

# Re-evaluating the role of refugee integration factors for building more equitable allocation algorithms

CLARA STRASSER CEBALLOS\*, LMU Munich, Germany

MARCUS NOVOTNY\*, LMU Munich, Germany

CHRISTOPH KERN, LMU Munich, Germany and Munich Center for Machine Learning (MCML), Germany

Numerous studies in the social sciences have examined how individual and location-level characteristics influence refugees' integration outcomes. A more recent, smaller body of computational research has developed algorithmic tools that aim to improve refugee integration by optimizing matching to resettlement locations based on predicted outcomes. These tools, which are piloted in a number of countries, raise a number of concerns. This includes, first, their reliance on a narrow set of individual-level predictors – most of which are protected attributes under global anti-discrimination laws – overlooking valuable insights from migration studies that may improve predictive accuracy. Second, they guide refugee placement decisions without assessing group fairness, potentially reinforcing existing inequalities. Against this background, we draw on comprehensive refugee panel data from Germany and study the economic integration of refugees through the lens of predictive modeling. Specifically, we develop prediction models that integrate and test a wide range of integration factors from migration research. We then compare our extended model configurations with existing refugee-location matching algorithms, and evaluate group model performance to assess generalizability and fairness. Overall, we highlight the importance of integrating insights from migration studies into the development of algorithmic decision-making tools to improve their reliability and promote fair outcomes across diverse groups.

Keywords: Refugees, integration, predictive modeling, algorithmic matching, fairness

## Reference Format:

Clara Strasser Ceballos, Marcus Novotny, and Christoph Kern. 2025. Re-evaluating the role of refugee integration factors for building more equitable allocation algorithms. In *Proceedings of Fourth European Workshop on Algorithmic Fairness (EWAf'25)*. Proceedings of Machine Learning Research, 6 pages.

## 1 Introduction

In 2015 and 2016, the European Union (EU) received more than 2.4 million asylum applications, the majority of which were submitted in Germany [14]. This brought the integration of refugees to the forefront of both political and academic debate. Integration, a complex and multi-dimensional process with disputed definitions [9], can be understood as a two-way process in which both refugees and members of the host society adapt to one

---

\*Both authors contributed equally to this research.

---

Authors' Contact Information: Clara Strasser Ceballos, LMU Munich, Munich, Germany, Clara.StrasserCeballos@stat.uni-muenchen.de; Marcus Novotny, LMU Munich, Munich, Germany, Marcus.Novotny@stat.uni-muenchen.de; Christoph Kern, LMU Munich, Munich, Germany and Munich Center for Machine Learning (MCML), Munich, Germany.

---

This paper is published under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International (CC-BY-NC-ND 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

EWAf'25, June 30–July 02, 2025, Eindhoven, NL

© 2025 Copyright held by the owner/author(s).

another across six key dimensions: economic, social, psychological, political, linguistic and navigational [10, 19]. Economic integration of refugees, often measured by employment outcomes, have been a primary focus of scholarly research [18]. In particular, numerous studies have sought to understand the relationship between individual and location-level characteristics and labour market chances of refugees. Key findings highlight the multifaceted nature of integration, influenced not only by common socio-demographic factors like sex [22], but also by a wide range of individual and location-level factors like health [23], accommodation type [18], asylum procedure duration [21], integration course access [20], regional voting behavior [27], and local attitudes toward migrants [3].

Another, more recent line of research has focused on improving the integration of refugees by designing tools that algorithmically assign newly arrived refugees to resettlement locations, with the overarching aim of maximizing integration chances [2, 5]. These tools operate in two layers: (1) a *prediction* layer, in which individual-level characteristics (e.g., age and sex) of refugees are used to predict their integration success, currently limited to employment probabilities, across potential resettlement locations, and (2) a *matching* layer, in which these refugees are matched to presumably optimal locations according to the predictions and existing capacity constraints [2, 5]. This algorithmic approach contrasts with traditional resettlement processes, where allocation decisions are driven uniquely by capacity constraints or (quasi-)random assignment mechanisms, such as proportional distribution keys [4, 28, 31], rather than focusing on refugee-specific outcomes. Two such tools, GeoMatch and Annie™ Moore [2, 5], are currently being piloted in the U.S., Switzerland and the Netherlands [17, 25]. While these approaches have reported promising results, predicting to improve refugee employment rates by roughly 40% in the U.S., and 73% in Switzerland [5], they also have clear limitations. These include using only a very limited set of predictors and not assessing the group fairness of the predictions. In particular, academic research on these tools – including fairness-related studies [6, 15] – has primarily focused on the *matching* layer [1, 11]. However, since predictions play a crucial role in shaping downstream matching outcomes, we highlight the role of studying fairness at the *prediction* layer.

*Contribution.* Our contribution is twofold, taking a critical look at the limitations of algorithmic decision-making tools to predict refugee integration. First, we conduct a literature review of empirical migration studies on refugee economic integration to develop more holistic predictive models that incorporate these insights. This is particularly important as the extensive research on integration factors remains largely underutilized in current prediction tools such as GeoMatch and Annie™ Moore, which rely on only nine to eleven (protected) attributes as predictors. Second, we evaluate our extended predictive models in terms of performance and group fairness, comparing them against one of the prominent algorithmic matching tools: GeoMatch.

## 2 Preliminary Results: An Overview of Integration Factors

We conducted a literature review of studies that quantitatively examine the economic integration of refugees using the IAB-BAMF-SOEP Survey of Refugees in Germany – a rich, representative dataset that enables high-quality quantitative research [27]. From over 40 relevant studies, we selected ten recent papers for detailed comparison.

Table 1 summarizes the variables included and the methodologies employed in these studies. To highlight key differences in variable selection between migration research and predictive modeling, we also include the predictors used in current applications of the GeoMatch and Annie™ Moore tools. We categorize all variables into three conceptual levels: *pre-arrival*, *post-arrival*, and *location*. Pre-arrival variables refer to individual-level

characteristics that are fixed or acquired before migration (e.g., country of origin and education level), and are generally observable or collectible at the time of arrival. In contrast, post-arrival variables capture characteristics that occur or develop after arrival (e.g. asylum status and social contacts). Lastly, location-level variables capture attributes of the initial refugee resettlement location (e.g., unemployment rate and voting behavior).

	Migration Studies										Prediction Tools	
	Schilling & Stillman [27]	Mendola et al. [23]	Aksoy et al. [3]	Kanas & Kosyakova [20]	Hanafi & Marouani [18]	Tjaden & Spörlein [30]	Kosyakova et al. [22]	Meyer & Winkler [24]	Salikutluk & Menke [26]	Kosyakova & Brenzel [21]	GeoMatch [5]	Annie™ Moore [2]
<i>Pre-Arrival</i>												
Children			✓	✓	✓	✓	✓	✓	✓	✓		✓
Disability		✓										✓
Health status			✓				✓	✓	✓			✓
Education level	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓
Intention to work		✓								✓		
Language level			✓				✓			✓	✓	✓
Migration support from family/friends			✓	✓			✓			✓		
Religious affiliation			✓								✓	
Socio-demographics	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Traumatic experiences				✓			✓					
Values and norms							✓		✓			
Work experience				✓	✓		✓		✓	✓		
<i>Post-Arrival</i>												
Accommodation type	✓	✓			✓		✓	✓		✓		
Asylum status		✓		✓		✓	✓		✓	✓		
Integration course participation			✓	✓				✓	✓			
Language course participation							✓	✓	✓	✓		
Language level					✓		✓	✓	✓			
Social contact natives				✓			✓	✓	✓			
Social contact non-natives (co-ethics)					✓		✓					
Social contact non-natives (other)					✓		✓	✓				
Support job seeking			✓							✓		
Xenophobia perception		✓					✓					
Other				✓	✓		✓		✓			
<i>Location</i>												
Attitudes towards migrants			✓	✓								
GDP per capita	✓					✓		✓				
Population density	✓			✓			✓					
Share of foreigners	✓		✓	✓		✓	✓	✓		✓		
Unemployment rate			✓	✓			✓	✓		✓		
Voting behavior	✓		✓	✓		✓						
Other	✓					✓	✓	✓				

**Table 1. In contrast to current implementations of GeoMatch and Annie™ Moore, empirical migration research considers a wider range of integration factors at the pre-arrival, post-arrival, and location level.** The table shows ten recent papers studying the economic integration of refugees in Germany using the IAB-BAMF-SOEP survey. Papers are sorted by publication year from left to right. The operationalization of economic integration varies across papers. Socio-demographic characteristics include age at arrival, sex, family status, and country of origin. A check mark (✓) indicates that the study includes the variable.

The overview highlights the limited set of predictors included by GeoMatch and Annie™ Moore. The tools rely exclusively on *pre-arrival* characteristics, most of which are considered protected attributes under global anti-discrimination laws, including age, sex, family status, country of origin, language, and religious affiliation [29]. In contrast, migration studies extend far beyond this narrow scope, and incorporate a rich set of *post-arrival* and *location* variables, which often have strong significant effects in their respective studies. For instance, Aksoy et al. [3] show that local unemployment decreases refugees' likelihood of employment, whereas positive local attitudes towards migrants increases economic integration. Notably, the review highlights the rich but underutilized body of

empirical evidence available from explanatory studies that can be used to improve predictive models of refugee integration and potentially inform refugee-location matching tools.

### 3 Analytical Strategy

*Data.* We draw on data from the IAB-BAMF-SOEP Survey of Refugees in Germany, which is integrated into the German Socio-Economic Panel [8, 13]. The survey collects representative information on refugees who have arrived in Germany since January 2013 by drawing random samples from the Central Register of Foreigners (AZR). We use the survey to construct information for all working age refugees who arrived between 2014 and 2017. For information on the resettlement locations (i.e. the federal states in Germany), we use a variety of data sources, including data from the Federal Statistical Office [12] and the Federal Employment Agency [16].

*Predictors and target.* We expand upon the predictor sets used in existing algorithmic refugee-location matching tools by incorporating all theoretically motivated variables identified in our literature review of migration studies (see Table 1). As existing models mostly rely on pre-arrival factors, we can explicitly evaluate whether continuously updated post-arrival and location variables may enhance predictive accuracy and fairness.

Our analysis focuses on a single target variable representing the economic dimension of integration: long-term employment, defined as whether a refugee is employed within three years of arrival. This operationalization follows the GeoMatch approach used in the Swiss context, acknowledging that, as in Switzerland, employment rates among refugees in Germany are typically low within the first few years of arrival [5, 7]. We, however, highlight that integration is a multifaceted concept with multiple dimensions next to economic integration.

*Methodology.* We reconstruct the prediction layer of GeoMatch to generate employment predictions for newly arriving refugees. Specifically, we train models on data from individuals who arrived in Germany in 2014–2015 and generate predictions for those who arrived in 2016–2017. We specify several models in sequence: a benchmark model that replicates GeoMatch’s set of predictors, followed by models that successively incorporate (1) additional pre-arrival variables identified in migration studies, (2) post-arrival variables, and finally, (3) location-level variables.

*Evaluation and preliminary findings.* We evaluate model performance using standard metrics to determine which explanatory variables enhance predictive power. To evaluate fairness, we assess group-level metrics such as Equalized Odds and Equal Opportunity across protected characteristics (e.g., sex, country of origin and religious affiliation). Finally, we compare each extended model to the benchmark to examine improvements in both performance and fairness. Preliminary results indicate that relying solely on a limited set of pre-arrival predictors – as done by GeoMatch – leads to low predictive performance and unequal outcomes across protected groups. Expanding the set of predictors, even within the pre-arrival category, improves predictive accuracy and yields notable fairness gains, particularly in reducing disparities by sex as measured through Equal Opportunity metrics.

### References

- [1] Avidit Acharya, Kirk Bansak, and Jens Hainmueller. 2022. Combining Outcome-Based and Preference-Based Matching: A Constrained Priority Mechanism. *Political Analysis* 30, 1 (Jan. 2022), 89–112. <https://doi.org/10.1017/pan.2020.48>
- [2] Narges Ahani, Tommy Andersson, Alessandro Martinello, Alexander Teytelboym, and Andrew C. Trapp. 2021. Placement Optimization in Refugee Resettlement. *Operations Research* 69, 5 (Sept. 2021), 1468–1486. <https://doi.org/10.1287/opre.2020.2093>

- [3] Cevat Giray Aksoy, Panu Poutvaara, and Felicitas Schikora. 2023. First Time around: Local Conditions and Multi-Dimensional Integration of Refugees. *Journal of Urban Economics* 137 (Sept. 2023), 103588. <https://doi.org/10.1016/j.jue.2023.103588>
- [4] BAMF. 2024. Initial Distribution of Asylum Seekers (EASY). <https://www.bamf.de/EN/Themen/AsylFluechtlingsschutz/AblaufAsylverfahrens/Erstverteilung/erstverteilungnode.html>.
- [5] Kirk Bansak, Jeremy Ferwerda, Jens Hainmueller, Andrea Dillon, Dominik Hangartner, Duncan Lawrence, and Jeremy Weinstein. 2018. Improving Refugee Integration through Data-Driven Algorithmic Assignment. *Science* 359, 6373 (Jan. 2018), 325–329. <https://doi.org/10.1126/science.aao4408>
- [6] Kirk Bansak and Linna Marten. 2024. Algorithmic Decision-Making, Fairness, and the Distribution of Impact: Application to Refugee Matching in Sweden. *SOFI Working Papers in Labour Economics* 6/2024, Stockholm University, Swedish Institute for Social Research. (2024).
- [7] Herbert Brücker. 2024. Arbeitsmarktintegration von Geflüchteten: Verbesserte institutionelle Rahmenbedingungen fördern die Erwerbstätigkeit. *IAB* (2024).
- [8] Herbert Brücker, Nina Rother Rother, and Jürgen Schupp. 2017. IAB-BAMF-SOEP Befragung von Geflüchteten 2016. Studiendesign, Feldergebnisse sowie Analysen zu schulischer wie beruflicher Qualifikation, Sprachkenntnissen sowie kognitiven Potenzialen. *IAB Forschungsbericht* 13 (2017). <https://doi.org/10.5684/soep.core.v38eu>
- [9] Stephen Castles, Maja Korac, Ellie Vasta, and Steven Vertovec. 2003. *Integration: Mapping the Field*. Home Office. Research, Development and Statistics Directorate.
- [10] European Commission. 2025. Integration - European Commission. [https://home-affairs.ec.europa.eu/networks/european-migration-network-emn/emn-asylum-and-migration-glossary/glossary/integration\\_en](https://home-affairs.ec.europa.eu/networks/european-migration-network-emn/emn-asylum-and-migration-glossary/glossary/integration_en).
- [11] David Delacretaz, Scott Kominers, and Alexander Teytelboym. 2024. Data and Code for: Matching Mechanisms for Refugee Resettlement. <https://doi.org/10.3886/E191062V1>
- [12] DESTATIS. 2025. *GENESIS-Online*. Retrieved 2025-03-06 from <https://www-genesis.destatis.de/datenbank/online>
- [13] DIW. 2024. IAB-BAMF-SOEP-Befragung Geflüchteter 2022, Daten der Jahre 2016-2022. <https://doi.org/10.5684/soep.iab-bamf-soep-mig.2022>
- [14] Eurostat. 2024. Asylum Applicants by Type - Annual Aggregated Data. <https://ec.europa.eu/eurostat/web/migration-asylum/asylum/database>. <https://doi.org/10.2908/TPS00191>
- [15] Daniel Freund, Thodoris Lykouris, Elisabeth Paulson, Bradley Sturt, and Wentao Weng. 2023. Group Fairness in Dynamic Refugee Assignment. In *Proceedings of the 24th ACM Conference on Economics and Computation (EC '23)*. Association for Computing Machinery, New York, NY, USA, 701. <https://doi.org/10.1145/3580507.3597758>
- [16] Bundesagentur für Arbeit. 2025. *Statistiken*. Retrieved 2025-03-06 from <https://statistik.arbeitsagentur.de/DE/Navigation/Statistiken/Statistiken-Nav.html>
- [17] Global Compact on Refugees. 2024. *GeoMatch: Connecting People to Places Using Artificial Intelligence*. Retrieved 2024-11-05 from <http://globalcompactrefugees.org/good-practices/geomatch-connecting-people-places-using-artificial-intelligence>
- [18] Cyrine Hannafi and Mohamed Ali Marouani. 2023. Social Integration of Syrian Refugees and Their Intention to Stay in Germany. *Journal of Population Economics* 36, 2 (April 2023), 581–607. <https://doi.org/10.1007/s00148-022-00913-1>
- [19] Niklas Harder, 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 (2018), 11483–11488. <https://doi.org/10.1073/pnas.1808793115>
- [20] Agnieszka Kanas and Yuliya Kosyakova. 2023. Greater Local Supply of Language Courses Improves Refugees' Labor Market Integration. *European Societies* 25, 1 (Jan. 2023), 1–36. <https://doi.org/10.1080/14616696.2022.2096915>
- [21] Yuliya Kosyakova and Hanna Brenzel. 2020. The Role of Length of Asylum Procedure and Legal Status in the Labour Market Integration of Refugees in Germany. *SozW Soziale Welt* 71, 1-2 (Aug. 2020), 123–159. <https://doi.org/10.5771/0038-6073-2020-1-2-123>
- [22] Yuliya Kosyakova, Zerrin Salikutluk, and Jörg Hartmann. 2023. Gender Employment Gap at Arrival and Its Dynamics: The Case of Refugees in Germany. *Research in Social Stratification and Mobility* 87 (Oct. 2023), 100842. <https://doi.org/10.1016/j.rssm.2023.100842>
- [23] Daria Mendola, Anna Maria Parroco, and Paolo Li Donni. 2024. We Made It to Germany ... and Now? Interdependent Risks of Vulnerability for Refugees in a High-Income Country. *Journal of Ethnic and Migration Studies* 50, 4 (Feb. 2024), 1059–1079. <https://doi.org/10.1080/1369183X.2023.2206009>

- [24] Franziska Meyer and Oliver Winkler. 2023. Place of Residence Does Matter for Educational Integration: The Relevance of Spatial Contexts for Refugees' Transition to VET in Germany. *Social Sciences* 12, 3 (March 2023), 120. <https://doi.org/10.3390/socsci12030120>
- [25] OPSI. 2020. *Annie™ MOORE (Matching for Outcome Optimization and Refugee Empowerment) - Observatory of Public Sector Innovation*. Retrieved 2024-11-06 from <https://oecd-opsi.org/innovations/annie/>, <https://oecd-opsi.org/innovations/annie/>
- [26] Zerrin Salikutluk and Katrin Menke. 2021. Gendered Integration? How Recently Arrived Male and Female Refugees Fare on the German Labour Market. *Journal of Family Research* 33, 2 (Sept. 2021), 284–321. <https://doi.org/10.20377/jfr-474>
- [27] Pia Schilling and Steven Stillman. 2024. The Impact of Natives' Attitudes on Refugee Integration. *Labour Economics* 87 (April 2024), 102465. <https://doi.org/10.1016/j.labeco.2023.102465>
- [28] SEM. 2024. Verteilung der Asylsuchenden auf die Kantone. <https://www.bj.admin.ch/sem/de/home/asyl/asylverfahren/nationale-verfahren/verteilung-kantone.html>.
- [29] Jan Simson, Alessandro Fabris, and Christoph Kern. 2024. Lazy Data Practices Harm Fairness Research. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*. Association for Computing Machinery, New York, NY, USA, 642–659. <https://doi.org/10.1145/3630106.3658931>
- [30] Jasper Tjaden and Christoph Spörlein. 2023. How Much Do “Local Policies” Matter for Refugee Integration? An Analytical Model and Evidence from a Highly Decentralized Country. *International Migration Review* (Oct. 2023), 01979183231205561. <https://doi.org/10.1177/01979183231205561>
- [31] U.S. Department of State. 2024. Reception and Placement. <https://www.state.gov/refugee-admissions/reception-and-placement/>.