

Why Do People Move?

A Data-Driven Approach to Identifying and Predicting Gender-Specific Aspirations to Migrate

*Daniel Halim
Suneha Seetahul*



WORLD BANK GROUP

Gender Global Theme &
East Asia and the Pacific Region

April 2023

Abstract

Work-related migration has many potential drivers. While current literature has outlined a theoretical framework of various “push-pull” factors affecting the likelihood of international migration, empirical papers are often constrained by the scarcity of detailed data on migration, especially in developing countries, and are forced to look at few of these factors in isolation. When detailed data is available, researchers may face arbitrary choices of which variables to include and how to sequence their inclusion. As male and female migrants tend to face occupational segregation, the

determinants of migration likely differ by gender, which compounds these data challenges. To overcome these three issues, this paper uses a rich primary household survey among migrant communities in Indonesia and employs two supervised machine-learning methods to identify the top predictors of migration by gender: random forests and least absolute shrinkage and selection operator stability selection. The paper confirms some determinants established by earlier studies and reveals several additional ones, as well as identifies differences in predictors by gender.

This paper is a product of the Gender Global Theme and the Office of the Chief Economist, East Asia and the Pacific Region.. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://www.worldbank.org/prwp>. The authors may be contacted at dhalim@worldbank.org and sseetahul@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Why Do People Move? A Data-Driven Approach to Identifying and Predicting Gender-Specific Aspirations to Migrate*

Daniel Halim

World Bank

Suneha Seetahul

World Bank

Keywords: Migration, gender, machine-learning

JEL codes: F22, J16

* We thank Andrew Mason, Elizaveta Perova, Tobias Pfitze, Woubet Kassa, Emcet Oktay Tas, Jessica Ludwig, and seminar participants at the World Bank East Asia & Pacific Chief Economist's Office Brown Bag and Africa Fellowship BBL for many helpful discussions and suggestions. All errors remain our own.

Data was collected by World Bank East Asia & Pacific Gender Innovation Lab and was funded through the Umbrella Facility for Gender Equality in partnership with the Australian Department of Foreign Affairs and Trade. The Umbrella Facility for Gender Equality (UFGE) is a multi-donor trust fund administered by the World Bank to advance gender equality and women's empowerment through experimentation and knowledge creation to help governments and the private sector focus policy and programs on scalable solutions with sustainable outcomes. The UFGE is supported with generous contributions from Australia, Canada, Denmark, Finland, Germany, Iceland, Latvia, the Netherlands, Norway, Spain, Sweden, Switzerland, United Kingdom, United States, and the Bill and Melinda Gates Foundation.

The findings, interpretations, and conclusions expressed in this report do not necessarily reflect the views of the Board of Executive Directors of the World Bank or the governments they represent.

Contact information: World Bank Group, 1818 H Street, NW, Washington, DC 20433. Email: dhalim@worldbank.org; sseetahul@worldbank.org.

1. Introduction

Between 2010 and 2018, the share of global migrants increased by 25 percent, with 9 in 10 international migrants crossing borders due to economic reasons (World Bank, 2019). While these numbers suggest that migration increasingly becomes an important strategy for individuals to improve their livelihoods, economists are often puzzled why there are not as many migrants (e.g. Bryan et al., 2014). Standard economic theory suggests that migrants move from regions with high labor supplies and low demand to regions with lower labor supply will higher demand (Harris and Todaro, 1970). Yet, episodes of migration have not reflected the predictions of equilibrium-based migration theory, leading migration research to consider other potential factors such as social interactions (Radu, 2008). Moreover, migration is an important poverty alleviation tool with significant potential welfare gains. By doubling the number of young immigrants from developing countries to high income countries, the estimated global annual income gain could be USD 1.4 trillion (World Bank, 2018). Migration can benefit both migrants (e.g., via access to better jobs) and their family members (e.g., via remittances). Recent studies estimate that remittances have a positive effect on economic growth in developing countries that have less advanced financial markets (Sobiech, 2019). However, in 2018, only 3.5 percent of the global population were migrants. Understanding the drivers of migration is important not only to shed light on this puzzle, but also because migration can be an important engine for growth and development—offering substantial gains for the migrants, and both the sending and destination countries.

One of these potential gains can be in the form of expanding women’s employment opportunities. In 2019, only 33.6 percent of women in lower-middle income countries participated in the labor force (ILOSTAT, 2019). With an important and increasing share of migrants being women (Fleury, 2016), gender can play a distinct role in the migration decision and process. Women and men are subject to different social norms, which have repercussions on their likelihood and experience of migration (Bastia & Haagsman, 2020). However, the migration literature has long been gender-blind (Donato & Gabaccia, 2015) and existing studies that focused on women’s migration focused on non-labor drivers of migration, such as, marriage and family reunification (Ghosh, 2009). But women are also likely to migrate for economic reasons. For instance, Afsar (2011) shows that in India, many women migrate to find overseas employment in order to escape poverty, and Botea et al. (2018) find that in Lesotho women migrate to cope with economic shocks. Studies also show that single, higher educated women (Richter and Taylor, 2008), and those with social networks have higher likelihood of migration as documented by Fleury (2016). Understanding gender-specific barriers or determinants to migrate may

help policy makers design coherent domestic labor market and migration policies to maximize development gains.

A large and growing body of literature documents numerous push and pull factors affecting the likelihood of international migration—from economic gains to social networks and to prevailing social norms in sending and destination countries. Nevertheless, many empirical papers are often constrained by the scarcity of detailed data on migration and potential drivers for migration. A large body of the empirical literature instead evaluates specific channels such as education (e.g. Docquier et al., 2020) or social networks (Cairn and Smyth, 2011). In the rare instances in which rich data is available in developing countries, researchers may face arbitrary choices of which variables to include and how to sequence their inclusion in the analysis. For example, in estimating the black-white wage gap in the United States, Gelbach (2016) demonstrates that the choice of covariates to include first led to different interpretations of what factors matter in explaining the wage differential and by how much.

To overcome this issue and systematically compare drivers of migration for women and men, we leverage rich household data on international migration from Indonesia and implement a data-driven approach to identify and rank the predictors of migration, separately for women and men. We analyze 301 potential determinants of migration with 5,131 men and 8,187 women in our sample. We use two different supervised machine-learning (ML) algorithms to identify the determinants of migration: (i) first, using a Random Forest algorithm to identify the top predictors based on the importance indicator, and (ii) second, using LASSO Stability Selection to identify the predictors with the greatest impacts on the variance by computing the rate of appearance of these top predictors in multiple iterations of LASSO. These methods have two key advantages in comparison to theory-driven empirical models: (1) they select variables that have important predictive powers; and (2) they do not require that we specify a prediction model beforehand which would require subjective choices—the models are indeed generated in a data-driven way which is a “guarantee against overfitting and specification hunting” (Cengiz et al., 2021).

Since we only have data on past migration, we use interest in migration as our main outcome variable. Although interest in migration is likely to include larger shares of individuals than actual migration, unlike actual migration flows, migration intentions are not plagued by self-selection in who can overcome mobility barriers and those who cannot (Fouarge & Ester, 2007; Gubert & Senne, 2016). Moreover, interest to migrate and subsequent migration are positively correlated (Creighton, 2013;

Milasi, 2020; van Dalen & Henkens, 2013), as confirmed by our data.² Both of these reasons make interest to migrate an appealing variable to use for our purpose.

The two ML models highlight interesting patterns across gender. We find a large set of common predictors for women and men, such as local labor market characteristics and individual economic characteristics (e.g., income). The models also indicate gender-specific predictors. For instance, knowing a migrant who is a domestic worker is an important predictor of interest to migrate for women but not for men. Our models confirm drivers widely documented in previous literature (e.g., social networks) and less explored factors (e.g., access to communication) as predictors for interest in migration. The predictive models have encouragingly low rates of misclassification although sensitivity levels remain relatively low. We compare and discuss the contribution of the two proposed methods in terms of explanatory and predictive power of the models.

Our main contributions are threefold. First, we systematically evaluate numerous determinants of migration previously documented in the literature and other plausible determinants with supervised ML algorithms, ranking predictors based on their degree of importance and predictive power. This approach helps us shed some light on the puzzle of why there is an underinvestment in migration as an investment tool. Second, we add to the increasing number of applied economics studies employing ML algorithms to build predictive models (see for instance Cengiz et al. (2021) who study the impact of minimum wages on labor market outcomes; Gao et al. (2020) who identify nutritionally vulnerable households; Jayachandran et al. (2021) who identify the best survey questions to measure women's agency and Lentz et al. (2019) who provide food security crises prediction tools). We add the analysis of microeconomic determinants of migration to this set of studies and contrast our findings to the predictions of a theory-based empirical model of migration. To our knowledge, this is the first study looking at microeconomic determinants of migration with ML algorithms. Third, we systematically contrast women and men's determinants for migration.

The rest of this paper is organized as follows. Section 2 briefly reviews the literature on the drivers of migration. In Section 3 we describe the rich primary data on migration and our two supervised-ML methods. Section 4 presents the results. Section 5 presents robustness checks and proposes a discussion. Section 6 compares the data-driven and theory-driven approaches. Finally, section 7 concludes.

² Further discussed in Section 6.

2. Drivers of migration

Studies on migration from developing countries have long been debated and tested. In a seminal study aiming to analyze a comprehensive set of factors of migration, Lee (1966) classifies the different factors into four groups: push factors (linked to the origin country), pull factors (linked to the destination country), intervening obstacles, and personal factors. Moreover, standard models of migration showed that rational agents moved from regions in which labor was abundant and wages low to regions in which labor was scarcer and wages high (Harris and Todaro, 1970). The new economics of migration focused on the societal dimension of the phenomenon, in an attempt to reconcile structuralist and neoclassical approaches of migration. Since then, the literature has evolved to reflect the much more complex reality of migration and has shifted a large part of its focus to developing countries (De Haas, 2010). Despite the large set of studies on migration that exist today, ranking the determinants of migration remains a difficult task. Moreover, whether these determinants are gendered is not systematically addressed in the literature.

Migration models focusing on human capital show that an individual's decision to migrate is guided by the comparison of the returns to human capital in the destination and origin countries (see for instance (Dustmann et al., 2011)). Kahanec & Fabo (2013) point out that generally, the level of education and employment status are not important drivers for youth migration inside the European Union, but individuals with completed education are more likely to report intentions to stay abroad more than five years and less likely to report permanent migration intentions. Van Mol (2016) finds that higher educated youths have a higher likelihood of migrating when they face unemployment in the European context, in line with Bartolini et al. (2017). In developing countries, Dibeh et al. (2019) and Docquier et al. (2020) show that education level is an important driver of migration in the Middle East and North Africa region. Similar evidence is found for the Pacific (Gibson & McKenzie, 2011).

The desire to migrate can also be influenced by institutional contexts, living standards and culture where the destination country becomes more attractive than the origin country. Macroeconomic factors such as slow economic growth, weak job creation, lack of meritocracy and upward mobility are strong push factors especially for highly educated individuals (Bartolini et al., 2017; de Grip et al., 2010; Dibeh et al., 2019). Moreover, perceptions of corruption in governments, discontentment with the political situation and undersupplied public services provision can also increase young people's intention to emigrate (Etling et al., 2020; Van Mol, 2016). Naudé (2010) finds that among 45 Sub-Saharan African countries, the main determinants of migration between 1965 and 2005 were armed conflict and lack of employment opportunities.

Studies also show the important contribution of social networks on the desire and decision to migrate (Boyd, 1989). Krissman (2005) calls for a larger conception of what constitute the migrant's network, by including potential employers in the destination countries and labor smugglers, in order to understand international migration. Access to information about living abroad through siblings and friends increases intention to migrate (Cairns & Smyth, 2011) but social network from the home country decreases it, namely when there are strong links with family and local community. For instance, using data from Rwanda, Blumenstock et al. (2019) find that the probability of leaving home decreases proportional to the size of the home network. Bryan et al. (2014) point out that migration is a risky investment. Inasmuch as social networks could reduce the noise and perception of risks, their model suggests that firsthand, personal experience of migration is needed to overcome the information friction. They argue that this explains why there are fewer migrants than expected despite huge potential returns from migration.

Age is also an important determinant of migration, with younger people having a stronger incentive to migrate than older individuals (Cairns & Smyth, 2011; Kahanec & Fabo, 2013; Milasi, 2020). One of the reasons why age is such an important driver of migration is related to an individual's familial/marital status which affects the likelihood of migration, namely, not having partner ties and childbearing (Smith & Floro, 2020; Van Mol, 2016). Milasi (2020) shows that women are less likely than men to desire to migrate, but that the probability to plan to migrate does not significantly depend on gender. He explains this by the fact that the actual planning of migration is guided by traditional drivers (e.g. income and education) that do not necessarily depend on gender whereas the desire to migrate in itself is more influenced by gender-related factors (Milasi, 2020; Ruysen & Salomone, 2018).

Studies find that poverty and deprivation do not affect migration in the same way for women and men (Fleury, 2016). Smith and Floro (2020) find that among women from 84 low- and middle-income countries, food insecurity was a constraint rather than a push to migrate, while this relationship was not significant for men. Indeed, male migration increases with home or business ownership whereas these same factors tend to decrease female migration (Donato, 1993). Nevertheless, there is evidence of a positive education selection into migration for women (see for instance Richter and Taylor, 2008). Moreover, Heering et al. (2004) show that in Morocco, intention to migrate is higher when women have low job satisfaction and they suggest that this intention signals their modernity in comparison to more conservative Moroccan women. Studies also show that structural gender inequalities and discrimination can increase women's migration (Fleury, 2016). Indeed, discriminatory practices (e.g.,

early marriage, female genital mutilation, gendered social stigmas and gender-based violence) are likely to prompt migration (Ferrant et al., 2014).

Gender is a crucial part of understanding the processes of migration among Indonesians, and the recent history of migration is likely to have both a discouraging effect (fear of ill-treatment) or an encouraging effect (better opportunities than informal labor) for women. Indonesia is one of the largest origin countries of migrants in the world. Opportunities to work abroad attract Indonesians because of a lack of domestic labor demand and higher potential earnings in destination countries (World Bank, 2017). Concerning liquidity constraints, using data from Indonesia, Bazzi (2017) shows that positive rainfall and price of rice stocks are associated with larger international migration flows. He shows that this phenomenon is linked to the role of liquidity constraints in the decision-making process of migration. Emigration was largely female before the introduction of migration restrictions towards multiple countries in the Middle-East, North- and East-Africa and Pakistan (Makovec et al., 2018). Indeed, female migrants were facing risks throughout their migration journey, leading the Indonesian government to implement these moratoriums as an attempt to address the cases of ill-treatment (ranging from physical and sexual abuse, forced labor and unpaid wages) facing Indonesian women working abroad (World Bank, 2017). Makovec et al. (2018) show that migration restrictions towards Saudi Arabia increased the share of informal employment in Indonesia, formal labor markets being unable to absorb the excess labor supply.

3. Data and methods

3.1. Data

This study draws from household data collected for the impact evaluation of the Government of Indonesia's *DESMIGRATIF* program. The *DESMIGRATIF* program aims to provide a holistic approach to promote safe international migration from Indonesia, based on the four pillars of: (i) better migration information and services, (ii) productive enterprises, (iii) community parenting, and (iv) financial co-operatives for potential migrants, return migrants, and families of migrants. Because of the multidimensional nature of the program, the baseline survey—used in this study—contains rich household data. The survey was carried out in October 2018 and covers 176 villages, spanning across 6 provinces and 58 districts. The data is not nationally representative. Selected villages have higher shares of international migrants, consistent with the Ministry of Manpower's intention to allocate the

DESMIGRATIF program in areas with higher migrant concentrations, which tend to be more rural.³ The sample includes 13,318 households, with an average of 43 households per village.

The survey is conducted in two parts. A first questionnaire is administered to the most knowledgeable individual regarding household affairs. This questionnaire covers a roster of household members with basic demographic, education, and employment participation questions; household agricultural and non-agricultural enterprises, assets (including home conditions, e.g. material used for walls of the residence), access to savings and loans, receipts of social assistance programs, economic shocks; and proxy questions about current and return migrants' migration experiences (henceforth, referred to as the "household survey"). The second part of this survey randomly selects one eligible individual aged between 18 to 40 in each household to be interviewed in more details (henceforth, referred to as the "individual survey").⁴ It complements the education and employment information from the household survey, with additional questions covering literacy skills (in national and foreign languages), job search, and hours of work. It also includes detailed questions on interest to migrate, intended occupation and destination country, personal migration experience, social network pertaining to other migrants, interactions with migrant brokers, expectations of salaries, costs, and work conditions in destination countries, and perceptions of risks and returns from migration. We use most of the questions in the household and individual surveys combined, and we either include the questions as-is, recode them as dummy indicators, or define new variables from multiple questions.⁵ We take a broad view of what can be a potential determinant for migration to explore other potential determinants that have not been established in prior literature or that are not commonly collected in other surveys, such as individual perceptions of success and means to achieve success. We use virtually all variables collected in the survey.⁶ We end up with 301 continuous or binary potential determinants that will be fed in as inputs to the ML algorithms. The full list of potential determinant variables used in our analysis is presented in Appendix Table 1.

Our sample is composed of 5,131 men and 8,187 women. The individuals here are what we consider as "potential migrants". Due to the random selection of an individual within the household, the

³ Of the households in our survey, 96.1 percent reside in rural areas, compared to only 44 percent of households residing in rural areas in a recent nationally representative socioeconomic survey (SUSENAS 2019-2020).

⁴ Ideally, each household should have a corresponding individual representative. However, in practice, the household and individual surveys have to be conducted on two separate occasions due to the length of the survey. There are a few attritions (0.1 percent), where the randomly selected individual cannot be followed-up. Since the interest to migrate question is asked in the individual (not household) survey, we can only use the balanced sample with both household and individual surveys completed (N = 13,318).

⁵ Each category of categorical variables is recoded as separate binary variables.

⁶ We exclude irrelevant questions or questions with no interpretable meaning for our analysis. This includes questions such as the time and day of the interview.

individual selected may be on various spectrums of likelihood to migrate: some may have migrated in the past, and others have not. Our analysis is conducted separately for men and women; therefore, the results are not affected by the gender imbalance. The average age of respondents is 29.8 years and the average education, measured by years of schooling is 9.4, with similar levels for women and men (9.3 and 9.7 respectively). Of the female respondents, 81 percent are married compared to only 53 percent of the male respondents.

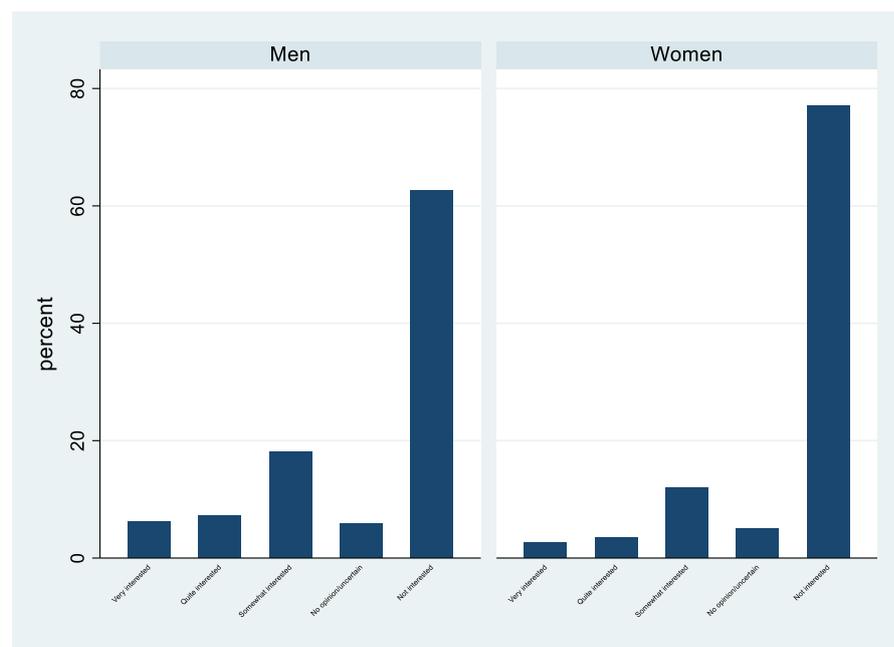
Table 1. Sample description

Variable	Whole sample		Women		Men		Min	Max
	Percent		Percent		Percent			
Female	61%		N.a.		N.a.		N.a.	N.a.
Married	70%		81%		53%		N.a.	N.a.
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.		
Age	29.76	6.61	30.20	6.43	29.06	6.83	18.00	40.00
Years of schooling	9.42	3.52	9.28	3.52	9.65	3.50	0.00	18.00
Number of children living with Respondent	0.79	1.15	0.97	1.22	0.49	0.98	0.00	7.00
Total work income (in millions, IDR)	7.4	21	4.4	15	12	27	0	960

Source: Authors' calculations from DESMIGRATIF (2018)

This study aims to predict individuals' interest in migration, captured by a self-declared response to the question of "How interested are you in working abroad?" This question is in a 5-points Likert scale format, ranging from "not interested" to "very interested" (see Figure 1).

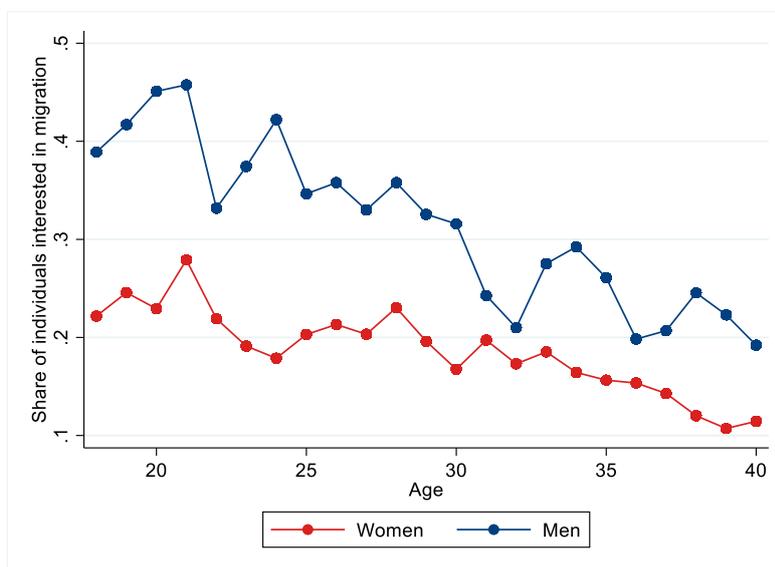
Figure 1. Interest in migration in Likert scale by gender



Source: Authors' calculation.

We recoded this question to a dummy indicator of “interested in migration”, which takes a value of 1 if “very”, “quite”, or “somewhat interested”, and 0 otherwise. In our data, 23 percent of individuals (18 percent of women and 31 percent of men)⁷ declare that they are interested in migrating. This represents a larger share of individuals than the estimated migration rates from Indonesia. The World Bank estimated the stock of international migrants from Indonesia at 9 million in 2016, which represents 7 percent of the total labor force (World Bank, 2017).⁸ Figure 2 shows the share of individuals interested in migration by gender and age. Women are generally less interested in migration than men and there seems to be a decreasing trend of interest in migration with age for both groups.

Figure 2. Interest in migration by gender and age



Source: Authors’ calculations from DESMIGRATIF (2018)

3.2. Methods

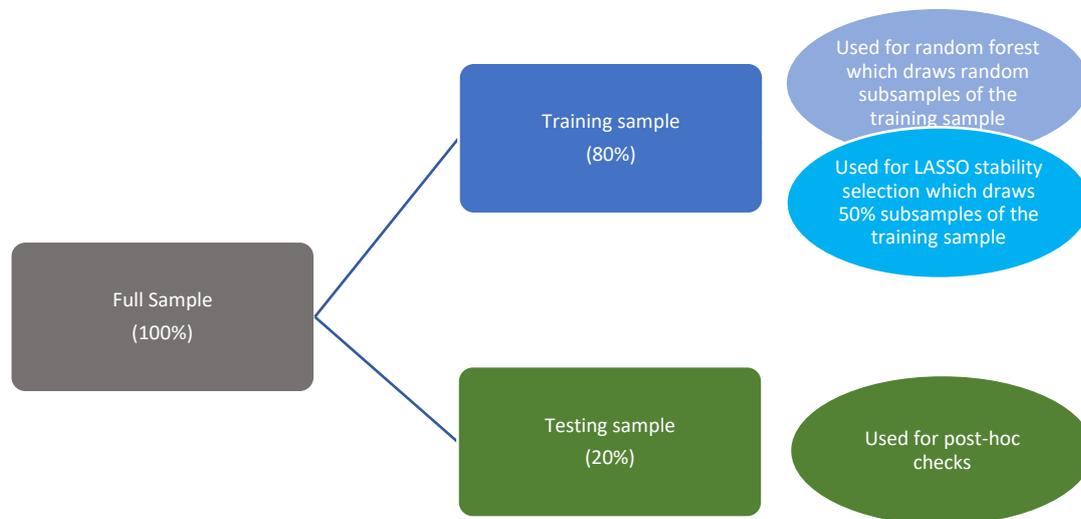
This study aims to leverage rich survey data with household characteristics, roster of household members’ employment participation, and individual preferences and perceptions of migration to identify and rank the determinants of interest in migration for women and men. However, standard econometric methods are not best suited for this task because it runs the risk of overfitting with the

⁷ This difference is statistically significant at the 1% level.

⁸ The official estimated number of *documented* emigrants from Indonesia in 2018 was 283,640 which represents approximately 0.2 percent of the workforce (BP2MI, 2019). The official number relies on proper registrations, which could severely underestimate the actual number of migrants if many migrants go through *undocumented* routes (e.g. gain employment overseas while entering the destination country with a tourist visa, which is legally allowed in some countries).

inclusion of 301 potential determinants as independent variables. The specified model will lack degrees of freedom due to the high-dimensionality of the data. We therefore turn to machine-learning methods that mitigate these concerns. The mechanics of machine-learning find patterns in the data while simultaneously deal with overfitting (i.e. by ensuring that the model is least affected by noise in the data) through different methods such as bootstrap aggregation or regularization (Storm et al., 2020).

Figure 3. Training and testing samples



Source: Authors' illustration.

This paper uses two supervised machine-learning methods to identify the most relevant variables to predict interest in migration, separately for the female and male samples. The first method, Random Forest (Breiman, 2001), provides us with a ranked list of predictors, based on an importance indicator. Some predictors with low correlation with the dependent variable can still have a high importance. Random Forest can choose a variable with a coefficient close to zero (in magnitude) in an OLS regression if omitting this variable would increase the predictive power of the model. Consequently, the top predictors chosen by Random Forest are important from a prediction accuracy point of view but may not all be the most informative as to why men and women are interested in emigrating. In order to also understand this explanatory component, we complement our analysis with a second method, LASSO stability selection, which allows the detection of variables with high coefficient magnitudes. These two methods allow us to create two complementary lists of variables that we can compare in terms of prediction accuracy and economic meaningfulness.

We conduct the analysis in two steps. First, we train the ML models on training samples, separately for women and men, containing 80 percent of randomly selected observations from each sex group. In the second step, we test the performance of the models in testing samples containing the remaining 20 percent of each sex group.

3.2.1. Random Forest

The Random Forest algorithm (Breiman, 2001) is a decision-tree based method. A decision tree consists of successive partitioning of the data: by splitting the sample in two more homogenous groups for each independent variable used in the tree in order to predict a dependent variable (see Figure 4 for an illustration). A Random Forest is an aggregation of multiple decision-trees and can be used to predict binary or continuous variables. The predicted value is calculated from all individual decision trees.

Figure 4. Classification trees and data partitioning

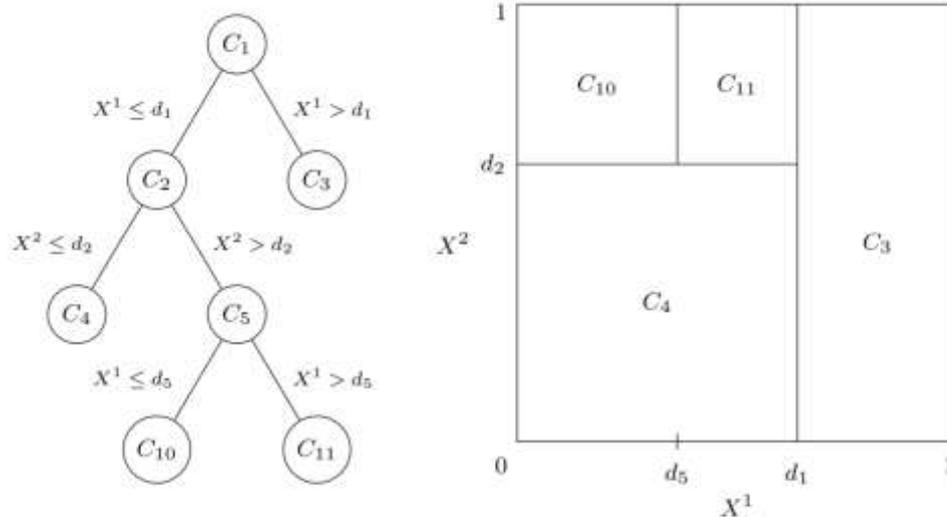


Fig. 2.2 Left: a classification tree to predict the class label corresponding to a given x . Right: the associated partition in the explanatory variables' space, here the unit square (C_1 , C_2 , and C_5 do not appear because they are not associated with leaves)

Source: Schonlau (2020).

To ensure that the model is not prone to overfitting, which is a common issue among tree classifiers (i.e. prediction algorithms based on classification trees), Random Forest uses bootstrap aggregation (or, commonly referred to as “bagging”) which randomizes *which* observations are used in each tree. Random Forest also randomizes the number of variables used in each tree (or, commonly referred as “feature randomness”). Because Random Forest randomizes the subsample of data (pertaining to *both* observations and variables) included in each tree, it can measure an out-of-bag error by testing

the prediction accuracy of each tree in the remaining sample. The decision of which variable will be used to split the data (i.e constitute a *node*) (see Figure 4) and how the node will be split is based on the maximization of the entropy criterion value.⁹

Breiman (2001) shows that the prediction accuracy measured by the out-of-bag error decreases as the number of trees increases. We therefore implement 1,000 bootstrap replications. In Section 5.1, we discuss the robustness of our results to different selections of: (i) bootstrap replications and (ii) the share of observations assigned to training vs. testing samples.

Once all the trees composing the random forest have been generated, the predicting variables can be ranked by importance score. This score measures the increase in the misclassification rate in the forest when the observed values of a given variable are randomly permuted in the out-of-bag samples (Genuer et al., 2010).

3.2.2. LASSO Stability Selection

The second type of supervised machine learning method used in this paper is LASSO Stability Selection. This algorithm aims to detect which variables have the highest probability of being selected by LASSO (Meinshausen & Bühlmann, 2010; Jayachandran et al., 2021), which is a method commonly used to select a reduced set of variables to predict a dependent variable.

LASSO (Least Absolute Shrinkage and Selection Operator) is a model selection method, which aims to balance between goodness of fit and parsimony. It imposes a constraint to the coefficient of the penalty function (the regularization parameter λ), which causes some coefficients to shrink towards zero. In so doing, LASSO keeps variables that are most associated with the dependent variable and drops the least associated ones.

In a nutshell, LASSO Stability Selection performs multiple repetitions of LASSO. The procedure, first, draws a 50 percent subsample without replacement from the training sample, then runs the LASSO algorithm on the complete list of independent variables. To be consistent with the Random Forest method, we choose to run 1,000 replications. We can compute the rate of appearance of each variable being selected in each LASSO run from the 1,000 repetitions. This yields the probability that the variable is chosen by LASSO as a predictor of the dependent variable (in our case, it is the interest to migrate). We can rank predictors in a descending rate of appearance.

⁹ The entropy criterion is given by the formula: $E = -\sum_{i=1}^c p_i \cdot \log(p_i)$ with c being the number of unique classes and p_i the prior probability of each given class (Schonlau, 2020).

3.2.3. Comparison, cross-validation and explanatory variables of the models

To gauge further insights into what the predictive models would suggest as positive and negative determinants of migration, and the strength of association of each determinant, we run a logistic regression of the dependent variable on a reduced set of top predictors identified by each method. The two methods rank predictors based on different indicators: importance in Random Forest and rate of appearance in LASSO Stability Selection. However, they do not provide guidance as to how to define the minimum cutoffs for either the importance or rate of appearance indicators, nor how to define the number of top predictors. In each instance, we need to decide on an arbitrary cutoff of what counts as “top predictors.” To be consistent, we define the same cutoff of 30 highest ranked variables as top predictors for each method. While this is an arbitrary cutoff, Section 5.2 checks the robustness of our results with different cutoffs of top predictors and suggests that 30 is a reasonable number. Note that since we run each method separately for the female and male samples, we can have different top predictors for each gender. This allows us to contrast the predictive and explanatory abilities of each method for each gender.

We run the following three logistic regressions to contrast the two methods:

$$y_{ig} = \alpha_{0g} + \alpha_{1g} RFPredictors_{ig} + \mu_{ig} \quad (1)$$

$$y_{ig} = \beta_{0g} + \beta_{1g} LSSPredictors_{ig} + \omega_{ig} \quad (2)$$

$$y_{ig} = \gamma_{0g} + \gamma_{1g} CommonPredictors_{ig} + \gamma_{2g} RFPredictors_{ig} + \gamma_{3g} LSSPredictors_{ig} + \varepsilon_{ig} \quad (3)$$

where y_{ig} is a dummy indicator for interest in migration for individual i of gender g . $RFPredictors_{ig}$ and $LSSPredictors_{ig}$ are vectors of top 30 predictors as identified by Random Forest (RF) and LASSO Stability Selection (LSS) methods, respectively, for each gender g . The third equation includes the combined set of top 30 predictors identified by both methods, which can be parsed into these three distinct groups: (i) overlapping predictors identified by both methods, $CommonPredictors_{ig}$, (ii) predictors identified exclusively by RF $RFPredictors_{ig}$, and (iii) predictors identified exclusively by LSS, $LSSPredictors_{ig}$. Our main coefficients of interest are α_{1g} , β_{1g} , and γ_{2g} which inform whether each factor is positively or negatively associated with interest in migration.

This completes the first step of our analysis using the randomly selected 80% training subsample (for each gender). Now, we turn to the second step of the analysis using the remaining 20% testing subsample.

We assess the prediction accuracy of our models by running the three logistic regressions (Equation 1-3) and calculating misclassification rates (as described in Figure 5) with the 20% testing sample. The

rationale for doing this step of the analysis with the testing sample is to make sure we are checking the quality of the models with data that was not used to create the model. This allows to check if the model is overfitted. We use two distinct misclassification indicators—accuracy and sensitivity—to compare the predictive power of each model, defined as:

$$Accuracy = \frac{True\ Negatives + True\ Positives}{N}$$

$$Sensitivity = \frac{True\ Positives}{True\ Positives + False\ negatives}$$

where *True Negatives* counts the number of correctly predicted “zeros” and *True Positives* counts the number of correctly predicted “ones”. *N* is the total number of observations in the testing sample.

Figure 5. Prediction accuracy classification

True values Predicted values	0	1
0	True negative	False negative
1	False positive	True positive

Source: Authors

4. Results

Section 4.1 presents the top predictors from Random Forest and LASSO Stability Selection and identifies the predictors that are common for women and men or gender-specific. Section 4.2 then presents logistic regressions of interest in migration to gauge the sign and statistical significance of the top predictors with the dependent variable.

4.1. Top predictors of interest in migration

Table 2 presents the top predictors identified by Random Forest and Table 3 presents the top predictors identified by LASSO Stability Selection. Appendix Table 2 presents the descriptive statistics of these top predictors.

Table 2. Top predictors of interest in migration identified by Random Forest

Predictors	Importance rank from Random Forest		Appearance in 1000 LASSO (in percent)	
	Women	Men	Women	Men
Common predictors for women and men				
Household asset score	1	2	0.6	1.3
Monthly expected wages abroad relative to village mean income	2	1	10.4	62.8
Village Male Labor Force Participation	3	3	88.5	99.8
Village Female Labor Force Participation	4	4	47.8	2.4
Relative cost of migration	5	5	11.3	9.1
Household non-agricultural income	6	6	4.3	25
Age	7	7	98.1	100
Percentage of remittances Respondent would send home if migrated	8	9	47.5	13.6
Knowledge of documents required to work abroad	9	12	60.7	76.5
Lack of knowledge about Indonesian migrants' expected salary in 6 main destination economies (Malaysia; Saudi Arabia; Taiwan, China; Singapore; Hong Kong SAR, China; Brunei Darussalam)	10	16	100	99.9
Highest male education in the household (Years of schooling)	11	21	3	6.3
Proportion of male adults in the household	12	13	2	13.7
Proportion of female adults in the household	13	15	3.8	25.6
Area of land owned by the household	14	17	3	13.1
Index of communication	15	24	12.2	37.7
Highest female education in the household (Years of schooling)	16	14	64.6	5.9
Lack of knowledge about Indonesian migrants' type of work in 6 main destination economies (Malaysia; Saudi Arabia; Taiwan, China; Singapore; Hong Kong SAR, China; Brunei Darussalam)	17	18	34.8	94.8
Years of schooling	18	20	0.1	20.5
Household agricultural income	19	22	2.1	27.6
Respondent total work income (logged)	20	8	0.5	85.6
PM's income from main job (logged)	21	10	25.7	1.6
Share of underemployment in household	22	23	12.1	5.8
Respondent non-agricultural income	23	11	49.6	36.5
Household enterprise operating period	24	25	17.6	2.4
Share of PM's income on total income	25	19	6.2	7.4
Total profit from all household enterprises	26	26	27.7	29
Predictors for women only				
Number of children living with Respondent	27		2.0	
Length of migration of latest returning migrant in the household	28		78.1	
Have friends/acquaintances working: Taiwan, China; Hong Kong SAR, China; Singapore	29		100.0	
Type of work friend/acquaintances are doing abroad: domestic worker	30		88.9	
Predictors for men only				
Access to informal credit		27		1.9
Number of jobs of Respondent		28		25.7
Perceptions of success: Most successful person in the village is someone who has a business		29		8.2
Household benefitted from Rice for the poor or non-cash food assistance program		30		1.9

Source: Authors' calculations from DESMIGRATIF (2018). Note: The rank represents the importance of each variable from Random Forest prediction models implemented separately for women and men with 1000 iterations and 327 variables. The rate of appearance in LASSO Stability Selection is obtained from implementing 1000 LASSO iterations with the 327 variables and calculating the occurrence of each variable in each LASSO iteration.

Table 3. Top predictors of interest in migration identified by LASSO Stability Selection

Predictors	Appearance in 1000 LASSO (in percent)		Importance rank from Random Forest	
	Women	Men	Women	Men
Common predictors for women and men				
Lack of knowledge about Indonesian migrants' expected salary in 6 main destination economies (Malaysia; Saudi Arabia; Taiwan, China; Singapore; Hong Kong SAR, China; Brunei Darussalam)	100.0	99.9	10	16
Respondent doesn't know where to register if interested in working abroad	99.8	98.9	108	106
Perceptions of success: Most successful person in the village is international migrant	99.7	100.0	56	81
Worked abroad in past 5 years	98.9	98.5	73	103
Age	98.1	100.0	7	7
Applied to work abroad in the past 5 years	91.0	100.0	144	121
Literacy in foreign language other than English	90.2	64.7	103	115
Type of work friend/acquaintances are doing abroad: Factory worker	89.6	97.4	76	39
Reason why did not travel abroad despite applying: for family reasons other than caring for children	88.6	64.4	153	201
Village Male Labor Force Participation	88.5	99.8	3	3
Not married	82.7	60.2	87	76
Unemployment within past 12 months	80.4	97.6	93	94
Length of migration of latest returning migrant in the household	78.1	73.9	28	56
English literacy (reading)	75.0	71.2	62	71
No knowledge of number of friends/acquaintances working abroad	73.1	58.1	52	46
Respondent has been contacted by broker in past 5 years	67.7	72.7	135	138
Knowledge of documents required to work abroad	60.7	76.5	9	12
When last abroad, obtained information about opportunities through radio, newspaper, internet	60.2	82.5	121	232
Predictors for women only				
Have friends/acquaintances working in: Taiwan, China; Hong Kong SAR, China; Singapore	100.0		29	
Where Respondent applied to work abroad: broker/sponsor	96.0		139	
Type of work friend/acquaintances are doing abroad: domestic worker	88.9		30	
Reason why latest returning migrant worked abroad: to follow the success of other migrant workers	84.1		185	
Number friends/acquaintances working abroad: more than 10	78.1		94	
Married	76.6		91	
Respondent is a wage worker	75.0		150	
Highest female education in the household (Years of schooling)	64.6		16	
Not looking for employment because currently awaiting call-back from employer	62.2		214	
Number of female international migrants in household	62.1		110	
When Respondent last worked abroad, obtained information about opportunities through social network	61.7		105	
How did latest returning migrant learn about opportunity to work abroad: from family	58.6		141	
Predictors for men only				
Number of male national migrants in the household		96.4		47
Lack of knowledge about Indonesian migrants' type of work in 6 main destination economies (Malaysia; Saudi Arabia; Taiwan, China; Singapore; Hong Kong SAR, China; Brunei Darussalam)		94.8		18
How did latest returning migrant learn about opportunity to work abroad: found information on their own		87.1		212
Main reason why latest returning migrant returned to Indonesia: contract not renewed		86.5		153
Respondent total work income (logged)		85.6		8
PM's migrant network faced issues abroad: long work hours, heavy workload		81.0		215
Madurese ethnicity		80.5		145
Respondent has friends/acquaintances currently working abroad		72.9		51
Reason why current migrant is working abroad: to earn higher income than possible in Indonesia		71.9		172
No knowledge of the risks of working abroad without all the necessary documents		65.5		179
Elementary school diploma		65.0		92
Monthly expected wages abroad relative to village mean income		62.8		1

Source: Authors' calculations from DESMIGRATIF (2018). Note: The rank represents the importance of each variable from Random Forest prediction models implemented separately for women and men with 1000 iterations and 327 variables. The rate of appearance in LASSO Stability Selection is obtained from implementing 1000 LASSO iterations with the 327 variables and calculating the occurrence of each variable in each LASSO iteration.

Tables 2 and 3 show that only a few variables (8 for women and 7 for men) are commonly identified as top 30 predictors by both methods. Moreover, 4 of these predictors (in bold in the following lists) are common for women and men. For women, the predictors identified by both Random Forest and LASSO Stability Selection are:

- **Age**
- Have friends/acquaintances working in Taiwan, China; Hong Kong SAR, China; Singapore
- Highest female education in the household (Years of schooling)
- **Knowledge of documents required to work abroad**
- **Lack of knowledge about expected salaries in 6 main destination economies (Malaysia; Saudi Arabia; Taiwan, China; Singapore; Hong Kong SAR, China; Brunei Darussalam)**
- Duration of migration spell of the latest returning migrant in the household
- Type of work friends/acquaintances are doing abroad: domestic worker
- **Average Male Labor Force Participation in the village**

For men, the predictors identified by both methods are:

- **Age**
- **Knowledge of documents required to work abroad**
- **Lack of knowledge about expected salaries in 6 main destination economies (Malaysia; Saudi Arabia; Taiwan, China; Singapore; Hong Kong SAR, China; Brunei Darussalam)**
- Lack of knowledge about Indonesian migrants' type of work in 6 main destination economies (Malaysia; Saudi Arabia; Taiwan, China; Singapore; Hong Kong SAR, China; Brunei Darussalam)
- Expected monthly income abroad relative to village mean income
- Total work income (logged)
- **Average Male Labor Force Participation (LFP) in the village**

Since these predictors are cross-validated by both methods, they can be considered as the most relevant determinants of interest in migration in our sample. This cross-validation only considers the top 30 predictors in terms of rate of Importance from Random Forest and rate of appearance in LASSO Stability Selection. Considering larger lists would allow to identify other common predictors but these will mechanically have relatively lower predictive and explanatory relevance. Moreover, it is interesting to note that depending on the criteria used (i.e. importance in Random Forest and rate of appearance in LASSO Stability Selection), the lists of predictors identify very different predictors. Indeed, some predictors can appear on top of the list with one given method and have less relevance with the other method as shown when comparing the importance rank (from Random Forest) and the rate of appearance (from LASSO Stability Selection) in each table of predictors. For these reasons, we will analyze both lists, keeping in mind that the underlying idea of the importance criteria is to choose variables with a high predictive power whereas the rate of appearance in LASSO Stability Selection is

based on coefficient sizes and the method eliminates variables that are least associated with the outcome.

Tables 2 and 3 also show that there are both common and gender-specific predictors for interest in migration. Among the variables cross validated by both methods, four are common to both women and men (in bold) and the remaining predictors are specific to each gender group.

In the complete list of top Random Forest predictors presented in Table 2, we can see that 26 out of the 30 variables are common for women and men. This implies that if both groups have similar values for the given variables, they will have the closer levels of interest in migration. However, average interest in migration for women is statistically significantly different from that of men. This suggests that women and men may not share the same sample characteristics. Indeed, among the common predictors, many **individual income** variables are identified (i.e. respondent's total work income, respondent's non-agricultural income and the share of the respondent's income relative to total household income). Appendix Table 2 shows statistically significant differences in average values for women and men, with women systematically having lower levels of income compared to men.

Moreover, the common list of predictors also includes **household-level demographic characteristics**,¹⁰ **economic characteristics**,¹¹ **knowledge about migration**,¹² **migration income expectations**,¹³ and **the index of communication**. Interestingly, **local labor market characteristics** (average female and male LFP in the village) also appear as predictors for both women and men. Four **gender-specific factors**—covering **family structure** and **social network's migration experience**—provide interesting information on women's determinants of interest in migration: number of days since the latest returning migrant in the household returned to Indonesia; number of coresident children; whether respondent has migrant friends/acquaintances working in Taiwan, China; Hong Kong SAR, China; or Singapore; and duration of migration spell of the latest returning migrant in the household.

¹⁰ Specific indicators under this category are proportion of male and female adults in the household and the highest male and female education levels in the household.

¹¹ Specific indicators under this category are household agricultural and non-agricultural income and household asset score.

¹² Specific indicators under this category are the knowledge of which documents are required to work abroad and the lack of knowledge on Indonesian migrants' type of work and salaries in 6 common destination economies specifically mentioned in the survey - i.e., Malaysia; Saudi Arabia; Taiwan, China; Singapore; Hong Kong SAR, China; and Brunei Darussalam.

¹³ Specific indicators under this category are expected monthly wages abroad relative to village mean income; expected cost to migrate abroad relative to expected income from migrating; and expected share of income that an individual would remit back home if they migrate.

LASSO Stability Selection identifies 18 common predictors for women and men and 12 gender-specific predictors. The list of variables for both gender groups shows interesting predictors related to **past migration or migration planning experience**,¹⁴ **social network's migration experience**,¹⁵ **local labor market characteristics and experience**,¹⁶ and **knowledge about migration**.¹⁷ The **gender-specific predictors** include social network and labor market characteristics variables for both women (e.g. having friends or acquaintances working abroad as domestic workers or having more than 10 friends or acquaintances working abroad) and men (e.g. number of male domestic migrants in the household).

¹⁴ Specific indicators under this category are worked or applied to work abroad in the past 5 years; reason why did not travel abroad despite applying: for family reasons other than caring for children; when last worked abroad; obtained information about migration opportunities through radio, newspaper, internet.

¹⁵ Specific indicators under this category are type of work friends/acquaintances are doing abroad: factory worker; duration of migration spell of the latest returning migrant in the household; no knowledge of number of friends/acquaintances working abroad.

¹⁶ Specific indicators under this category are average Male Labor Force Participation in the village; unemployment within past 12 months.

¹⁷ Specific indicators under this category are lack of knowledge about expected salaries in 6 main destination economies – Malaysia; Saudi Arabia; Taiwan, China; Singapore; Hong Kong SAR, China; and Brunei Darussalam; knowledge of documents required to work abroad.

Table 4. Logistic regressions of interest in migration by gender

	Random Forest Predictors - Women	Random Forest Predictors - Men	LSS Predictors - Women	LSS Predictors - Men	All Predictors - Women	All Predictors - Men
Determinants validated by both methods						
Age	-0.036*** (0.006)	-0.033*** (0.006)	-0.033*** (0.006)	-0.043*** (0.007)	-0.028*** (0.007)	-0.044*** (0.007)
Village Male Labor Force Participation	-1.529*** (0.458)	-3.190*** (0.501)	-1.543*** (0.407)	-2.350*** (0.445)	-1.304*** (0.480)	-2.691*** (0.534)
Knowledge of documents required to work abroad	0.051*** (0.012)	0.053*** (0.012)	0.018 (0.013)	0.019 (0.014)	0.015 (0.013)	0.017 (0.014)
Lack of knowledge about Indonesian migrants' expected salary in 6 main destination economies (Malaysia; Saudi Arabia; Taiwan, China; Singapore; Hong Kong SAR, China; Brunei Darussalam)	-0.756*** (0.103)	-0.578*** (0.106)	-0.710*** (0.083)	-0.470*** (0.110)	-0.568*** (0.108)	-0.455*** (0.114)
Highest female education in the household (Years of schooling)	-0.077*** (0.028)	-0.015 (0.010)	-0.050*** (0.012)		-0.079*** (0.029)	-0.011 (0.011)
Length of migration of latest returning migrant in the household	0.000*** (0.000)		0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Monthly expected wages abroad relative to village mean income	-0.017** (0.007)	-0.009** (0.004)		-0.015*** (0.005)	-0.018*** (0.007)	-0.016*** (0.005)
Lack of knowledge about Indonesian migrants' type of work in 6 main destination economies (Malaysia; Saudi Arabia; Taiwan, China; Singapore; Hong Kong SAR, China; Brunei Darussalam)	-0.169 (0.164)	-0.701*** (0.160)		-0.408** (0.170)	-0.320* (0.172)	-0.407** (0.175)
Respondent total work income (logged)	0.028** (0.011)	-0.009 (0.009)		-0.009 (0.006)	0.021* (0.012)	-0.017* (0.010)
Have PM's friends/acquaintances working: Taiwan, China; Hong Kong SAR, China; Singapore	0.786*** (0.082)		0.511*** (0.085)		0.511*** (0.089)	
Type of work friend/acquaintances are doing abroad: domestic worker	0.304***		0.098		0.082	

	(0.078)	(0.083)	(0.084)	
Determinants validated by only one method				
Household asset score	-0.060 (0.040)	0.019 (0.041)	-0.056 (0.042)	0.020 (0.044)
Village Female Labor Force Participation	-0.185 (0.335)	0.617* (0.345)	-0.405 (0.352)	0.436 (0.375)
Relative cost of migration	-0.015 (0.014)	-0.002 (0.010)	-0.018 (0.015)	-0.008 (0.010)
Household non-agricultural income	0.000** (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Percentage of remittances Respondent would send home if migrated	-0.002* (0.001)	-0.001 (0.001)	-0.002 (0.001)	-0.001 (0.001)
Highest male education in the household (Years of schooling)	-0.009 (0.010)	-0.025 (0.028)	-0.002 (0.010)	-0.025 (0.030)
Proportion of male adults in the household	-0.223 (0.262)	0.146 (0.217)	-0.246 (0.273)	-0.020 (0.238)
Proportion of female adults in the household	0.059 (0.248)	-0.252 (0.260)	-0.258 (0.270)	-0.304 (0.280)
Area of land owned by the household	-0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)	0.000 (0.000)
Index of communication	0.517*** (0.129)	0.382*** (0.131)	0.283** (0.139)	0.217 (0.143)
Years of schooling	0.027 (0.027)	0.052** (0.026)	0.032 (0.028)	0.022 (0.029)
Household agricultural income	0.000 (0.000)	-0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)
PM's income from main job (logged)	-0.043*** (0.011)	-0.002 (0.008)	-0.023* (0.012)	0.010 (0.009)
Share of underemployment in household	0.058 (0.076)	-0.070 (0.080)	0.084 (0.080)	-0.052 (0.085)
Respondent non-agricultural income	-0.000*** (0.000)	-0.000 (0.000)	-0.000*** (0.000)	-0.000 (0.000)
Household enterprise operating period	-0.000	-0.000	-0.000	0.000

	(0.000)	(0.000)		(0.000)	(0.000)
Share of PM's income on total income	0.002	-0.119		-0.062	0.047
	(0.156)	(0.131)		(0.164)	(0.142)
Total profit from all household enterprises	-0.000*	-0.000***		-0.000	-0.000**
	(0.000)	(0.000)		(0.000)	(0.000)
Number of children living with Respondent	-0.026			-0.004	
	(0.034)			(0.035)	
Access to informal credit		0.008			0.026
		(0.065)			(0.069)
Number of jobs of Respondent		-0.024			0.027
		(0.069)			(0.076)
Perceptions of success: Most successful person in the village is someone who has a business		0.096			0.066
		(0.065)			(0.069)
Household benefitted from Rice for the poor or non-cash food assistance program		-0.014			0.012
		(0.067)			(0.072)
Respondent doesn't know where to register if interested in working abroad			-0.576***	-0.305***	-0.548***
			(0.106)	(0.106)	(0.107)
Perceptions of success: Most successful person in the village is international migrant			0.416***	0.547***	0.389***
			(0.076)	(0.092)	(0.077)
Respondent has worked abroad in past 5 years			0.291*	0.477***	0.305*
			(0.157)	(0.097)	(0.159)
Respondent has applied to work abroad in the past 5 years			0.911**	1.269***	1.031***
			(0.363)	(0.210)	(0.367)
Literacy in foreign language other than English			0.205**	0.227*	0.219**
			(0.100)	(0.120)	(0.101)
Type of work friend/acquaintances are doing abroad: Factory worker			0.197**	0.218***	0.212**
			(0.093)	(0.084)	(0.094)
Not married			0.163	0.126	0.157
			(0.175)	(0.093)	(0.178)
Unemployment within past 12 months			0.266***	0.256**	0.212**
			(0.087)	(0.109)	(0.091)

English literacy (reading)	0.301*** (0.083)	0.199** (0.087)	0.295*** (0.085)	0.193** (0.094)
No knowledge of number of friends/acquaintances working abroad	-0.518*** (0.090)	-0.620 (0.417)	-0.513*** (0.091)	-0.613 (0.419)
Respondent has been contacted by broker in past 5 years	0.252** (0.120)	0.337*** (0.120)	0.270** (0.121)	0.352*** (0.120)
When Respondent last abroad, obtained information about opportunities through radio, newspaper, internet	0.540** (0.242)	0.711** (0.338)	0.553** (0.245)	0.762** (0.342)
Where Respondent applied to work abroad: broker/sponsor	1.144** (0.546)		0.998* (0.549)	
Reason why latest returning migrant worked abroad: to follow the success of other migrant workers	0.458 (0.303)		0.472 (0.307)	
Number friends/acquaintances working abroad: more than 10	0.162 (0.102)		0.175* (0.103)	
Married	-0.302** (0.152)		-0.342** (0.157)	
Respondent is a wage worker	-0.346*** (0.113)		-0.285** (0.129)	
Not looking for employment because currently awaiting call back from employer	0.648* (0.349)		0.656* (0.354)	
Number of female international migrants in household	0.070 (0.120)		0.059 (0.123)	
When Respondent last abroad, obtained information about opportunities through social network	0.453** (0.228)		0.465** (0.230)	
How did latest returning migrant learn about opportunity to work abroad: from family	0.343* (0.179)		0.342* (0.182)	
Number of male national migrants in the household		0.309*** (0.068)		0.317*** (0.069)

How did latest returning migrant learn about opportunity to work abroad: found information on their own				1.209** (0.485)		1.247** (0.485)
Main reason why latest returning migrant returned to Indonesia: contract not renewed				0.266 (0.187)		0.264 (0.189)
Respondent's migrant network faced issues abroad: long work hours, heavy workload				0.908*** (0.285)		0.878*** (0.290)
Madurese ethnicity				-0.529*** (0.144)		-0.529*** (0.149)
Respondent has friends/acquaintances currently working abroad				-0.092 (0.423)		-0.093 (0.425)
Reason why current migrant is working abroad: to earn higher income than possible in Indonesia				0.613*** (0.191)		0.601*** (0.195)
No knowledge of the risks of working abroad without all the necessary documents				-0.438** (0.187)		-0.455** (0.188)
Elementary school diploma				-0.206** (0.095)		-0.212* (0.121)
Reason why Respondent did not travel abroad despite applying: for family reason other than caring for children				1.743* (1.049)		1.712 (1.052)
Constant	1.454*** (0.410)	2.827*** (0.438)	1.302*** (0.439)	2.585*** (0.601)	1.551*** (0.469)	2.900*** (0.668)
Observations	8,186	5,093	8,059	5,049	8,059	5,011

Source: Authors' calculations from DESMIGRATIF (2018). Note: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

4.2. A data-driven logistic model of interest in migration

Using predictors identified in Section 4.1., we implement logistic regressions of interest in migration described in equations (1) to (3). Table 4 shows the estimation results for separate female and male sub-samples. These results allow us to analyze the sign, statistical significance, and magnitude of the predictors.

Among cross-validated determinants identified by both RF and LSS, Table 4 shows that most of them have statistically significant relationships with interest in migration. Age has a negative effect on willingness to migrate, indicating that younger individuals have a higher interest in migration than older ones, echoing findings from Milasi (2020). Moreover, the lack of knowledge on Indonesian migrants' type of work and salaries in 6 common destination countries are negatively correlated to the interest in migration. For both women and men, the signs of coefficients are negative. Indeed, Table 4 shows that men are less likely to be interested in migration by 44 percent¹⁸ and 51 percent, respectively, when they have no knowledge of the salary and type of occupation that Indonesian migrants do in those 6 countries. Women are 51 percent less likely to want to migrate when they have no knowledge of Indonesian migrants' salaries. Knowledge of documents required to work abroad is only statistically significant in the Random Forest specification, showing a positive (albeit small) relationship with interest in migration. It is worth noting that our results do not indicate causal directions. For example, knowledge about migration processes may boost interest in migration, but it is also possible that only people who are interested in migration have sought knowledge on migration. Further research is needed to identify the causal pathways.

The gender-specific factors among the cross-validated determinants indicate that women who have friends or acquaintances working in Taiwan, Hong Kong, Singapore, which are high-income countries geographically close to Indonesia, are more likely to be interested in migration. Having friends or acquaintances working as domestic workers is also positively related to migration interest for women. These results show the importance of social networks and existence of work opportunities abroad as pull factors for women and the gendered aspects of migration, in terms of destination countries and occupations.

The negative sign of expected monthly wages relative to village mean income is surprising. Indeed, we would expect a positive relationship between expected wages abroad and interest in migration. Since this indicator is a ratio, the negative sign can be explained by negative association of the numerator (expected monthly wages) with interest in migration, or a positive association of the denominator

¹⁸ Odds ratios are calculated by exponentiating the coefficients presented in Table 4.

(village mean income) with interest in migration. Further explorations shows that the numerator does not have the expected negative association. Indeed, people who have no interest in migration declare very high values of expected monthly wages abroad, reaching up to IDR 500 million (or around USD 34,700).¹⁹ These unrealistic wages declared by those uninterested in migration—potentially driven by lack of knowledge about salaries abroad—may explain the negative and statistically significant coefficient. Moreover, the variance of the expected monthly wages also shows higher noise for those who are not interested in migration in comparison to those who are interested.

Most determinants have the same sign relationship with interest in migration for both women and men sub-samples. Note however that this is not the case for average female LFP in the village which has no statistically significant relationship with interest in migration for women but a negative one for men (statistically significant at the 10 percent level). Average male LFP in the village on the other hand is statistically significantly and negatively associated with interest in migration for both women and men. Concerning men, these results suggest that the lack of local opportunities may be an incentive to migrate. However, women's LFP being positively related to migration interest for men calls for further research as it could reflect the phenomenon where men exit from the labor market when it gets saturated with women and switch to another labor market.

Concerning the liquidity constraints hypothesis, impeding migration as shown by Bazzi (2017), our results show that total individual income is statistically significantly associated with an increased interest in migration for women, but not for men, for whom the coefficient is negative (not statistically significant in the Random Forest Predictors regression and statistically significant at the 10 percent level in the All Predictors regression). However, income from main job is negatively associated with interest in migration for women. These results suggest that after accounting for the income from the main job, the additional income from extra jobs may provide women with more liquidity and increases their interest in migration.

Among the predictors identified by only one of the methods, the index of communication is an interesting determinant, which to our knowledge has not received much attention in the literature. This index represents access to various means of telecommunication technology. Table 4 shows that it has a positive relationship with interest in migration. Similar to knowledge about migration, the direction of this relationship is unclear and calls for further research. Access to technology may expand one's horizon and awareness of different migration opportunities, processes, and risks and rewards from migration. Alternatively, it may expand the social networks circle, allowing exchanges of ideas

¹⁹ Using an exchange rate of 1 USD = 14,380 IDR (per Dec 31, 2018).

and experiences with migrants abroad. For example, Aker, Clemens, Ksoll (2011) find that randomizing training to use simple phones among adult students increased their likelihood of seasonal migration. The perception of success is another interesting determinant for migration interest. Indeed, individuals who believe that the most successful person in the village is an international migrant are more likely to be interested in migration.

We find that past migration experience is a strong determinant for interest in migration. Respondents who have worked abroad in the past five years are more likely to want to migrate. This can either suggest: (a) path dependency, where former migrants are more likely to do another spell of migration, or (b) potential dissatisfaction from the local labor market upon return.

Finally, the results indicate many statistically significant social network variables, suggesting that knowing migrants who are either family or friends increase interest in migration, echoing findings from other countries (eg. Richter and Taylor, 2008). Moreover, our results confirm the relevance of including labor smugglers in the scope of social network related to migration, following Krissman (2005). Indeed, respondents who have been contacted by a broker are significantly more likely to be interested in migration than their peers.

5. Robustness checks and discussion

5.1. Model performance

Table 5 presents performance indicators of the logistic models implemented using the top predictors identified by Random Forest and LASSO Stability Selection. These models are estimated on the testing sample (i.e. 20 percent of observations not included in the training of the models and the determination of the top predictors). The table compares the accuracy and sensitivity of the logistic models which respectively indicate the share of correctly predicted cases and the share of correctly predicted cases taking the value of 1.

Table 5. Performance indicators of the logistic regression models

Model	Accuracy	Sensitivity
Female logistic regression model with top 30 Random Forest predictors	81.64%	16.67%
Male logistic regression model with top 30 Random Forest predictors	72.00%	26.33%
Female logistic regression model with top 30 LASSO Stability Selection predictors	82.52%	20.97%
Male logistic regression model with top 30 LASSO Stability Selection predictors	70.57%	15.46%

Source: Authors' calculations from DESMIGRATIF (2018).

Table 5 shows an encouraging rate of correctly classified cases (accuracy rates ranging from 70.57 to 81.64 percent). However, it is important to note that the sensitivity measure, which is relevant in cases where the dependent variable is less likely to take the value of 1 compared to 0, shows that overall, the logistic models, have a sensitivity level ranging from 15.46 to 26.33 percent. These levels, although quite low, are similar to Sansone (2019) who uses LASSO to identify predictors that are then used in logistic regressions to measure school dropout. In their study, the sensitivity levels of the logistic models range from 23 to 28 percent. For women, the performance of the logistic models using LASSO Stability Selection predictors is better than one using Random Forest predictors, with higher levels of accuracy and sensitivity. For men, however, the reverse is true. Considering separate women and men subsamples is therefore a good option to choose the best model for prediction purposes. The low levels of sensitivity also show that aspirations to migrate may be difficult to predict, even with a large set of explanatory variables.²⁰

Table 6. Performance indicators of the Random Forest models

Model	Out-of-bag error	Accuracy	Sensitivity
Female Random Forest model using all variables	17.6%	19.21%	0.32%
Male Random Forest model using all variables	17.55%	31.74%	6.25%
Female Random Forest model using top 30 predictors	30.04%	19.15%	0.96%
Male Random Forest model using top 30 predictors	29.99%	32.5%	6.31%

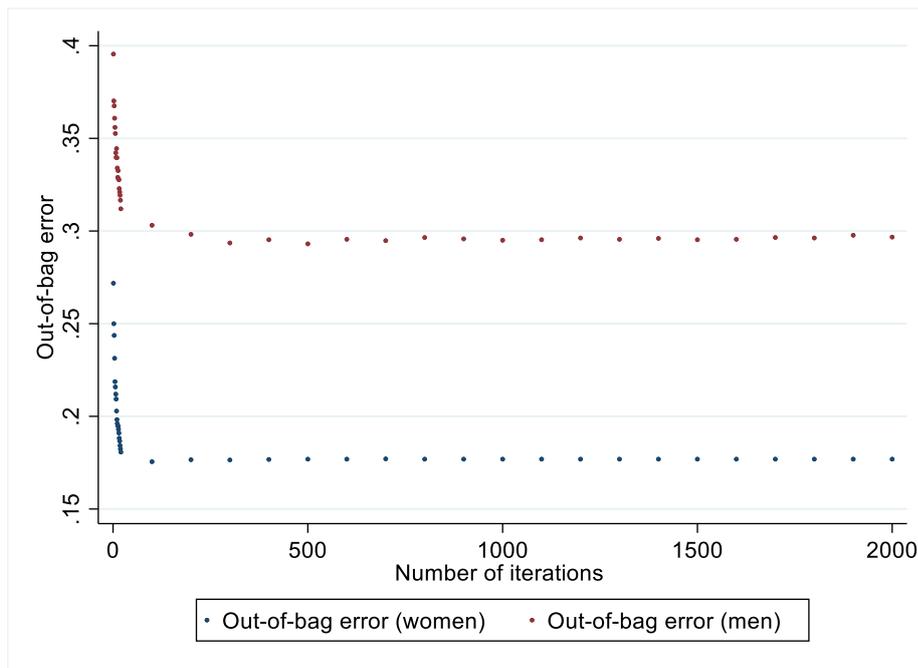
Source: Authors' calculations from DESMIGRATIF (2018).

In addition, we check the robustness of our results using Random Forests' internal error measurement. This built-in feature is not available for the LASSO Stability Selection algorithm. Table 6 shows the performance of the Random Forest models using accuracy and sensitivity indicators as well as the out-of-bag error. The accuracy and sensitivity are calculated in the testing sample whereas the out-of-bag error is calculated during the bootstrap aggregation process of the Random Forest, in the training sample. The table compares the performance of Random Forest using the complete list of 301 variables as well as restricted models with only the top 30 predictors. The table shows that the

²⁰ Note that we also tested the accuracy and sensitivity levels of logistic estimations with top predictors obtained from logistic LASSO instead of linear LASSO. These results (available upon request) show no improvement in prediction accuracy.

Random Forest performs better for women compared to men, with out-of-bag errors of approximately 17 percent, compared to the male Random Forest models where the out-of-bag errors are approximately 30 percent. The performance of the Random Forest models with all 301 variables is similar to the restricted one with only top 30 predictors, suggesting high predictive power of the top 30 predictors.

Figure 6. Out-of-bag error by number of iterations



Source: Authors' calculations from DESMIGRATIF (2018). Note: the out-of-bag errors are obtained from separate Random Forest models in which different number of iterations are included. All the models contain the 301 variables.

However, the predictive power of the Random Forest as suggested by this internal error measurement is much lower than suggested earlier in Table 5, using logistical regressions. Indeed, the sensitivity levels are around zero for women and approximately 6 percent for men, which mean that the model does not predict well who is interested in migrating in the testing sample. Note that the sensitivity level is 100 percent in the training sample. This poor performance could be due to the model being over-trained or the small testing sample. To gauge which reason is potentially driving our poor performance, we compared the sensitivity levels for: (i) different numbers of iterations and (ii) with a different split of testing and training samples (50 percent of randomly selected observations in each group). The performance of the models do not improve when implementing these changes (Appendix Table 3). We also implemented Random Forest models with different depth (number of variables) and

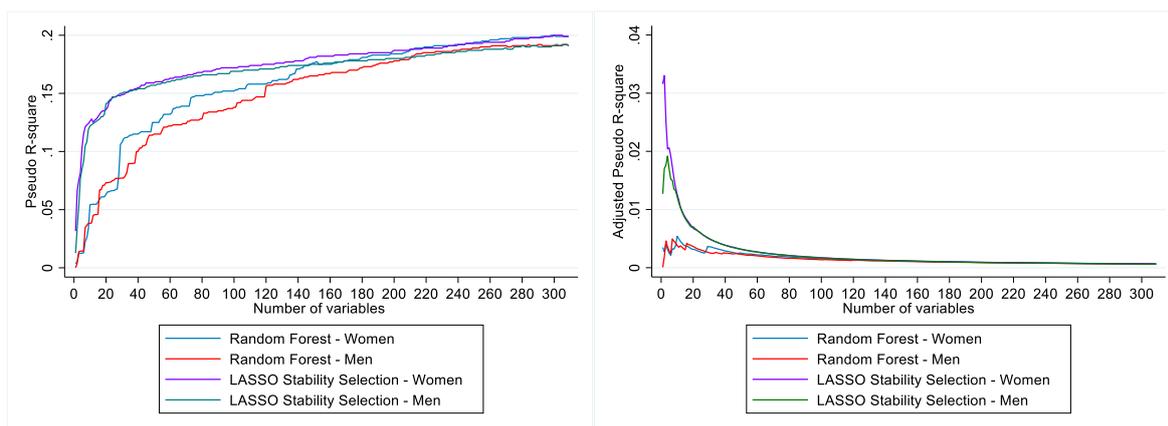
leaf sizes (number of leaves) and the models do not improve.²¹ Finally, we check the relationship between the number of iterations (trees) and the out-of-bag error. Figure 6 shows that the out-of-bag error reduces with the number of iterations and stabilizes for both female and male samples from approximately 100 iterations. Therefore, implementing more iterations could not lead to lower out-of-bag errors since we already implement 1,000 iterations in our base specification.

Overall, the logistic regression models have a stronger predictive power than the Random Forest model. This implies that if provided with the right list of predictors, one could use simple logistic regression framework to predict aspirations to migrate. Moreover, the method presented in this study may be better suited for explanatory than predictive purposes and further research on the best suited predictive method should be conducted. Nevertheless, it is possible that migration interest is structurally hard to predict and that no large predictive gains will be possible even with alternative models.

5.2. Number of predictors

In this study, we choose to retain only the top 30 predictors from Random Forest and LASSO Stability Selection. Considering that we have 301 variables, more top predictors could be analyzed. However, each additional predictor has a lower relevance on the margin. Moreover, since we use these predictors in a regression framework, we mitigate the chance of bias from multicollinearity by limiting the number of predictors. Hence, we argue that 30 is a reasonable number.

Figure 7. Explanatory power of models with different number of predictors



Source: Authors' calculations from DESMIGRATIF (2018)

Figure 7 shows the pseudo R-squared and adjusted pseudo R-squared from logistic regressions of interest in migration where independent variables are sequentially included based on their

²¹ Results available upon request.

importance ranking (decreasing order) from Random Forest and rate of appearance ranking (decreasing order) from LASSO Stability Selection. The concavity of the pseudo R-square curves indicate that there is no substantial gain from additional predictors (beyond our lists of 30 predictors) from each method. The inflection points for adjusted R-square are around 10 to 15 variables, lower than our base specification with 30 variables, which also indicates that we are not losing from the data by capping it at 30 variables.

Interestingly, the maximum value of the R-squared never exceeds 0.19 for men and women, which also suggest that a large part of the variance in the data remains unexplained, even when adding the complete list of 301 variables in the logistic regression. This argument complements the discussion on predictive power of the model in Section 5.1.

5.3. Migration interest and actual migration

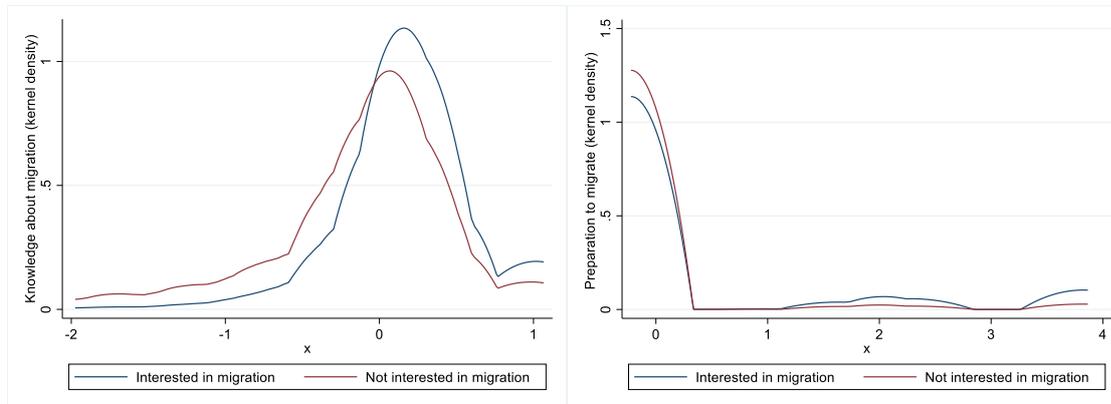
The positive correlation between interest in migration and actual migration has been shown in the literature (Creighton, 2013; Milasi, 2020; van Dalen & Henkens, 2013). Interest in migration illustrates aspirations that are less impacted by physical mobility barriers than actual migration. The relevance of studying interest in migration is therefore to overcome the selection bias of only observing migrants. For instance, if such predictions are aimed at better targeting potential migrants to provide them with information on safe migration processes, using actual past migration to do so would suffer from important endogeneity biases, blurring the profile of potential migrants. An interesting addition to our analysis would be to observe the extent to which individuals who are interested in migration are able to materialize this aspiration and whether the predictors of interest in migration perform well to predict actual migration.

In order to look further into the relationship between interest in migration and more materialized migration plans, we compute two indices aggregating different variables for each family of outcomes following Kling, Liebman, and Katz (2007): (i) knowledge about migration process index and (ii) preparation to migrate index.²²

²² The knowledge about the migration process index is constructed using these variables: don't know where to look for information on work abroad (reverse coded); don't know where to register for work abroad, if interested (reverse coded); don't know which documents are required to work abroad (reverse coded); don't know the risks associated with working abroad without all the necessary documents (reverse coded); know whether migrant friends/acquaintances faced any problem in the process of migration; don't know what type of problem their friends/acquaintances faced (reverse coded).

Preparation to migrate index is constructed using these variables: completed a health test before migrating in the past; don't know about health test required by the official migration process (reverse coded); completed a training before migrating in the past; don't know about training required by the official migration process (reverse coded); don't know the content of the training (reverse coded); participated in pre-departure sessions; don't know about any pre-departure sessions required by the official migration process (reverse coded).

Figure 8. Kernel density plots of knowledge and preparation about migration by interest in migration



Source: Authors' calculations from DESMIGRATIF (2018)

Figure 8 shows kernel density plots of the indices by groups of interest to migrate. We can see that knowledge about migration seems to be higher for those interested in migrating. The preparation to migrate plot however shows a weaker relationship with interest in migration: those who are not interested in migration have a higher density for lower scores of the index and those who are interested in migration have slightly higher densities for higher scores.

Moreover, a phone survey conducted in 2020 among 3,224 respondents of our sample provides us with information on migration for a subset of individuals. Indeed, 55 potential migrants from the sample migrated. First of all, there is a 10 percent (significant at 1 percent level) correlation between interest in migration and actual migration. This correlation level should however be interpreted with caution considering that there is a potential negative selection of migrants in the sample of the survey.

6. Comparing the data-driven approach and theory-driven approach

To verify the relevance of machine-learning methods, we implement a logistic regression model of interest in migration using variables identified in the literature. We draw a list of variables from Milasi (2020), which estimates the determinants for the desire to migrate among young people from 139 countries. Milasi (2020) groups the determinants of migration in the following categories: Individual Characteristics (female, age, have kids, urban, poor health, education), Wealth (lack of food or shelter, difficulty living on present income, no improvement in living standards), Network (no opportunity to meet people, no help from friends/relatives), Context (discontent with local amenities, discontent with local education system, no trust in national government, corruption in government), Labor

Market (worsening economy, bad time to find a job, not getting ahead by working hard, unemployed, out of workforce). We use variables that are similar or that we believe represent good proxies of the ones used in Milasi (2020). A few of the variables cannot be proxied by any of the variables we have in our data. In such cases, we exclude these variables from the regression. The variables available in our data overlap quite well with the specification of Milasi (2020) in terms of push factors, to the exception of social networks. Our data only contains information on social networks abroad. For this reason, we replace the network variables by the number of acquaintances abroad. Moreover, we do not have any variables that can capture context, except for village-level FLFP and MLFP, which provide some information on local labor market opportunities.

Interest in migration can therefore be estimated by the logistic model presented in Equation 4.

$$y_{ig} = \beta_{0g} + \beta_{1g} LiteraturePredictors_{ig} + \delta_{ig} \quad (4)$$

Where *LiteraturePredictors* is a vector that includes age, having a child dummy variable, years of schooling, household asset score, whether the household benefits from a social program dummy, number of friends/acquaintances abroad, labor market status, village male and female LFP rates.

The predictive power of this model, shown in Table 7 is smaller than the one obtained from the ML determinants model.²³ This result suggest a better predictive power of the ML models but note that this result can be due to the fact that there are fewer variables included in this model than in the ML models.

Table 7. Performance indicators of the literature-driven models

Model	Accuracy	Sensitivity
Female logistic regression with variable list derived from the literature	80.84%	2.56%
Male logistic regression with variable list derived from the literature	68.65%	18.13%

Source: Authors' calculations from DESMIGRATIF (2018).

²³ The estimation results are presented in Appendix Table 4.

7. Conclusion

This study uses Random Forest and LASSO Stability Selection to identify the top determinants of interest in migration among Indonesian women and men. Random Forest relies on an importance criterion, choosing variables with a high predictive power, while LASSO Stability Selection uses the rate of appearance criterion based on coefficient sizes association of the independent variables with the outcome variable. We find a large set of common predictors for women and men, such as local labor market characteristics and individual economic characteristics (e.g. income). The models also identify gender-specific predictors. For instance, knowing Indonesian factory workers abroad is a predictor for both women and men whereas knowing Indonesian domestic workers abroad is only a predictor for women. Our models confirm drivers widely documented in previous literature (e.g. social networks) and less explored factors (e.g. access to communication) as predictors for migration aspiration. For policy makers, understanding the gender-specific drivers of migration in a given country and being able to rank them can provide valuable information. Not only does it allow to establish a profile of potential migrants who need to be targeted for specific policies such as information on safe migration, but it also allows to assess whether policy makers should address the “push” factors that lead people to emigrate and encourage local labor market participation, or encourage migration instead. An interesting extension of this research is to be able to use the list of identified predictors with other datasets that may not contain as much information as available in our rich household survey data. This feature would be particularly interesting for policy makers aiming at identifying potential migrants using readily available household surveys. Further research is needed to improve the prediction accuracy of the models. The validity of our identified variables list is likely time- and country-specific, depending on the development level of the region and historical, political and social characteristics of countries. The results presented in this study should therefore be read in light of these considerations.

References

- Bartolini, L., Gropas, R., & Triandafyllidou, A. (2017). Drivers of highly skilled mobility from Southern Europe: escaping the crisis and emancipating oneself. *Journal of Ethnic and Migration Studies*, 43(4), 652–673.
- Bastia, T., & Haagsman, K. (2020). Gender, migration and development. In T. Bastia & R. Skeldon (Eds.), *Routledge Handbook of Migration and Development*. Routledge.
- Bazzi, S. (2017). Wealth Heterogeneity and the Income Elasticity of Migration. *American Economic Journal: Applied Economics*, 9(2), 219–255.
- Bryan, G., Chowdhury, S. and Mobarak, A.M. (2014), Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh. *Econometrica*, 82: 1671-1748.
- Blumenstock, J., Chi, G., & Tan, X. (2019). Migration and the Value of Social Networks (CEPR Discussion Papers, Issue 13611). C.E.P.R. Discussion Papers. <https://doi.org/DOI>:
- Botea, I., Chakravarty, S., & Compernelle, N. (2018). Female Migration in Lesotho: Determinants and Opportunities. In *Policy Research Working Paper (Issue 8307)*. World Bank, Washington, DC.
- Boyd, M. (1989). Family and Personal Networks in International Migration: Recent Developments and New Agendas. *The International Migration Review*, 23(3), 638–670.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 5–32.
- Cairns, D., & Smyth, J. (2011). I wouldn't mind moving actually: Exploring Student Mobility in Northern Ireland. *International Migration*, 49(2), 135–161.
- Cengiz, D., Dube, A., Lindner, A., & Zentler-Munro, D. (2021). Seeing Beyond the Trees: Using Machine Learning to Estimate the Impact of Minimum Wages on Labor Market Outcomes.
- Cerrutti, M., & Massey, D. S. (2001). On the auspices of female migration from Mexico to the United states. *Demography*, 38(2), 187–200.
- Creighton, M. J. (2013). The role of aspirations in domestic and international migration. *Social Science Journal*, 50(1), 79–88.
- de Grip, A., Fouarge, D., & Sauermann, J. (2010). What affects international migration of European science and engineering graduates? *Economics of Innovation and New Technology*, 19(5), 407–421.
- De Haas, H. (2010). Migration and development: A theoretical perspective. *International Migration Review*, 44(1), 227–264.
- Dibeh, G., Fakh, A., & Marrouch, W. (2019). Labor market and institutional drivers of youth irregular migration in the Middle East and North Africa region. *Journal of Industrial Relations*, 61(2), 225–251.
- Docquier, F., Tansel, A., & Turati, R. (2020). Do Emigrants Self-Select Along Cultural Traits? Evidence from the MENA Countries. *International Migration Review*, 54(2), 388–422.
- Donato, K. M., & Gabaccia, D. (2015). *Gender and International Migration*. Russell Sage Foundation.
- Etling, A., Backeberg, L., & Tholen, J. (2020). The political dimension of young people's migration intentions: evidence from the Arab Mediterranean region. *Journal of Ethnic and Migration Studies*, 46(7), 1388–1404

- Ferrant, G., Tuccio, M., Loiseau, E., & Nowacka, K. (2014). The role of discriminatory social institutions in female South-South migration (Issue April). www.genderindex.org
- Fleury, A. (2016). Understanding Women and Migration: A Literature Review. KNOMAD Working Paper Series, February, 48. [http://www.knomad.org/docs/gender/KNOMAD Working Paper 8 final_Formatted.pdf](http://www.knomad.org/docs/gender/KNOMAD%20Working%20Paper%208%20final_Formatted.pdf)
- Fouarge, D., & Ester, P. (2007). Factors determining international and regional migration in Europe (European Foundation for the Improvement of Living and Working Conditions 2007).
- Gao, C., Fei, C. J., McCarl, B. A., & Leatham, D. J. (2020). Identifying vulnerable households using machine-learning. *Sustainability (Switzerland)*, 12(15), 1–18.
- Ghosh, J. (2009). Migration and Gender Empowerment: Recent Trends and Emerging Issues. *Human Development Research Paper*, 04.
- Gibson, J., & McKenzie, D. (2011). The microeconomic determinants of emigration and return migration of the best and brightest: Evidence from the Pacific. *Journal of Development Economics*, 95(1), 18–29.
- Gubert, F., & Senne, J.-N. (2016). Is the European Union attractive for potential migrants? An investigation of migration intentions across the world. *OECD Social, Employment and Migration Working Papers*, 188, 39.
- Heering, L., van der Erf, R., & van Wissen, L. (2004). The role of family networks and migration culture in the continuation of Moroccan emigration: a gender perspective. *Journal of Ethnic and Migration Studies*, 30(2), 323–337.
- Jayachandran, S., Biradavolu, M., Cooper, J., & Health, P. (2021). Using machine learning and qualitative interviews to design a five-question survey module for women ' s agency. *March*.
- Kahanec, M., & Fabo, B. (2013). Migration strategies of crisis-stricken youth in an enlarged European Union. *Transfer: European Review of Labour and Research*, 19(3), 365–380.
- Krissman, F. (2005). Sin Coyote Ni Patrón: Why the “Migrant Network” Fails to Explain International Migration. *International Migration Review*, 39(1), 4–44.
- Lee, E. S. (1966). A Theory of Migration. *Demography*, 3(1), 47–57. <https://www.jstor.org/stable/2060063>
- Lentz, E. C., Michelson, H., Baylis, K., & Zhou, Y. (2019). A data-driven approach improves food insecurity crisis prediction. *World Development*, 122, 399–409.
- Makovec, M., Purnamasari, R. S., Sandi, M., & Savitri, A. R. (2018). Intended versus unintended consequences of migration restriction policies: evidence from a natural experiment in Indonesia. *Journal of Economic Geography*, 18(4), 915–950.
- Meinshausen, N., & Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4), 417–473.
- Milasi, S. (2020). What Drives Youth’s Intention to Migrate Abroad? Evidence from International Survey Data. *IZA Journal of Development and Migration*, 11(1).
- Moretto, M., & Vergalli, S. (2008). Migration dynamics. *Journal of Economics*, 93(3), 223–265.

- Naudé, W. (2010). The determinants of migration from sub-saharan african countries. *Journal of African Economies*, 19(3), 330–356.
- Ruysen, I., & Salomone, S. (2018). Female migration: A way out of discrimination? *Journal of Development Economics*, 130(October 2017), 224–241.
- Sansone, D. (2019). Beyond Early Warning Indicators: High School Dropout and Machine Learning. *Oxford Bulletin of Economics and Statistics*, 81(2), 456–485.
- Sjaastad, L. A. (1962). The Costs and Returns of Human Migration. *The Journal of Political Economy*, LXX(5), 80–93.
- Smith, M. D., & Floro, M. S. (2020). Food insecurity, gender, and international migration in low- and middle-income countries. *Food Policy*, 91(January 2019), 101837.
- Sobiech, I. (2019). Remittances, finance and growth: Does financial development foster the impact of remittances on economic growth? *World Development*, 113, 44–59.
- Storm, H., Baylis, K., & Heckelei, T. (2020). Machine learning in agricultural and applied economics. *European Review of Agricultural Economics*, 47(3), 849–892.
- van Dalen, H. P., & Henkens, K. (2013). Explaining emigration intentions and behaviour in the Netherlands, 2005–10. *Population Studies*, 67(2), 225–241.
- Van Mol, C. (2016). Migration aspirations of European youth in times of crisis. *Journal of Youth Studies*, 19(10), 1303–1320.
- World Bank. (2017). Indonesia' s Global Workers Juggling Opportunities & Risks. In *The World Bank*.
- World Bank. (2018). *Moving for Prosperity: Global Migration and Labor Markets*. Washington, DC: World Bank.
- World Bank. (2019). *Leveraging Economic Migration for Development: A Briefing for the World Bank Board* (Issue September). World Bank.

Appendix

Appendix Table 1. List of independent variables

Access to cooperative credit

Access to formal credit

Access to informal credit

Active occupied

Age

Amount of money spent by highest-earning current migrant to access employment opportunity

Amount respondent was expected to pay to migrate

At least one of the highest-earning current migrant's children has been required to repeat a year of school after left in Indonesia by migrant

Became unemployed in last 12 months

Broker offered respondent a position in: Hong Kong SAR, China; Taiwan, China; Japan; the Republic of Korea or Singapore

Broker offered respondent a position in: Malaysia or Brunei Darussalam

Broker offered respondent a position in: Saudi Arabia, Kuwait, Qatar, Oman, UAE or Bahrain

Broker offered respondent a position in: USA

Country in which respondent's friends/acquaintances are working: Hong Kong SAR, China; Taiwan, China; Japan; the Republic of Korea; Singapore

Country in which respondent's friends/acquaintances are working: Malaysia, Brunei Darussalam

Country in which respondent's friends/acquaintances are working: Saudi Arabia, Kuwait, Qatar, Oman, UAE, Bahrain

Country in which respondent's friends/acquaintances are working: USA

Days since latest returning migrant came back

Destination of highest earning current migrant (Hong Kong SAR, China; Taiwan, China; Japan; the Republic of Korea; Singapore)

Destination of highest earning current migrant (Malaysia, Brunei Darussalam)

Destination of highest earning current migrant (Saudi Arabia, Kuwait, Qatar, Oman, UAE, Bahrain)

Destination of highest earning current migrant (USA)

Destination of latest returning migrant (Hong Kong SAR, China; Taiwan, China; Japan; the Republic of Korea; Singapore)

Destination of latest returning migrant (Malaysia, Brunei Darussalam)

Destination of latest returning migrant (Saudi Arabia, Kuwait, Qatar, Oman, UAE, Bahrain)

Destination of latest returning migrant (USA)

Diploma (elementary)

Diploma (higher secondary)

Diploma (lower secondary)

Diploma (tertiary)

Divorced or widowed

Does this household benefit from, or has it benefited from social protection programs (Kartu Sehat, ASKESKIN, JAMKESMAS, KIS, BPJS or JKN)?

Does this household benefit from, or has it benefited from, the BSM (Assistance for Poor Students) or KIP (Smart Indonesia Card) programs?

Does this household benefit from, or has it benefited from, the JSLU (Social Welfare for the Elderly) program?

Does this household benefit from, or has it benefited from, the PKH (Family of Hope) program?

Does this household benefit from, or has it benefited from, the PKSA (Social Welfare for Children) program?

Does this household benefit from, or has it benefited from, the Rice for the Poor (RASKIN) and Non-Cash Food Assistance programs?

Does this household benefit from, or has it benefited from, the Social Welfare for the Disabled program?

Does this household have, or has it had, a Declaration of Poverty (SKTM)?

Economic shock in last two years: bankruptcy of a household member

Economic shock in last two years: death of Hoh or main provider

Economic shock in last two years: failed harvest

Economic shock in last two years: serious illness of Hoh or main provider

Ethnicity (Flores)

Ethnicity (Javanese)

Ethnicity (Madurese)

Ethnicity (Sundanese)

Ethnicity (Timor)

Female

Female child in household (below 3)

Female child in household (between 4 and 6)

Female child in household (between 7 and 11)

Full time employment

Has the respondent's friends/acquaintances faced any problems abroad

Highest female education in the household (years of schooling)

Highest male education in the household (years of schooling)

Highest monthly wage among current migrants in the household

Highest-earning current migrant in household has left children below 18 in Indonesia

Household involved in the production of agricultural products over the past 12 months

Household asset score

Household does not use electricity

Household enterprise employs at least one individual

Household frequently communicates with highest earning current migrant

Household has cooperative savings

Household informal savings

Household involved in agri-production for subsistence

Household member has a savings bank account

household member use internet to communicate with highest earning current migrant

Household's non-agricultural income

Household's non-work income

How happy are the children of highest earning current migrant to speak with them? (score 1-5)

How highest-earning learned about job opportunity abroad: don't know

How highest-earning learned about job opportunity abroad: found information on their own

How highest-earning learned about job opportunity abroad: from broker/sponsor

How highest-earning learned about job opportunity abroad: from company/employer

How highest-earning learned about job opportunity abroad: from family

How highest-earning learned about job opportunity abroad: from friends/neighbor

How highest-earning learned about job opportunity abroad: from migrant labor placement agency

How highest-earning learned about job opportunity abroad: from private agent

How highest-earning learned about job opportunity abroad: from schools/ skills education institution

How highest-earning learned about job opportunity abroad: other

How latest returning migrant gained employment abroad: domestic agent

How latest returning migrant gained employment abroad: employment placement agency

How latest returning migrant gained employment abroad: informal pathways

How latest returning migrant learned about job opportunity abroad: don't know

How latest returning migrant learned about job opportunity abroad: found information on their own

How latest returning migrant learned about job opportunity abroad: from broker/sponsor

How latest returning migrant learned about job opportunity abroad: from company/employer

How latest returning migrant learned about job opportunity abroad: from family

How latest returning migrant learned about job opportunity abroad: from friends/neighbor

How latest returning migrant learned about job opportunity abroad: from migrant labor placement agency

How latest returning migrant learned about job opportunity abroad: from private agent

How latest returning migrant learned about job opportunity abroad: from schools/ skills education institution

How latest returning migrant learned about job opportunity abroad: other

How many people living near respondent have worked abroad without all necessary documents: 1 to 9

How many people living near respondent have worked abroad without all necessary documents: don't know

How many people living near respondent have worked abroad without all necessary documents: more than 10

How many people living near respondent have worked abroad without all necessary documents: none

How much money highest-earning current migrant from the household sent over the past 12 months

How respondent obtained information about opportunities abroad for past migration: broker/sponsor

How respondent obtained information about opportunities abroad for past migration: formal institutions

How respondent obtained information about opportunities abroad for past migration: friends/family

How respondent obtained information about opportunities abroad for past migration: media and internet

How respondent obtained information about opportunities abroad for past migration: private labor placement agency

In the past, respondent has applied to work abroad through: broker/sponsor

In the past, respondent has applied to work abroad through: friends/family

In the past, respondent has applied to work abroad through: internet

In the past, respondent has applied to work abroad through: job fair

In the past, respondent has applied to work abroad through: labor office

In the past, respondent has applied to work abroad through: private labor placement agency

In the past, respondent has applied to work abroad through: school

Index of communication

Index of transportation

Issues respondent's friends/acquaintances faced: don't know

Issues respondent's friends/acquaintances faced: physical or psychological violence

Issues respondent's friends/acquaintances faced: recruitment/positioning deception

Issues respondent's friends/acquaintances faced: threats

Issues respondent's friends/acquaintances faced: wage issues

Issues respondent's friends/acquaintances faced: work hours issues

Job type: Government employee

Job type: Family worker, unpaid

Job type: Freelance worker, agricultural

Job type: Freelance worker, non-agricultural

Job type: Private sector employee

Job type: self-employed

Job type: Working, with support of non-permanent/non-paid labor

Job type: Working, with support of permanent/paid labor

Knowledge of documents required to work abroad (score)

Length of migration of latest returning migrant in the household

Literacy Bahasa (read)

Literacy Bahasa (write)

Literacy in a foreign language other than English

Literacy in Arabic (read)

Literacy in Arabic (write)

Literacy in English (read)

Literacy in English (write)

Looked for employment by contacting friends and family

Looked for employment by trying to start a new business

Looked for employment by using formal channels (contacting labor office or potential employer)

Main reason why highest earning current migrant chose destination country: ease of language and visa and proximity to Indonesia

Main reason why highest earning current migrant chose destination country: social network

Main reason why highest earning current migrant chose destination country: higher wages

Main reason why highest earning current migrant chose destination country: religious reasons

Main reason why highest earning current migrant moved: don't know

Main reason why highest earning current migrant moved: other

Main reason why highest earning current migrant moved: to earn a higher income than possible in Indonesia

Main reason why highest earning current migrant moved: to follow success of former migrant workers

Main reason why highest earning current migrant moved: to gain experience working abroad

Main reason why highest earning current migrant moved: unemployment

Main reason why highest earning current migrant moved: unemployment

Main reason why latest returning migrant chose destination country: ease of language and visa and proximity to Indonesia

Main reason why latest returning migrant chose destination country: social network

Main reason why latest returning migrant chose destination country: higher wages

Main reason why latest returning migrant chose destination country: religious reasons

Main reason why latest returning migrant moved: to gain experience working abroad

Main reason why latest returning migrant moved: don't know

Main reason why latest returning migrant moved: invited by others economy

Main reason why latest returning migrant moved: other

Main reason why latest returning migrant moved: to earn a higher income than possible in Indonesia

Main reason why latest returning migrant moved: to follow success of former migrant workers

Main reason why latest returning migrant moved: to fulfill/improve the family economy

Main reason why latest returning migrant moved: to prepare for needs that require large costs

Main reason why latest returning migrant moved: unemployment

Male child in household (below 3)

Male child in household (between 4 and 6)

Male child in household (between 7 and 11)

Married

Means to success: Find employment in Jakarta/another major Indonesian

Means to success: Complete a high level of education

Means to success: Find employment abroad

Means to success: Start own business

Method highest-earning current migrant in the household used to send remittances: formal

Method highest-earning current migrant in the household used to send remittances: informal

Monthly expected wages if migrated relative to the village mean income
Monthly profit of household business
Most successful person in the village: Able to donate much money
Most successful person in the village: Able to holiday abroad
Most successful person in the village: Able to undertake religious pilgrimages (hajj, umroh) travel to holy cities (Jerusalem, Vatican, etc.)
Most successful person in the village: Has a big house
Most successful person in the village: Has a business
Most successful person in the village: Has a motorcycle
Most successful person in the village: Has an automobile
Most successful person in the village: Has permanent employment
Most successful person in the village: Wealthy
Muslim religion
No knowledge of Indonesian migrants' salary in 6 main destination countries
No knowledge of Indonesian migrants' type of work in 6 main destination countries
No knowledge on where to register if interested in working abroad
Not looking for employment because awaiting busy season
Not looking for employment because awaiting call-back from employer
Not looking for employment because currently student
Not looking for employment because dealing with household
Not married
Number of children of highest-earning current migrant in household
Number of children of respondent living elsewhere
Number of children of respondent living with respondent
Number of female current international migrants in the household
Number of female return international migrants in the household
Number of female return national migrants in the household
Number of jobs
Number of male current international migrants in the household
Number of male return international migrants in the household
Number of male return national migrants in the household
Number of minutes of communication with highest-earning current migrant in the past week
Number of respondent's friends/acquaintances who worked abroad in the past five years: 1 to 9
Number of respondent's friends/acquaintances who worked abroad in the past five years: more than 10
Number of respondent's friends/acquaintances who worked abroad in the past five years: no knowledge
Number of respondent's friends/acquaintances who worked abroad in the past five years: none
Number of respondent's friends/acquaintances working abroad: 1 to 9
Number of respondent's friends/acquaintances working abroad: more than 10
Number of respondent's friends/acquaintances working abroad: no knowledge
Number of times respondent has worked abroad in the past five years
Own a house
Proportion of female adults in the household
Proportion of female elderly in the household
Proportion of male adults in the household
Proportion of male elderly in the household
Reason why latest returning migrant came back: contract not renewed
Reason why latest returning migrant came back: goals achieved
Reason why latest returning migrant came back: personal reasons

Reason why latest returning migrant came back: physical or psychological abuse
Reason why latest returning migrant came back: wage problems
Reason why latest returning migrant came back: workplace issues
Relationship of respondent with household head: child
Relationship of respondent with household head: household head
Relationship of respondent with household head: son or daughter in law
Relationship of respondent with household head: spouse
Relative cost of migration
Remittances received from highest-earning migrant in household has been used for: buying electronics/furniture/ motorbike/automobile
Remittances received from highest-earning migrant in household has been used for: buying/building or revating a house
Remittances received from highest-earning migrant in household has been used for: daily needs
Remittances received from highest-earning migrant in household has been used for: education
Remittances received from highest-earning migrant in household has been used for: purchasing jewelry/gold for savings
Remittances received from highest-earning migrant in household has been used for: repaying loans for migrant work
Remittances received from highest-earning migrant in household has been used for: repaying other loans
Remittances received from highest-earning migrant in household has been used for: savings
Remittances received from highest-earning migrant in household has been used for: venture capital
Respondent's agricultural income
Respondent communicated with broker: broker visited respondent's home
Respondent communicated with broker: respondent visited broker's home
Respondent has applied to work abroad in the past five years
Respondent has been contacted by a broker in the past five years
Respondent has cattle as savings
Respondent has formal savings
Respondent has friends or acquaintances working abroad
Respondent is part of the household's most profitable enterprise
respondent works in agriculture
Respondent's non-agricultural income
Respondent's non-work income
Risks of working abroad that respondent knows of: don't know
Risks of working abroad that respondent knows of: physical or psychological violence
Risks of working abroad that respondent knows of: recruitment/positioning deception
Risks of working abroad that respondent knows of: threats
Risks of working abroad that respondent knows of: wage issues
Risks of working abroad that respondent knows of: work hours issues
Salary offered by broker
Share of income respondent would send as remittances if migrated
Share of remittances on total household income
Share of respondent's income on total household income
Share of underemployed individuals among household adults
Since how long has the household had a business?
Total amount of remittances received in past 12 months from all current migrants in household
Total area of land owned by the household (in square meters)
Total income from all jobs (logged)
Total income from main job (logged)

Type of work of highest-earning current migrant in household: Agricultural/miner/construction
Type of work of highest-earning current migrant in household: Care
Type of work of highest-earning current migrant in household: Domestic work
Type of work of highest-earning current migrant in household: Manufacturing
Type of work of highest-earning current migrant in household: professional worker
Type of work of highest-earning current migrant in household: Services
Type of work of latest returning migrant in household: Agricultural/miner/construction
Type of work of latest returning migrant in household: Care
Type of work of latest returning migrant in household: Domestic work
Type of work of latest returning migrant in household: Manufacturing
Type of work of latest returning migrant in household: professional worker
Type of work of latest returning migrant in household: Services
Type of work respondent's friends/acquaintances are doing abroad: Agricultural/miner/construction
Type of work respondent's friends/acquaintances are doing abroad: Manufacturing
Type of work respondent's friends/acquaintances are doing abroad: professional worker
Type of work respondent's friends/acquaintances are doing abroad: Care
Type of work respondent's friends/acquaintances are doing abroad: Domestic work
Type of work respondent's friends/acquaintances are doing abroad: No knowledge
Type of work respondent's friends/acquaintances are doing abroad: Services
Underemployed
Unemployed because awaiting busy season
Unemployed because awaiting next harvest
Unemployed because starting a new business
Village Female Labor Force Participation
Village Male Labor Force Participation
Wage worker
Where respondent would register if interested in working abroad: Formal channels
Where respondent would register if interested in working abroad: Private labor placement agency
Where respondent would register if interested in working abroad: broker/sponsor
Where respondent would register if interested in working abroad: internet
Why respondent did not migrate when last applied: currently in process
Why respondent did not migrate when last applied: does not meet the requirements
Why respondent did not migrate when last applied: not permitted by spouse/parents
Why respondent did not migrate when last applied: unable to pass health test
Works in household enterprise
Years of schooling

Appendix Table 2. Descriptive statistics of top 30 predictors

	Whole sample		Women		Men		Min	Max
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.		
Interest in migration	0.23	0.42	0.18	0.38	0.32	0.46	0	1
Household asset score	0.00	1.00	0.01	1.01	-0.01	0.99	-3.79	3.43
Monthly expected wages abroad relative to village mean income	6.23	67.51	6.25	85.76	6.20	8.10	0.055	6255.2
Relative cost of migration	2.37	2.82	2.31	2.53	2.46	3.23	0	121.66
Village Male Labor Force Participation	0.77	0.08	0.77	0.08	0.76	0.08	0.553	1
Household non-agricultural income	2697905	6272964	2560774	5969511	2918100	6726431	0	3.05E+08
Village Female Labor Force Participation	0.45	0.12	0.45	0.12	0.45	0.12	0.22	0.82
Age	29.75	6.61	30.20	6.43	29.03	6.83	18	40
Percentage of remittances PM would send home if migrated	22.96	26.83	22.83	26.97	23.16	26.60	0	100
Knowledge of documents required to work abroad	3.56	2.62	3.62	2.64	3.48	2.59	0	10
Proportion of male adults in the household	0.32	0.17	0.26	0.14	0.41	0.18	0	1
Highest male education in the household (Years of schooling)	8.71	3.99	7.84	4.15	10.09	3.27	0	18
Proportion of female adults in the household	0.34	0.15	0.37	0.16	0.29	0.13	0	1
Lack of knowledge about Indonesian migrants' expected salary in 6 main destination economies (Malaysia; Saudi Arabia; Taiwan, China; Singapore; Hong Kong SAR, China; Brunei Darussalam)	0.62	0.39	0.63	0.39	0.61	0.40	0	1
Area of land owned by the household	2062	12423	2037	14272	2107	8664	0	1000000
Index of communication	0.00	0.30	-0.03	0.30	0.04	0.30	-0.359	0.685
Lack of knowledge about Indonesian migrants' type of work in 6 main destination economies (Malaysia; Saudi Arabia; Taiwan, China; Singapore; Hong Kong SAR, China; Brunei Darussalam)	0.29	0.28	0.29	0.28	0.29	0.28	0	0.667
Highest female education in the household (Years of schooling)	9.06	3.69	9.71	3.32	8.03	4.00	0	18
Years of schooling	9.42	3.52	9.28	3.52	9.66	3.50	0	18
Household agricultural income	505490	1705840	467637	1610518	566273	1847190	0	6.40E+07
PM total work income (logged)	8.33	7.81	5.99	7.50	12.10	6.77	0	20.68
Share of underemployment in household	0.38	0.41	0.37	0.41	0.39	0.41	0	1
PM non-agricultural income	809939	2127694	520271	1846460	1275068	2443693	0	7.50E+07
PM's income from main job (logged)	7.50	7.79	5.44	7.35	10.81	7.31	0	20.11
Household enterprise operating period	33.18	85.24	33.28	84.45	33.01	86.50	0	1188
Share of PM's income on total income	0.33	0.31	0.25	0.28	0.45	0.31	0	1
Total profit from all household enterprises	596276	3314454	616223	3960360	564246	1854844	0	3.00E+08
Days since latest returning migrant came back	90.44	321.09	87.89	315.63	94.52	329.65	0	2069
Number of children living with PM	0.78	1.15	0.97	1.21	0.49	0.98	0	9
Have friends/acquaintances are working: Taiwan, China; Hong Kong SAR, China; Singapore	0.28	0.45	0.29	0.45	0.27	0.45	0	1
Length of migration of latest returning migrant in the household	186	779	183	777	191	782	0	13755
Number of jobs of PM	0.67	0.64	0.50	0.59	0.93	0.63	0	4
Perceptions of success: Most successful person in the village is someone who has a business	0.52	0.50	0.49	0.50	0.55	0.50	0	1
Number of friends and acquaintances who worked abroad and returned to Indonesia in past 5 years: 1 to 9	0.41	0.49	0.39	0.49	0.45	0.50	0	1
Household benefitted from Rice for the poor or non-cash food assistance program	0.49	0.50	0.49	0.50	0.50	0.50	0	1

Source: Authors' calculations from DESMIGRATIF (2018)

Appendix Table 3. Accuracy and sensitivity checks of Random Forest

Sensitivity analysis (female models)				Sensitivity analysis (male models)			
<i>Sample splitting: train 80% - test 20%</i>							
Number of iterations	OOB error	Accuracy	Sensitivity	Number of iterations	OOB error	Accuracy	Sensitivity
1000	0.18	0.8	0	1000	0.3	0.7	0.07
900	0.18	0.8	0	900	0.3	0.8	0.07
800	0.18	0.8	0	800	0.3	0.8	0.06
700	0.18	0.8	0	700	0.3	0.8	0.06
600	0.18	0.8	0	600	0.3	0.8	0.07
500	0.18	0.8	0	500	0.3	0.8	0.06
400	0.18	0.8	0	400	0.3	0.8	0.07
300	0.18	0.8	0	300	0.3	0.8	0.07
200	0.18	0.8	0	200	0.3	0.8	0.08
100	0.18	0.8	0	100	0.3	0.8	0.08
<i>Sample splitting: train 50% - test 50%</i>							
Number of iterations	OOB error	Accuracy	Sensitivity	Number of iterations	OOB error	Accuracy	Sensitivity
1000	0.18	0.83	0	1000	0.3	0.69	0.04
900	0.18	0.83	0	900	0.3	0.69	0.04
800	0.18	0.83	0	800	0.3	0.7	0.06
700	0.18	0.83	0	700	0.3	0.7	0.06
600	0.18	0.83	0	600	0.3	0.7	0.06
500	0.18	0.83	0	500	0.3	0.7	0.06
400	0.18	0.83	0	400	0.3	0.69	0.05
300	0.18	0.83	0	300	0.3	0.69	0.07
200	0.18	0.83	0	200	0.3	0.7	0.06
100	0.18	0.83	0	100	0.3	0.71	0.09

Source: Authors' calculations from DESMIGRATIF (2018)

Appendix Table 4. Literature-driven model of determinants of migration

VARIABLES	(1) women-training sample	(3) men- training sample
age	-0.034*** (0.006)	-0.038*** (0.007)
No child	0.200** (0.093)	0.066 (0.095)
Years of schooling	-0.021* (0.011)	0.053*** (0.011)
Household Asset Score	0.068* (0.041)	0.014 (0.042)
Social Program	-0.013 (0.079)	-0.023 (0.085)
Number of friends/acquaintances abroad	0.071*** (0.006)	0.033*** (0.004)
Employment status (ref: unemployed)		
Inactive	0.055 (0.165)	0.183 (0.158)
Employed	-0.214 (0.167)	-0.272* (0.139)
Village MLFP	-2.390*** (0.495)	-3.646*** (0.548)
Village FLFP	-0.448 (0.356)	0.666* (0.369)
Constant	1.501*** (0.452)	2.300*** (0.472)
Observations	6,563	4,091

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Source: Authors' calculations from DESMIGRATIF (2018)