NORCE Norwegian Research Centre AS
www.norceresearch.no

The Feasibility of Using Data-Driven Algorithmic Recommendations for Refugee Placement in Norway

Authors:

Jeremy Ferwerda (IPL, Dartmouth College), Henning Finseraas (NTNU), Dag Arne Christensen (NORCE)

Rapport 15-2022, NORCE Health & Social Sciences

Report title	The Feasibility of Using Data-Driven Algorithmic Recommendations for Refugee Placement in Norway
Project No	[Prosjektnummer]
Institution	NORCE Health & Social Sciences
Client	IMDI
Classification:	Open
Report No.	15-2022
ISBN	978-82-8408-224-0
No. of pages	21
Date of publ.:	[Publiseringsmnd.]
CC-licence	CC BY 4.0
Citation	[Sitering]
Keywords	[Stikkord]

Summary

A growing body of research suggests that refugees' initial settlement area can have a long-run impact on subsequent integration outcomes. As a result, matching refugees and asylum seekers to initial locations where they are likely to succeed holds the potential to improve their labor market integration. In this report we focus on the GeoMatch algorithm, which is a recommendation tool that provides settlement officers with data-driven location recommendations for incoming refugees and asylum seekers. Leveraging machine learning on historical data, the tool predicts labor market outcomes for individuals across possible settlement areas. A flexible allocation algorithm then provides location recommendations for each family unit while taking capacity constraints into account. Drawing on administrative data from Statistics Norway and incorporating a set of realistic constraints, we find that using GeoMatch recommendations could improve refugees' monthly earnings by up to 55% over baseline. The report ends with a discussion of how the tool can be implemented in the Norwegian context.

Innhold

Figures

1. Abstracts

A growing body of research suggests that refugees' initial settlement area can have a long-run impact on subsequent integration outcomes. As a result, matching refugees and asylum seekers to initial locations where they are likely to succeed holds the potential to improve their labor market integration. In this report we focus on the GeoMatch algorithm, which is a recommendation tool that provides settlement officers with data-driven location recommendations for incoming refugees and asylum seekers. Leveraging machine learning on historical data, the tool predicts labor market outcomes for individuals across possible settlement areas. A flexible allocation algorithm then provides location recommendations for each family unit while taking capacity constraints into account. Drawing on administrative data from Statistics Norway and incorporating a set of realistic constraints, we find that using GeoMatch recommendations could improve refugees' monthly earnings by up to 55% over baseline. The report ends with a discussion of how the tool can be implemented in the Norwegian context.

Sammendrag (Norwegian)

Nyere forskning tyder på at flyktningers opprinnelige bosettingssted kan påvirke integreringsresultater på lang sikt. Det innebærer at en bedre match av flyktninger/asylsøkere og bosettingssted har potensiale til å forbedre arbeidsmarkedsintegrasjonen. I denne rapporten studerer vi dette ved hjelp av GeoMatch-algoritmen, som er et anbefalingsverktøy som gir bosettingsansvarlige datadrevne lokaliseringsanbefalinger for innkommende flyktninger og asylsøkere. Ved å utnytte maskinlæring på historiske data, predikerer verktøyet arbeidsmarkedsresultater for enkeltpersoner på tvers av mulige bosettingsområder. En fleksibel allokeringsalgoritme gir deretter plasseringsanbefalinger for hver familieenhet samtidig som det tas hensyn til kapasitetsbegrensninger. Ved å trekke på administrative data fra Statistisk sentralbyrå og inkorporere et sett med realistiske begrensninger, finner vi at bruk av GeoMatch-anbefalinger kan forbedre flyktningers månedlige inntekter med opptil 55% over baseline. Rapporten avsluttes med en diskusjon om hvordan verktøyet kan implementeres i norsk kontekst.

2. Introduction

Improving refugees' labor market integration is a key policy goal. One policy tool available to Norwegian policy makers with respect to improving integration consists of optimizing the choice of the initial settlement location for refugees and asylum seekers. In the first report in this project, we interviewed settlement officers and found that they often feel they lack information to make well-informed settlement decisions. They have therefore asked for tools to improve the decision-making process. In this second report, we demonstrate the potential of using an algorithmic recommendation tool (GeoMatch) to meet this request.

The GeoMatch technology is developed by the Immigration Policy Lab (IPL), an academic organization with branches at Stanford University and ETH Zurich that focuses on innovation in immigration policy. The underlying approach, which is outlined in Bansak et al. (2018), uses historical data on settlements and labor market outcomes to predict labor market integration in each settlement location, and then applies an allocation algorithm to generate locational recommendations for individual cases. If implemented, settlement officers will be able to use the GeoMatch tool to receive data-driven location recommendations for incoming cases, which can be considered wholistically as part of the existing decision process.

Using historical administrative data, this report conducts a retrospective impact evaluation to assess the potential of the GeoMatch tool. Below we explain in detail how we have conducted the test, the results, and associated limitations.

We predict substantial gains in refugee earnings and employment from using GeoMatch. The evaluation using historical data suggests that labor market earnings could be improved by approximately 11 000 to 17 000 NOK per month, depending on the restrictions applied in the allocation process. As explained below, these predictions are based on settlement in labor market regions and not in municipalities, as municipalities settle too few cases to provide accurate predictions based on historical data.

Given these promising results, we are confident that the GeoMatch tool could be implemented in Norway. However, we suggest that the recommendation tool should be tested in a small scale, and on a systematic manner on new arrivals before being fully implemented.

3. Data driven algorithmic recommendations

IPL has developed an algorithm called GeoMatch that identifies synergies between refugees' background characteristics, their settlement location, and their integration outcomes (e.g. employment, earnings, or other available outcomes) and then uses these patterns to generate optimal matches for incoming individuals and families.

The first step uses machine learning on historical data to predict incoming cases' economic integration at each possible location, based on the outcomes of similar, previously placed cases, and then recommends the location where the probability of successful integration is highest. The matching algorithm then incorporates constraints based on pre-existing location-specific quotas so that only a specific number of individuals is recommended to be settled in each location within a particular assignment period (i.e. month or year). If available, additional case-level constraints can be implemented as needed (such as matching individuals with health conditions to areas with hospitals).

The matching algorithm is written in the programming language R, but can be combined with user-friendly front-end software that allows program supervisors to input information about refugee cases, receive placement recommendations for each case, and decide whether to accept or override those recommendations. The Swiss State Secretariat for Migration is currently using the GeoMatch tool to help their placement officers assign asylum seekers to Swiss Cantons in the context of a large-scale pilot test.

Rich data availability on refugees and asylum seekers in Norway suggests that algorithmically assisted recommendations could be a viable opportunity for improving integration outcomes. In this section, we outline the retrospective evaluation applied to the Norwegian context. All code we have used to produce the results is reported as a zipped file attached to the delivery of the report.

3.1. Data

Our main data source is administrative data from Statistics Norway. We combine information from several registers to create panel data of the complete population of refugees and asylum seekers. The data includes variables such as country of origin, age, gender, initial municipality of residence, and labor market outcomes. The appendix provides a full list of the variables we employ.

In addition, we rely on publicly available data on municipal characteristics that we use as predictors of labor market integration. This includes the local unemployment level, median earnings for immigrants (constructed from admin data), and the age composition of the population. Finally, to calculate constraints, we use IMDi's data on actual historical assignments, as well as information on how many refugees and asylum seekers municipalities decided to settle in each year.

3.2. Target population

The target population is comprised of refugees and asylum seekers who are granted residence permits and are settled in a municipality through the regular settlement procedure. In building our predictive models, we only consider data and outcomes for adults (18 years or older). We exclude family reunion arrivals since such these individuals are sent to pre-determined locations within Norway. We further exclude a small number of refugees that had such severe health issues that they received disability benefit within one year after arrival.

3.3. Modeling

The modeling approach is based on the methodology developed in Bansak et al. (2018) and Bansak and Paulson (2022). Using the data sources described above, we merge the historical data for the target population's background characteristics, economic outcomes, and geographic locations. Using supervised machine learning on the merged data (the "training dataset"), we fit separate models across each geographic location in order to determine the characteristics of individuals who are most likely to succeed within each location. After predicting integration outcomes at each location, we derive location rankings for each individual, which are then aggregated to the family unit level (i.e., maximizing average expected outcomes for the full family unit). Next, an algorithm recommends a specific location for each family unit. Since not all refugees can be placed in their highest ranked location due to capacity constraints, the algorithm seeks to optimally match cases to locations in a manner that improves outcomes across the full cohort of refugees.

The specific algorithm we use is a dynamic optimal algorithm (Bansak and Paulson 2022), which further takes into account that incoming refugees are being settled on a rolling basis when placement decisions are made, i.e. they are not all settled at once. The recommendations returned in the retrospective impact evaluation are thus assigned in the sequential order that cases were actually processed in. To avoid too many top ranked positions being recommended to those placed early in an assignment window (given that yearly quotas are not yet filled), the dynamic algorithm estimates whether it is likely that later arrivals will be a better fit for a particular settlement location. If so, it will take this into account when allocating early arrivals in order to improve outcomes across all incoming refugees. An additional aspect of the algorithm focuses on load balancing – in other words, it prioritizes a relatively even distribution of refugees across locations over time to prevent overwhelming a particular location.

Additional details about the specific modeling techniques and procedures used can be found in the referenced papers, as well as the Technical Appendix.

3.4. Target geography

Selecting the appropriate target unit of geography requires balancing three imperatives: (1) generating sufficient options to provide the algorithmic procedure with as much inter-geographic variation as possible, (2) ensuring that each geographic option is associated with a sufficient amount of historical data such that accurate and effective predictive models can be trained, and (3) identifying levels or regions of geography that are administratively compatible with the underlying goals and procedures.

On this basis, we determined that the best target unit of geography in the Norwegian context is the labor market region. Municipality of arrival is too disaggregated to confidently predict outcomes because many municipalities have settled few refugees and asylum seekers, which implies that predictions will be imprecise. We also considered using (the historical 19) counties as the unit of assignment, but concluded that the geographic level is too aggregated to offer consistently reliable predictions of labor market success. Instead, we used labor market regions as unit of assignment. There are different versions of labor market regions, but we relied on Bhuller's (2009) 46 regions, which are based on commuting patterns, to derive labor market regions that better reflect true labor markets than e.g. the regional categorization used by Statistics Norway. As a result, we generate separate predictive models for each of the 46 labor market regions, and the algorithmic procedure determines the optimal labor market region for each refugee (i.e. the labor market region where they have highest chance of success, subject to the constraints we specify).

Municipalities are nested within labor market regions. Therefore, to determine the capacity of a labor market region, we aggregate the number of settlements to the labor market region by summing the municipal-level quotas. In a real-world implementation, the algorithm could also be modified to perform a two-stage recommendation; first assigning to a labor market region, and then suggesting a municipality with open capacity.

3.5. Target and additional outcomes

Since the data-driven algorithmic recommendations are designed to optimize an outcome of interest, choosing an appropriate target outcome is critical. In keeping with the core objectives of facilitating the economic integration and success of refugees, we focus on indicators of economic success. Further, in choosing the specific outcome variable to feed into the algorithm for optimization, we must balance several considerations. The first is the goal of helping individuals achieve long-term and lasting success. The second is the need to take shorter-term information into account, so that the algorithmic methods are resilient and can adapt to changes in economic conditions over time with higher velocity. Finally, the third is the need for an outcome with sufficient variation across individuals and locations for algorithmic recommendations to be viable from a statistical and mathematical perspective.

For these reasons and because our target population will likely be enrolled in the introductory program the first two years after arrival, we choose to measure our outcomes three years after assignment. We study two outcomes. Our main outcome is total labor market earnings three years after arrival (earnings) as recorded in the administrative data. The second outcome is an equivalent of full time employment. Here we follow Havnes and Mogstad (2011) and define full time employment as having earnings that are four times above the basic unit (grunnbeløpet) in the social insurance system. Together, earnings and full time employment provides a good overview of labor market integration. Both outcomes are scaled by month of arrival since the outcomes are measured at the close of the calendar year, which means that early and late arrivals have different length of residence when the outcomes are measured.

Although the selection of labor market earnings in the third year represents a balance of considerations, note that the underlying approach is flexible and can incorporate alternate measures of labor market success if needed.

3.6. Predictors

The background characteristics used as predictors in the modeling procedure are used to predict individual-level outcomes across potential locations. Hence, it is important to include all relevant predictors that have a plausible link to the target outcome. In addition, since the models are designed to be applied in a real-world implementation to refugees and asylum seekers before they have been assigned, only pre-assignment variables can be employed as predictors.

Based on data availability, the factors we include as predictors are gender, education, marriage status, children, household size, immigration category, indicators for the largest ancestry countries, and indicators for what continent the ancestry country is located in. In addition, we include the month and year of assignment as predictors, since there might be seasonal patterns in arrivals and integration outcomes. Finally, we include municipal level variables (measured during the year of arrival) that may be important predictors of integration: median earnings of immigrants in the municipality of assignment (calculated from the administrative data), municipal unemployment level, and share of young people in the municipality. Since these data points are available on a more rapid basis than earnings outcomes in the administrative data, they also make the forecasting performed by the algorithm more resilient, as it can pick up changes in local economic conditions that are not yet visible within the outcomes measured in the administrative data.

Additional information on variable coding can be found in the Technical Appendix.

3.7. Time Period

The latest earnings year available in the current administrative data is 2018. Since our outcome is measured after three years, we focus our retrospective evaluation on the 2015 cohort to ensure that the evaluation mirrors a real-world implementation as closely as possible (in which outcomes would be evaluated three years after arrival). We note, however, that the results are similar when using prior years as the test cohort. In a full implementation, the underlying models would be updated each year as soon as the latest earnings data become available. In combination with including recent economic data at the municipal level (see section above), this enables the algorithm to flexibly adjust to changing economic conditions.

3.8. Assignment and Constraints

After the predictive models have been trained on historical data, they can be used to predict the expected success of new arrivals at each possible location. In theory, this information could be used to assign each incoming individual to the best possible location. However, as explained above, the algorithmic recommendations must also respect the constraints and restrictions that exist under the status quo assignment procedure. There are several key types of assignment constraints that we take into account in the evaluation.

The first set of constraints are related to family structure. All individuals in the same family must be placed together in the same location. The second set relate to location capacity constraints. That is, cases must be recommended to be placed across labor market regions according to pre-determined capacity and proportionality guidance at the municipal level. To mirror these constraints, we assign the exact number of cases to each labor market region that actually were settled in that region during each month of 2015. In other words, our evaluation permits cases to be moved to other locations, but holds the total number of cases settled within each location constant.

As an illustration to demonstrate the flexibility of the algorithm we also relax this constraint and allow for some flexibility in the number of refugees recommended to be placed in each labor market region, in a manner proportional to the initial political decision at the municipality level (using annual data on political settlement decisions, as reported by IMDi). This increased flexibility results in higher potential gains, because the algorithm is free to recommend more cases to locations which may be associated with higher integration outcomes for a particular case.

According to existing procedures for the assignment of immigrants, there exist some “hard criteria" that pre-determine the location where a case will be placed. Beyond the family reunification constraints mentioned above, these hard criteria often relate to medical issues. A real-world implementation of the procedure would also take these hard criteria into account. Our information on health status is, however, limited in the administrative data so we cannot fully take this into account. However, as explained above, we remove the few arrivals that were on disability insurance as early as the year after arrival to avoid biasing the evaluation.

4. Results

Figure 1 evaluates how well the machine learning models were able to predict individual earnings using historical data on the synergies between individuals' background characteristics and locations. The dotted diagonal line visualizes a perfect prediction, i.e. the mean predicted outcome from the model is identical to individuals' actual income. The blue line plots the predicted outcome for each decile of the earnings variable, for which the distribution is shown at the bottom of the figure. When the blue line is below the dotted line the models overestimate an individual's earnings, while the models underestimate it when the blue line is above the line.

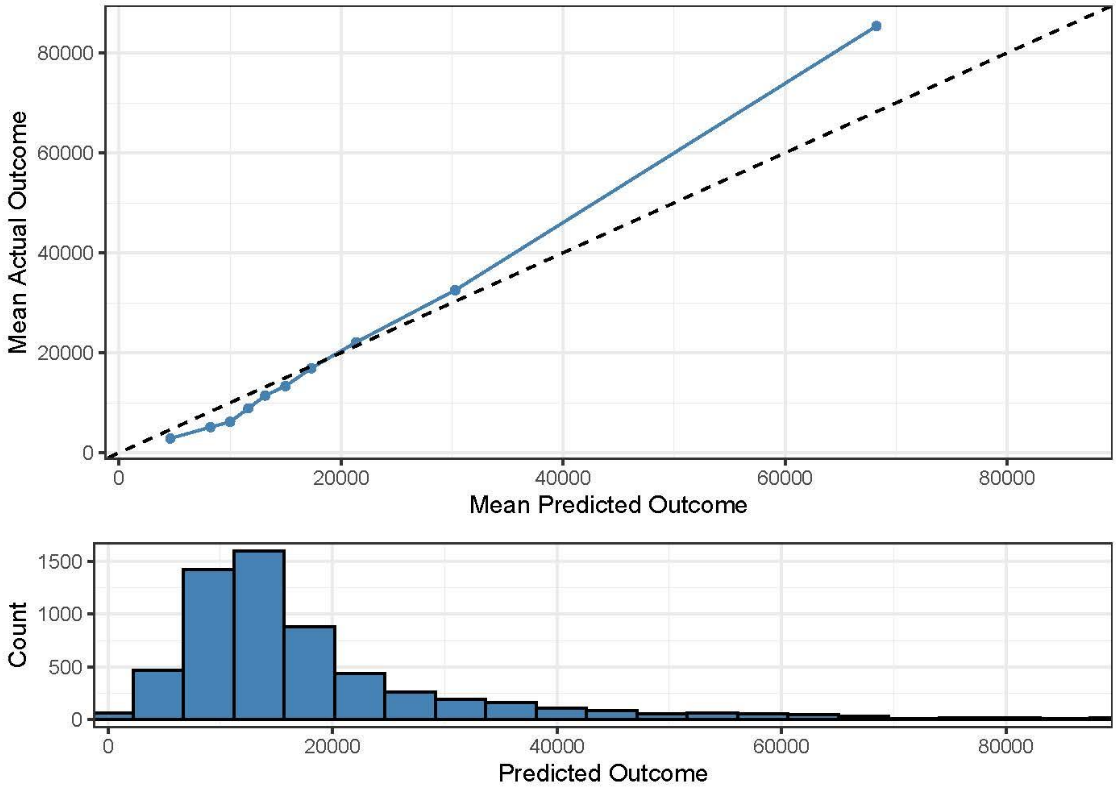

Figure 1: Prediction Accuracy for Earnings Outcome, 2015 Cohort

Although it is difficult to perfectly predict incomes due to idiosyncratic factors, the figure demonstrates that the model predictions are very close to the actual outcome for most of the earnings distribution. The only exception is the top decile (individuals who are at the 90th percentile or greater in terms of actual earnings), where earnings are underestimated. This suggests that we lack the information in the data to identify these relatively extreme cases, probably because they have abilities, experience, or other human capital that we do not observe. However, we note that the model is still able to predict which types of individuals are likely to do well (i.e. it provides a high earnings prediction for this individual), and the prediction model is very accurate in terms of actual NOK earned for about 90 percent of the population. The remaining uncertainty in the estimates is taken into account when presenting expected earnings in the following figures.

Figure 2 (next page) shows the predicted gains from applying the matching algorithm, with the restriction that regions cannot settle more refugees than they actually did in a given month in the historical data. The red bar shows the average earnings that we observe in the data, which amounts to about 20 000 NOK per month. We label this as average monthly earnings, but note that a more precise definition is annual earnings scaled by the number of months after the expected completion of the introductory program. The blue bar is the prediction based on following the GeoMatch recommendations, with the lines indicating the uncertainty in the estimate due to modeling error (discussed on the previous page). It is important to note that this estimate is based on counterfactual recommendations applied to historic data, and thus does not necessarily signify how the algorithm would perform on a prospective basis. Rather, it reflects the best available estimate based on modeling assumptions. The results suggest that following the recommendations would lead to average earnings of 31 000 NOK per month, which implies a difference of approximately 11 000 NOK, which represents a 55% increase over the baseline in the historical data.

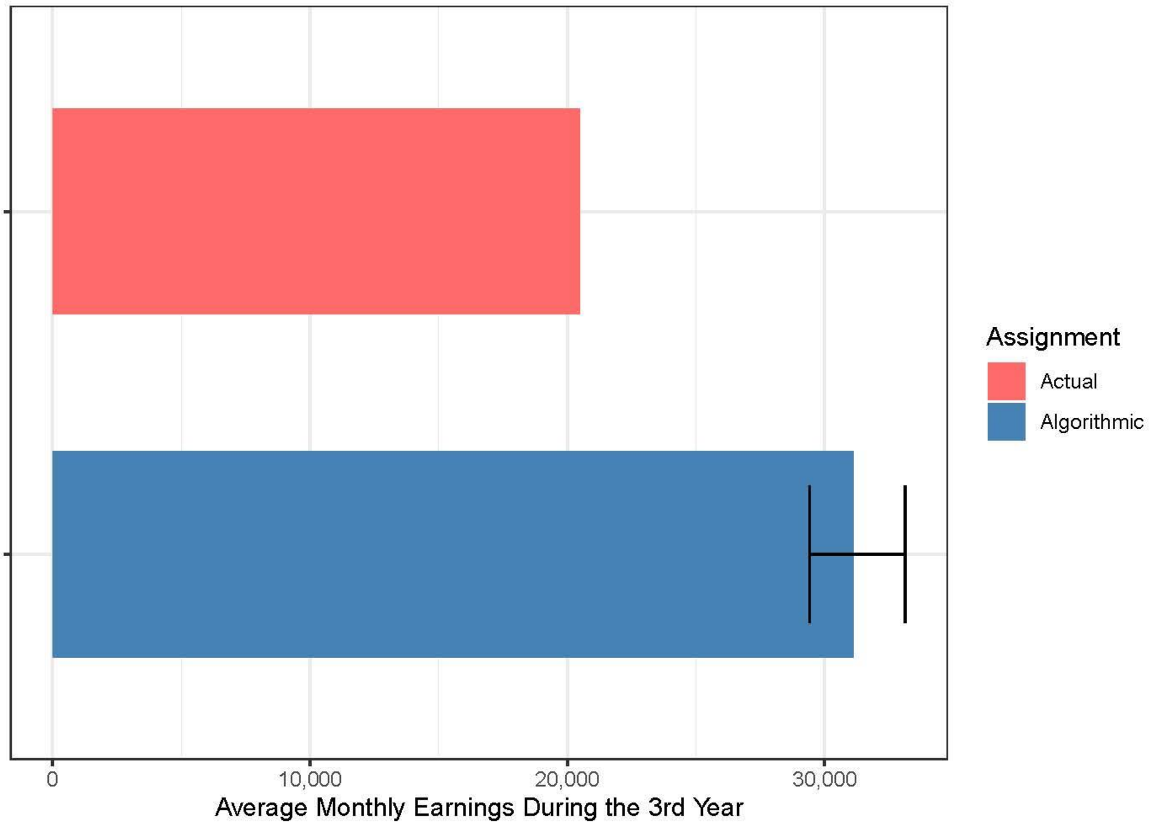

Figure 2: Potential Gains of Using Algorithmic Recommendations on Labor Market Earnings. Restriction: Regions cannot settle more refugees than they actually did in a given month.

Figure 3 replaces the earnings variable with the estimate of full time employment. The figure shows that employment increases from about 15 percent in the actual data to about 30 percent in the counterfactual scenario in which GeoMatch was used. This suggests that a large share of gains happen around the threshold we use to define full time employment.

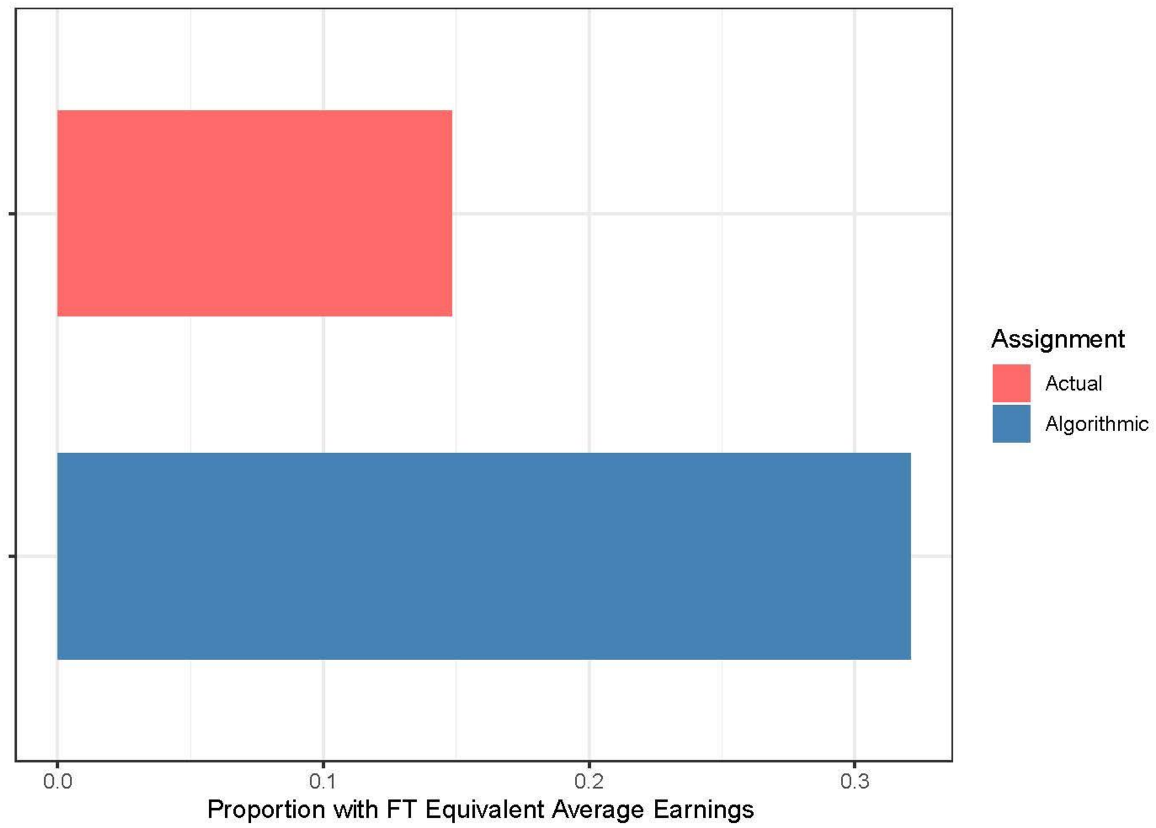

Figure 3: Potential Gains of Using Algorithmic Recommendations -- Full-time Employment using Earnings Thresholds. Restriction: Regions cannot settle more refugees than they actually did in a given month.

Next, we examine gains across different sociodemographic groups to ensure that no groups would be potentially disadvantaged by the algorithm. As we see in Figure 4, gains are not restricted to a small number of groups. While there is variation —such as gains being larger for high educated and smaller for families with children—none of these groups stand out as having expected losses from the recommendations or very small gains.

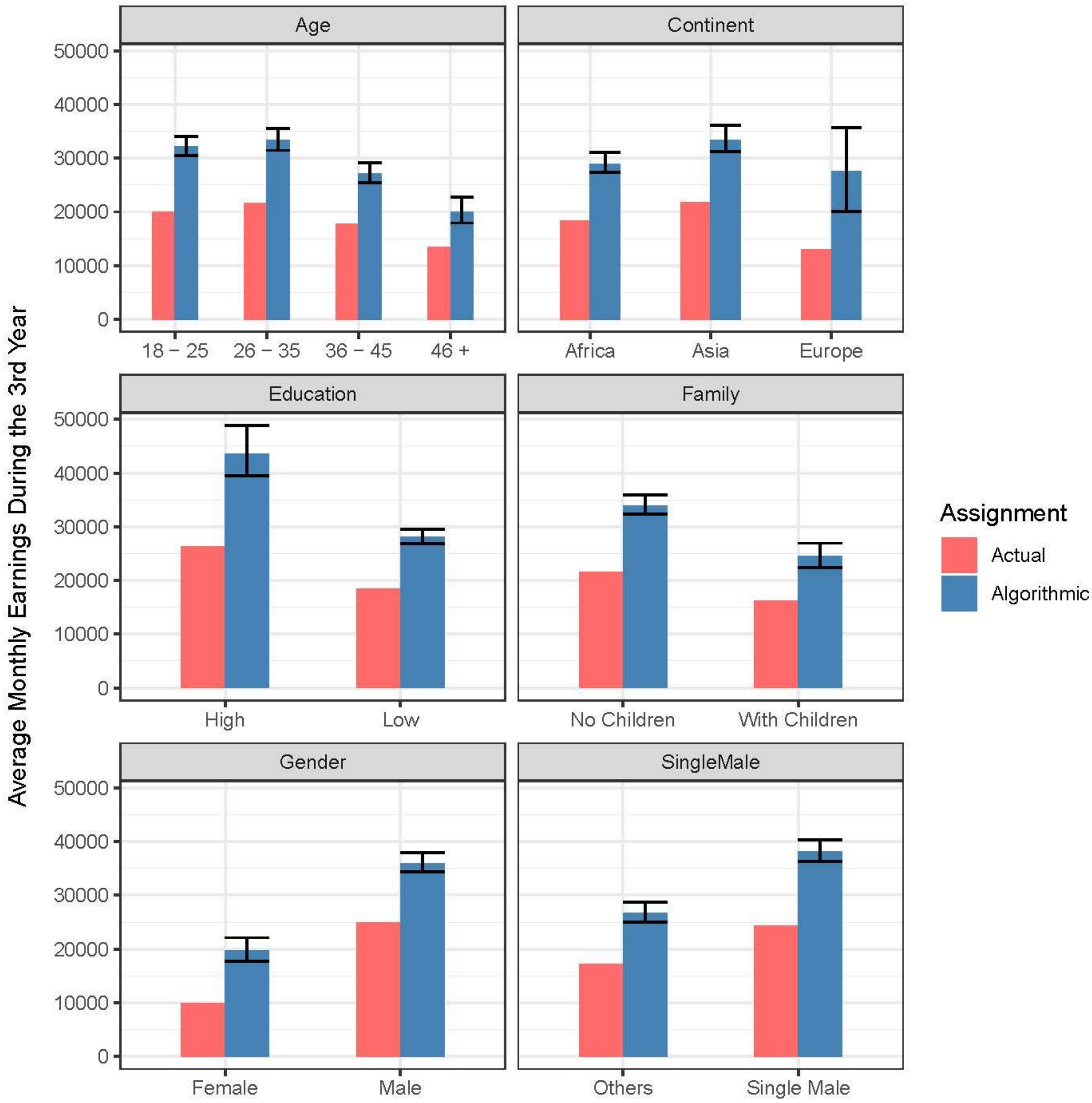

Figure 4: Potential Gains of Algorithmic Recommendations on Labor Market Earnings by Groups. Restriction: Regions cannot settle more refugees than they actually did in a given month.

One potential concern with following algorithmic recommendations is that it may result in uneven flows to optimal labor market regions early in the year or month. However, the load balancing aspect of this algorithm takes this into account by prioritizing distribution across regions within a particular assignment window. This is visible in Figure 5, which shows that the algorithm distributes recommendations across labor market regions and over time in a realistic manner (note that the large number of recommendations to Gudbrandsdalen in the middle of the year reflects a spike in the monthly historical data).

Figure 5: Flow of Refugees over the year by Labour Market Region. Restriction: Regions cannot settle more refugees than they actually did in a given month.

4.1. Flexibility

Finally, in Figures 6 and 7 we illustrate the flexibility of the algorithm by relaxing the constraint that we precisely match historical settlements in 2015. Instead, we add some flexibility to the number of assignments, proportional to the political decisions made by municipalities. As mentioned above, this is a less restrictive condition, because if the algorithm can deviate from the number of historical assignments, the algorithm has more flexibility to settle refugees in their best ranked region. Another way of thinking about this change is that it illustrates the potential gains associated with placing more weight on the considerations of refugees and asylum seekers rather than the considerations of municipalities (see discussion in Seeberg et al. 2020).

Since the restriction we apply now is less constraining, the potential gains increase, albeit not dramatically (from about 10 000 NOK in Figure 2 to about 17 000 NOK in Figure 6). Figure 7 further shows that this gain can be achieved without compromising the flow across regions. Moreover, we see that the algorithm still settles refugees across all regions, meaning that a less restrictive constraint does not violate the policy goal of settling refugees across the whole country.

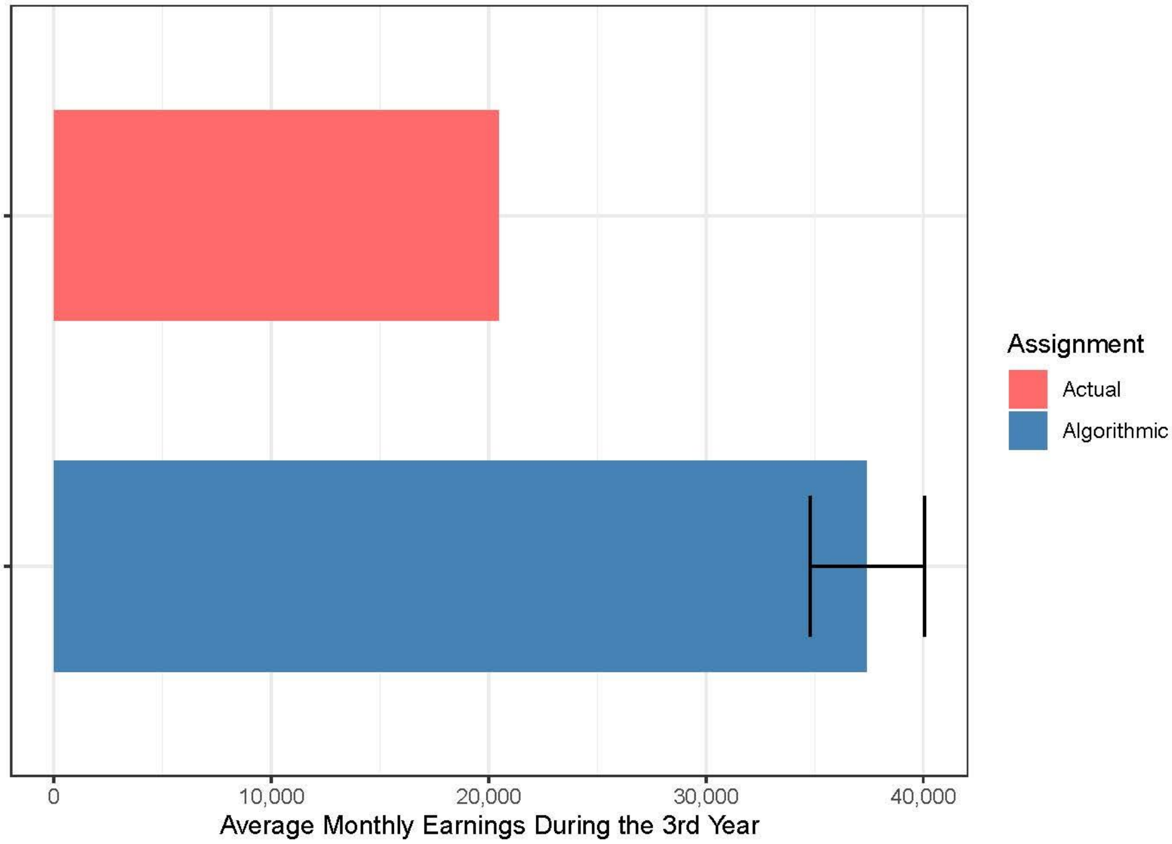

Figure 6: Potential Gains of Algorithmic Recommendations on Labor Market Earnings. Restriction: Regions cannot settle more refugees than the municipalities in the region decided to settle in a given month.

Figure 7: Flow of Refugees over the year by Labour Market Region. Restriction: Regions cannot settle more refugees than the municipalities in the region decided to settle in a given year (divided into monthly quotas).

5. Discussion

We next discuss concerns about targeted settlements raised by Seeberg et al. (2020). Part of their discussion is framed around to what extent settlement policy should put more weight on the preferences of refugees versus the preferences of municipalities. This is an interesting academic and normative discussion, but given the institutional framework in Norway, the preferences of the municipalities are nevertheless likely to prove important going forward. However, as they point out, the IPL has developed an algorithm that makes it possible to take refugee preferences into account (Bansak et al. 2022) if IMDi are asked to do so. Thus, implementing an algorithmic recommendation tool does not necessarily limit policy flexibility with respect to this trade-off.

Another concern raised by Seeberg et al. (2020) regards the transparency of algorithmic settlement. We agree that this is relevant concern, however, we are not sure the current system provides more transparency with respect to why a refugee is settled in a particular municipality. In principle, the refugee can be provided with the set of background characteristics that was used to predict integration success in particular regions and thereby provide an explanation for the placement decision.

It is also important to stress again that GeoMatch is a recommendation rather than an automatic placement tool. The tool is designed to leverage Norway's rich administrative data to provide insight into potential patterns that caseworkers may not otherwise be able to observe. However, all recommendations can be overridden by caseworkers, and the recommendations are thus best viewed as an additional informational input into the existing process rather than a replacement.

Finally, Seeberg et al. (2020) raise concerns regarding the possibility of finding optimal matches, which is driven by the presumption that gains will be concentrated in few municipalities which cannot settle all incoming cases. As explained above, the algorithm is optimal in the sense that it distributes cases to maximize overall gains. If only a few areas are optimal for refugees and asylum seekers, the total gains will be small, but this is an empirical question. In practice, we find that there is wide variation in the labor market regions that are optimal for refugees with different background characteristics.

Another issue relates to the possibility that the available data is not sufficient in terms of background characteristics or too sparse in terms of numbers to make good predictions. In this report, we have dealt with the sparsity concern by allocating refugees at the regional rather than the municipal level. As shown by Figure 1, this level of aggregation results in good predictions. This further suggests that the prediction is not hampered by a lack of pre-arrival background characteristics in the data. However, we note that the results would still need to be tested in a realized setting, and the level of aggregation may have to be adjusted if flows of refugees and asylum seekers significantly decline in the future.

5.1. Implementation

We conclude this report by outlining how the algorithm could be prospectively implemented by IMDi to provide location recommendations for new arrivals.

Although the R code provided in this report will allow IMDi to train the underlying machine learning models on historical data, the algorithm would need to be provided with information on the background characteristics of incoming cases, rather than historical data, to provide recommendations on an ongoing basis. IPL has full-time support staff that can simplify this process, and have experience in setting up a front-end interface that is designed to make GeoMatch as accessible as possible to caseworkers.

There are two options for this interface. The first is by paralleling the approach already implemented for Swiss caseworkers. In Switzerland, the trained models (from administrative data) and the GeoMatch algorithm are hosted on a secure web server in Switzerland. When making location decisions, caseworkers access a secure web interface where they manually enter the individual-level background characteristics used by the model (gender, age, family size, children etc. for all adults in the case) using drop down menus. They are then provided with a location recommendation for the case, which can be incorporated into their existing decision-making process. Alternatively, caseworkers can upload a formatted spreadsheet to rapidly process incoming cases in batch form.

The second option would consist of an internal implementation at IMDi, which is a process currently being pursued by some other countries implementing GeoMatch. In this approach, IPL would access the administrative data to train machine learning models once per year, after additional outcomes are observed in the administrative data. These models would be delivered to IMDi and could be hosted, along with the code for the algorithmic recommendation, on IMDi's servers. IMDi would then set up an internal data feed in which information on new arrivals would be sent to the algorithm. This fully hosted solution would not require manual entry or uploads by caseworkers, but would require some IT effort on the part of IMDi, albeit facilitated by IPL's in-house software developer.

Regardless of the approach used, the GeoMatch tool should be rigorously tested before a full scale implementation. The estimates described in these reports are potential gains based on historical data, and would need to be carefully tested in a real-world context. The ideal approach is to run a pilot which would provide an experimental test of the efficiency of the algorithm. This can be done by randomly allocating cases to either the standard process or to a process where GeoMatch provides a recommendation that the case worker either follows or overrules. Information on (a) whether the standard process or GeoMatch was used, and (b) whether GeoMatch's recommendation was followed, could be merged with administrative data from Statistics Norway or with survey data to examine if cases which followed the algorithm's recommendations have a better integration trajectory than cases assigned according to the status quo process. Following evaluation and potential adjustment to the algorithm, GeoMatch could then be used for the full batch of incoming cases as desired by caseworkers.

6. References

Acharya, Avidit, Kirsk Bansak, and Jens Hainmueller (2022). Combining Outcome-Based and Preference-Based Matching: A Constrained Priority Mechanism. *Political Analysis* 30(1): 89–112.

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

Bansak, Kirk and Elisabeth Paulson (2022). “*Outcome-Driven Dynamic Refugee Assignment with Allocation Balancing*” https://arxiv.org/abs/2007.03069.

Bhuller, Manudeep Singh (2009). *Inndeling av Norge i arbeidsmarkedsregioner*. Oslo, SSB 2009/24.

Fasani, Francesco, Tommaso Frattini and Luigi Minale (2021). “Lift the Ban? Initial Employment Restrictions and Refugee Labour Market Outcomes.” *Journal of European Economic Association* 19(5), 2803–2854.

Hainmueller, Jens, Dominik Hangartner and Duncan Lawrence (2016). “When lives are put on hold: Lengthy asylum processes decrease employment among refugees.” *Science Advances* 2(8):e1600432.

Harder, Niklas, Lucila Figueroa, Rachel M Gillum, Dominik Hangartner, David D Laitin and Jens Hainmueller (2018). “Multidimensional measure of immigrant integration.” *Proceedings of the National Academy of Sciences* 115(45):11483–11488.

Havnes, Tarjei and Magne Mogstad (2011). “No Child Left Behind: Subsidized Child Care and Children's Long-Run Outcomes.” *American Economic Journal: Economic Policy* 3(2): 97-129.

Marbach, Moritz, Jens Hainmueller and Dominik Hangartner (2018). “The long-term impact of employment bans on the economic integration of refugees.” *Science Advances* 4(9):eaap9519.

Seeberg, Marie Louise, Berit Aasen, Justyna Bell, Christer Hyggen and Ida Tolgensbakk (2020). Strategisk mottaksplassering og treffsikker besetting: tidlige grep for integrering av flyktninger i kommune. NOVA Rapport 2/2020.

7. Appendix

7.1. Data Sources

Admin registers used: Population, earnings, education, residency, household, civil status.

7.2. Variables

The following variables were used in modeling or prediction.

- *Male* (admin data). Binary. 1
- *Age at arrival* (admin data). Continuous. 1.
- *Low education at arrival* (admin data). Binary indicator of low/missing education. 1
- *High education at arrival* (admin data). Binary indicator of university education. 1
- *Married at arrival* (admin data). Binary indicator of marriage/cohabitation. 1
- *Children at arrival* (admin data). Binary indicator of having children. 1
- *Country of birth* (admin data). Set of binary indicators for largest ancestry country. 1
- *Household size* (admin data). Continuous. 1
- *Quota refugee* (admin data). Binary indicator of arrival category. 1
- *Continent*. Binary indicators of whether ancestry continent is Africa or Asia. 1
- *Median earnings of immigrants in the municipality* (own calculations from admin data). Continuous. 1
- *Unemployment rate in the municipality* (Fiva et al 2020). Continuous. 1
- *Share of the population younger than school age* (Fiva et al. 2020). Continuous. 1
- *Full time employed three years after arrival* (admin data). Binary indicator of having earnings that are four times the basic unit in the social insurance system. 2
- *Earnings three years after arrival* (admin data). Earnings adjusted by month of arrival. 2

1 = Variable used to train machine learning models

2 = Variable used as outcome

7.3. Modeling

Machine Learning Models

We use an individual's background characteristics to predict their probability of employment and earnings three years after assignment. However, given the specific characteristics that lead to a high probability of employment plausibly differ across labor market regions, we model each location separately.

To generate our models, we employ the supervised machine learning method called stochastic gradient boosted trees, which is implemented via a customized gradient boosting machine (gbm) package within R, a language and environment for statistical computing.

7.4. Tuning and Model Diagnostics

In our implementation of gradient boosted trees, we used validation to determine the optimal tuning parameters. Specifically, we determine the ideal number of boosting iterations (trees) as well as the ideal interaction depth.

This process works as follows. First, we choose a specific interaction depth. With this parameter fixed, we fit models over a sequence of boosting iterations (normally 1-700 trees). For each model in sequence, we use cross validation to measure the root mean square error (RMSE). To avoid potentially choosing a local minimum, if the model that minimized RMSE was within 100 trees of the maximum number of trees considered, we would re-run the model but increase the number of boosting iterations by 500. We repeat this process as many times as necessary. In the end, we record the length (number of trees) in the boosted sequence that minimized CV error. We also record the CV error value (the RMSE).

We then repeat the above process for the next interaction depth we consider. In the end, we have a set of tuned models for each interaction depth. Among these models, we choose the interaction depth value that lead to the lowest overall CV error.

The set of interaction depths we considered were:

- interaction.depth: 3-8

Additionally, we set the following gbm parameters:

- shrinkage: .01
- nminobsinnode: 5
- bag.fraction : 0.5