mathbf{R} :

$$\mathbf{R} = \begin{bmatrix} \dot{x}_{11} & \cdots & \dot{x}_{1m} & \cdots & \dot{x}_{1p} \\ \vdots & & \vdots & & \vdots \\ \dot{x}_{r1} & \cdots & \dot{x}_{rm} & \cdots & \dot{x}_{rp} \\ \vdots & & \vdots & & \vdots \\ \dot{x}_{n_R1} & \cdots & \dot{x}_{n_Rm} & \cdots & \dot{x}_{n_Rp} \end{bmatrix} = \begin{bmatrix} \vec{x}_1 \\ \vdots \\ \vec{x}_r \\ \vdots \\ \vec{x}_{n_R} \end{bmatrix}$$

where n_R denotes the number of new refugee arrivals.

The matrix \mathbf{R} corresponds to the 2016 Q3 refugees in this application.

4. For all refugees in \mathbf{R} and all resettlement locations, estimate $P(\dot{y}_r = 1 | \vec{x}_r, \dot{w}_r = l)$ as follows.

For $r = 1, \dots, n_R$:

For $l = 1, \dots, k$:

Estimate $P(\dot{y}_r = 1 | \vec{x}_r, \dot{w}_r = l)$ by applying l th model in Θ to \vec{x}_r :

$$\hat{P}(\dot{y}_r = 1 | \vec{x}_r, \dot{w}_r = l) = \hat{\theta}_l(\vec{x}_r) \equiv \pi_{rl}$$

Arrange the π_{rl} into a vector, $\vec{\pi}_r = [\pi_{r1}, \dots, \pi_{rk}]$.

5. Produce a matrix of predicted probabilities, with rows corresponding to new refugees and columns corresponding to resettlement locations, as follows.

Arrange vectors $\vec{\pi}_r$ into rows of the matrix \mathbf{M} :

$$\mathbf{M} = \begin{bmatrix} \vec{\pi}_1 \\ \vdots \\ \vec{\pi}_r \\ \vdots \\ \vec{\pi}_{n_R} \end{bmatrix} = \begin{bmatrix} \pi_{11} & \cdots & \pi_{1l} & \cdots & \pi_{1k} \\ \vdots & & \vdots & & \vdots \\ \pi_{r1} & \cdots & \pi_{rl} & \cdots & \pi_{rk} \\ \vdots & & \vdots & & \vdots \\ \pi_{n_R1} & \cdots & \pi_{n_Rl} & \cdots & \pi_{n_Rk} \end{bmatrix}$$

This is the final modeling stage output.

We follow Bansak et al. (2018) and use boosted trees (Friedman et al., 2009, Friedman, 2001) to estimate $\hat{\theta}_l(\vec{x}_r)$ in step 2(b). See Bansak et al. (2018) for more details on the selection criteria and model performance metrics leading to the choice of boosted trees. Specifically, we use stochastic gradient boosted trees (bag fraction of 0.5) with a binomial deviance loss function (Friedman, 2002, Friedman et al., 2009), which we implemented in R using the `gbm` package (Ridgeway, 2017). Tuning parameter values, including the interaction depth, learning rate, and number of boosting iterations (the early stopping point) are selected via cross-validation within the training data for each location-specific model.

We use the following predictors: *Free case*, *Speaks English*, *Age at arrival*, *Male*, *Education* (ordered variable differentiating between no/unknown education, less than secondary, secondary, technical/professional, and university), *Country of origin* (one binary variable for each of the largest origin groups including Burma, Iraq, Bhutan, Somalia, Afghanistan, Democratic Republic of Congo, Iran, Eritrea, Ukraine, Syria, Sudan, Ethiopia, and Moldova), *Year of arrival*, and *Month of arrival*.

D.7 Mapping to Case-Level

Since the assignment of refugees typically takes place at the level of the case (typically a family), we need to map the refugee-level predicted probabilities from the modeling process to a case-level metric. For each case-location pair, we apply the mapping function to the refugee-location predicted probabilities for all refugees belonging to that case, yielding a single value for that case-location pair. This results in a new matrix (\mathbf{M}^* matrix) with the same number of columns (locations) as previously but now as many rows as cases rather than refugees.

Formally, let $i = 1, \dots, n$ denote the refugee case, with a total of n cases, where $n \leq n_R$. The mapping process then proceeds as follows:

1. Perform mapping of individual predicted probabilities to case-level metric as follows.

For $i = 1, \dots, n$:

For $l = 1, \dots, k$:

Let $\tilde{\pi}_{il} = \{\pi_{rl} \mid \forall r \in i\}$. (That is, $\tilde{\pi}_{il}$ is the set of all π_{rl} for the l th location and refugees belonging to the i th case.)

Compute $\gamma_{il} = \psi(\tilde{\pi}_{il})$ where ψ is a predetermined mapping function.

Arrange the γ_{il} into a vector, $\vec{\gamma}_i = [\gamma_{i1}, \dots, \gamma_{ik}]$.

2. Produce a matrix containing the case-level metric for all case-location pairs, with rows corresponding to cases and columns corresponding to resettlement locations, as follows.

Arrange vectors $\vec{\gamma}_i$ produced in step 1 into rows of the matrix \mathbf{M}^* :

$$\mathbf{M}^* = \begin{bmatrix} \vec{\gamma}_1 \\ \vdots \\ \vec{\gamma}_i \\ \vdots \\ \vec{\gamma}_n \end{bmatrix} = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1l} & \cdots & \gamma_{1k} \\ \vdots & & \vdots & & \vdots \\ \gamma_{i1} & \cdots & \gamma_{il} & \cdots & \gamma_{ik} \\ \vdots & & \vdots & & \vdots \\ \gamma_{n1} & \cdots & \gamma_{nl} & \cdots & \gamma_{nk} \end{bmatrix}$$

This is the final mapping stage output.

In step 1, the function ψ must be specified. In our application, we employ the mean for both the predicted probabilities of employment and the predicted probabilities of out-migration (see Bansak et al. (2018) for alternative choices).

D.8 Final Construction of Outcome Scores and Preference Ranks

The \mathbf{M}^* matrix pertaining to the predicted probabilities of employment directly provides the outcome scores for use in the mechanism. However, the \mathbf{M}^* matrix pertaining to the predicted probabilities of out-migration must be further transformed to provide the (inferred) preference ranks. Specifically, for each row (case), we rank locations such that the location with the lowest (highest) average probability of out-migration is ranked first (last), producing a preference rank vector for each case.

E Tables

Table S1: Descriptive Statistics for United States Refugee Sample

	Mean	SD
Male	0.53	0.50
Speaks English	0.42	0.49
Age:		
18-29	0.44	0.50
30-39	0.28	0.45
40-49	0.16	0.37
50+	0.11	0.31
Education:		
None/Unknown	0.18	0.39
Less than Secondary	0.39	0.49
Secondary	0.21	0.41
Advanced	0.10	0.30
University	0.12	0.33
Origin:		
Burma	0.23	0.42
Iraq	0.20	0.40
Bhutan	0.13	0.34
Somalia	0.11	0.31
Afghanistan	0.07	0.25
Other	0.26	0.44
Employed	0.23	0.42

Sample consists of refugees of working age that were resettled by one of the largest resettlement agencies and arrived in the period from quarter 1, 2011 to quarter 3, 2016. N = 33,782.

F Figures

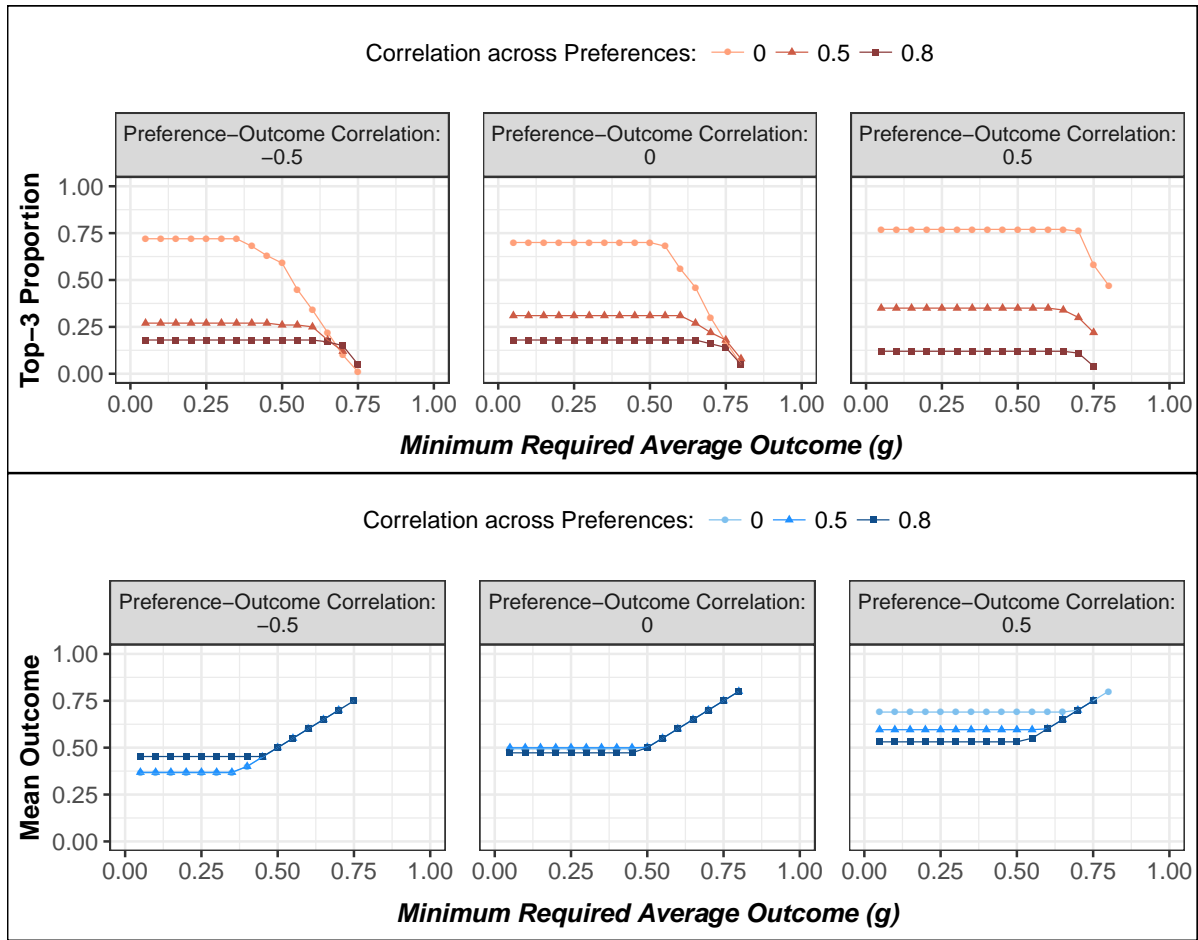


Figure S1: Results from applying our assignment mechanism to simulated data (without truncated preferences) that varies the correlations between location preference and integration outcome vectors and the correlations between preference vectors across families. This figure shows the results of the same simulations as in the main text Figure 1, except that the simulated families' preference rank vectors were not truncated in the simulations illustrated here. Upper panel shows the average probability that a family was assigned to one of its top three locations. Lower panel shows the realized average integration outcomes, i.e. the average projected probability of employment. $N = 100$.

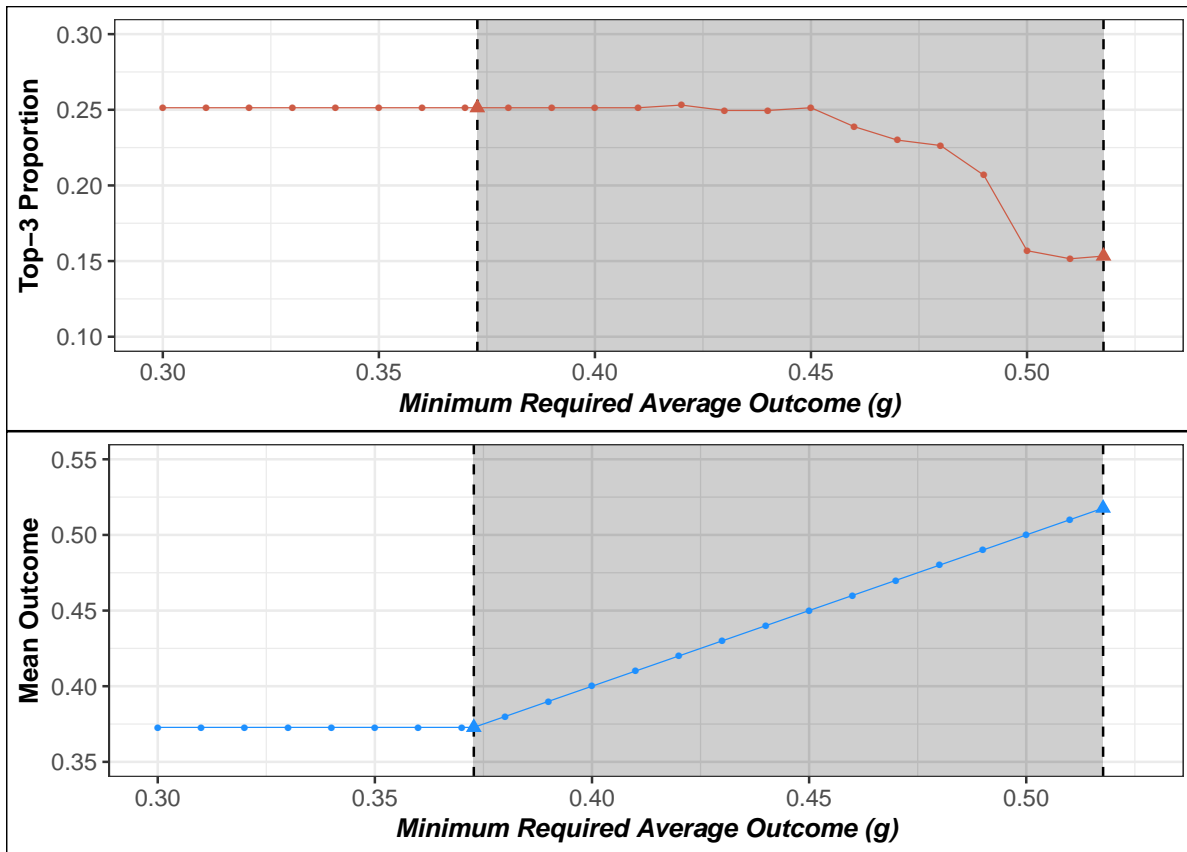


Figure S2: Results of applying the assignment mechanism to refugee families in the United States (without truncated preferences) for various specified thresholds for the expected minimum level of average integration outcomes (g). This figure shows the results of applying the mechanism to the same data as in the main text Figure 3, except that the families' preference rank vectors were not truncated in the application illustrated here. Upper panel shows the average probability that a refugee got assigned to one of their top three locations. Lower panel shows the realized average integration outcomes, i.e. the average projected probability of employment. $N = 561$ families who arrived in Q3 of 2016.

References

- ABDULKADIROĞLU, A., P. A. PATHAK, AND A. E. ROTH (2009): “Strategy-proofness versus efficiency in matching with indifferences: Redesigning the NYC high school match,” *American Economic Review*, 99, 1954–78.
- ABDULKADIROĞLU, A. AND T. SÖNMEZ (2003): “School choice: A mechanism design approach,” *American economic review*, 93, 729–747.
- ANDERSSON, T. AND L. EHLERS (2016): “Assigning refugees to landlords in Sweden: Stable maximum matchings,” Tech. rep., Working Paper, Lund University.
- ÅSLUND, O. AND D.-O. ROTH (2007): “Do when and where matter? Initial labour market conditions and immigrant earnings,” *The Economic Journal*, 117, 422–448.
- BANSAK, K., J. FERWERDA, J. HAINMUELLER, A. DILLON, D. HANGARTNER, D. LAWRENCE, AND J. WEINSTEIN (2018): “Improving refugee integration through data-driven algorithmic assignment,” *Science*, 359, 325–329.
- BERTSEKAS, D. P. AND P. TSENG (1994): *RELAX-IV: A faster version of the RELAX code for solving minimum cost flow problems*, Massachusetts Institute of Technology, Laboratory for Information and Decision Systems Cambridge, MA.
- BURKARD, R., M. DELL’AMICO, AND S. MARTELLO (revised reprint, 2012): *Assignment problems*, Society for Industrial and Applied Mathematics.
- COMMISSION, E. (2016): “Action Plan on the integration of third country nationals,” *Brussels, 7.6.2016 COM(2016) 377 final*.
- DAMM, A. P. (2014): “Neighborhood quality and labor market outcomes: Evidence from quasi-random neighborhood assignment of immigrants,” *Journal of Urban Economics*, 79, 139–166.
- DELACRÉTAZ, D., S. D. KOMINERS, AND A. TEYTELBOYM (2016): “Refugee resettlement,” Tech. rep., Working Paper, University of Melbourne.
- FERNÁNDEZ-HUERTAS MORAGA, J. AND H. RAPOPORT (2015): “Tradable refugee-admission quotas and EU asylum policy,” *CESifo Economic Studies*, 61, 638–672.
- FRIEDMAN, J. H. (2001): “Greedy function approximation: a gradient boosting machine,” *Annals of statistics*, 29, 1189–1232.
- (2002): “Stochastic gradient boosting,” *Computational Statistics & Data Analysis*, 38, 367–378.
- FRIEDMAN, J. H., T. HASTIE, AND R. TIBSHIRANI (2009): *The elements of statistical learning, 2nd ed.*, Springer.

- GÖLZ, P. AND A. D. PROCACCIA (2018): “Migration as Submodular Optimization,” *arXiv preprint arXiv:1809.02673*.
- HANSEN, B. B. AND S. O. KLOPFER (2006): “Optimal full matching and related designs via network flows,” *Journal of computational and Graphical Statistics*, 15, 609–627.
- KUHN, H. W. (1955): “The Hungarian method for the assignment problem,” *Naval Research Logistics (NRL)*, 2, 83–97.
- LINDERT, J., M. G. CARTA, I. SCHÄFER, AND R. F. MOLLICA (2016): “Refugees mental health? A public mental health challenge,” *The European Journal of Public Health*, 26, 374–375.
- MILGROM, P. R. AND S. TADELIS (2018): “How Artificial Intelligence and Machine Learning Can Impact Market Design,” Tech. rep., National Bureau of Economic Research.
- MORAGA, J. F.-H. AND H. RAPOPORT (2014): “Tradable immigration quotas,” *Journal of Public Economics*, 115, 94–108.
- MOUSA, S. (2018): “Boosting Refugee Outcomes: Evidence from Policy, Academia, and Social Innovation,” *Academia, and Social Innovation (October 2, 2018)*.
- MUNKRES, J. (1957): “Algorithms for the assignment and transportation problems,” *Journal of the society for industrial and applied mathematics*, 5, 32–38.
- RIDGEWAY, G. (2017): “Package ‘gbm,’” Tech. rep., Comprehensive R Archive Network (CRAN), <https://cran.r-project.org/web/packages/gbm/gbm.pdf>.
- ROTH, A. E. (2018): “Marketplaces, markets, and market design,” *American Economic Review*, 108, 1609–58.
- SATTERTHWAITE, M. A. AND H. SONNENSCHNEIN (1981): “Strategy-proof allocation mechanisms at differentiable points,” *The Review of Economic Studies*, 48, 587–597.
- UNITED NATIONS (2018): “Report of the United Nations High Commissioner for Refugees,” Tech. rep.