


RESEARCH ARTICLE

Predicting mobility aspirations in Lebanon and Turkey: a data-driven exploration using machine learning

Simon Ruhnke¹  and Ramona Rischke²

¹Berliner Institut für Empirische Integrations- und Migrationsforschung, Humboldt-University of Berlin, Berlin, Germany

²Migration Department, German Centre for Empirical Integration- & Migration Research, Berlin, Germany

Corresponding author: Simon Ruhnke; Email: simon.ruhnke@hu-berlin.de

Received: 17 November 2023; **Revised:** 17 July 2024; **Accepted:** 29 July 2024

Keywords: aspirations-capability framework; (im)mobility aspirations; machine learning; migration theory; Lebanon; Turkey; Syria; refugees; prediction; Random Forest

Abbreviations: Lasso, Lasso-regression; RF, Random Forest Algorithm; Step, Backwards stepwise regression

Abstract

The aspirations-ability framework proposed by Carling has begun to place the question of who aspires to migrate at the center of migration research. In this article, building on key determinants assumed to impact individual migration decisions, we investigate their prediction accuracy when observed in the same dataset and in different mixed-migration contexts. In particular, we use a rigorous model selection approach and develop a machine learning algorithm to analyze two original cross-sectional face-to-face surveys conducted in Turkey and Lebanon among Syrian migrants and their respective host populations in early 2021. Studying similar nationalities in two hosting contexts with a distinct history of both immigration and emigration and large shares of assumed-to-be mobile populations, we illustrate that a) (im)mobility aspirations are hard to predict even under ‘ideal’ methodological circumstances, b) commonly referenced “migration drivers” fail to perform well in predicting migration aspirations in our study contexts, while c) aspects relating to social cohesion, political representation and hope play an important role that warrants more emphasis in future research and policymaking. Methodologically, we identify key challenges in quantitative research on predicting migration aspirations and propose a novel modeling approach to address these challenges.

Policy Significance Statement

We identify a large gap between respondents expressing strong considerations to migrate and those with concrete plans. This pattern anticipates widespread involuntary immobility among our sample and calls for policies that create safe pathways for migration and address the underlying predictors of (im)mobility aspirations. Our findings suggest that among those predictors, rarely considered societal factors such as discrimination, social cohesion, and political representation are important in (im)mobility decision-making, particularly among Syrian refugees, and should be included alongside traditional economic considerations in any asylum strategy aiming to promote “durable solutions.” Gender, meanwhile, does not emerge as an important predictor, suggesting that the observed male bias in actual migration stems from restrictive circumstantial factors rather than an independent female preference for immobility.

1. Introduction

While public debates and policy attention are often directed at migratory movements and migration governance, shedding light on mobility-related decision-making of migrants often rests on individual-level data on (im)mobility aspirations.¹ Despite (im)mobility aspirations being notably different from actual moves, focusing on (im)mobility aspirations is a worthwhile goal from both a scientific and policy perspective. Aspirations usually precede actual moves and have been shown to correlate with migration flows on an aggregate level (Tjaden et al., 2019), such that an understanding of their determinants helps to shed light on “root causes” of migration and to anticipate migration dynamics. In addition, aspirations—not only to move, but also to stay—are often formed under conditions of uncertainty and at essential crossroads in people’s life and tend to have long lasting effects (e.g., Czaika, 2021). (Im)mobility aspirations further influence both current well-being and behavior and are thus essential in understanding the individual as well as social welfare effects of migration (Aslany et al., 2021).

On a global scale, migration and changes in (im)mobility aspirations are relatively rare events in many populations rendering it a methodologically challenging subject of study. In 2020, the international migrant stock stood at 3.6% of the world population in 2020 (IOM GMDAC, 2023), a share that was shown to remain relatively stable in the past (Czaika and de Haas, 2014, Fransen and de Haas, 2021). The number of those expressing a desire to permanently leave their countries of residence was an estimated 16% among adults in 2021 (Gallup, 2023). This latter figure is slightly higher than it was when last reported in 2010, when it stood at 14% and exhibited a substantial gap to the share of those making active plans to leave (8%) (Gallup, 2012). At the same time, for countries affected by violent conflicts, both theory and empirical evidence suggest substantially heightened mobility aspirations and migration pressure,² yet also suggest substantial barriers once an initial movement of those capable of leaving after outbreaks of violence fulfills their aspirations. Overall, mobility aspirations can be very high in the extremes, and both migratory moves and migration aspirations are prone to producing unbalanced data. Hence, while understanding migratory movements and predicting migration aspirations remains high on the political agenda, the underlying data patterns are methodologically challenging.

Thus, our contribution to addressing these empirical challenges is thus threefold. First, we use novel survey data collected in Lebanon and Turkey in 2020/2021 that was specially designed to study individual (im)mobility decision-making. These data allow us to analyze multiple indicators of individual (im)mobility aspirations as well as a broader spectrum of potential individual- and household-level factors than other studies on migration aspirations that rely on secondary data collected for different purposes. Second, we use rigorous model selection and develop a machine learning approach to facilitate the full use of this rich data source to optimize the overall prediction performance given the described challenges, rather than investigating the statistical significance of a preselected set of predictors. Third, a data-driven comparison across Syrian refugee and host populations in two mixed-migration countries that are among the main host countries for Syrian refugees allows us to reflect on the importance of their respective contexts in understanding the determinants of mobility aspirations.

Our findings suggest a relative scarcity of concrete plans that stands in contrast to impactful public and political discourses in Europe that seem to at least partly rest on the assumptions of a high prevalence of immanent migration plans among and beyond our Syrian study population and across time. Further, even when considering a large variety of potential determinants, some aspects of individual migration decision-making cannot be modeled reliably in our data. This we interpret to result from the complexity of aspiration formation as well as the relative rarity of migration aspirations and concrete migration plans.

¹ Throughout the paper, we opt for the use of (im)mobility aspirations rather than the term migration aspirations that is more commonly found in the literature, to capture the notion of immobility as being one end of the immobility-mobility spectrum and to expand our discussion beyond the mobility bias ascribed to said literature.

² Considering the countries that have produced the largest absolute numbers of international refugees in the past decade (Syria, Afghanistan, and Ukraine), the estimated shares of cross-border forced migration ranges between 14% and 23% of the population, with internal migration of similar magnitudes. These estimates are based on figures provided by the UNHCR Global Trends Report 2022 (6.5 m. of Syrian, 5.7 m. of Afghan, and 5.7 m. of Ukrainian refugees) as well as the IOM GMDAC.

Our findings further substantiate that dominant factors associated with (im)mobility aspirations vary in their composition and relative importance across our country samples, and across different groups within each geography (here: Syrians, and host populations, respectively³). Among the similarities across samples, we establish the relevance of some factors that are relatively rarely elicited in general population surveys, such as hope and indicators of social cohesion. At the same time, and contrary to our expectations, gender does not appear as an important independent predictor of (im)mobility aspirations in either group or country sample.

2. Background

2.1. Study contexts

Both Lebanon and Turkey are key asylum countries for the people displaced from the civil war unfolding in neighboring Syria since 2011, jointly hosting over 5 million Syrian refugees (UNHCR, 2022a, 2022b). At the time of data collection, Lebanon was already in the midst of multiple devastating and accumulating political and economic crises (World Bank, 2021), threatening the livelihood of both Syrians and Lebanese residents (e.g., Rischke and Talebi, 2021). Despite an estimated quarter of the population in Lebanon being made up of Syrian refugees, the country lacks a formal asylum policy (Geha and Talhouk, 2019), and its government maintains that it is not a country of asylum but rather a transit station for onward mobility (Janmyr, 2016). As a result, the responsibility of providing education, social, and healthcare services falls largely to underfunded international and non-governmental organizations. While challenging in its own right, the situation of Syrians in Turkey during the survey period was somewhat more stable. The designation of Temporary Protection Status grants them free access to schooling and healthcare services (Yıldırım et al., 2019), while speaking to an equal reluctance to provide durable solutions. In addition, a prohibition to freely relocate within Turkey and lacking access to the formal labor market and sustainable livelihoods still puts the refugee population at a more precarious position than much of the Turkish host population (Ruhnke, 2021), thus also possibly informing plans for secondary migration (Ilcan et al., 2018). In sum, due to the challenging economic and political situations in both study contexts, we would a priori expect a higher prevalence of mobility aspirations among Syrians relative to their respected host populations and overall higher aspirations to stay for both host and Syrians in Turkey as compared to crisis-struck Lebanon.

2.2. Literature and theoretical framework

Much of the scholarly literature on individual (im)mobility aspirations has built on the aspirations-ability framework introduced by Carling (2002) and later advanced by Carling and Schewel (2018) and de Haas (2021), aspirations-capability framework. This theoretical framework presents migration as a two-step process and is in line with general psychologically rooted theories of staged decision-making processes such as the Theory of Planned Behavior (Ajzen, 1991). In the first step, mobility aspirations are formed before— Influenced by structural- and individual-level factors determining the ability of a person to fulfill their aspirations—the aspirations turn into (im)mobility outcomes in the second step (Carling, 2002). Notably, this framework applies to different (im)mobility forms, that is, to the continuum between forced and voluntary forms of (im)mobility, as well as to groups that differ in their previous migration experience (e.g., Syrians and “hosts” in Turkey and Lebanon).⁴ In this study, we follow the scholarly tradition that, while acknowledging that it is neither a necessary nor sufficient condition for actual mobility, emphasizes the importance of individual mobility aspirations and its determinants in understanding the nature of both human mobility and immobility, including the degree to which it is voluntary or forced.

³ We use the term “host population” to indicate non-Syrian nationals in our sample. In Lebanon, these were mainly Lebanese, and in Turkey, people originating from Turkey. While doing so for the sake of brevity, we acknowledge the shortcomings of binary categories such as “refugees” and “hosts.”

⁴ Previous migration experiences are one factor that can influence (im)mobility aspirations and that enters our empirical model as a predictor.

2.2.1. Determinants of mobility aspirations

(Im)mobility decision-making processes are widely acknowledged to be complex (e.g., Willekens, 2021): Different opportunity structures, both to stay and to leave, interact with potential needs (e.g., Czaika, 2021) and motives as well as individual and household characteristics (see Figure 1). The formation of individual (im)mobility aspirations can be considered as one outcome of this process. Different areas of considerations and clusters of factors overlap and interact with each other at the level of individuals and households, which speaks to the difficulty of identifying the role and relative importance of individual factors. Varying opportunity structures on the societal level, that is, the extent to which human mobility and in particular relocation and the crossing of borders is enabled or inhibited for different groups in society (e.g., differentiated migration governance regimes for refugees, “irregular migrants” and “ordinary citizens”) add another layer of complexity that varies across time and space.

In what follows, we take these migration regimes and other societal and governmental factors largely as given. Informed by our interest in shedding light on individual decision-making that was guiding our data collection, we focus on individual- and household-level determinants with a focus on *current* (im)mobility aspirations. However, we do argue that individual aspirations also reflect household decision-making processes since individuals take intra-household dynamics into consideration when forming their preferences and aspirations.

Individual and household characteristics put forward in the literature comprise demographic factors including transnational family structures, socioeconomic factors, personal values and attitudes, involvement in the local community, multidimensional well-being indicators, and past migration experiences. In what follows, we will provide an overview of the theoretical considerations behind the different groups of potential determinants of migration aspiration that we will include in our analysis (for a complete overview, see Appendix Table A1).⁵

Demographic factors including family structures and international networks. In general, mobile populations are assumed to be fairly young (e.g., Aslany et al., 2021), which is reflected in many migration studies relying on data that is collected only among young age groups. In addition, migration and (im)mobility decision-making is assumed to be a gendered process, with men being more prone to migrate compared to women (Ibid).

Transnational family structures can speak to motives (e.g., family reunification or keeping families from separating) as well as opportunity structures, that is, capabilities, to migrate (through international

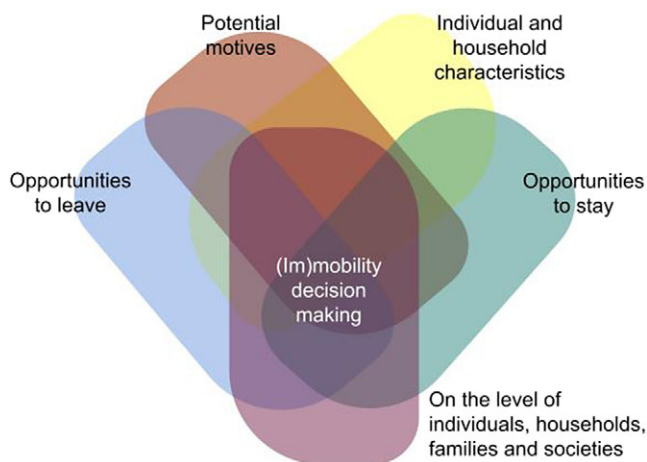


Figure 1. Complexity of (im)mobility decision-making.

⁵ While a detailed review of general migration theories is beyond the scope of this article, for excellent overview articles see, e.g., Massey et al., 1993; de Haas, 2021.

networks) but also to stay (e.g., in case of care responsibilities for remaining family members, looking after immovable assets).

Socioeconomic factors. Both at the individual and household levels, socioeconomic factors speak to the capabilities of fulfilling individual (im)mobility preferences—either having sufficient resources to move if desired, or to stay, without the perceived need to search livelihood opportunities elsewhere or to spread household income generation risks across territorial boundaries. Individuals with good paying, high-status jobs (or at least non-precarious jobs) are *ceteris paribus* assumed to have fewer motives to leave for economic reasons, while having greater financial means to form concrete migration plans or turn them into migration if they wish so. Through presumably more stable work, these job profiles also increase the chances that individuals are socially well-embedded compared to day laborers, for instance. Home ownership and actively being in education can be considered other “ties that bind” such that both factors are expected to *ceteris paribus* be negatively associated with migration aspirations.

Receiving aid as a main source of household income speaks both to the poor economic conditions of a respondent’s household, and to being part of larger support networks. Poverty at the household level is expected to spur general considerations or preferences for leaving; however, the extent to which concrete plans are formed and actions taken is not only expected to depend on financial capabilities, but also the relative role of individuals within households. In other words, depending on their age, gender, current labor market participation, and education attendance, some individuals are expected to choose or be chosen to migrate in intra-household bargaining processes.

Personal values, attitudes, and assessments. Personal values and attitudes are expected to influence both (im)mobility decision-making as well as destination choices. For instance, it is commonly assumed that potential migrants’ preferences for destinations tend to be those where predominant cultural values reflect the values of potential migrants themselves. This should be particularly so for permanent location changes.

Values that are frequently considered in the existing studies concern gender norms and religious values (e.g., van Dalen et al., 2005; Docquier et al., 2020). This may speak to the preference formation of migrants as much as it speaks to narratives surrounding migration deterrence in destination countries, in particular, the fear in some host societies of holding different values than immigrants. Other attitudes we consider concern risk-taking preferences and believing that “fate is in one’s own hand.” This is based on the notion that migration is a process that inherently carries uncertainties, which individuals with certain risk-taking preferences (e.g., Goldbach and Schlüter, 2018; Kiriscioglu and Ustubici, 2023) and/or a sense of self-efficacy are more likely to embrace (e.g., Ajzen, 1991) or cope with compared to others.

Involvement in the local community. Community involvement is associated both with emotional connections to the local community (i.e., to the neighborhood or city of residence) or to the current country of residence and with network connections that individuals have and can make use of to cope with challenges and hardships. Overall, we would assume different aspects of community involvement to increase the likelihood of immobility aspirations. A negative “change in community belonging,” as well as expressing “feeling like an outsider” more frequently, we would expect to be *ceteris paribus* positively associated with mobility aspirations.

Multidimensional well-being. In our data collection, we cover several dimensions of multidimensional well-being that include overall life satisfaction, health, discrimination experiences, neighborhood characteristics, a sense of political representation, and hope. In general, indicators associated with a high relative well-being are assumed to increase aspirations to stay and reduce aspirations to move, respectively.

Health is operationalized by the current self-rated health, changes in the same, as well as the PHQ score, which is a psychological indicator for mental health burden over the past two weeks. Ill health as well as a deterioration of health, particularly among those who perceive difficulties in seeing a doctor, are expected to feed into the motivation to move, while potentially negatively impacting the capabilities to do so.

Discrimination experiences in the current place of residence are expected to increase mobility aspirations. We cover different forms of discrimination—specifically based on nationality and religion—and how it materializes as lived experiences, specifically as verbal threats, and/or physical violence.

The quality of one's neighborhood is expected to influence mobility aspirations such that those in poor quality neighborhoods are *ceteris paribus* expected to aspire to move, at least to a different neighborhood. To the extent that neighborhood quality approximates economic well-being, we also expect it to be associated with income-related capabilities. Identifying as "belonging to the majority group" within the local neighborhood—however, "majority group" is being defined by respondents—is expected to be associated with relative welfare benefits. We consider this indicator to proxy the availability of within-group support structures.

The degree of perceived political representation measures the extent to which respondents agree that they are represented by the government in their country of current residence. We assume that this factor is positively associated with aspirations to stay. Finally, the assessment that the current country of residence is on a path for a better future (i.e., hope)—a phrasing that is suitable because both Lebanon and Turkey during the time of our surveys were characterized by (severe) economic and sociopolitical crises—is important because it directly speaks to livelihood opportunities to stay, and potential motives to leave.

Past migration experiences. Past migration experiences—both of the respondents themselves as well as by family members—are assumed to increase the capabilities to move for the reason of reducing the uncertainties and increasing the chances of having established network structures abroad. Having return migrants among family members could both increase network structures, or—depending on the experiences of return migrants and the information they provide, reduce migration aspirations (e.g., Auer and Schaub, 2023). At the same time, to the extent that having contact to return migrants is associated with reduced remittances received, it might add additional financial motives to move.

In general, to the extent that barriers to move are internalized, we would expect that lacking capabilities to overcome these barriers are not only affecting plans to migrate but also general considerations to move, whereas desires to leave (or stay) should be more independent of capabilities to leave (or stay).

2.2.2. Empirical approaches

The empirical literature on the formation of mobility aspirations that informed and was informed by the theoretical considerations outlined above is as vast and interdisciplinary as the field of migration research itself. But based on a comprehensive review by Aslany et al. (2021), we identify four common patterns in this heterogeneous literature: First, falling in line with the traditional "mobility bias" (Schewel, 2020), most studies focus on aspirations to move, that is, migration aspirations, as the key outcome of interest, rather than aspirations to stay. Second, all of the studies reviewed by Aslany et al. (2021) use regression analysis to identify significant determinants of migration aspirations. This reliance on regression analysis results in a third common pattern: Depending on the authors discipline, theory of interest, and available data, most analyses include a small number of variables of interest, their most salient covariates, and a number of common control variables (what we will later refer to as the "greatest hits" of predictors). The focus of these empirical studies is to robustly identify a significant relationship between the variables of interest and migration aspirations. Thus, the emphasis lies on the regression coefficients and *p*-values of a relatively small set of variables, rather than the predictive performance or explanatory value of the overall regression model (e.g., the R^2). A notable recent advancement in this literature is the work by Carling et al. (2023), who expand this regression-based approach to 42 potential determinants of migration aspiration across 28 different local contexts. While the study's primary aim remains similar to that of the previously described literature, it provides a useful point of comparison for the latter part of our analysis (see Section 4.3).

The present study should not be understood as standing in opposition to this regression-based literature and its aims. This modus of inquiry has proven highly productive and undoubtedly advanced our understanding of the determination of individual migration aspirations. Nonetheless, in this study, we choose an altogether different approach that breaks with all four of the common patterns in the empirical literature on mobility aspirations. Rather than investigating whether a particular theory about the formation of migration aspiration is supported by our data, we want know whether, given a rich set of

factors suggested across several theories and schools of thought as determinants of migration aspiration, we can actually predict who does or does not want to migrate and how to best do so.

3. Data and empirical strategy

The goal of our analysis is threefold: We seek to a) identify among novel and more established approaches an appropriate method capable of predicting (im)mobility aspirations given the unique empirical challenges this endeavor poses and the data at hand, b) measure the reliability of this method across multiple measure of (im)mobility aspiration, and c) identify those household- and individual-level characteristics most important in this prediction in our study contexts.

For this analysis, we use survey data collected in Lebanon and Turkey between September 2020 and February 2021 as part of the TRANSMIT research project. The surveys each aimed at collecting representative samples of the Syrian population as well as a sample of the host population living in the same neighborhoods. In the absence of reliable registry data, the surveys employ stratified area sampling and random walk techniques and are conducted via computer-assisted face-to-face interviews (CAPI, for details, s. [Supplementary Material](#)). The total sample size is 2.732 in Turkey and 2.500 in Lebanon.

3.1. Dependent variables

In the quantitative migration literature, a large variety of surveys, indicators, and proxy variables has been used to analyze migration aspirations, that we understand here as **considerations, plans, or intentions** to migrate. As a consequence, there is no uniform way in which “migration aspirations” have been operationalized or even conceptualized. There is, however, a growing understanding that the way in which migration aspirations are operationalized indeed matters for understanding (im)mobility aspirations, and that different operationalizations capture different decision-making processes that are not necessarily expected to follow similar logics (e.g., Carling and Schewel, 2018; Carling, 2019). These different decision-making processes are linked, for instance, to the formation of individual preferences, the preparedness or necessity to leave, or the likelihood thereof (Carling and Schewel, 2018).

Reviewing questions related to migration aspirations from more than 50 quantitative surveys, Carling (2019) differentiates the mindset, action, and conditionality inherent in survey items. He concludes that surveys should include complementary questions in order to reflect different aspects of (im)mobility aspirations, which we concur and which is precisely what we had been doing as part of our original survey data collections in the TRANSMIT research project.⁶ In addition, it is recommended to disclose the exact formulation of survey items to reflect on the scope and limitation of given survey questions (Carling and Mjelva, 2021), which we turn to next.

In general, in our data collections, we have defined and iterated to respondents our definition of *migration* as changing their place of residence for more than 3 months.

3.1.1. Considerations to move abroad

Respondents were asked about their considerations to move to another country in the following way:

How much, if at all, are you considering to move to another country to live (for more than 3 months)?

Please rate on a scale from 0 to 10 where 0 stands for “I don’t want to move at all” and 10 stands for “I really want to move.” Any number in between is valid, too.

According to Carling (2019), “considerations” as a concept stand out because it enquires about a fact (have you considered or not) rather than an attitude, the latter of which can easily change. At the same time, considerations would blend awareness about migration as a possible course of action and the evaluation of that consideration as something desirable (Ibid). This is important to note because individuals who have

⁶ The research project “Transnational Perspectives on Migration and Integration” is a BMFSFJ-funded joint research project.

internalized barriers to move (or to stay) are unlikely to perceive migrating (or staying) as a feasible course of action (e.g., Czaika et al., 2021; Appadurai, 2004 on the role of a “capacity to aspire”). Put differently, to some degree, aspirations themselves inherently reflect capabilities.

3.1.2. *Having concrete plans to move away*

Respondents were asked about their concrete plans to move in the following way:

Have you made concrete plans to move away from your current place of residence within the next 12 months? (move > 3 months). If so: Where do you plan to move?

In line with Carling and Mjelva (2021), we argue that “concrete plans” to migrate—within the current country of residence or abroad—reflect another stage in the decision-making process, namely being closer compared to “general considerations” to taking specific actions, such as sharing plans with others, seeking support or applying for visas. Following this chain of argumentation, having concrete plans implies previous considerations to move, whereas general considerations can be observed independent from concrete plans to move. Accordingly, this question was administered only to those respondents who report above zero considerations to move.

3.1.3. *Desire to stay or live somewhere else*

Presuming severely limited opportunities for international mobility among parts of the target population in our survey contexts (e.g., due to movement restrictions, visa requirements, and poverty), we asked our respondents a hypothetical question about their permanent location of choice in the same way as used by the Gallup World Polls:

Ideally, if you had the opportunity, would you like to move permanently to another country, or would you prefer to continue living in this country? Where would you want to permanently live?

The knowledge interest behind this question was not only to gather information about individual preferences to migrate or to stay in a fashion comparable to existing large-scale data, but to shed some light on longer-term intentions to return or stay for the group of Syrians in our sample, independent from current material constraints. Carling (2019) points out that, due to the balanced nature of mobility options provided, this question is well suited to assess respondent’s relative desirability of (im)mobility, regardless of their feasibility or actual pursuit.

3.1.4. *Operationalization*

Our dependent variables, concrete **plans**, and the **desire to stay** are binary. Since responses to the “consideration to move”-item are clustered around the values 0 and 10, and to make analysis and interpretation consistent between dependent variables, we opted to dichotomize this variable as well. The considerations indicator is set as 1 if respondents’ degree to which they are considering to move is reported to be the highest value of 10, thus capturing **strong considerations to move**.⁷ The concrete plans indicator is equal to 1 if respondents report that they have concrete plans to move, while the desire to stay indicator equals 1 if they named their respective country of residence as the country they would want to live permanently.

3.2. *Independent variables*

Based on the existing literature on migration decision-making (see Section 2.2), we identify 47 plausible individual- and household-level determinants of respondent’s migration aspirations. All variables and

⁷Results for specifications of the strong considerations indicator at different cut-off points can be found in the Supplementary Materials (see Supplementary Figures S2 and S3).

their respective definitions are displayed in Appendix Table A1. Missing values in the independent variables are imputed using the Random Forest (RF) algorithm (Breiman, 2001) implemented through the `missForest` package in R (Stekhoven and Bühlmann, 2012). RF is a nonparametric prediction algorithm that can process both continuous and categorical variables and performs comparatively well in high-dimensional datasets.

3.3. Empirical challenges and method selection

The breadth of potential determinants we aim to capture as well as the distribution of migration aspirations, pose 3 key challenges we need to account for in choosing our modeling approach: 1. overfitting and multicollinearity, 2. unbalanced data, and 3. interpretability of model outputs. In the following, we consider three popular modeling strategies and how they are commonly employed to deal with these three challenges. Two of them, backward stepwise regression (“Step” hereafter) and Lasso-regression (“Lasso” hereafter; s. Tibshirani, 1996), are parametric estimation techniques that allow us to remain within the familiar regression analysis framework so commonly used in the empirical literature on mobility aspirations. Third, the RF algorithm (s. Breiman, 2001) is a nonparametric modeling approach specifically popular in the field of machine learning to deal with complex prediction tasks like the one we outlined here. Due to the iterative nature of the decision-tree concept that underlies the RF, it is able to model interactions of independent variables without prior definitions of these interactions needed and thus bears the potential to capture the complex, and overlapping structure of the formation of migration aspirations (Figure 1).

3.3.1. Overfitting and multicollinearity

When working with a large number of independent variables (i) (relative to the sample size), a common risk that is particularly pronounced in traditional regression modeling techniques is the risk of **overfitting** the model to the given data. This results in a good model fit within the given data, but poor external validity of the model. A large i further introduces the risk of **multicollinearity** among the independent variables. Stepwise regression and Lasso Regression deal with these risks by systematically excluding those regressors that do not sufficiently contribute to either the model fit (Step) or the maximization of a penalized Likelihood Function (Lasso). The RF in turn is a so-called ensemble estimation method that only considers a random subset of k for each iteration of the estimation process (so-called “bagging”) and then aggregates results across all such instances. This approach has been shown to mitigate the risk of both overfitting and multicollinearity (Breiman, 2001).

3.3.2. Unbalanced data

This implies that relatively little information about the characteristics of observations with a positive outcome ((im)mobility aspirations) is available to build an estimation model, even if the overall sample size is relatively large. This small *effective* sample size further increases the risk of overfitting described above. It also means that a model that exclusively predicts a negative outcome will achieve a reasonably high prediction accuracy, while being incapable of successfully predicting a positive outcome. It is thus not suitable to assess the performance of our final model based on the “simple” prediction accuracy, but rather the Kappa score, which accounts for the unbalanced distribution of dependent variables (see details below). Additionally, since the RF algorithm draws a subset of the data for each iteration of its estimation, it allows to account for the unbalanced data by systematically oversampling the rare outcome for each iteration. The two regression approaches (Step and Lasso) lack such “built-in” adjustment options.

3.3.3. Interpretability

Finally, to draw conclusions on the specific associations between our predictors and outcome variables, we need to ensure that the output our models produce are **interpretable**. Here, the two regression approaches appear to have a clear advantage, as they provide model coefficients familiar to any social scientist. Data scientists, on the other hand, have developed numerous machine learning approaches to

deal with overfitting, unbalanced data and other data constraints. Yet, unlike statistical inference, machine learning is primarily focused on optimizing the predictive performance of a given model, often resulting in “black box” models that do not allow for an intuitive interpretation of the relationships between independent variables and the outcome of interest. Unlike many of these approaches, however, the RF also provides a way to look into the box and identify the association between independent and dependent variables, discussed in Section 3.2.4.

To choose the best performing method among the three approaches considered while avoiding the fallacy of overfitting our data, we perform cross-validations by splitting each sample into a test (80%) and training (20%) sample.⁸ We do so separately for each sample population and outcome variable. Due to the unbalanced nature of the distribution of migration aspirations in our samples, we evaluate the performance of the three considered modeling approaches based on Cohen’s Kappa or the Kappa score metric (Cohen, 1960). In the case of a rare outcome, more intuitive metrics such as prediction accuracy can be misleading, as simple “guessing” based on the outcome distribution can result in a reasonably high accuracy, without the model adding any substantive predictive value. The Kappa score evaluates predictive performance relative to such accuracy by (conditional) chance alone and thus provides a metric of the added predictive value of the model that accounts for the underlying outcome distribution. In the case of a binary cross-validation, it can be calculated as:

$$K = \frac{2 \cdot (TP \cdot TN - FP \cdot FN)}{(TP + FP) \cdot (FP + TN) + (TP + FN) \cdot (FN + TN)},$$

where TP = True Positive, TN = True Negative, FP = False Positive, and FN = False Negative (Chicco et al., 2021).

Once the best performing method is identified, the Kappa score is also used to assess the performance of different model specifications of the method deemed most suitable.

The results of these cross-validations (see Section 4.2) imply that the RF algorithm is best suited to predict (im)mobility aspiration in our samples. As a consequence, we will focus on the RF results for identifying most important predictors among the variables entering our models.

3.4. Variable importance

Following Breiman (2001), we assess the relative importance of each independent variable entering the RF using permutation-based variable importance scores. These are calculated based on the mean decrease in out-of-bag (OOB) prediction accuracy of the RF model if the information contained in the respective variable was to be effectively removed from the estimation by randomly altering the observed value. The score is calculated as the difference between the estimation error in the permuted data and the estimation error in the original data. Thus, the further the removal of a specific variable decreases the model performance, the higher its importance score. Results are scaled to 100 and should thus be compared within each model but not across separate models. The directionality of the association between predictors and the variable of interest (including nonlinear effects) can be gauged based on partial-dependence plots (PDPs) that can be found in the Supplementary Material in Supplementary Figure S1. PDPs visualize the dependent variable in relation to changes in the independent variable holding all other variables at their observed values (Greenwell, 2017).⁹

⁸Data splits are proportional to the relative frequency of the respective dependent variable, that is, training and testing samples include the same share of positive outcomes.

⁹The PDP for a given predictor x and a modeling function f is constructed by 1) forcing x to be equal to a constant value $i \in \{\min(x), \dots, \max(x)\}$ across all observations in the training data, 2) performing prediction using f and the thus altered data, 3) averaging the predicted outcome across all observations in the altered training data, 4) plotting the i against the average predicted outcome, and 5) repeating the process for all values of i . Unlike, for example, coefficients in OLS regression, the nominal values of the outcome variable displayed in the PDP do not have an intuitive “real-world” interpretation. Instead, interpretation should be restricted to the directionality of the association with the outcome variable, that is, if the PDP is upward sloping x is positively associated with the outcome variable and negatively if the PDP is downward sloping.

All calculations are performed in R version 4.2.2. Model estimation and selection is performed separately for each outcome-sample combination using the *caret* package (Kuhn, 2020).

4. Results

4.1. Descriptive results

Table A2 in Appendix displays the distribution of the three indicators of (im)mobility aspirations we studied for each of the four sample-country pairings. Consistent with the expectations outlined in Section 2.1: The Syrian sample displays a higher prevalence of mobility aspirations and a lower desire to stay than their respective host populations, with the exception of concrete plans to move in Lebanon. Here, despite displaying substantially higher average consideration to move, Syrian respondents are less likely to have any concrete plans, thus hinting at a lower ability to act on their considerations compared to the Lebanese population. Generally, the substantial gap in prevalence between strong considerations to move, and actual plans to do so stands out across the two countries with only 5% of Syrians and 3.4% of hosts reporting concrete plans.

It should further be noted that 44% of Syrian respondents in Turkey expressed a desire to permanently stay in Turkey, given a (hypothetically) unrestricted choice.

In general, respondents in Lebanon display higher mobility aspirations and a lower rate of desiring to stay than their respective counterparts in Turkey. The sheer magnitude of these aspirations to leave Lebanon is surprising and indicative of the unprecedented crises the country is facing.

4.1.1. Method selection

The Kappa scores from the 80/20 cross-validation for the Step, Lasso, and RF models are consistently highest for the RF approach, for all sample and outcome combinations, with the exception of the desire to stay among Syrians in Turkey (see Figure 2), where the RF, however, still performs reasonably well. We conclude that, overall, the RF is best suited to model different aspects of (im)mobility aspirations, given the specific data at hand and the described methodological challenges. The remaining results presented in this study will thus be based on the RF approach.

In assessing the predictive performance of different RF model specifications using the Kappa score in the OOB sample,¹⁰ we contrast our full model using all 47 independent variables, and rerun the model based on the variables identified by the data-led algorithm as the top 20, top 10, top 5 most important variables, respectively, and separately for each outcome variable of interest. In addition, we use the “greatest hits” of commonly analyzed migration determinants as identified by the systematic literature review of Aslany et al. (2021).¹¹ These include age, gender, marital status, socioeconomic status (proxied here by household economic status, see also Table A1), employment status and other employment-/activity-related variables (proxied here by employment as a day laborer, see also Table A1), as well as educational attainment. The results can be seen in Figure 3

First, we note that in **Lebanon**, *general considerations to move* can be predicted relatively well,¹² particularly among the Syrian population. *Concrete plans* cannot be predicted well, which results from relatively low numbers of individuals who concretely plan to move (see Section 4.1). It is noteworthy that the RF algorithm, while the best performing among the modeling approaches considered and designed to allow processing unbalanced data, remains unable to render a reliable estimate of concrete plans. The *desire to stay* in Lebanon, the hypothetical preference that is unbalanced for the sample of Syrians and

¹⁰ Since the RF employs bagging for each tree, the predictive performance of each tree can be evaluated on those observations not included in the training of the tree and then averaged across all trees. This OOB prediction thus follows the same principle as the cross-validation employed for model selection and evaluates the model performance only on data that has not been “seen” by the model.

¹¹ We consider as “greatest hits” variables all those that have been used in at least 20 of the reviewed studies.

¹² There is no agreement as to which cut-off values of the Kappa score correspond to which quality in predictions. However, 0.6 is often considered to indicate a good, 0.4 a moderate/acceptable and below 0.2 a bad performance (Cohen, 1960).

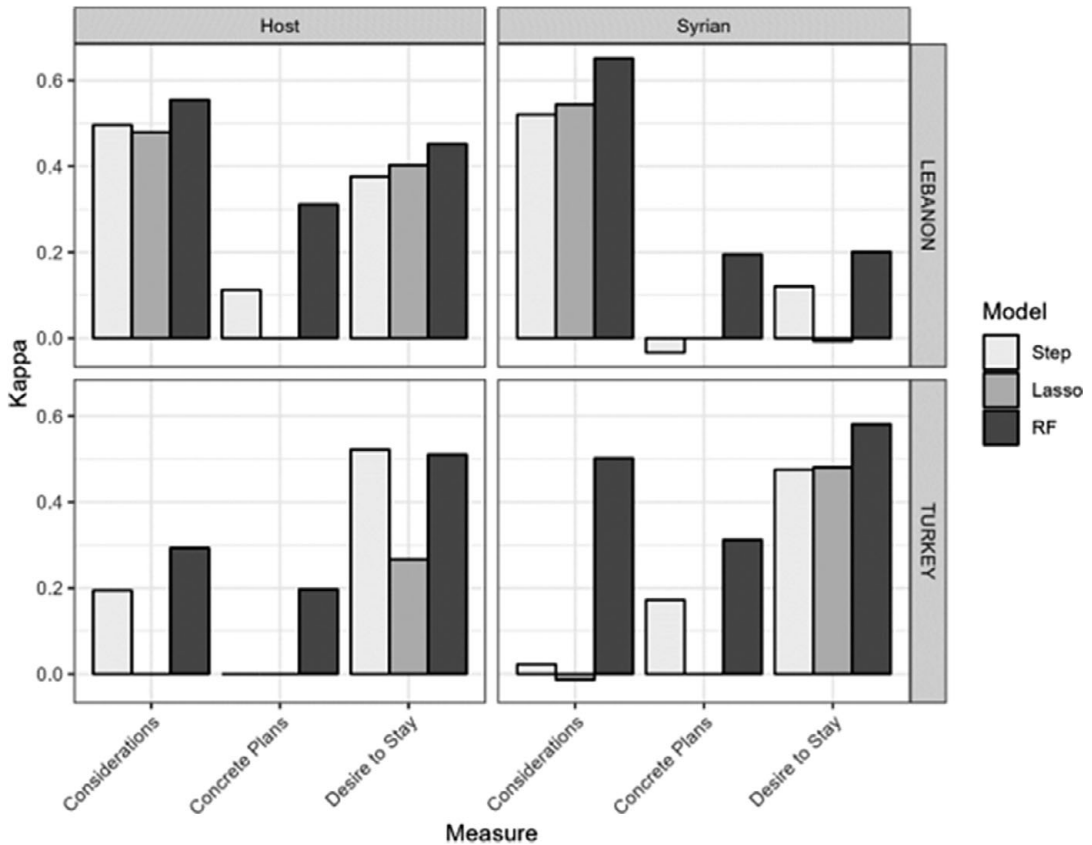


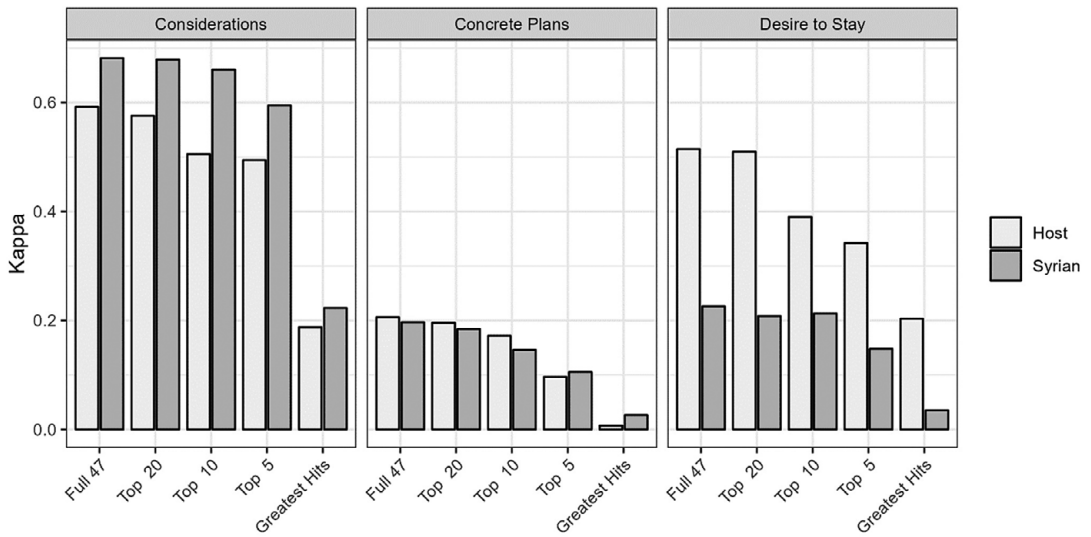
Figure 2. Kappa scores (im)mobility aspirations based on 80/20 cross-validation of Step, Lasso, and RF models (Note: Higher scores indicate a better predictive performance. Passing the threshold of 0.4 is assumed to indicate acceptable results, whereas a threshold of 0.6 and higher would indicate a good model fit).

much more balanced for the host population, can be decently predicted for the latter but not the former group. In **Turkey**, on the other hand, we find that for the host population that displays generally low migration aspirations, only the desire to stay can be predicted reasonably well. For the Syrian population, we find the *desire to stay* and *general considerations* to be modeled with acceptable prediction performance. Overall, the differences that we see in the ease of predicting our outcomes of interest are strong and are also present among population groups that a priori were expected to show substantial similarities in their (im)mobility decision-making (Syrians in Lebanon and Turkey, respectively). Hence, while it is well established that context matters in understanding mobility aspirations—which our findings below substantiate—on a methodological level, it was contrary to our expectations to find these stark differences in the *ease of predicting individual migration aspirations* in our study contexts.

Looking at the outcome variables that can be predicted reasonably well, another result stands out, which is that the “greatest hits” of independent variables as established in the literature thus far, perform very poorly in our study context.

How pivotal the unbalanced nature of (im)mobility aspirations are for our ability to predict them reliably can be seen in the Kappa scores for different cut-offs of the “considerations to move” indicator reported in the Supplementary Materials (Supplementary Figure S2). The more balanced the distribution of the binary split (i.e., more toward the “equal 10” cut-off for Lebanon and more toward “equal 0” for Turkey) the more reliably the outcome variable can be predicted.

Lebanon



Turkey

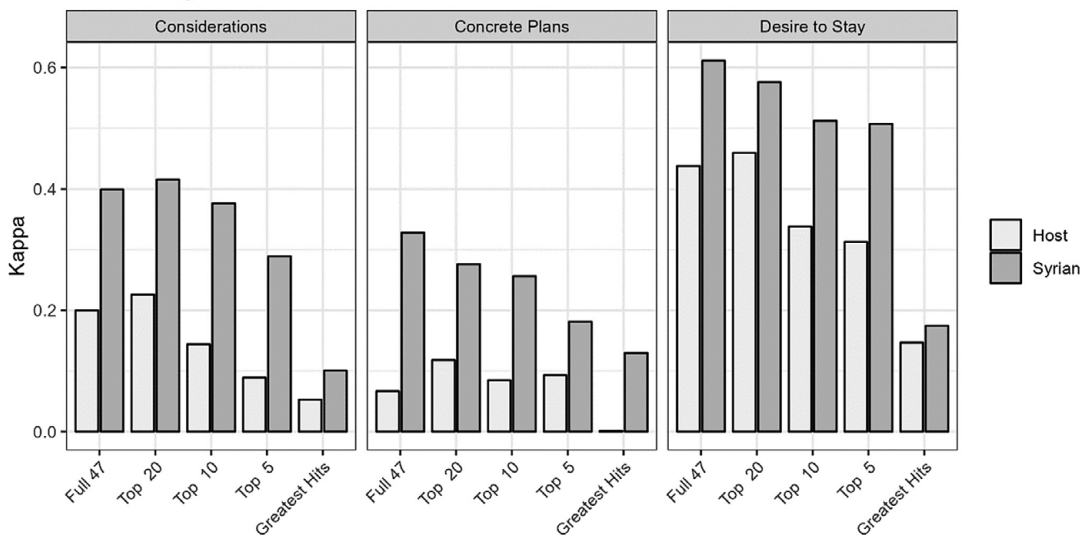


Figure 3. Kappa scores (im)mobility aspirations based on out-of-bag sample of Random Forest estimation (Note: Higher scores indicate a better predictive performance. Passing the threshold of 0.4 is assumed to indicate acceptable results, whereas a threshold of 0.6 and higher would indicate a good model fit.)

4.2. Variable importance

In this section, unless stated otherwise, we will restrict the discussion of the most important predictors to those outcome-sample combinations for which our model performs reasonably well. For Lebanon, we will mainly focus on considerations to move for both Syrian and host samples and on the desire to stay among the host population. For Turkey, we will focus on the desire to stay for both samples and considerations to move among Syrians.

All importance scores are shown in Figure 4. For each outcome variable, the graph displays the top 20 predictors as derived from the full model using all 47 predictors. While the aim of this section is to

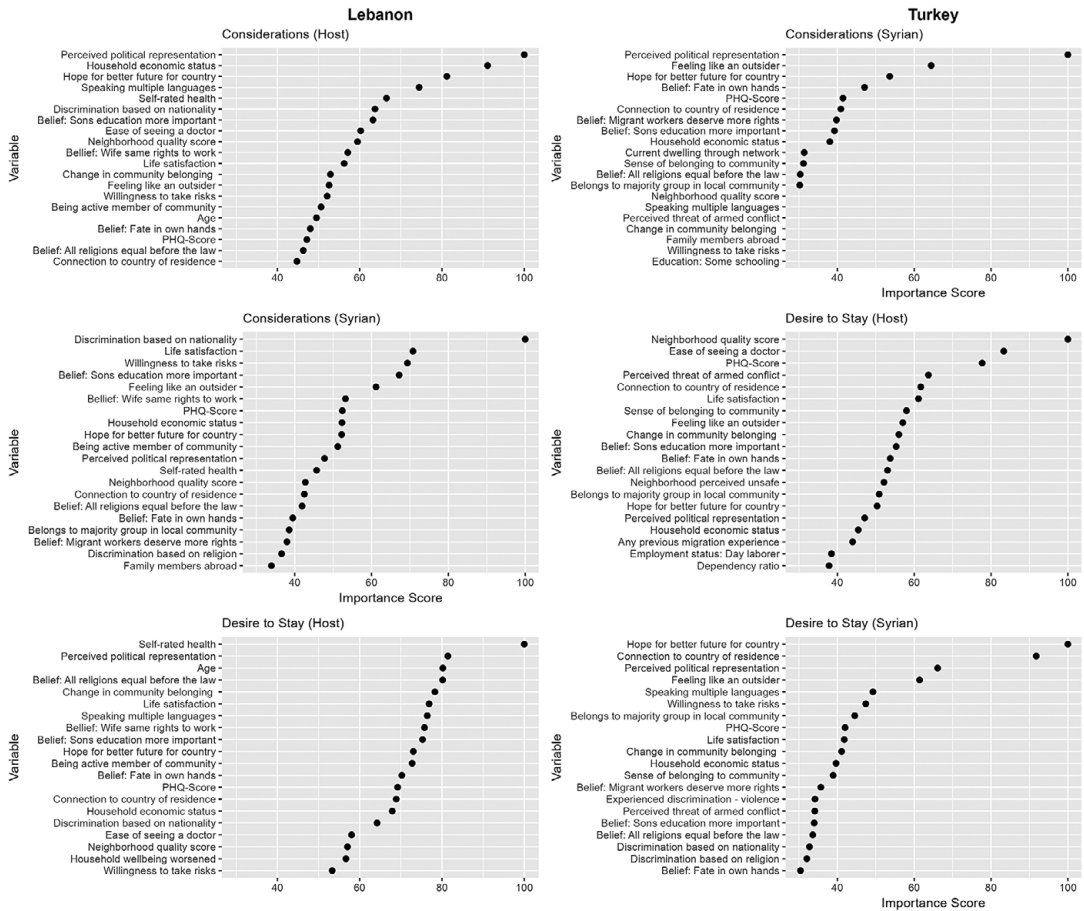


Figure 4. Permutation based importance scores derived from Random Forest for Syrian and host population in Lebanon (a) and Turkey (b) (Note: Importance Scores are scaled to 100.).

discuss the most important predictors in our samples, considering that the prediction performance for the top 20 variables is very close to the full model (see Kappa scores above), the graphs also reflect the range of variables driving the overall prediction performance.

4.2.1. Lebanon

Comparing the most important factors among the Syrian and host populations in **Lebanon** presented in Figure 4 underlines that their formation of migration aspirations appears to differ. While for the host population, hope in the future of the country, household economic status, as well as political representation have the highest importance score when it comes to considerations to move, for the Syrian population discrimination experiences are by far the most important factor, followed by life satisfaction. In the context of Lebanon, where the political situation has been increasingly unstable, and the country faces one of the worst economic crises in modern history (World Bank, 2021), considerations over the overall political and economic state of the country seem strongly tied to individual migration aspirations. At the same time, with several crises accumulating, discrimination of Syrian and other predominantly noncitizens has become more prevalent (Majed et al., 2021) and—as our findings emphasize—an important factor associated with migration aspirations.

In addition to factors speaking to the macrostructural situation, such as hope in a positive future for the country of stay and the sense of being politically represented, the most important factors associated with

the desire to stay among the host population include individual factors, such as age, health, and language proficiency. A look at the PDPs in [Supplementary Figure S1](#) reveals that the absence of these individual resources (i.e., youth, health, and language skills) is, on average, associated with a higher probability to express a desire to stay. This suggests that even given the hypothetical phrasing of this question, individuals may already take their ability to move into considerations when expressing their ideal location choice, thus possibly reflecting acquiescent immobility¹³ rather than true voluntary immobility (Schewel, 2020).

Among the variables included in the overall weakly predictive “greatest hits” indicators, it is the household economic situation and the age (among the host population) that are included in the 20 most important predictors of the RF Model. The household economic situation speaks to motives and capabilities to both stay and leave. Looking at the PDPs ([Supplementary Figure S1](#)), we find that in Lebanon, more poverty is associated with higher considerations to move, which is consistent with income-generating motives dominating and in line with evidence speaking to the impact of the devastating economic conditions in the country (Majed et al., 2021).

4.2.2. Turkey

Among Syrian respondents, it is the connection to the country of residence as well as hope in the future of the country of residence that is by far the strongest factor associated with Syrians’ desire to stay in Turkey—a desire that is prevalent among 44% of respondents. This is followed by other indicators related to social cohesion dynamics and social inclusion, namely, the extent of “feeling like an outsider” and the perceived degree of “political representation.” The remaining factors identified include (other) indicators of multidimensional well-being, such as life satisfaction, a sense of self-efficacy and (mental) health as well as risk attitudes and language skills. This relative importance of perceived sociocultural connection to and life satisfaction in the host country (Özkan et al., 2021; Üstübici and Elçi, 2022), as well as risk perceptions and attitudes for the formation of (im)mobility aspiration of Syrians (Kiriscioglu and Ustübici, 2023), aligns with recent evidence from Turkey.

For the Turkish host population, the neighborhood quality and the access to a medical doctor, stand out in predicting the desire to stay. Furthermore, poor mental health indicators are associated with a lower desire to stay, suggesting that an overall frustration with respondents’ living situation may be driving migration aspirations among the host population. Notably, indicators of social connection and embeddedness appear to play an important role in predicting the desire to stay for both Syrian and host samples in Turkey.

None of the demographic factors, nor international networks is identified as being among the top 20 most important variables. Among the socioeconomic factors, it is again the household economic situation that matters, and the ability to speak multiple languages, which may speak to the ability of Syrians to speak Turkish or any other language. In the case of Syrian respondents in Turkey, it is respondents from better-off households that are more likely to want to stay (albeit with diminishing slope), which is consistent with the opportunities to stay in the sense of a lesser need to find economic opportunities elsewhere.

4.2.3. Comparative perspectives

Notwithstanding the differences we see across Syrians and host populations in Lebanon and Turkey when it comes to the most important predictors of (im)mobility aspirations, our data suggest some interesting similarities when considering the full range of variables depicted in [Figure 4](#). Similarities include the role of variables related to social cohesion dynamics, such as community involvement and belonging, variables related to health, including indicators of psychological stress, as well as the role of personal values and attitudes. That is, these variables—in different orders and magnitudes of their underlying importance score—are listed in [Figure 4](#) for both the Syrian and host populations (for a graphic overview,

¹³ This concept refers to people who do not aspire to migrate and at the same time are unable to do so.

see [Supplementary Figure S5](#) in the Supplementary Material). The same holds for the household economic situation and an overall indicator of general life satisfaction—the first being among the “greatest hits” indicators, and the latter not having made the list of greatest hits indicators but being collected in a number of large-scale surveys.

The absence of the gender variable from our list of the most important predictors may seem rather surprising. In none of the outcome-sample combinations, we consider does the information whether a respondent is female or not prove to be among the 20 most important variables in predicting respondent’s (im)mobility aspirations. This result stands out because gender, besides age, is the single most commonly analyzed potential determinant of (im)mobility aspirations and is frequently found to have a significant correlation with the latter (Aslany et al., 2021; Debray et al., 2023). Yet, gender roles in migration are a) highly dependent on individual’s sociocultural context, so regional differences are to be expected and b) can function as a proxy for other individual characteristics and mechanisms that studies account for differently. Üstübüci et al. (2021) for instance found male Syrians in Turkey to be more likely to aspire to migrate than their female counterparts. Similarly, Dibeh et al. (2018) find male Lebanese youth to be more likely to express migration aspirations. Yet, both studies control only for basic demographic and economic characteristics (Dibeh et al., 2018; Üstübüci et al., 2021). Carling et al. (2023), meanwhile, controlled for a large set of sociocultural factors and “root causes” and found gender to be not or only weakly associated with migration aspirations in several of the study regions, including Turkey (Carling et al., 2023).

Among the displaced Syrian population, actual migration into Europe, on the other hand, displays a pronounced gender bias (e.g., Spörlein et al., 2020). Our findings suggest that this female migration gap may emerge not at the aspiration but at the ability stage of the two-step migration process (Carling and Schewel, 2018).

Another noteworthy absence includes variables related to past migration experiences that signal international networks. This finding is consistent with a stepwise formation of migration aspirations, where opportunity structures matter to a lesser degree at the stage of forming general considerations to move (ibid.). It could also result from a narrow definition of family and migration networks for our questionnaire only enquiring the whereabouts of a limited number of family members and friends.

5. Discussion

5.1. Methodological and conceptual reflections

Among the findings of this study that stand out are that Individual-level (im)mobility aspirations are hard to predict. Even here where we collected data in populations with an overall relatively high prevalence of migration aspirations (esp. Lebanese and Syrian samples), captured a broad spectrum of potential predictors identified in the literature, and employed an approach specifically selected for its ability to handle this type of modeling task, predictive performance can only be rated “good” or “moderate” for a subset of the outcomes and populations analyzed. It stands to be argued that this combination of somewhat ideal circumstances is rarely given in studies modeling (im)mobility aspirations, and thus the overall explanatory value of the regression models commonly used in the literature may be relatively low. This conclusion seems especially plausible when considering the poor predictive performance of the greatest hits, that is, those variables found in most models of individual migration aspiration (and incidentally, many less specialized population surveys).

This is not to say that the conclusions about the relationship between migration aspirations and the respective determinants of interest drawn from the existing literature are without merit. Rather, we posit that despite the knowledge gained from these existing studies on migration aspirations, we should not overstate our collective understanding of how migration aspirations are formed on the individual level. Our findings once again show that these formation processes are highly complex and heterogeneous, echoing similar arguments made regarding the limited ability to reliably forecast international migration flows (Arango, 2000; Brücker and Siliverstovs, 2006; de Valk et al., 2022).

Further insights can be gleaned from the list of the most important variables identified in our analysis. First, the relative ranking of the most important predictors varies between our different populations and national contexts, including between Syrians residing in Lebanon and those in Turkey, of whom, *a priori*, one could have expected a higher degree of similarity. At the very least, this should serve as a reminder to make ample use of country fixed effects when conducting cross-sectional analyses of (im)mobility aspirations. But even more than that, it reemphasizes the fact that the processes by which (im)mobility aspirations arise are highly context- and population-specific, and any model of (im)mobility decision-making ought to take such contextual factors into account (Carling et al., 2023). In our study context, it seems likely that the acute crisis setting in Lebanon, and the relative precarity of the Syrian population in either country as they consider secondary migration are key factors behind the observed differences in the most important predictors of (im)mobility aspirations.

Second, across all subsamples we analyzed, our list of the most important predictors each include factors such as sense of belonging, political representation, or hope for one's country of residence, of which are not commonly included in analyses of (im)mobility aspirations and even less commonly included in general population surveys. While the assertions about the context specificity of (im)mobility aspirations rendered above certainly hold for our findings as well, these types of societal factors that Bekaert (2021, p.40) referred to as "soft factors" should merit a closer look for studies in other contexts and populations as well. The relative importance of hope in a positive development of the country of residence and political representation, in particular, demonstrate that the formation (im)mobility aspirations is an innately forward-facing process that is tied to individuals' broader life and societal aspirations. Our results thus lend further support to a growing literature aiming to conceptualize the different temporalities informing individual and collective thinking on migration (Carling and Collins, 2018; Baas and Yeoh, 2019; Müller-Funk and Fransen, 2023).

Given the importance of political representation we identify, future research ought to investigate how shifts in political momentum in a country or the denial of such shifts (e.g., the stalled revolutionary momentum in Lebanon in 2019 or the defeat of the opposition candidate in the 2023 presidential elections in Turkey) impact (im)mobility aspirations. Further investigation is also needed to understand which level of government (local, region, national) is most salient in individuals' migration decision-making and whether they inform different forms of mobility (e.g., internal vs. international migration).

5.2. Policy implications

For decades, policymakers and researchers have attempted to predict migration, in order to anticipate and manage changing population dynamics (de Valk et al., 2022). The need for well-informed policy responses to emerging migration needs have only grown since then. However, so has the number of critiques warning of a limited ability to conduct such migration forecasting reliably. To this day, migration forecasting remains limited by the lack of comprehensive data on migration flows, the sheer number of potential determinants, and their heterogeneity across different regional contexts (Arango, 2000; Brücker and Siliverstovs, 2006; de Valk et al., 2022). Following the understanding of migration as a two-step process, our analysis shows these limitations in the predictability of migration flows also extend to the formation of (im)mobility aspirations. The process by which individuals arrive at the conviction, wish, or intention, to move or stay is highly complex and context-dependent. We thus have to warn against simplified narratives that may convince policymakers that knowing about a country's demographic structure or economic conditions may allow for a reliable anticipation of (im)mobility intentions among its population. Nonetheless, we believe that our analysis also allows for a number of insights regarding the specific context it is drawn from.

In our descriptive analysis, we observe a substantial gap between the number of respondents strongly considering to migrate and those that have concrete plans to do so. As the formation of concrete plans presuppose an expectation to be able to act on them, this gap is one indicator of a high prevalence of involuntary immobility in our study populations (Schewel, 2020). This in turn calls for policy action to support safe pathways for these populations to act on their mobility aspirations. To the extent that the expression of

strong considerations serves instrumental purposes of expressing dissatisfaction with the current circumstances, supporting access to local integration as a durable solution can also help close this gap.

As commonly assumed in the policy discourse, we observe a relatively low desire to stay in their current country of residence among our Syrian samples compared to their respective host population. But our descriptive results also demonstrate a considerable variability in immobility aspirations within the Syrian population depending on their national context. Among Syrians in Lebanon, faced by both a rampant economic crisis as well as social and political exclusion, immobility aspirations are largely absent. Meanwhile, a large number of Syrians in Turkey would prefer to stay there if given free choice. This latter tendency contrasts with an impactful narrative in the European policy discourse at the time that seemed to assume a near universal desire for secondary migration to Europe among Syrians in Turkey if not constrained by border restriction. A fact-based understanding of such intra-country variations in (im)mobility aspirations is a key in anticipating and addressing future migration-related needs.

Regarding the potential mechanisms underlying these descriptive patterns, the role of economic resources as the primary factor driving migration aspirations in the regional context of this study appears to be overstated in the public and policy discourse. While the economic situation of the respondent's household consistently ranks among the 20 most important predictors of (im)mobility aspirations, it does so behind factors such as experiences of discrimination, sense of belonging to the local community, and perceived political representation. The importance of these latter factors and relative unimportance of economic factors is consistent with recent evidence on (im)mobility aspirations based on the Gallup World Poll (Debray et al., 2023) and the MIGNEX project (Carling et al., 2023) and has direct implications for contemporary migration and asylum management policy.

One cornerstone of global and particularly European asylum policies is to provide economic support to displaced populations in the countries neighboring the population's country of origin. While such interventions aim at addressing humanitarian needs, they are also expected to reduce "root causes" of (onward) forced migration. The EU-Turkey statement of March 2016 is a prime example for this dimension of migration management. The deal saw the EU commit 6 billion euros in support of the refugee population in Turkey, in exchange for a shutting down of irregular migration from Turkey toward Europe (Haferlach and Kurban, 2017), and it has been seen by some policymakers as a blueprint for similar agreements (Ruhnke, 2021). Recent initiatives by the EU to establish migration agreements with Morocco, Tunisia and Egypt have followed a similar pattern of providing economic support in exchange for more stringent border controls and expected reduction in migration toward Europe. What the relative importance of societal factors for the formation of (im)mobility aspirations suggests is that such externalization approaches are unlikely to have a lasting impact on people's intention to migrate. Even when assuming that the economic stimuli entailed in the aforementioned deals reach (potential) migrant populations, they alone will not dissuade individuals from aspiring to migrate as long as deficits in factors such as social integration, political representation, and discrimination are not addressed as well. This sustained aspiration in combination with the increased border enforcement also part of the recent migration agreements could result in a high number of involuntarily immobile people and the establishment of ever more precarious modes of migration.

Our analysis has further shown respondents' gender to not be an important independent predictor of individual mobility aspirations in our study contexts. Given the observed underrepresentation of women among the Syrian migrants in Europe (Spörlein et al., 2020), these findings suggest that involuntary immobility may be particularly pronounced among those identifying as female.

5.3. *Limitations*

Our study is not without limitations. First, the choice of models that enter our model selection process is by no means comprehensive. The constantly evolving field of machine learning has produced a large number of competing modeling approaches that arguably would have achieved a better predictive performance than the RF approach we presented here. Yet, given the constraints imposed by the underlying data, we expect these performance increases to be minor and to not warrant the often considerably higher degree of

complexity and less intuitive interpretation. We instead chose to restrict ourselves to models that have seen some use in social science research.

Second, some evidence suggests that variable importance scores derived from the RF models may be biased toward continuous and categorical variables with a higher number of categories (Strobl et al., 2007). In other words, the importance of variables with only a few categories (such as binary variables) may be systematically underestimated. This could, for instance, explain the surprising absence of the binary gender variable among the most important predictors discussed above. To account for this possibility, we opted to use permutation-based importance scores, which Strobl et al. (2007) demonstrated to be less susceptible to this kind of bias. As a robustness check, we also repeated our analysis of considerations to move using the conditional RF approach suggested by Strobl et al. (2008) as an unbiased alternative to the conventional RF.¹⁴ The conditional importance scores derived from this analysis are depicted in [Supplementary Figure S4](#) in the Supplementary Material. The gender variable remains absent from the 20 most important variables, providing further evidence that our inference on this variable is not driven by model-induced bias.

Third, as described above, the survey data used in this analysis were sampled without reliable registry data to serve as a sampling frame. While the area sampling method employed was designed to capture as representative a sample of the Syrian population as was possible given these constraints, we cannot claim to use nationally representative data, especially regarding the host population, which was not the primary sampling focus of the project from which our data are drawn.

6. Conclusion

In this study, we explored a new methodological territory in the quantitative investigation of mobility decision-making. Using an RF machine learning algorithm, we set out to investigate how far and which of the numerous potential determinants proposed in the broad and interdisciplinary literature on migration aspirations can predict individual (im)mobility outcomes in the mix-migration contexts of Lebanon and Turkey. Our findings broaden the evidence base demonstrating that the formation of (im)mobility aspirations is a highly complex process that varies considerably between different population and geographic contexts. While we see our findings in part as a call to humility regarding our collective ability to generalize on and model migration decision-making, they also open up interesting new avenues for future research and policy. For our study context, we show that in addition to more commonly discussed determinants of mobility such as age and economic deprivation, aspects such as social cohesion, political representation, and hope play an important role in predicting mobility aspirations and warrant thorough consideration both in future migration research as well as policymaking.

Supplementary material. The supplementary material for this article can be found at <http://doi.org/10.1017/dap.2024.32>.

Data availability statement. The data that support the findings of this study are available upon request from the corresponding author. At the time of publication, a longitudinal study from which the data is drawn is still ongoing. Therefore, access to the data has to remain restricted due to data protection and privacy concerns until the study concludes. Upon conclusion of the study in 2024, replication data and code will be made available.

Acknowledgments. The authors would like to thank Lidwina Gundacker and her team for their incredible data processing efforts to generate scientific user files from the raw survey data. The authors would also like to thank Hans Lüder and three anonymous reviewers for very constructive feedback on earlier versions of this manuscript. The authors extend their gratitude to the active participants of paper discussion formats at the DeZIM-Institute, notably the migration department, conference participants at the Homecoming Conference of the Göttinger School of Development Economics, and the 9th Annual Conference on Migration and Diversity at the WZB. The authors are also very grateful to Herbert Brücker, Nader Talebi, and TRANSMIT colleagues at BIM for their pivotal roles in the data collection. The authors are also indebted toward the respondents who shared their experiences and aspirations with the survey team.

¹⁴ The Conditional Forest Model is considerably more computationally intensive than the conventional RF model and relies on subsampling without replacement which makes the adjustment for the unbalanced data structure we performed for the RF not feasible. We thus opted against applying this approach to our entire analysis.

Author contribution. Conceptualization of the study, interpretation of results, manuscript writing, and revisions: S.R. & R.R. Data analysis performed by S.R. All authors approve the final submitted draft.

Funding statement. This work was supported by the German Federal Ministry for Family Affairs, Senior Citizens, Women and Youth (grant No. 3920405WZB).

Competing interest. The authors declare none.

References

- Appadurai A** (2004) The Capacity to Aspire: Culture and the Terms of Recognition. In Rao V, Culture WM and **Public Action**. The World Bank. Washington, DC. USA In Culture and Public Action.
- Ajzen I** (1991) The theory of planned behavior. *Organizational Behavior and Human Decision Processes* 50(2), 179–211.
- Arango J** (2000) Explaining migration: A critical view. *International Social Science Journal* 52(165), 283–296. <https://doi.org/10.1111/1468-2451.00259>.
- Aslany M, Carling J, Mjelva MB and Sommerfelt T** (2021) *Systematic Review of Determinants of Migration Aspirations. QuantMig Project Deliverable D2.2*. Southampton: University of Southampton.
- Auer D and Schaub M** (2023) Returning from greener pastures? How exposure to returnees affects migration plans. *World Development* (169), 106291. <https://doi.org/10.1016/j.worlddev.2023.106291>.
- Baas M and Yeoh BS** (2019) Introduction: Migration studies and critical temporalities. *Current Sociology*, 67(2), 161–168. <https://doi.org/10.1177/0011392118792924>
- Bekaert E, Constant AF, Foubert K and Ruysen I** (2021) *Longing for Which Home: Evidence from Global Aspirations to Stay, Return or Migrate Onwards (SSRN Scholarly Paper 3929195)*. Social Science Research Network. <https://doi.org/10.2139/ssrn.3929195>.
- Breiman L** (2001) Random forests. *Machine Learning* 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Brücker H and Siliverstovs B** (2006) On the estimation and forecasting of international migration: How relevant is heterogeneity across countries? *Empirical Economics* 31(3), 735–754. <https://doi.org/10.1007/s00181-005-0049-y>.
- Carling J** (2002) Migration in the age of involuntary immobility: Theoretical reflections and Cape Verdean experiences. *Journal of Ethnic and Migration Studies* 28(1), 5–42. <https://doi.org/10.1080/13691830120103912>.
- Carling J** (2019) *Measuring migration aspirations and related concepts* (MIGNEX Background Paper, p. 45). Available at <https://www.mignex.org/publications/measuring-migration-aspirations-and-related-concepts>.
- Carling J and Collins F** (2018) Aspiration, desire and drivers of migration. *Journal of Ethnic and Migration Studies* 44(6), 909–926. <https://doi.org/10.1080/1369183X.2017.1384134>.
- Carling J and Mjelva MB** (2021) *Survey Instruments and Survey Data on Migration Aspirations. QuantMig Project Deliverable D2.1*. Southampton: University of Southampton.
- Carling J and Schewel K** (2018) Revisiting aspiration and ability in international migration. *Journal of Ethnic and Migration Studies* 44(6), 945–963. <https://doi.org/10.1080/1369183X.2017.1384146>.
- Carling J, Hagen-Zanker J, and Rubio M** (2023) The multi-level determination of migration processes (MIGNEX Background Paper). Peace Research Institute Oslo. www.mignex.org/d061.
- Chicco D, Warrens MJ and Jurman G** (2021) The Matthews correlation coefficient (MCC) is more informative than Cohen’s kappa and brier score in binary classification assessment. *IEEE Access* 9, 78368–78381. <https://doi.org/10.1109/ACCESS.2021.3084050>.
- Cohen J** (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>.
- Czaika M and de Haas H** (2014) The globalization of migration: Has the world become more migratory? *International Migration Review* 48(2), 283–323. <https://doi.org/10.1111/imre.12095>.
- Czaika M, Bijak J, and Prike T** (2021). Migration Decision-Making and Its Key Dimensions. *The ANNALS of the American Academy of Political and Social Science*, 697(1), 15–31. <https://doi.org/10.1177/000271622111052233>
- de Haas H**(2021) A theory of migration: The aspirations-capabilities framework. *Comparative Migration Studies* 9(1), 8. <https://doi.org/10.1186/s40878-020-00210-4>.
- de Valk HAG, Acostamadiedo E, Guan Q, Melde S, Mooyaart J, Sohst RR and Tjaden J** (2022) How to predict future migration: Different methods explained and compared. In Scholten P (ed.), *Introduction to Migration Studies: An Interactive Guide to the Literatures on Migration and Diversity*. Cham, Germany: Springer International Publishing, pp. 463–482. https://doi.org/10.1007/978-3-030-92377-8_28.
- Debray A, Ruysen I and Schewel K** (2023) The aspiration to stay: A global analysis. *International Migration Review* 01979183231216087. <https://doi.org/10.1177/01979183231216087>.
- Dibeh G, Fakh A and Marrouch W** (2018) Decision to emigrate amongst the youth in Lebanon. *International Migration* 56(1), 5–22. <https://doi.org/10.1111/imig.12347>.
- Docquier F, Tansel A and Turati R** (2020) Do emigrants self-select along cultural traits? Evidence from the MENA countries. *International Migration Review* 54(2), 388–422. <https://doi.org/10.1177/0197918319849011>.
- Fransen S and de Haas H** (2021) Trends and patterns of global refugee migration. *Population and Development* 48(1), 97–128.
- Gallup** (2012). Available at <https://news.gallup.com/poll/152951/Nearly-Million-Worldwide-Planning-Migrate-Soon.aspx>.

- Gallup** (2023). Available at <https://news.gallup.com/poll/468218/nearly-900-million-worldwide-wanted-migrate-2021.aspx> (last access each 2024 October 07).
- Geha C and Talhouk J** (2019) From recipients of aid to shapers of policies: Conceptualizing government–United Nations relations during the Syrian refugee crisis in Lebanon. *Journal of Refugee Studies* 32(4), 645–663. <https://doi.org/10.1093/jrs/fey052>.
- Goldbach C and Schlüter A** (2018) Risk aversion, time preferences, and out-migration. Experimental evidence from Ghana and Indonesia. *Journal of Economic Behavior & Organization* 150, 132–148. <https://doi.org/10.1016/j.jebo.2018.04.013>.
- Greenwell BM** (2017) pdp: An R package for constructing partial dependence plots. *R Journal* 9(1), 421.
- Haferlach L and Kurban D** (2017) Lessons learnt from the EU-Turkey refugee agreement in guiding EU migration partnerships with origin and transit countries. *Global Policy* 8(S4), 85–93. <https://doi.org/10.1111/1758-5899.12432>.
- Ilan S, Rygiel K and Baban F** (2018) The ambiguous architecture of precarity: Temporary protection, everyday living and migrant journeys of Syrian refugees. *International Journal of Migration and Border Studies* 4(1–2), 51–70. <https://doi.org/10.1504/IJMS.2018.091226>.
- IOM GMDAC** (2023) Migration Data Portal, Total number of international migrants at mid-year 2020 (accessed 8 September 2023).
- Janmyr M** (2016) Precarity in exile: The legal status of Syrian refugees in Lebanon. *Refugee Survey Quarterly* 35(4), 58–78. <https://doi.org/10.1093/rsq/hdw016>.
- Kiriscioglu E and Ustubici A** (2023) At least, at the border, I am killing myself by my own will”: Migration aspirations and risk perceptions among Syrian and afghan communities. *Journal of Immigrant & Refugee Studies*, 1–15. <https://doi.org/10.1080/15562948.2023.2198485>.
- Kuhn M** (2020) *Caret: Classification and Regression Training. R package version 6.0-85*. [Computer software]. Available at <https://CRAN.R-project.org/package=caret>.
- Majed R, Wazze S and Chahine M** (2021) *Migration Aspirations amongst Syrian Refugees amidst the Financial and Political Crisis in Lebanon* (Research Report). Friedrich Naumann Foundation for Freedom. <https://doi.org/10.13140/RG.2.2.31432.44807>.
- Massey DS, Arango J, Graeme H, Kouaouci A, Pellegrino A and Taylor JE** (1993). Theories of international migration: A review and appraisal. *Population and Development Review* 19 (3), 431–466.
- Müller-Funk L and Fransen S** (2023) I will return strong”: The role of life aspirations in refugees’ return aspirations. *International Migration Review* 57(4), 1739–1770. <https://doi.org/10.1177/01979183221131554>.
- Özkan Z, Ergün N and Çakal H** (2021) Positive versus negative contact and refugees’ intentions to migrate: The mediating role of perceived discrimination, life satisfaction and identification with the host society among Syrian refugees in Turkey. *Journal of Community & Applied Social Psychology* 31(4), 438–451. <https://doi.org/10.1002/casp.2508>.
- Rischke R and Talebi N** (2021) Lebanon at a critical juncture - Perspectives of Syrians and Lebanese in Lebanon 2019-2021 (May 14, 2021). Available at <https://ssrn.com/abstract=3848223>; <https://doi.org/10.2139/ssrn.3848223>.
- Ruhnke S** (2021) *The EU-Turkey Deal as a successful Blueprint?* (Policy Brief #1; MERGE X TRANSMIT Data Brief). Humboldt-Universität zu Berlin. Available at <https://www.projekte.hu-berlin.de/en/merge/publications/data-brief-the-eu-turkey-deal-as-a-successful-blueprint>.
- Schewel K** (2020) Understanding immobility: Moving beyond the mobility bias in migration studies. *International Migration Review* 54(2), 328–355. <https://doi.org/10.1177/0197918319831952>.
- Spörlein C, Kristen C, Schmidt R and Welker J** (2020) Selectivity profiles of recently arrived refugees and labour migrants in Germany. *Soziale Welt* 71(1–2), 54–89. <https://doi.org/10.5771/0038-6073-2020-1-2-54>
- Stekhoven DJ and Buehlmann P** (2012) MissForest—Non-parametric missing value imputation for mixed-type data. *Bioinformatics* 28(1), 112–118.
- Strobl C, Boulesteix A-L, Kneib T, Augustin T and Zeileis A** (2008) Conditional variable importance for random forests. *BMC Bioinformatics* 9(1), 307. <https://doi.org/10.1186/1471-2105-9-307>.
- Strobl C, Boulesteix A-L, Zeileis A and Hothorn T** (2007) Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics* 8(1), 25. <https://doi.org/10.1186/1471-2105-8-25>.
- Tibshirani R** (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 58(1), 267–288.
- Tjaden J, Auer D and Laczko F** (2019) Linking migration intentions with flows: Evidence and potential use. *International Migration* 57(1), 36–57. <https://doi.org/10.1111/imig.12502>.
- UNHCR** (2022a) Syria Regional Refugee Response Türkiye, UNHCR Operational Data Portal Refugee Situations. Available at <https://data.unhcr.org/en/situations/syria/location/113> (accessed 4 November 2022).
- UNHCR** (2022b) Syria Regional Refugee Response Lebanon, UNHCR Operational Data Portal Refugee Situations Available at <https://data.unhcr.org/en/situations/syria/location/71> (accessed 4 November 2022).
- Üstübcü A and Elçi E** (2022) Aspirations among young refugees in Turkey: Social class, integration and onward migration in forced migration contexts. *Journal of Ethnic and Migration Studies* 48(20), 4865–4884. <https://doi.org/10.1080/1369183X.2022.2123433>.
- Üstübcü A, Kiriscioglu E, and Elçi E** (2021). Migration and Development: Measuring migration aspirations and the impact of refugee assistance in Turkey (ADMIGOV Deliverable 6.1; p. 68). Amsterdam Institute for Social Science Research. <https://dare.uva.nl/search?identifier=9c9f46a5-6839-429c-b813-2c1e9b941e78>
- van Dalen HP, Groenewold G and Schoorl JJ** (2005) Out of Africa: What drives the pressure to emigrate? *Journal of Population Economics* 18(4), 741–778.
- Willekens F** (2021) The emigration decision process foundations for modelling. In *QuantMig Project Deliverable D2.3*. Southampton: University of Southampton.

World Bank (2021) Lebanon Sinking (To the Top 3) (Lebanon Economic Monitor (LEM)). World Bank. Available at <http://documents.worldbank.org/curated/en/394741622469174252/pdf/Lebanon-Economic-Monitor-Lebanon-Sinking-to-the-Top-3.pdf>.

Yıldırım CA, Komsuoğlu A and Özekmekçi İ (2019) The transformation of the primary health care system for Syrian refugees in Turkey. *Asian and Pacific Migration Journal* 28(1), 75–96. <https://doi.org/10.1177/0117196819832721>.

Appendix

Table A1. Determinants of (im)mobility aspirations^a

Demographic factors including family structures		
Age ^a	age	Numeric
Gender ^a	female	[0,1]: no, yes
Marital status ^a	married	[0,1]: unmarried, married
Number of children	nchild	Numeric
Dependency ratio	dep_ratio	(num. household members age 15 to 64)/ household size
Number of family members living abroad	fam_abroad	numeric. Family includes all children, parents, spouses and siblings of respondent.
Family member in same city	fam_cit	[0,1]: no, yes
<i>Socioeconomic factors</i>		
Household level		
Home ownership	own_dwell	[0,1]: no, yes
Household economic status ^a	hhecon	[1,...,6]: money not enough for food, ..., afford anything
Household well-being worsened in last 2 years	wellbeing_lag	[0,1]: no, yes
Aid among main income sources (past 4 weeks)	aid	[0,1]: no, yes
Access to Wi-Fi	own_wifi	[0,1]: no, yes
Current dwelling through social contacts	netw_dwell	[0,1]: no, yes
Current job through social contacts	netw_empl	[0,1]: no, yes
Individual level		
Employment status ^a	empl_d	[0,1,2]: unemployed, employed, day labor
Predominantly manual labor	empl_man	[0,1]: no, yes
Educational attainment ^a	edu	[1,..., 5]: "never attended", "some school", "high school certificate", "attended university", "Bachelors or higher"
Speaking more than one language	lang_multi	[0,1]: no, yes
<i>Values and attitudes</i>		
Fate in own hands	locuscontrol	[1,..., 7]: strongly agree, ..., strongly disagree
Willingness to take risks	risk	[1,..., 7]: strongly risk taking, ..., strongly risk avoiding
All religious groups equal before the law	val_relrights	[1,..., 7]: strongly agree, ..., strongly disagree
Migrant workers deserve more rights	val_migrrights	[1,..., 7]: strongly agree, ..., strongly disagree

(Continued)

Table A1. Continued

Demographic factors including family structures		
Age ^a	age	Numeric
Sons education more important	<i>gender_childededu</i>	[1,..., 7]: strongly agree, ..., strongly disagree
Wife same rights to work	<i>gender_work</i>	[1,..., 7]: strongly agree, ..., strongly disagree
Connection to country of residence	<i>ipl1</i>	[1, ..., 4]: very close, ..., no connection at all
Perceived threat of armed conflict	<i>riskconflict</i>	[1,..., 7]: strongly agree, ..., strongly disagree
<i>Community involvement</i>		
Sense of belonging to community	<i>comm_belong</i>	[1,..., 7]: strongly agree, ..., strongly disagree
Change in community belonging over past 2 years	<i>comm_chbelong</i>	[1,..., 7]: much stronger, ..., much weaker
Being active member of community	<i>comm_active</i>	[1,..., 7]: strongly agree, ..., strongly disagree
Frequency of feeling like an outsider	<i>ipl2</i>	[1, ..., 5]: never, ..., always
<i>Multidimensional well-being</i>		
Degree of perceived political representation	<i>repres</i>	[0, ..., 10]: not represented at all, ..., very well represented
Hope that country of residence is on path for better future	<i>hope</i>	[1,..., 7]: strongly agree, ..., strongly disagree
Life satisfaction		
Life satisfaction	<i>lifesat</i>	[1,...,7]: very satisfied,..., very dissatisfied
Life satisfaction worsened in last 2 years	<i>lifesat_lag</i>	[0,1]: no, yes
Health		
Current self-rated health	<i>health</i>	[1,...,5]: very good,..., very poor
Self-rated health worsened in last 2 years	<i>health_lag</i>	[0,1]: no, yes
PHQ score ^b		
Ease of seeing a doctor	<i>phq_score</i>	Numeric
	<i>ipl3</i>	[0, ..., 6]: very difficult, ..., very easy
Discrimination experiences		
Experienced verbal discrimination in last 2 years	<i>discr_verb</i>	[0,1]: no, yes
Experienced physical violence due to discrimination in last 2 years	<i>discr_phys</i>	[0,1]: no, yes
Experienced discrimination based on nationality in last 2 years	<i>discr_nat</i>	[1, ..., 5]: never, ..., very often
Experienced discrimination based on religion in last 2 years	<i>discr_rel</i>	[1, ..., 5]: never, ..., very often
Neighborhood characteristics		
Neighborhood quality score ^c	<i>neib</i>	numeric
Neighborhood perceived unsafe	<i>neib_safe</i>	[1,..., 7]: strongly agree, ..., strongly disagree
Belongs to majority group in local community	<i>comm_major</i>	[0,1]: no, yes

(Continued)

Table A1. *Continued*

Demographic factors including family structures		
Age ^a	age	Numeric
Migration experiences		
Any previous migration experience	<i>mig_gen</i>	[0,1]: no, yes
International migration experience	<i>mig_int</i>	[0,1]: no, yes

Source: Own elaboration.

^aVariables included in the “greatest hits” specification.

^bAverage degree of agreement with neighborhood quality with regards to dwelling maintenance, trash management and environmental hazards, [1, ..., 7]: strongly agree, ..., strongly disagree.

^cSum score of the 8-item Patient Health Questionnaire (PHQ-8) [0, ..., 24]: no risk of depression, ..., high risk of depression.

Table A2. Summary statistics, dependent variable^a

	Lebanon		Turkey		Overall	
	Host (N=1246)	Syrian (N=1254)	Host (N=1307)	Syrian (N=1425)	Host (N=2553)	Syrian (N=2679)
<i>Strong considerations</i>						
no	836 (67.1%)	539 (43.0%)	1250 (95.6%)	1228 (86.2%)	2086 (81.7%)	1767 (66.0%)
yes	410 (32.9%)	715 (57.0%)	49 (3.7%)	176 (12.4%)	459 (18.0%)	891 (33.3%)
Missing	0 (0%)	0 (0%)	8 (0.6%)	21 (1.5%)	8 (0.3%)	21 (0.8%)
<i>Concrete plans</i>						
no	1163 (93.3%)	1179 (94.0%)	1222 (93.5%)	1292 (90.7%)	2385 (93.4%)	2471 (92.2%)
yes	70 (5.6%)	64 (5.1%)	16 (1.2%)	69 (4.8%)	86 (3.4%)	133 (5.0%)
Missing	13 (1.0%)	11 (0.9%)	69 (5.3%)	64 (4.5%)	82 (3.2%)	75 (2.8%)
<i>Desire to stay</i>						
no	801 (64.3%)	1183 (94.3%)	143 (10.9%)	798 (56.0%)	944 (37.0%)	1981 (73.9%)
yes	437 (35.1%)	67 (5.3%)	1163 (89.0%)	627 (44.0%)	1600 (62.7%)	694 (25.9%)
Missing	8 (0.6%)	4 (0.3%)	1 (0.1%)	0 (0%)	9 (0.4%)	4 (0.1%)

^aIn Turkey, we note that more respondents refuse to provide information about any concrete plans they may have which may signal a higher level of distrust.