The Migration Propensity Index: A tool to anticipate migration decisions in Central America

Francisco Ceballos[*]
Manuel A. Hernandez[*]
Miguel Almanzar[+]
Susana Constenla[~]
Álvaro Espinoza Hernandez[^]

Abstract

International migration has grown rapidly over the past two decades, at an annual rate of 2.4%, prompting increased interest in identifying the root causes of outmigration and the population groups more likely to emigrate. However, anticipating migration is a complex task, as the decision to emigrate is often determined by multiple push and pull factors that are typically interrelated and are not always directly observable. This study proposes the Migration Propensity Index (MPI), a novel approach to indirectly estimate a household's propensity or probability to permanently migrate internationally. The central idea is to identify and keep track of a reduced set of household-level indicators that are strongly correlated with the (latent) decision of individuals to migrate abroad. Taken together and converted into an index, the combined indicators reflect the likelihood that one or more individuals from a given household will emigrate. The MPI is concise, easy to implement, and statistically rigorous, and avoids asking direct, sensitive questions about migration attempts or intentions, which can be prone to refusals and underreporting. We introduce the general calibration methodology —which relies on an out-of-sample cross validation procedure using panel data— and discuss its application to Guatemala and Honduras, in addition to relevant aspects around field implementation and potential extensions. The MPI can assist policymakers to monitor the current and future migration sentiment across populations of interest and to target economic and social policies aimed at improving local living conditions and economic opportunities.

Keywords: International migration; Scoring index; Migrants; Guatemala; Honduras

Acknowledgments: We thank the valuable research assistance of Cynthia Paz and María Lucía Berrospi, and O&M Consultores and Raíces Colectivo for outstanding data collection efforts. The study was funded by the United States Agency for International Development (USAID) and One CGIAR's Fragility, Conflict, and Migration Initiative. The funding sources had no role in data collection, data analysis, or report writing, the conclusions of the study are the responsibility of the authors and do not necessarily reflect the views of the International Food Policy Research Institute (IFPRI), One CGIAR, UNICEF, USAID, or the United States Government.

Authors' affiliations: [*] International Food Policy Research Institute; [+] UNICEF; [~] UC Berkeley; [^] Colorado State University

1. Introduction

International migration has grown rapidly over the past two decades, at an annual rate of 2.4%. In 2020, it was estimated that 281 million international migrants existed across the globe, equivalent to 3.6% of the world population (United Nations, 2020). A multidimensional phenomenon, migration is determined by a wide set of factors, including push factors, encouraging people to move out of their current location, and pull factors, attracting people to move into a new location (Rubenstein, 2017). These factors are generally grouped into four categories: economic factors (e.g., job opportunities, wages); environmental factors (e.g., food availability, weather); social factors (e.g., availability of services, quality of life); and safety/cultural factors (e.g., political stability, crime). Among these potential factors, The World Bank (2018) emphasizes the role of employment opportunities, wage differentials, and distance (whether physical or cultural) as important global factors shaping the observed international migration patterns. In the case of migration from the Northern Triangle (El Salvador, Guatemala, and Honduras) to the United States, additional factors include family reunification, vulnerability to natural disasters, and crime and insecurity (Cohn et al., 2017; CAI, 2019; Congressional Research Service, 2019a & 2019b; National Immigration Forum, 2019; Ruiz Soto et al., 2021).[1]

Cross-border migration tends to stir political controversy and yield strongly opposing perspectives among the main stakeholders, such as migrants themselves, their families and other actors in the communities that they leave behind, and different actors in the destination communities (Usher, 1977). For successful migrants, migration can result in higher incomes, increased access to services, and improved livelihoods, including better education and nutrition for their children (FAO, 2018); these benefits, however, can take time to materialize, leaving migrants in vulnerable positions (Mattoo, Neagu, and Ozden, 2005). In addition, migration generally occurs at great peril, with many migrants experiencing extreme hardships, violence, and even losing their lives in the journey (Donnelly and Hagan, 2014; Isacson, Meyer, and Morales, 2014). In terms of the families that migrants leave behind, remittances can constitute an important source of income, allowing them to invest more in education and housing and attain an overall better quality of life (Yang and Martinez, 2006; Adams and Cuecuecha, 2010). On the other hand, migration may entail an increased work burden for family members staying behind, such as in agricultural, home-maintenance, and child-rearing activities (Singh, 2015) and a brain drain for the overall communities of origin (Docquier and Marfouk, 2006). Moreover, migration may bring about an array of psychological costs for both migrants and their families (Torres and Wallace, 2013; Torres et al., 2016). The effect on the destination communities is ambiguous and significant dissent on the net effects of migration can be observed. On one hand, immigration is regarded as a threat to local livelihoods, potentially crowding-out local labor supply, disrupting customs, or negatively impacting the safety of communities (Galor and Stark, 1991; Batut and Schneider-Strawczynski, 2022; Ortega and Verdugo, 2022). On the other hand, migration can actively contribute to progress and the transformation of societies, both as a current and as a historical phenomenon (de Haan, 1999; FAO, 2018; Diodato, Morrison, and Petralia, 2022).

Identifying migration drivers requires a comprehensive and integrated analysis as migration decisions are generally influenced by many factors at the individual, household, local, regional, national, and

[1] A study from Creative Associates International (2019) identifies economic (unemployment, especially among youth; household economic hardship), transnational ties (having a relative living abroad; receiving remittances), and victimization (crime victim or relative/friend killed) as the three main drivers shaping international migration from the Northern Triangle.

international level. These factors are typically interrelated, vary over time, often reinforce one another, and are not always directly observable, adding complexity to the analysis. For instance, the decision to emigrate is not necessarily an individual decision but is likely made together with other family members (Stark, 1991). Migration may also range from voluntary to forced migration, where the conditions in the origin countries become untenable (due to, for example, conflict or other safety and security issues or lack of food and climate extremes), forcing vulnerable households to leave (FAO, 2018; Jessoe et al., 2018; de Brauw, 2019; Mahajan and Yang, 2020; Clemens, 2021; Hernandez et al., 2023). Similarly, the potential opportunities of an individual to emigrate may depend on his network outside the community and abroad (Munshi, 2003; McKenzie, 2006; Bertoli and Ruyssen, 2018.).[2]

This paper proposes the Migration Propensity Index (MPI), a tool to estimate the probability of at least one person in a household migrating in the near future. In particular, we describe a general methodology to calibrate the index, with the goal of obtaining an objective measure —at the household level— of the probability that individuals from that household will emigrate abroad within the next 12 to 24 months. The MPI intends to overcome the limitations of self-reported migration or migration intentions, which rely on sensitive questions that can render such measures subjective, inaccurate, and difficult to monitor over time. In contrast, the MPI:

1. Avoids sensitive direct questions about attempts and intentions to emigrate, which are prone to refusals or underreporting or may have undesired consequences if they are repeatedly asked over time to the same group of people;[3]
2. Is easy to implement and can be calculated in the field by relying on a concise set of simple, non-invasive questions that are easy to collect;
3. Is statistically rigorous as it is derived using advanced statistical methods, that account for the timing of migration decisions and the potential factors correlated with these decisions, and tested through cross-validation procedures focusing on the out-of-sample predictive power of the index; and
4. Can aid both monitoring and targeting purposes, including monitoring (direct or indirect) beneficiaries of programs and/or targeting social and economic programs to new households at risk to emigrate.

The MPI is similar, in essence, to the Progress Out of Poverty/Poverty Probability Index (PPI), originally developed by the Grameen Foundation and currently implemented by Innovations for Poverty Action (IPA) in several countries.[4] The PPI relies on ten specific household characteristics and asset ownership questions to estimate the likelihood that a household is living below the poverty line (Schreiner, 2013). Such a tool allows for targeting of social programs by field staff in real time, eliminating the need for lengthy, in-depth income or expenditure surveys or more complex poverty mapping exercises.[5] Naturally, the increased simplicity and immediacy of the approach could come at the cost of reduced accuracy.

[2] See Skeldon (1997) for an extensive review of the determinants of migration.

[3] For example, repeated sensitive questions to beneficiaries of a certain program every 3 or 4 months may result in people considering migration more seriously or opting out of the program.

[4] https://www.povertyindex.org/.

[5] Another proxy indicator is the CASHPOR housing index, which measures housing living conditions based on a reduced set of simple questions and was originally implemented in India to prioritize microcredit among poor households. See, e.g., https://spm.opportunity.org/the-impact.

Understanding migration decisions, however, imposes different and additional challenges given the nature and potential factors driving these decisions. In particular, the decision to emigrate should be regarded as a latent (hidden) decision process that can be correlated with multiple factors, some of which ultimately push an individual to emigrate; and these conditions or circumstances that trigger someone to emigrate may (or may not) periodically change over time. Accordingly, we modify the PPI approach to model the latent emigration decision, with the goal of obtaining a reliable, simple, concise, and statistically sound migration scoring index that can help donors, policy makers, and program implementers better approximate and track the probability of emigration among the population of interest.

We estimate the index for the case of Guatemala and Honduras, relying on rich and extensive panel datasets collected in the past decade among large samples of households in regions prone to migration. Guatemala and Honduras are part of the Northern Triangle, the region exhibiting the highest emigration rates in Latin America. For instance, recent estimates suggest that around 265,000 people left the Northern Triangle region annually during each of the last five years, most of them to the United States (Congressional Research Service, 2019b). The resulting models show a large importance of pull factors (in the form of network variables), in addition to demographic characteristics related to the age and education of young adults in the household and some key asset and economic indicators. Moreover, the indices for Guatemala and Honduras share a common underlying set of variables proxying for the decision to migrate —even though both calibrations were conducted independently over two separate large samples of households—, providing strong evidence for the methodology's effectiveness.

While this paper develops a methodology to identify households more likely to migrate, it does this without implying or qualifying migration as a negative phenomenon or one that should be curtailed. Rather, we argue that migration is an ultimate and costly decision for migrants, who would be less likely to uproot themselves from their families and communities and embark on a perilous journey if there were more labor and economic opportunities at home. More broadly, the decision to emigrate could reflect an adaptation strategy to support livelihoods, build resilience, and protect against fragility (Hernandez et al., 2023). Thus, the aim of the MPI is to provide a useful policy tool that can summarize a set of key factors predicting this decision —without attempting to individually or causally identify them— and allow for the targeting of policies that can improve the social and economic environment at the regions or communities of origin that show a higher likelihood of emigration.

The remainder of the paper is organized as follows. Section 2 describes the general methodology for calibrating the MPI, including the statistical methods and validation procedures used. Section 3 presents the two MPI calibrations performed to predict international migration in Guatemala and Honduras, including their data sources, their derivation, the variables selected by the model, a comparison between the two resulting models, and common concerns while fielding the MPI questionnaires. Section 4 discusses additional topics of interest, including considerations around field implementation of the MPI, an application of the index for geographic targeting using Census microdata from Guatemala, and a preliminary, fully-out-of-sample validation exercise of the MPI-Guatemala. Section 5 provides concluding remarks.

2. Methodology

This section describes the general methodology followed to derive the Migration Propensity Index. The overall goal is to derive a simple, concise, and statistically-rigorous index based on easily-measurable, reliable household characteristics and conditions that can predict with reasonable confidence whether an individual in a household will emigrate or not in the near future.

Given that the target users of this index are donors, policy makers, and program implementers and that the index is basically meant for monitoring and targeting purposes as close to real time as possible, the index must be *simple*. In order to be readily adopted, practitioners must be able to grasp the inner workings of the index and trust its capacity to predict a phenomenon they already know reasonably well in qualitative terms. Complex statistical or machine learning models with non-linear effects and cross-interactions between variables are generally more difficult to understand and require a large amount of training data. In this regard, the MPI is based on commonly-used statistical regression techniques and focuses on key dimensions identified in the literature that could likely affect the propensity of an individual to emigrate.

In addition, information gathering in remote, rural areas is costly and faces numerous challenges. Conducting long household surveys in person may be cost-prohibitive in many cases. The index must thus be *concise*. The MPI comprises twelve variables carefully selected to maximize its predictive capacity based on currently available data. Moreover, these variables must be *easy to measure* through direct, non-invasive questions and with minimal reporting bias. For instance, directly asking individuals about their intentions to emigrate may possibly lead to underreporting —out of fear or mistrust—, or to refusals to respond. The MPI must use indicators that can be measured with ease in the field and indirect enough to encourage honest (unbiased) responses. It must also contemplate the possibility of asking these questions over the phone, which can significantly reduce costly transportation and logistics involved in sending a survey team to the field.

Lastly, if the index is to be used with confidence to guide policy and programmatic targeting, it must be *statistically rigorous*. Since the decision to emigrate is dynamic in nature —a decision made today based on uncertain benefits and costs in the future— and since migration is generally followed by an influx of remittances into the migrant's household, model fit and likely confounding effects (endogeneity problems) must be carefully considered to obtain a rigorous index with a high predictive power. The MPI circumvents endogeneity problems by relying on panel household surveys, which avoids confounding from incoming remittances allowing to relate a household's *ex ante* characteristics with their *ex post* decision to emigrate. Furthermore, to avoid model overfitting issues, the MPI is constructed and evaluated using cross-validation techniques, which split the working sample into two subsamples: one subsample for model fitting (training subsample) and the other subsample for evaluating the model's predictive capacity (evaluating subsample).

The next subsections describe the methodology for the construction and validation of the proposed MPI. Wherever possible, we keep the discussion at a general level and refrain from delving into details specific to a particular calibration.

2.1. The timing problem

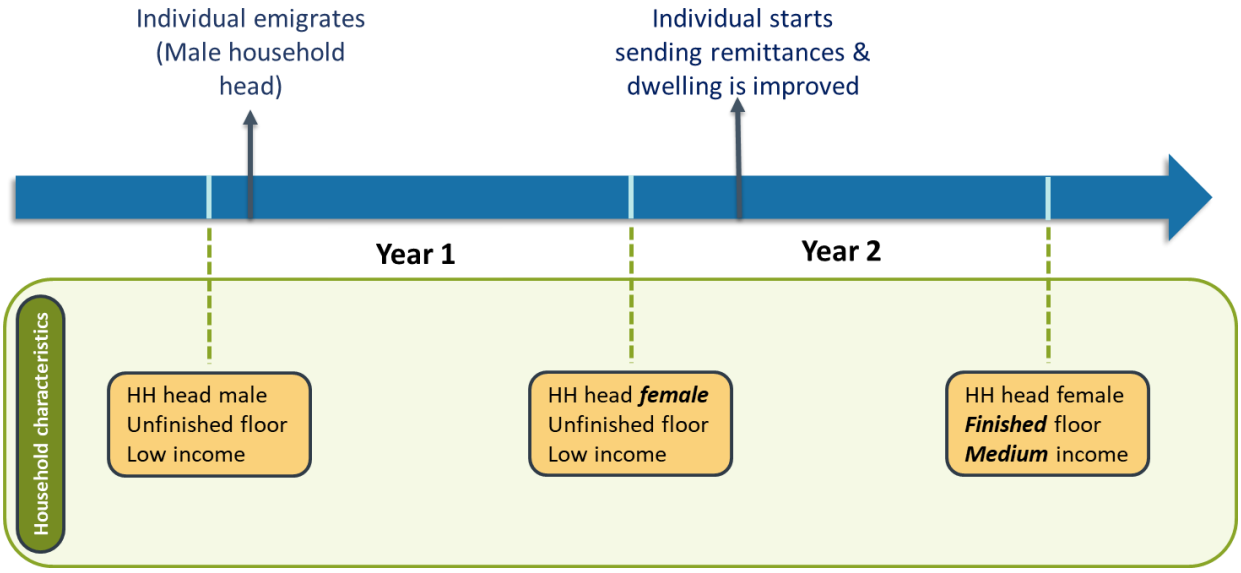
In deriving an index to predict migration, an analyst faces a potential confounding problem that stems from the fact that a household seeing one of its members migrate abroad tends to receive an influx of remittances shortly after. Such a difficulty justifies the use of longitudinal (panel) data when calibrating the MPI.

Figure 1 provides a hypothetical example to illustrate the nature of this problem. Early during year 1, the head of household emigrates. His spouse, now head of household, stays behind and about a year later starts receiving remittances from her husband, which she uses to finish the floor and do other improvements to the dwelling. Now, if one were to interview the household a single time at, say, the end of year 2, one would observe a medium income household, with finished floor and a female head of household. Further inquiring about recent migration at the household, one may erroneously associate these household characteristics to observing an individual in the household emigrate; that is, wrongly conclude that middle income households are the ones migrating and that economic conditions were not a large driver of the decision. In reality, however, at the point where the migration decision took place, the household was of low income, had unfinished floors, and had a male head of household. In other words, cross-sectional data would confound the drivers of migration with its effects.

One approach to circumvent this timing problem is to rely on longitudinal or panel data collected over two or more years across the same households. With such data in hand, an analyst could observe household characteristics before the migration decision took place, and correctly associate those with the decision. In our hypothetical example, this would imply relying on data on household characteristics collected at the beginning of year 1, and data on migration during the last year collected at the beginning of year 2. By observing the same household over time, panel data would allow a more precise identification of the drivers triggering the decision at the time it was made.

As such, all MPI calibrations rely on panel datasets following a sample of households over time and their decisions to migrate in the period in between. Certainly, while the use of panel data does not fully eliminate the timing problem (since migration in the household may go back beyond the most recent instance) or preclude other data issues, we think this is the best possible basis on which to construct a credible predictive tool for migration. In what follows, we assume that these types of data are available. In cases where they were not, it is still possible to carry out the methodology below using cross-sectional data, at the cost of potential confounding.

Figure 1. Advantage of using longitudinal over cross-sectional data



2.2. Preparing the migration decision indicator and the variable universe of push and pull factors

Once the main data source has been identified, the first phase of the index calibration is to construct an indicator variable for whether an individual in a given household has migrated abroad in the recent past. This indicator will serve as the main observed outcome that the model will attempt to predict out-of-sample based on a set of observable variables. The exact definition for this indicator depends on the survey frequency and specific survey question(s) asked. In the case of the Guatemala and Honduras MPIs, this indicator takes a value of 1 if someone in the household had emigrated since the previous survey, and 0 otherwise.

The next step is to construct a broad set of variables (or 'variable universe') available in the surveys that could potentially be correlated with the decision to migrate. These variables would arguably be related to common drivers of migration, including economic, environmental, social, and safety/cultural factors, with the migration literature described above being a good starting point to consider in this process. Some common variable categories include, e.g., household socioeconomic characteristics and composition (including age and education of different household members), dwelling materials, access to public services, assets holdings, landholdings and agricultural activities, income and expenditures including remittances, employment, participation in social programs, indicators of climate vulnerability or past experience of adverse climate events, history of migration in both the household and the broader community, perceptions of crime and insecurity, political participation, perceptions around corruption and governance, beliefs and attitudes, or geographic location, among others. This variable universe will serve as the main repository from which the model selection process draws variables to test their effectiveness in predicting the decision of an individual to migrate. It is important to construct an exhaustive list of all variables likely associated with the decision to migrate that can be obtained from the available panel data. Usually, though, the panel data at hand is unlikely to encompass all potentially

relevant dimensions or variables, so the resulting model should be viewed as a first attempt to construct an MPI, which can be refined in the future as more data become available.

Naturally, most of the variables considered will be related to local (push) factors. However, it is possible to observe some variables more closely related to external (pull) factors. For instance, in existing MPI calibrations we have considered whether the household declares to know anyone in their community that emigrated during the previous year or whether the household itself or other households in the community receive remittances, as proxies of the extent to which the community or household have experienced migration in the past. We interpret these variables as capturing migrant network effects through being connected to individuals that previously lived in the community and who have successfully migrated abroad, which could be an important pull factor by providing trusted information about the costs and benefits of migrating and prompting other individuals in the community to follow suit. In terms of broader pull factors, such as immigration policies or job opportunities abroad, we argue that, while important, they generally exert a blanket effect on attracting migrants, affecting all households in the origin country in a similar manner and thus acting as an overall (positive or negative) shifter to the probability to emigrate. As such, the MPI remains a valuable tool for identifying households most prone to emigrating within a given group, independent of the current external conditions.

2.3. Iterative process for variable selection

The calibration process relies on an out-of-sample cross-validation procedure to select the final model. This is a standard statistical procedure to assess the predictive performance of a model. The idea is to assess how the model will perform when applied to new data. To achieve this, we propose portioning the data into two subsamples: with 70% of the observations being used for estimating the model (i.e., model training); and 30% of the sample to test how well the resulting model predictions match with the actual observations of the outcome (i.e., model predictive performance). To minimize sampling bias, the procedure of splitting the sample is repeated 300 times.[6] In each of these repetitions, the coefficients of the model are estimated on different partitions with 70% of the observations and the metrics on the model's predictive accuracy are calculated over the remaining 30% of observations in the partition. For each metric, we average the measures obtained across the 300 partitions to derive an overall metric. We consider two metrics for the model's construction:[7]

Root mean square error (RMSE): A usual metric to evaluate the fit of a statistical model, which is constructed by squaring the difference between the model's predicted probability and the observed outcome (either 0 or 1), averaging these across all observations, and taking the square root. It measures how far the model predictions are from the actual occurrences.[8]

Concordance statistic (C-Stat): This metric is commonly used for binary-choice models since it measures the concordance between the observed outcome and the model's predicted probability for that outcome across all observations. A model that maximizes this metric exhibits the most internally coherent array of predicted probabilities, in the sense that observations with a positive outcome

[6] These partitions naturally maintain the full-sample proportions of households with and without a migrant during the last 12 months.

[7] We avoid using other standard metrics for the model construction, such as the correct classification rate or sensitivity and specificity measures, as these require converting the estimated probability outcomes into a binary regime prediction using a pre-determined (arbitrary) threshold. In the next section, we report all these performance measures for the final model selected.

[8] The RMSE ranges between 0 and 1. See Barnston (1992) for additional discussion on this measure.

(households having a migrant in our case) will have on average a higher predicted probability than those with a negative outcome.[9]

Since the dependent variable, whether an individual in the household recently emigrated, is binary, we must rely on binary or discrete choice statistical models. We work with a standard logistic regression model (Logit model), which allows to bound the outcome of the model to a probability range between 0 and 1.[10] Since the propensity to emigrate may vary according to local economic, social, and infrastructural conditions it is advisable to control for these aggregated effects, usually in the form of fixed effects at the most disaggregated level possible, depending on data limitations. Since rates of migration usually differ considerably between urban and rural areas, it is important to include an indicator variable to distinguish these. Finally, since working with panel data means that some households may be included twice in the working sample (i.e., when the household is interviewed across 2 or more survey rounds), standard errors should be clustered at the household level to account for the likely correlation in the observed outcomes within the same household.

In order to maximize the objectiveness and unbiasedness of the final selected model, we rely on a step-wise statistical procedure to select the variables with the highest predictive power from amongst the variable universe. We implement the following procedure on the 300 data partitions:

1. Consider every possible one-variable model, constructed by only including one single variable from the universe, together with the geographic fixed effects and the rural indicator.[11]
2. Estimate each of these one-variable models and calculate their out-of-sample predictive-power metric.
3. Select the one-variable model with the highest predictive power.
4. Add one additional variable (from the remaining universe of variables) to the selected one-variable model, to construct all possible two-variable models.
5. Estimate each of these two-variable models and calculate their out-of-sample predictive power metric.
6. Select the two-variable model with the highest predictive power.
7. Add one additional variable to the selected two-variable model (from the remaining universe of variables), to construct all possible three-variable models and repeat the selection process, adding one variable at a time and selecting the one with the highest predictive power in each round, until a 12-variable model is achieved.

The above provides an objective data-driven procedure to construct a statistically-sound model. However, in order to fulfill all of the objectives outlined in the Introduction, we allow for a degree of subjectivity in the process. Besides considering variables that are simple and easy to measure, we need to ensure that they are (1) relatively dynamic over time, (2) add value to the model, and (3) cover key dimensions around the decision to emigrate.

---

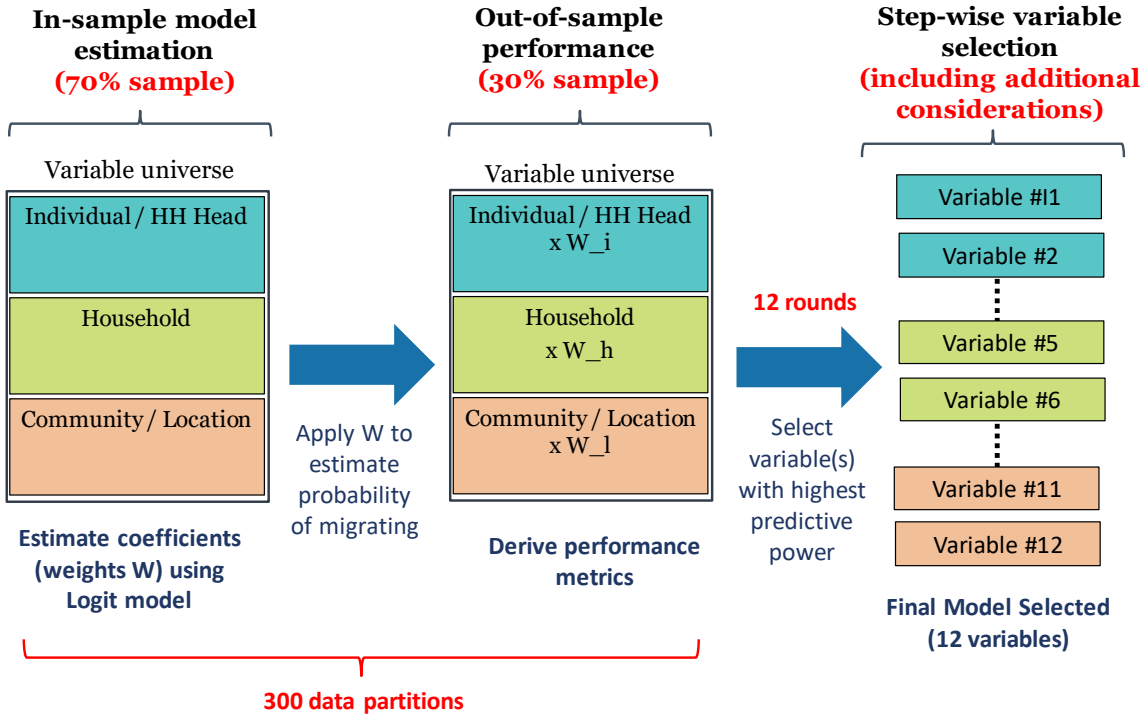
[9] The C-Stat ranges from 0.5 to 1. See Hosmer and Lemeshow (2000) for further details on the interpretation of the C-Stat.

[10] The Logit model is defined as $\mathrm{P}(Y=1|X)=\frac{\exp(X\beta)}{1-\exp(X\beta)}$ where $\mathrm{P}(Y=1|X)$ is the probability of observing an emigrant in a household (Y=1) conditional on a set of X factors. The parameters β are estimated via maximum likelihood estimation. For additional details see Greene (1993).

[11] The geographic administrative regions (e.g., department, state, region) fixed effects and rural indicator act as shifters and should be included in all step-wise regressions.

In terms of the first element, since one of the objectives of the index is to monitor the propensity of an individual in a household to emigrate, it is important that at least a subset of the selected variables likely change over time. For instance, the sex of the household head or his/her years of education would normally not change from one year to the next, while ownership of certain assets or exposure to natural shocks would. As for adding value to the model, we intend to avoid situations where two similar indicators (likely highly correlated) are jointly included. For example, if the model already includes a variable capturing whether any of the household members in a certain age group has completed 5 or more years of education, additionally including whether any member has completed 6 or more years of education would not add real value. Finally, it is advisable to ensure that most key dimensions around the decision to migrate are captured in the indicators proposed, including factors generally discussed in the broader migration literature. To achieve this, we propose following the general statistical procedure to select the optimal variable at each round but further checking if the new variable selected diverts from the objectives above; if so, this new variable is excluded, and the procedure is allowed to select the second-best variable for that round.[12] This process permits to have a final set of 12 variables that fulfill all the MPI objectives. Figure 2 summarizes the full methodology.

Figure 2. Outline of methodology to construct the Migration Propensity Index



2.4. Index construction and questionnaire calibration

Once the calibration process is complete, a set of 12 variables is obtained which, together with the 2 location variables, are used to estimate a logistic model for the entire sample. Each of these variables must then be transformed into individual questions to conform the MPI questionnaire, and their

[12] The estimation programming syntax in STATA for this process is available upon request from the authors.

coefficients in the final statistical model incorporated into simple scoring rules to be applied to each of the questions' answers to calculate the index. This section discusses this process. As noted earlier, the goal is to have a concise, yet reliable and rigorous index that can be easily calculated in real time or as soon as the information is collected.

In order to create the MPI questionnaire, the original instrument of the survey that was used to create the variable universe for calibration must be consulted. In some cases, the new questions may be straightforward, such as when inquiring, for instance, if the roof of the dwelling is of a specific material(s). In others, however, a new question must be constructed to simplify the MPI data collection process. For instance, survey information around the age and education level of household members usually originates from long household composition modules, where the enumerator inquires about each specific household member, and then proceeds to capture detailed information on a number of dimensions for each of these members. In the MPI questionnaire, however, it is important to inquire the required information one single time. In the calibrations for Guatemala and Honduras, for instance, the MPI questions are asked more generally in terms of whether the household has at least one member belonging to a given age group, gender, and educational attainment.

The scores for each question in the MPI questionnaire are, in turn, derived directly from the estimated statistical model. In particular, the logit coefficients for all selected variables are transformed to contemplate the unit of measurement of that variable, its potential range, and that the resulting total MPI score should preferably range between 0 and 100, such that the weight of each variable on the emigration probability is comparable to the weight it has on the overall score. The re-scaling of logit parameters to obtain the MPI scores depends on the logistic function and its predicted probabilities resulting from the estimated parameters of the model and the sample data, in addition to the particular choice of threshold for the different propensity groups. In the MPI calibrations, we have followed a procedure that sets the scoring rule in a way such that an MPI score of 50 (Threshold 1) is associated with an estimated probability of emigration of a household member of 0.01 (or 1%) and an MPI score of 70 (Threshold 2) corresponds to an estimated probability of 0.1 (or 10%).[13] Hence, thresholds of the MPI score have a meaning in terms of estimated probabilities of the logit regression model. It is important to bear in mind, however, that these probabilities are centered around the observed population-level migration rates and are usually low in absolute terms. Appendix B delineates a mathematical derivation of the re-scaling procedure while Appendix Table A.1 presents a lookup table that links MPI scores with the estimated probabilities stemming from the underlying logit model.

The resulting output is a short MPI questionnaire that includes 12 to 15 questions that can be applied to any household for which there is interest in determining their propensity to emigrate. Appendix Tables A.2 and A.3 show the resulting MPI questionnaires for Guatemala and Honduras.[14] The questions are intended to be simple, non-invasive, and easy to collect, without direct inquiries about intentions or actual attempts to migrate. After obtaining the answer to each question, the question's scoring rule indicates how to obtain the partial score contributed by that question.[15] Once all partial scores are derived from

[13] The thresholds could also follow a data-driven definition, i.e., selected following natural breaks in the effective probability of migration (estimated as the fraction of households around a given score that reported an effective migrant) among households in the sample. We preferred, however, to determine a specific value for the probability to emigrate (i.e. a value between 0 and 1).

[14] The details on their calibration are discussed in the following section.

[15] Weights are not exactly equal to the original model since some minor rounding is done to simplify and avoid multiple decimals in the scoring system.

the answers to all questions, the final MPI score is obtained from the sum of all partial scores. The total score can range from 0 to 100, where 100 represents the highest propensity or likelihood to emigrate. It is important to remember that the resulting score is associated to the probability that at least one member within the household will emigrate in the near future. Such a questionnaire can be easily implemented either in person or over the phone.

For ease of interpretation, the final MPI score (measured in a continuous scale) is categorized into one of three possible groups representing the likelihood to emigrate for the interviewed household, following a traffic-light labeling system:

- Low propensity of emigration (MPI score below Threshold 1): green household;
- Medium propensity of emigration (MPI score between Threshold 1 and Threshold 2): yellow household;
- High propensity of emigration (MPI score above Threshold 2): red household.

3. Calibrations

This section describes two initial calibrations of the MPI methodology, to predict international migration (typically to the United States) from Guatemala and Honduras, in addition to a comparison between both resulting models.

3.1. Guatemala

To derive the MPI for Guatemala, we work with three survey rounds collected among the same group of households between 2012 and 2014, as part of the impact evaluation of the *Plan Pacto Hambre Cero* (hereafter PPH0).[16] The survey was designed by IFPRI in coordination with the Secretariat of Food Security and Nutrition (SESAN) and implemented by the National Statistical Institute (INE). The targeted population of the PPH0 were households that have children under five years old, pregnant women, and/or women between 15 to 49 years old, and the final panel dataset is representative of this subpopulation within 60 selected municipalities.[17] The surveys were implemented during the months of November and December and collected information on household socioeconomic characteristics, household income sources, expenditures, and assets, dwelling characteristics, access to services and credit, participation in social programs, hygiene and family health, nutrition, and household food consumption. More details on the Guatemala dataset are described in Ceballos and Hernandez (2020).

Appendix Table A.4 shows the variable universe for the MPI calibration, which comprises 48 available variables for the 2,798 households in the panel. The total list is exhaustive of all variables likely associated with the decision to emigrate that are available in the PPH0 surveys, and that should be relatively easy to collect in future surveys. To capture whether a household had an instance of international migration, we rely on the following two questions, asked in each survey round:

[16] The PPH0 was a national strategy designed by the Government of Guatemala at the beginning of 2012 to address the problem of chronic malnutrition, acute malnutrition, and food insecurity with special emphasis on households with children under five years old, pregnant women and women of childbearing age (between 15 to 49 years old).

[17] In practice, this is close to the entire household population given that almost 89% of households in the listing had had at least one member in these targeted groups.

- Did any household member leave the community to seek work either temporarily or permanently during the last 12 months? 1.*Yes / 2.No*
- *[If "Yes"]* Where did they go? *(A community in…)*
    1. *Same municipality*
    2. *Same department*
    3. *Another department*
    4. *Guatemala City*
    5. *Another country*

We define a household as having had an international migrant during the previous 12 months if the answer to the first question above is "Yes" and the answer to the second question is "Another country".

Panel A in Table 1 lists the final 12 variables selected for the MPI, in the order in which they were selected based on the C-Stat metric (intended to maximize the model's coherence). The variables selected using the RMSE metric are also very similar to the ones reported below, providing additional support to the preferred model.

Peer and network variables stand out as important predictors of the probability that someone in a household emigrates. Recent emigration in the community, as captured by the community-level emigration during the last 12 months, has the largest effect, while a higher prevalence of longer-term emigrants in the community (defined as the share of households in the community who received remittances from abroad during the last 12 months) also has an important positive effect. Additionally, whether the household has an emigrant abroad (proxied by whether the household received remittances during the last 12 months) is an important pull factor positively influencing the index.

Some demographic characteristics also seem to be important decision factors. First, having any youth in the household (male or female members between 15 and 29 years of age) increases a household's MPI. This is consistent with much of the existing literature on migration, which finds youth as the main age group migrating abroad. In addition, education among this age group matters, with the propensity to emigrate increasing when at least one male member within the same age group has 5 years or more of completed education. In contrast, both education and age of the household head are negatively correlated with the decision to migrate. Hence, having some education and skills seems to be an important trigger factor for younger household members to emigrate (most likely the sons or grandsons of the household head) but operate in the opposite direction for household heads.[18] This finding highlights the importance of including variables that reflect intra-household dynamics in migration decisions.

Being a beneficiary of social protection programs (e.g., Bolsa Segura) also serves as an important indicator of a household having a lower propensity to emigrate, possibly signaling very poor households not having the overall means to migrate abroad. Similarly, households that report having made expenditures for home improvements during the last 12 months show a substantially lower propensity to emigrate. Finally, households with large landholdings (above 1 km^2 or 0.1 hectares), those growing a high-value crop (i.e. growing a crop other than maize or beans), or those not owning a television are associated to a lower probability to emigrate, albeit showing a weaker effect than the other selected variables.

[18] Certainly, households with better-educated heads may also proxy households with a higher income that are less likely to emigrate.

Table 1. Final variables selected for Migration Propensity Index

Panel A. Guatemala

#	Indicator	Effect on probability to migrate
1	Community-level emigration during the last 12 months	Positive
2	Community-level rate of migrants abroad	Positive
3	Youth aged 15-29 in household	Positive
4	Household is beneficiary of social protection programs	Negative
5	Household has a migrant abroad	Positive
6	Men aged 15-29 with five years of education or above	Positive
7	Education of household head	Negative
8	Household total agricultural land is above 1 km^2 or 0.1 hectares	Negative
9	Household incurred in home improvement in the last 12 months	Negative
10	Household owns a television	Positive
11	Grows a high-value crop	Negative
12	Age of household head	Negative
+	**Rural and region indicators**	**(Shifters)**

Panel B. Honduras

#	Indicator	Effect on probability to migrate
1	Community-level rate of migrants abroad	Positive
2	Household has received remittances (i.e., has a migrant abroad)	Positive
3	Household size	Positive
4	Men aged 15-29 with complete primary education or higher	Positive
5	Women aged 15-29 with at least one year of secondary education	Positive
6	Household head is male	Negative
7	Household owns agricultural land	Positive
8	Household owns livestock	Negative
9	Dwelling has finished roof (tiles/concrete)	Positive
10	Household owns a vehicle (car/motorcycle)	Positive
11	Household owns a television	Positive
12	Household experienced a drought event in the previous primera ag. season (SPI<-1)	Positive
+	**Rural and department indicators**	**(Shifters)**

3.2. Honduras

In the case of Honduras, we rely on a panel dataset collected by IFPRI for the impact evaluation of the USAID-ACCESO project (under the Feed the Future initiative of the US Government) across three survey rounds in 2012, 2013, and 2015. The surveys encompassed six departments in western Honduras where the activities under Feed the Future/USAID were taking place, i.e., Copán, Intibucá, La Paz, Lempira, Ocotepeque, and Santa Bárbara, and were designed to be representative at the department level. More details on the Honduras dataset are described in Almanzar et al. (2022). The surveys were implemented during Honduras' lean agricultural season, which spans the months from May to August, and included household socioeconomic characteristics, household income sources, employment, expenditures and assets, dwelling characteristics, access to services and credit, participation in social programs, hygiene and family health, nutrition, and household food consumption, among others. Additionally, to consider climate incidence, we included two variables calculated from external sources of whether the standardized precipitation index for any of the first (*primera*) agricultural season months was below -1 or -1.5 (i.e., one or one and a half standard deviation below the historic average).[19] All in all, the MPI calibration for Honduras uses 56 explanatory variables available for the 2,839 households in the panel (Appendix Table A.5).

To capture whether a household had an instance of international migration, we rely on differences between household members' lists between two consecutive survey rounds and on a follow-up question inquiring about the reason for which an individual that used to be a member of the household is no longer present, as follows:

- Why is [previous member] no longer a household member?
    1. *He/she passed away*
    2. *He/she was never a member; was in the roster by mistake*
    3. *He/she moved to another HH inside this village*
    4. *He/she moved to a different municipality*
    5. *He/she moved to Tegucigalpa*
    6. *He/she moved to a different department*
    7. *He/she moved to a different country (specify the country)*

We define a household as having had an international migrant since the previous survey round if there is at least one individual that belonged to the household during the previous survey round and is no longer a member of the household in this survey wave, and the answer to the question above was "He/she moved to a different country."

Panel B in Table 1 lists the final 12 variables selected for the MPI, in the order in which they were selected based on the C-Stat metric.

As with the MPI-Guatemala, peer and network variables are highly relevant predictors of the probability that someone in a household emigrates. A higher prevalence of longer-term migrants in the community (as captured by the share of households in the community who received remittances from abroad during the last 12 months) has the largest positive effect on the probability that one of the members of a given household emigrates. Additionally, whether the household has a migrant abroad (proxied by whether the household itself received remittances during the last 12 months) is another important pull factor positively influencing the propensity to migrate.

Key demographic characteristics also turn out to be important decision factors. First, a larger household size is associated with a higher propensity to migrate, as well as having youth with some level of education in the

[19] A household is considered to have suffered a drought event if for any of the months that comprise the last primera agricultural season (May to Semptember) prior to the moment of the survey the SPI (Standard Precipitation Index) was lower than -1 (in the case of the mild drought variable) or lower than -1.5 (in the case of the moderate drought variable). The monthly SPI was estimated using 20 years of historic precipitation data (2000-2020), at the municipality level, from the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS).

household (female or male members between 15 and 29 years of age with, respectively, at least one year of secondary education or complete primary). This is consistent with much of the existing literature on migration, which finds youth as the main age group migrating abroad, usually with a moderate level of education and skills. In contrast, being a male-headed household is negatively correlated with the decision to migrate.

In terms of assets, owning any amount of agricultural land, a vehicle, or a television, as well as having a finished roof, are all associated with a higher propensity to migrate, while owning livestock is associated with a lower propensity to do so, possibly signaling very poor households not having the required means to migrate abroad. Finally, households that experienced a mild drought event in the previous *primera* agricultural season, as captured by the standardized precipitation index being smaller than -1, are associated to a higher probability to emigrate.

It is worth noting that, unlike the case for Guatemala, the variable universe in the case of Honduras includes a large set of variables capturing different employment dimensions for the different members of the household, in terms of independent/dependent status, employment sectors, and various work dependency ratios. Despite such factors being normally mentioned in the literature as important determinants of migration, none of them were chosen by the variable selection algorithm. Of course, given that there is a high degree of correlation between household-level variables, the exclusion of employment indicators from the final index does not imply that these are unimportant, but rather that they did not add sufficient predictive power to the model, compared to other variables, to warrant their inclusion.

3.3. Comparison and model performance

Overall, and reassuringly, the set of variables that compose the MPI-Honduras is very similar to that for the calibrated MPI-Guatemala. First, both indices share variables proxying for pull factors —such as the rate of households with migrants abroad and whether the household itself has a migrant abroad—, as crucial indicators for the propensity of an individual in a household migrating in the near future. These particular pull factors account for roughly half of the indices' maximum scores (43% in the case of MPI-Honduras and 42% for MPI-Guatemala).

Second, both indices share variables relating to household demographics. Albeit a few differences, such as the MPI-Guatemala relying on age and years of education of the household head while the MPI-Honduras includes only gender of the household head, both indices identify the presence of individuals from the same age group (15 to 29 years of age) with some level of education (complete primary) as an important predictor of migration.

Third, both indices identify asset ownership (such as land or a television) and dwelling characteristics or recent expenditures in improving the household's dwelling as further predictors of the probability to emigrate. Nevertheless, the specific assets and dwelling characteristics are context-specific and do not necessarily coincide in the MPI of both countries.

It is important to keep in mind that the calibrations for both indices are fully data driven and, as such, completely agnostic of the variables selected in the other index. We thus see these similarities as strong evidence for the methodology's value in predicting potential migration in the context of Central America's Northern Triangle.

Panels A and B in Table 2 show several performance indicators of the MPI model calibrated in, respectively, Guatemala and Honduras. The first column in each panel reports the full in-sample metrics; that is, the metrics using the total working sample. The second and third columns report the training in-sample and out-of-sample metrics using the 70%-30% sample partitions from the cross-validation procedure described above; the reported measures are the corresponding averages and confidence intervals across the 300 replicated subsamples.

In Guatemala, the model shows a good predictive performance. In particular, the model's sensitivity or ability to correctly predict positive cases (i.e., households in the PPH0 sample with members who actually emigrated in the

following 12 months) is in the order of 44%. In other words, the model is able, by tracking 12 variables alone, to correctly identify roughly one in every two households where someone ends up emigrating. At the same time, the model's specificity or ability to correctly predict negative cases (i.e., households where members do not end up emigrating) is around 95%. That is, the model correctly identifies a non-emigrating household in 19 out of 20 cases. In Honduras, a similar predictive performance is obtained. The model's sensitivity is in the order of 63%, while the model's specificity or ability to correctly predict negative cases is around 80%. Thus, the model correctly identifies a non-migrating household in 4 out of 5 cases.

Table 2. In-sample and out-of-sample performance measures of calibrated models

Panel A. Guatemala

Measure	In-sample	Training In-sample		Out-of-sample	
	Full Sample	Mean	95% conf. interval	Mean	95% conf. interval
Root Mean Square Error	0.173	0.174	[0.165-0.185]	0.177	[0.154-0.197]
Concordance Stat.	0.864	0.866	[0.845-0.885]	0.835	[0.785-0.891]
Sensitivity	0.443	0.478	[0.406-0.541]	0.426	[0.298-0.566]
Specificity	0.948	0.945	[0.933-0.954]	0.944	[0.926-0.961]
Correct Classification Rate	0.930	0.928	[0.917-0.938]	0.925	[0.909-0.941]
Number of observations	2,713				
Number of households	1,961				
Number of migrants	97				

Panel B. Honduras

Measure	In-sample	Training In-sample		Out-of-sample	
	Full Sample	Mean	95% conf. interval	Mean	95% conf. interval
Root Mean Square Error	0.177	0.177	[0.171-0.183]	0.180	[0.165-0.193]
Concordance Stat.	0.789	0.794	[0.777-0.812]	0.760	[0.712-0.804]
Sensitivity	0.633	0.627	[0.575-0.674]	0.566	[0.461-0.676]
Specificity	0.796	0.800	[0.778-0.823]	0.798	[0.770-0.826]
Correct Classification Rate	0.791	0.795	[0.774-0.815]	0.790	[0.767-0.814]
Number of observations	5,262				
Number of households	2,839				
Number of migrants	180				

Note: The sensitivity (correct positive outcome rate), specificity (correct negative outcome rate), and correct classification rate (correct positive and negative outcomes rate) are calculated using a threshold of 0.1 to convert the estimated model probability into a binary prediction and compare them with the observed (actual) outcomes.

It is worth noting that the asymmetries between the predictive power of positive cases (i.e., correctly identifying an emigrant) and negative cases (i.e., correctly identifying a non-emigrant) are determined by the low prevalence of emigrants in the population or working sample: 3.5% of households reported an emigrant in our Guatemala sample (similar to the values reported in national surveys for Guatemala such as the ENCOVI) and 3.4% of households reported emigration in Honduras. When the prevalence of the modeled variable (emigration) is low, the predictive power of positive cases is lower as there are likely more false-positive outcomes due to a lower population (sample) probability; likewise, the predictive power of negative cases is higher as there are likely much less false-negative outcomes. In this line, a 44% sensitivity or positive predictive value for a 3.5% population-wise prevalence suggests a highly accurate metric (in relative terms).

Both the sensitivity and specificity metrics are very important for policymaking and targeting purposes. Correctly identifying households with future migrants is relevant for accurately targeting programs aimed at addressing the root causes of migration. Correctly identifying non-migrating households is pertinent to better allocate program resources, which can be used in more efficient ways or to achieve objectives other than deterring emigration. In this line, the correct classification rate is a metric that combines the two measures above, reporting the percentage of cases where the model's prediction accurately identifies the actual out-comes observed. This statistic indicates that around 93% of households are correctly classified by the model as either future emigrating or non-emigrating households in Guatemala, while in Honduras the model correctly classifies 79% of households according to this metric.

Overall, the final model obtained by following the MPI calibration methodology outlined above shows a better *ex post* fit in the case of Guatemala compared to Honduras. In particular, both the root mean square error and concordance statistic are slightly better in the MPI-Guatemala than in the MPI-Honduras.

4. Implementation and applications

This section discusses additional topics around the MPI, including some considerations for fielding the index, an application to geographic targeting using the microdata of the 2018 Population Census of Guatemala, and a preliminary, fully out of sample validation exercise conducted among a new surveyed sample of households in Guatemala.

4.1. Considerations for fielding the index

For illustration purposes, we first provide an example on how to apply the MPI. Appendix Table A.6 shows the index score calculation for a hypothetical household in Guatemala. The partial score for question b is calculated by dividing the answer in question ***b*** by that in question ***a***, and multiplying by 34. Similarly, the partial score for question ***d*** arises from the product of 5 and 1, for a total of 5. In general, one decimal is sufficient for calculating partial scores. The rest of the partial scores are taken directly from the number next to the answer given by the household to each of the remaining single choice questions. After summing up all partial scores, including that from the department where the household is located, the household's total score is 65.95, indicating that this is a household of medium migration propensity (since the score is between 50 and 70).

Generally speaking, the questions around the number of households in the community and the number of these with a recent migrant or currently receiving remittances are meant to provide rough household-level estimates for the proportion of households in the community with these characteristics, so as to avoid the need to interview all households in a community. Now, when the community under consideration is fairly large, this question could focus on a narrower definition of the community, such as those households with whom the interviewee normally interacts with (as this variable is mainly intended to proxy for potential peer or network effects in migration

decisions); or could be replaced altogether by community-level estimates of the corresponding proportion, in case these data were available. For instance, in a recent application in urban Honduras, we inquired about these estimates at the household's *barrio*, a much smaller administrative level similar to a small neighborhood. Overall, this indicator intends to capture a household's perception about the intensity of migration around them, and as such should remain valid regardless of the particular geographic definition or social group used in the question, as long as such definition remains local enough to reflect a household's day-to-day exposure to the sought dimensions around migration in their community.

In the case of the MPI-Honduras, the question on whether there was a month in the previous *primera* agricultural season when it rained considerably less than usual could be replaced by an indicator calculated from external sources. Also, the current questionnaire only includes scores for six departments (available in the calibration surveys), so when fielding the MPI in a different department one should use the score for the closest department (in the case of a neighboring department) or of the most similar in terms of migration outflows.

It is important to note that, while the proposed propensity categories may be sufficient to roughly determine a household's propensity to migrate, it is still possible to determine a specific value for the probability to do so from the underlying statistical model by using the lookup table in Appendix Table A.1. As stated above, however, these probabilities are centered around the observed population-level emigration rates and are thus low in absolute terms.

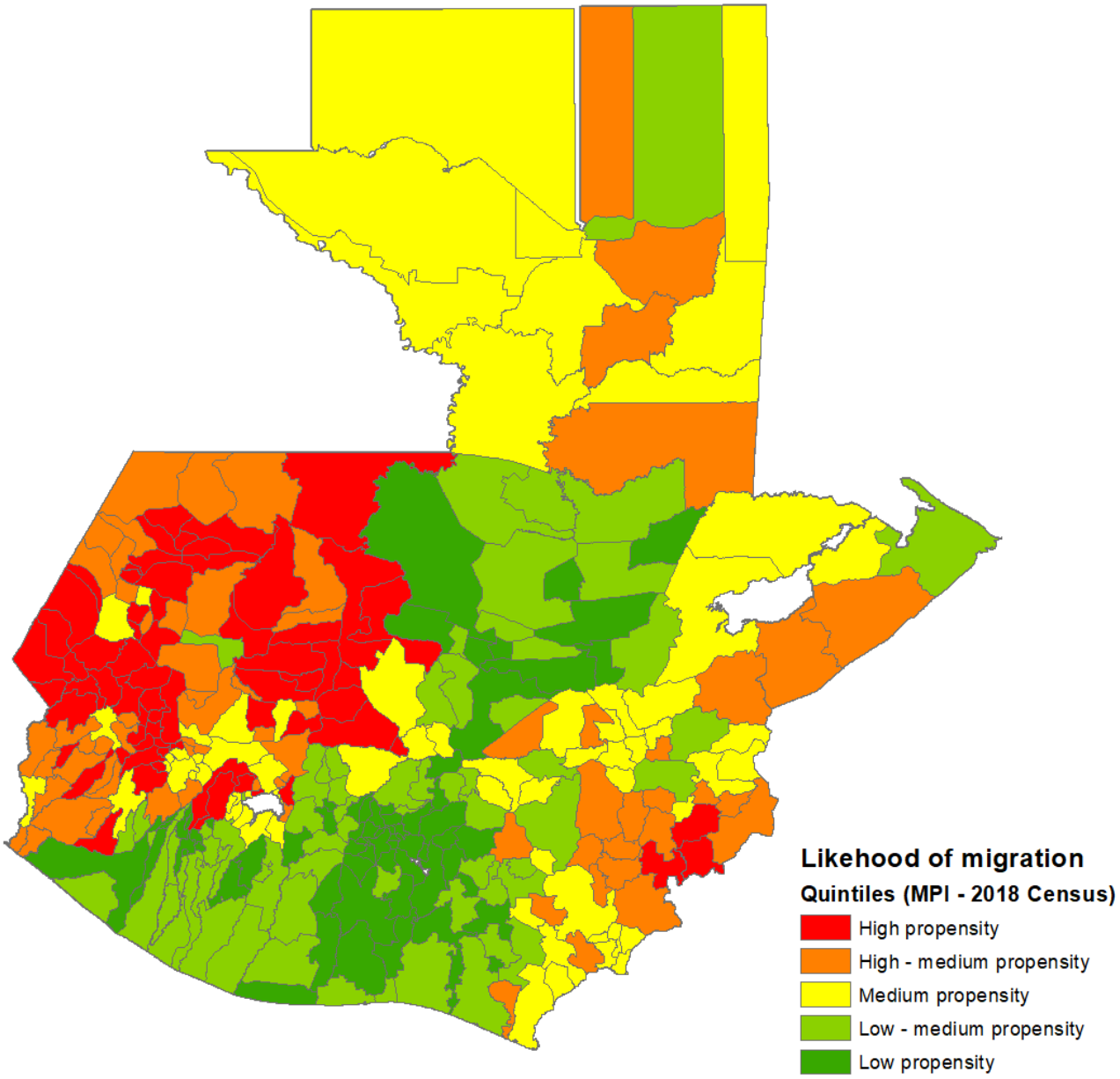
A number of additional tools are available online (or available from the authors upon request) for the MPI-Guatemala and MPI-Honduras, including the final questionnaires and scoresheets in Spanish and English, automated scoresheets to facilitate the index's calculations once answers to all questions are obtained from households, and other technical notes and materials.

4.2. MPI Aggregation using Guatemala Census

One of the most important applications of the MPI is that it can help prioritize areas for the targeting of development programs to improve the local environment and reduce the incentives to emigrate. Such targeting, however, is often constrained by the absence of representative survey data at the national level. To sidestep this challenge, we conducted an exercise to estimate the propensity to migrate at the municipality level based on the MPI using the microdata from the 2018 Population and Housing Census (INE, 2018). In particular, household-level estimates for the MPI were derived for the entire country and aggregated to obtain measures by municipality. Of the 14 variables needed for the MPI-Guatemala, 10 were available in the Census database, and we relied on the calibration survey data to impute the remaining 4 variables using their median value at the regional level (i.e. whether the household is a beneficiary of a government program, whether the household total agricultural land is above 1 km^2, whether the household incurred in home improvement in the previous 12 months, and whether the household grows a high value crop).

Such an exercise allows us to classify each of the 340 municipalities as of high, medium, or low likelihood of international migration, by considering the share of households in that municipality falling under each of these categories. Figure 3, for example, shows the spatial distribution of municipalities with a higher prevalence of households with a medium/high likelihood to emigrate, dividing municipalities into quintiles based on their percentage of households with a medium/high likelihood of emigration. These data are currently being used by USAID Guatemala and their implementing partners as an additional layer of information to determine priority areas in which to focus development programs.

Figure 3. Guatemala municipalities classified in terms of estimated MPI quintiles from 2018 Census



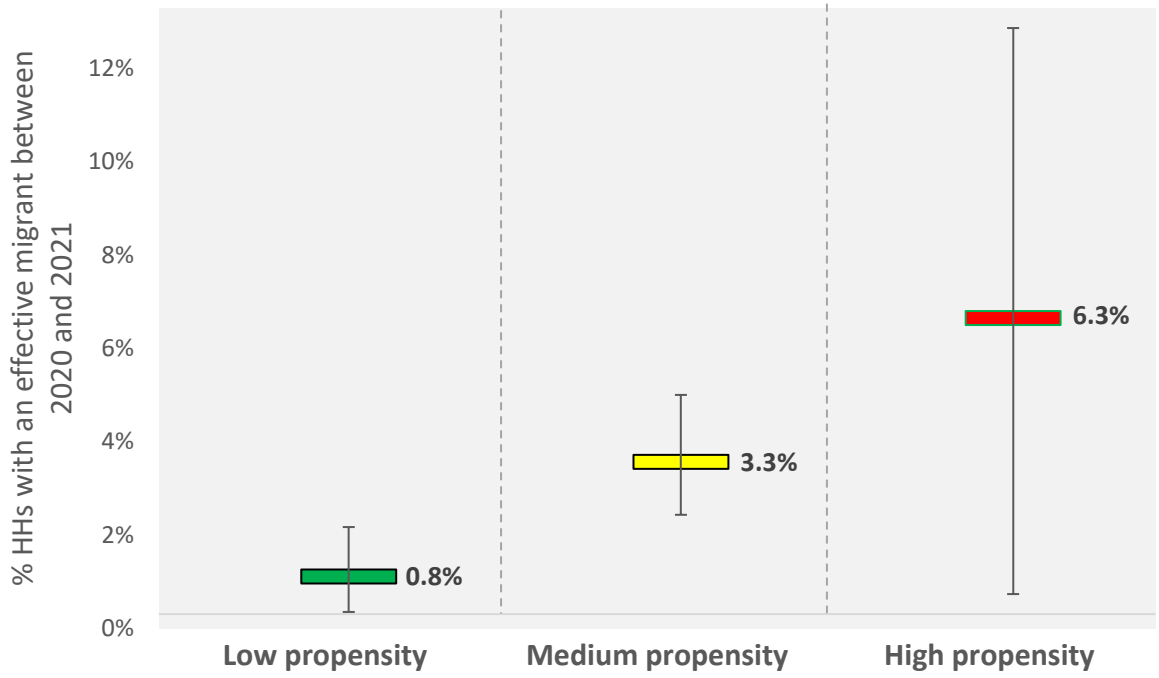
4.3. Preliminary validation of MPI-Guatemala using new data

This subsection presents a preliminary, fully out of sample validation of the Migration Propensity Index (MPI) for Guatemala, relying on two surveys conducted in 2020 and 2021 among a new sample of 1,170 households across the country as part of a migration monitoring exercise among beneficiaries of USAID projects. The panel dataset permits to associate effective *ex post* migration between 2020 and2021 to households' *ex ante* propensity to migrate, as captured during the first survey round in 2020. In particular, the answers to the MPI questionnaire in 2020 allow us to derive an MPI score and classify each household as being of low, medium, or high migration propensity.

Figure 4 plots the effective migration rate between 2020 and 2021 and the MPI propensity category in 2020. Households in a higher propensity group *ex ante* are considerably more likely to emigrate within the following 12 months than those in lower propensity groups. In particular, the share of households reporting an emigrant in 2021 is 0.8% among households classified of low propensity in 2020, 3.3% among households classified of medium propensity, and 6.3% among households of high propensity. The difference in migration rates between the low and medium propensity groups is statistically significant at the 95% level. However, while the observed migration rate for the high propensity group is almost double that in the medium propensity group, these are not statistically-significantly different at conventional levels due to the smaller number of households identified as being of high propensity in 2020 (63 households, compared to 735 and 372 for the medium and low propensity groups, respectively).

While these results are indicative of the MPI's effectiveness in predicting migration, the lack of additional data collection at baseline (beyond the MPI questionnaire) precludes us from answering important additional questions, which we discuss in further detail in the concluding section. An ongoing longitudinal data collection exercise with a new sample of households in Honduras will allow us to conduct a more complete validation of the MPI in the near future.

Figure 4. Effective migration between 2020 and 2021 by likelihood of emigration in 2020



Note: The likelihood of emigration is based on the Migration Propensity Index (MPI), estimated with data collected during the 2020 survey. The share of households (HHs) with an effective migrant between 2020 and 2021 captures the sample share of households in the 2021 survey in which the interviewee indicated that at least one household member registered in 2020 was no longer in the household in 2021 because he/she had migrated abroad. The analysis is based on a panel of 1,170 households interviewed in both survey waves.

5. Conclusions

The Migration Propensity Index (MPI) is a novel tool designed to provide donors, policy makers, and program implementers a way to measure and track the likelihood that individuals in a given household will emigrate in the near future.

This paper describes the overall MPI calibration methodology. The MPI is reliable —by avoiding sensitive, direct inquiries on migration intentions—, easy to implement —by relying on a concise and simple set of questions—, and statistically robust —constructed using advanced statistical methods and evaluated in terms of its out-of-sample predictive capacity. The MPI scoring system allows to classify households as of low, medium, or high migration likelihood or propensity to emigrate. The MPI can be applied to both monitor migration sentiment among households of interest and to identify or target households more prone to emigrate. The index questionnaire is simple and can be directly collected over the phone or embedded into a larger survey. Similarly, the non-invasive, indirect approach of the index allows its periodic application over time to the same population for monitoring purposes.

The paper proceeds to show two concrete MPI calibrations to predict cross-border migration in Guatemala and Honduras. In the final calibrated models, peer and network variables stand out as important predictors of the probability that someone in a household emigrates, in addition to demographic characteristics related to the age and education of young adults in the household and some key asset and economic indicators. Reassuringly, the resulting indices for Guatemala and Honduras share a common underlying set of variables proxying for the decision to migrate, particularly when it comes to pull factors and demographic variables. Since the calibrations for both indices are fully data driven and, as such, completely agnostic of the variables selected in the other index, we see these similarities as strong evidence for the methodology's effectiveness.

The paper describes important additional topics, including considerations around field implementation of the MPI, an application for geographic targeting using Guatemala's 2018 Population Census to estimate the MPI, and a preliminary, fully out of sample validation of the MPI-Guatemala that relies on a new and more recent sample of households, showing that the *ex ante* MPI score and classification into low-, medium- and high-propensity of emigration (measured at baseline) is generally predictive of *ex post* effective migration rates in the subsequent 12 months.

While this latter result is encouraging, some important questions remain around the validity of the MPI for predicting migration. For instance, the surveys used for the calibration of the index were limited in thematic scope and did not inquire about some aspects normally thought to affect the decision to migrate (such as crime, governance, or perceptions). How does the MPI fare against these other factors? Do the dimensions included in the MPI serve as a proxy of the missing dimensions? Are there important elements related to the decision to migrate that should be incorporated into the index in the future? In addition, considering that the surveys used to calibrate the MPI comprised mostly a rural sample, is the MPI effective at predicting migration in urban areas (where the decision to migrate may presumably be substantially different)? To respond to these, we are currently collecting rich data among a random sample of households in Honduras over three survey rounds, which we will use to formally test these aspects, and more generally the MPI's predictive capacity, fully out of sample and in a real-world environment.

All in all, while the MPI can serve as a useful tool to anticipate emigration decisions, the question remains on how to best make use of it. It is worth clarifying that we do not claim that the 12 indicators included in the MPI questionnaire are the only or even the main factors behind an individual's decision to migrate. The suite of MPI indices are just indirect measuring tools, empirically calibrated to track the probability of an individual within a household to emigrate in the near future. Similarly, these factors should not be interpreted as causal factors underlying migration decisions, but rather as observable markers that simply serve as proxies for identifying households with a higher likelihood to emigrate. In this sense, the MPI indicators are not meant to inform decisions on particular development interventions. Rather, the design of comprehensive programs and policies that tackle the multiple factors underlying migration decisions, providing economic opportunities in both rural

and urban settings, reducing crime and corruption, and overall raising the quality of life of individuals, remain the main avenue to develop resilient communities in which the incentives to emigrate are reduced.

Finally, the MPI should be seen as a flexible tool that can be calibrated for other settings or even adapted to approximate other migration decisions. In this regard, Ceballos, Hernandez and Espinoza Hernandez (2023) applied, for instance, the MPI methodology to predict permanent internal or domestic migration in both Guatemala and Honduras. Domestic migration is often considered to be a gateway to migrating abroad and relates to important development issues such as urbanization (and its effects on rural areas), the incomes of migrants and that of the households who stay behind, and potentially large changes in migrants' diets and health, among many others. The calibrated models for the i-MPI (internal-MPI) show similar goodness of fit as the ones for cross-border migration discussed above. The selected variables, however, differ in important ways to the ones for the international MPI, with a much lower reliance on pull factors related to cross-border migration, the inclusion of different demographic and asset ownership dimensions, and a higher importance of adverse weather variables on a household's propensity to migrate domestically.

References

Adams, R.H.J., and Cuecuecha, A. (2010). Remittances, Household Expenditure and Investment in Guatemala. World Development 38(11): 1626-1641.

Almanzar, M., Ceballos, F., Constenla, S., and Hernandez, M.A. (2022). Migration Propensity Index: Calibration for Honduras. IFPRI Technical Note (internal report). Washington, DC: International Food Policy Research Institute (IFPRI)

Barnston, A. (1992). Correspondence Among the Correlation [Root Mean Square Error] and Heidke Verification Measures; Refinement of the Heidke Score. Weather and Forecasting 7: 699-709.

Batut, C. and Schneider-Strawczynski, S., 2022. Rival guests or defiant hosts? The local economic impact of hosting refugees. Journal of Economic Geography, 22(2), pp.327-350.

Bertoli, S. and Ruyssen, I., 2018. Networks and migrants' intended destination. Journal of Economic Geography, 18(4), pp.705-728.

CAI (2019). Saliendo adelante: Why migrants risk it all. Creative Associates International, Washington. Retrieved from: https://www.saliendo-adelante.com/ (accessed November 2024).

Ceballos, F., and Hernandez, M.A. (2020). The Migration Propensity Index: An application to Guatemala. IFPRI Discussion Paper 1953. Washington, DC: International Food Policy Research Institute (IFPRI).

Ceballos, F., Hernandez, M., and Espinoza Hernandez, A. (2023). Internal Migration Propensity Index (i-MPI) for Guatemala and Honduras: A tool to estimate the household-level likelihood of domestic migration. Fragility, Conflict, and Migration Working Paper, CGIAR.

Clemens, M.A. (2021). Violence, development, and migration waves: Evidence from Central American child migrant apprehensions. Journal of Urban Economics 112: 103355.

Cohn, D., Passel, J., and Gonzalez-Barrera, A. (2017). Rise in U.S. Immigrants From El Salvador, Guatemala and Honduras Outpaces Growth from Elsewhere. Pew Research Center, December 7. Retrieved from: https://www.pewresearch.org/hispanic/2017/12/07/rise-in-u-s-immigrants-from-el-salvador-guatemala-and-honduras-outpaces-growth-from-elsewhere/ (accessed November 2024).

Congressional Research Service (2019a). Recent Migration to the United States from Central America: Frequently Asked Questions. CRS REPORT R45489, January 29. Retrieved from: https://fas.org/sgp/crs/row/R45489.pdf (accessed November 2024).

Congressional Research Service (2019b). Central American Migration: Root Causes and U.S. Policy. IN FOCUS IFIII5I, June 13. Retrieved from: https://fas.org/sgp/crs/row/IF11151.pdf (accessed November 2024).

Creatives Associates International (2019). Why Migrants Risk it All: What Drives Unauthorized Migration to the U.S.? What Shapes the Decision to Migrate? Mimeo, July.

de Brauw, A. (2019). Determinants of Migration Among Rural Youth Throughout the World. IFPRI Discussion Paper 1898.

de Haan, A. (1999). Livelihoods and Poverty: The Role of Migration-A Critical Review of the Migration Literature. The Journal of Development Studies 36(2): 1-47.

Diodato, D., Morrison, A. and Petralia, S., 2022. Migration and invention in the age of mass migration. Journal of Economic Geography, 22(2), pp.477-498.

Docquier, F., and Marfouk, A. (2006). International Migration by Education Attainment, 1990–2000. In International Migration, Remittances and the Brain Drain. Washington DC: World Bank, 151-199.

Donnelly, R., and Hagan, J.M. (2014). The Dangerous Journey: Migrant Smuggling from Mexico and Central America, Asia, and the Caribbean. In Hidden Lives and Human Rights in the United States: Understanding the Controversies and Tragedies of Undocumented Immigration, 71-106.

Food and Agriculture Organization – FAO (2018). The State of Food and Agriculture 2018: Migration, Agriculture and Rural Development. FAO, Rome.

Galor, O., and Stark, O. (1991). The Probability of Return Migration, Migrants' Work Effort, and Migrants' Performance. Journal of Development Economics 35(2): 399-405.

Greene, W.H. (1993). Econometric Analysis. New York: Macmillan Publishing Company.

Hernandez, M.A., Ecker, O., Läderach, P., and Maystadt, J.F. (2023). Forced migration, fragility, and policy responses for developing countries. Forthcoming Chapter in 2023 Global Food Policy Report, International Food Policy Research Institute, Washington DC.

Hosmer, D.W., and Lemeshow, S. (2000). Applied Logistic Regression (2nd Edition). John Wiley & Sons.

Instituto Nacional de Estadística – INE (2018). Herramientas de Datos del Censo 2018. Retrieved from: https://www.ine.gob.gt/censo-poblacion/ (accessed November 2024).

Isacson, A., Meyer, M., and Morales, G. (2014). Mexico's Other Border: Security, Migration, and the Humanitarian Crisis at the Line with Central America. Washington Office on Latin America, 2-44.

Jessoe, K., Manning, D.T., and Taylor, J.E. (2018). Climate Change and Labour Allocation in Rural Mexico: Evidence from Annual Fluctuations in Weather. The Economic Journal 128(608): 230-261.

Mahajan, P., and Yang, D. (2020). Taken by Storm: Hurricanes, Migrant Networks, and US Immigration. American Economic Journal: Applied Economics 12(2): 250-277.

Mattoo, A., Neagu, I.C., and Özden, C. (2005). Brain Waste. Educated Immigrants in the US Labor Market, World Bank Policy Research Paper 3581.

McKenzie, D. (2006). Beyond Remittances: The Effects of Migration on Mexican Households. In International Migration, Remittances and the Brain Drain. Washington DC: World Bank, 123-147.

Munshi, K. (2003). Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market. The Quarterly Journal of Economics 118(2): 549-599.

National Immigration Forum (2019). Push or Pull Factors: What Drives Central American Migrants to the U.S.? Retrieved from: https://immigrationforum.org/wp-content/uploads/2019/07/PushPull-Factors_Paper_Final.pdf (accessed November 2024).

Ortega, J. and Verdugo, G., 2022. Who stays and who leaves? Immigration and the selection of natives across locations. Journal of Economic Geography, 22(2), pp.221-260.

Rubenstein, J.M. (2017). The Cultural Landscape: An Introduction to Human Geography (12th Edition). Chapter 3: Migration. Pearson.

Ruiz Soto, A.G., Bottone, R., Waters, J., Williams, S., Louie, A., and Wang, Y. (2021). Charting a new regional course of action: The complex motivations and costs of Central American migration. Rome, Washington, DC, and Cambridge, MA: World Food Programme, Migration Policy Institute, and Civic Data Design Lab at Massachusetts Institute of Technology. Retrieved from https://www.wfp.org/publications/complex-motivations-and-costs-central-american-migration (accessed November 2024).

Schreiner, M. (2013). Simple Poverty Scorecard Poverty-Assessment Tool: Bangladesh. Mimeo, Microfinance Risk Management.

Singh, S. (2015). Increased Burden on Women and Male Out-Migration: An Analysis of Khul Gad Micro Watershed of the Kumoun Himalaya. International Journal of Social Work and Human Services Practice 3(3): 118-123.

Skeldon, R. (1997). Migration and Development: A Global Perspective. London: Routledge.

Stark, O. (1991). The Migration of Labor. Oxford, UK: Basil Blackwell.

Torres, J.M., Alcántara, C., Rudolph, K.E., and Viruell-Fuentes, E.A. (2016). Cross-Border Ties as Sources of Risk and Resilience: Do Cross-Border Ties Moderate the Relationship Between Migration-Related Stress and Psychological Distress for Latino Migrants in the United States? Journal of Health and Social Behavior 57(4): 436-452.

Torres, J.M., and Wallace, S.P. (2013). Migration Circumstances, Psychological Distress, and Self-Rated Physical Health for Latino Immigrants in the United States. American Journal of Public Health 103(9): 1619-1627.

United Nations (2020). Department of Economic and Social Affairs Population Division. International Migration Stock 2020. Retrieved from: https://www.un.org/development/desa/pd/content/international-migrant-stock (accessed November 2024).

Usher, D. (1977). Public Property and the Effects of Migration upon other Residents of the Migrants' Countries of Origin and Destination. Journal of Political Economy 85(5): 1001-1020.

Valentine, J.L., Barham, B., Gitter, S., and Nobles, J. (2017). Migration and the Pursuit of Education in Southern Mexico. Comparative Education Review 61(1): 000.

World Bank (2018). Moving for Prosperity: Global Migration and Labor Markets. Policy Research Report. The World Bank Group.

Yang, D., and Martinez, C. (2006). Remittances and Poverty in Migrants' Home Areas: Evidence from the Philippines. In International Migration, Remittances and the Brain Drain. Washington DC: World Bank, 81-121.

Appendix Table A.1. MPI score-to-probability lookup table

i-MPI score	Probability	Category	i-MPI score	Probability	Category	i-MPI score	Probability	Category
24 or under	< 0.001	LOW PROPENSITY	50	0.010	MEDIUM PROPENSITY	76	0.186	HIGH PROPENSITY
25	0.001		51	0.011		77	0.205	
26	0.001		52	0.013		78	0.225	
27	0.001		53	0.014		79	0.246	
28	0.001		54	0.016		80	0.269	
29	0.001		55	0.018		81	0.294	
30	0.001		56	0.020		82	0.319	
31	0.001		57	0.023		83	0.346	
32	0.001		58	0.026		84	0.373	
33	0.001		59	0.029		85	0.402	
34	0.001		60	0.032		86	0.431	
35	0.002		61	0.036		87	0.460	
36	0.002		62	0.041		88	0.490	
37	0.002		63	0.046		89	0.520	
38	0.002		64	0.051		90	0.550	
39	0.003		65	0.058		91	0.579	
40	0.003		66	0.064		92	0.608	
41	0.003		67	0.072		93	0.637	
42	0.004		68	0.080		94	0.664	
43	0.004		69	0.090		95	0.690	
44	0.005		70	0.100	HIGH PROPENSITY	96	0.715	
45	0.006		71	0.111		97	0.739	
46	0.006		72	0.124		98	0.761	
47	0.007		73	0.137		99	0.782	
48	0.008		74	0.152		100 or above	> 0.800	
49	0.009		75	0.168				

Appendix Table A.2. MPI-Guatemala questionnaire

	Question	Answer	Instructions	Score
a	Approximately how many households live in your community?	___ households		
b	In how many of these households has someone migrated out of the country in the last 12 months?	___ households	Calculate: (b / a) x 33	
c	How many of these households currently receive remittances?	___ households	Calculate: (c / a) x 14	
d	Does someone in your household currently receive remittances?	No	0	
		Yes	6	
e	How old is the head of your household?	___ years	Calculate: (95 - e) x 0.06	
f	How many years of completed education does the head of your household have?	___ years	Calculate: (17 - f) x 0.86	
g	Are there any men or women between 15 and 29 years old living in your household?	No	0	
		Yes	5	
h	Out of all the men between 15 and 29 years old in your household, has anyone completed 5 or more years of education?	No	0	
		Yes	5	
i	Is your household a beneficiary of Bolsa Segura?	No	10	
		Yes	0	
j	Does anyone in your household own more than 1,000 m2 (0.1 hectares) of agricultural land?	No	2	
		Yes	0	
k	Has your household incurred in any home improvement or expansion expenses during the last 12 months?	No	9	
		Yes	0	
l	Do you have a television in your home?	No	0	
		Yes	2	
m	Does anyone in your household cultivate agricultural crops other than maize or beans?	No	1	
		Yes	0	
n	Is your home located in a rural area?	No	0	
		Yes	8	
o	In which region is your home located?	Centro + Pacífico	15	
		Verapaces	19	
		Corredor Seco + Izabal	25	
		Altiplano Occidental	27	

Below 50	**Low propensity**	Between 50 and 70	**Medium propensity**	Above 70	**High propensity**	**Final score**	

Appendix Table A.3. MPI-Honduras questionnaire

	Question	Answer	Instructions	Score
a	Approximately how many households live in your community?	___ households		
b	In how many of these households has someone migrated out of the country in the last 12 months?	___ households	Calculate: (b / a) x 34	
c	Does any member of your household currently receive remittances?	No	0	
		Yes	9	
d	How many members does your household have in total?	___ people	Calculate: d x 1	
e	Of all the MEN between the ages of 15 and 29 in this household, has anyone completed the last grade of primary education or higher?	No	0	
		Yes	5	
f	Of all the WOMEN between the ages of 15 and 29 in this household, has anyone completed the first grade of secondary education or higher?	No	0	
		Yes	2	
g	Is the head of your household male?	No	6	
		Yes	0	
h	Does any member of the household own agricultural land (with or without official title)?	No	0	
		Yes	3	
i	Does the household own any type of livestock such as cows, bulls, chickens/hens, pigs, or goats/sheep?	No	2	
		Yes	0	
j	Is the roof of the house made of tiles or cement/concrete?	No	0	
		Yes	3	
k	Does any household member own a car or motorcycle?	No	0	
		Yes	5	
l	Do you have a television in your home?	No	0	
		Yes	5	
m	In the last primera agricultural season (May to September) was there a month in which it rained considerably less than usual?	No	0	
		Yes	11	
n	Is your home located in a rural area?	No	0	
		Yes	1	
o	In which department is your home located?	Copán	38	
		Intibucá	33	
		La Paz	30	
		Lempira	36	
		Ocotepeque	39	
		Santa Bárbara	32	

Below 50	**Low propensity**	Between 50 and 70	**Medium propensity**	Above 70	**High propensity**	**Final score**	

Appendix Table A.4. Variable universe for Guatemala's MPI

#	Name	Type	Definition	Timing
1	Gender of household head	Binary	One if the household head is male; zero if female	At time of the interview
2	Age of household head	Integer	Age of the household head in years	At time of the interview
3	Education of household head	Integer	Number of completed years of education of the household head	At time of the interview
4	Indigenous household	Binary	One if the interviewee considers him or herself to be indigenous; zero otherwise	At time of the interview
5	Household size	Integer	Number of household members	At time of the interview
6	Men aged 15-29 in household	Binary	One if the household contains at least one male member aged 15-29 years; zero otherwise	At time of the interview
7	Women aged 15-29 in household	Binary	One if the household contains at least one female member aged 15-29 years; zero otherwise	At time of the interview
8	Youth aged 15-29 in household	Binary	One if the household contains at least one (male or female) member aged 15-29 years; zero otherwise	At time of the interview
9	Men aged 30 or above in household	Binary	One if the household contains at least one male member aged 30 years or above; zero otherwise	At time of the interview
10	Women aged 30 or above in household	Binary	One if the household contains at least one female member aged 30 years or above; zero otherwise	At time of the interview
11	Adults aged 30 or above in household	Binary	One if the household contains at least one (male or female) member aged 30 years or above; zero otherwise	At time of the interview
12	Men aged 15-29 with five years of education or above	Binary	One if the household contains at least one male member aged 15-29 with five or more complete years of education; zero otherwise	At time of the interview
13	Women aged 15-29 with five years of education or above	Binary	One if the household contains at least one female member aged 15-29 with five or more complete years of education; zero otherwise	At time of the interview
14	Men aged 15-29 with primary education or above	Binary	One if the household contains at least one male member aged 15-29 with completed primary education or above; zero otherwise	At time of the interview
15	Women aged 15-29 with primary education or above	Binary	One if the household contains at least one female member aged 15-29 with completed primary education or above; zero otherwise	At time of the interview
16	Men aged 15-29 with 1 year of secondary education or above	Binary	One if the household contains at least one male member aged 15-29 with one or more complete years of secondary education; zero otherwise	At time of the interview
17	Women aged 15-29 with 1 year of secondary education or above	Binary	One if the household contains at least one female member aged 15-29 with one or more complete years of secondary education; zero otherwise	At time of the interview
18	Dwelling has finished walls	Binary	One if the dwelling walls are made of block or brick; zero otherwise	At time of the interview
19	Dwelling has finished ceiling	Binary	One if the dwelling ceiling is made of concrete or tiles; zero otherwise	At time of the interview
20	Dwelling has finished floors	Binary	One if the dwelling floors are made of cement, brick, ceramics, polished wood, or boards; zero otherwise	At time of the interview
21	Dwelling connected to water	Binary	One if the household obtains water for consumption from pipes in their dwelling or land; zero otherwise	At time of the interview
22	Dwelling connected to electricity	Binary	One if the dwelling is connected to an electric supply network; zero otherwise	At time of the interview
23	Dwelling connected to sewerage	Binary	One if the dwelling has a toilet connected to sewer of exclusive use for the household; zero otherwise	At time of the interview
24	Household total agricultural land	Continuous	Total hectares of agricultural land worked by a household, 0 if none	During the last 12 months

25	Household has agricultural land	Binary	One if the household works any agricultural land; zero otherwise	During the last 12 months
26	Household total agricultural land is above 1000m2	Binary	One if the household works agricultural land of 1000 square meters or above; zero otherwise	During the last 12 months
27	Produces a subsistence crop	Binary	One if the household cultivates maize or beans; zero otherwise	During the last 12 months
28	Produces a high-value crop	Binary	One if the household cultivates coffee, cardamom, a horticultural crop, or other crops apart from maize and beans; zero otherwise	During the last 12 months
29	Household owns a cellphone	Binary	One if the household owns a cellphone; zero otherwise	At time of the interview
30	Household owns a television	Binary	One if the household owns a television; zero otherwise	At time of the interview
31	Household owns a radio	Binary	One if the household owns a radio; zero otherwise	At time of the interview
32	Household owns a fridge	Binary	One if the household owns a fridge; zero otherwise	At time of the interview
33	Household owns a computer	Binary	One if the household owns a computer; zero otherwise	At time of the interview
34	Household has cable television	Binary	One if the household has a cable television subscription; zero otherwise	At time of the interview
35	Household owns houses or land	Binary	One if the household owns any real estate property; zero otherwise	At time of the interview
36	Household owns a vehicle	Binary	One if the household owns a vehicle; zero otherwise	At time of the interview
37	Household owns agricultural machinery or equipment	Binary	One if the household owns agricultural machinery or equipment; zero otherwise	At time of the interview
38	Household owns livestock	Binary	One if the household owns livestock; zero otherwise	At time of the interview
39	Household purchased durable goods during last 12 months	Binary	One if the household made any expenditures on new durable goods during the last 12 months; zero otherwise	During the last 12 months
40	Household purchased furniture during last 12 months	Binary	One if the household made any expenditures on new furniture during the last 12 months; zero otherwise	During the last 12 months
41	Household purchased vehicles during last 12 months	Binary	One if the household made any expenditures on new vehicles during the last 12 months; zero otherwise	During the last 12 months
42	Household made home improvement during last 12 months	Binary	One if the household made any expenditures on construction, expansion, or improvement of their dwelling during the last 12 months; zero otherwise	During the last 12 months
43	Household is beneficiary of Bolsa Segura	Binary	One if household is beneficiary of social program Bolsa Segura; zero otherwise	During the last 12 months
44	Household is beneficiary of Bono Seguro de Salud	Binary	One if household is beneficiary of social program Bono Seguro de Salud; zero otherwise	During the last 12 months
45	Household affected by a natural disaster	Binary	One if the household has been affected by a natural disaster (excess rain, flood, earthquake, volcano, pests, drought, frost,) during the last 3 years; zero otherwise	During the last 3 years
46	Household has a migrant abroad	Binary	One if the household received remittances during the last 12 months; zero otherwise	During the last 12 months
47	Community-level emigration during the last 12 months	Fraction [0-1]	Share of households in a household's community in which at least one member emigrated during the last 12 months	During the last 12 months
48	Community-level rate of migrants abroad	Fraction [0-1]	Share of households in a household's community that received remittances during the last 12 months	During the last 12 months

Appendix Table A.5. Variable universe for Honduras' MPI

#	Name	Type	Definition	Timing
1	Gender of household head	Binary	One if the household head is male; zero if female	At time of the interview
2	Age of household head	Integer	Age of the household head in years	At time of the interview
3	Education of household head	Integer	Number of completed years of education of the household head	At time of the interview
4	Indigenous household	Binary	One if the interviewee considers him or herself to be indigenous; zero otherwise	At time of the interview
5	Household size	Integer	Number of household members	At time of the interview
6	Men aged 15-29 in household	Binary	One if the household contains at least one male member aged 15-29 years; zero otherwise	At time of the interview
7	Women aged 15-29 in household	Binary	One if the household contains at least one female member aged 15-29 years; zero otherwise	At time of the interview
8	Youth aged 15-29 in household	Binary	One if the household contains at least one (male or female) member aged 15-29 years; zero otherwise	At time of the interview
9	Men aged 30 or above in household	Binary	One if the household contains at least one male member aged 30 years or above; zero otherwise	At time of the interview
10	Women aged 30 or above in household	Binary	One if the household contains at least one female member aged 30 years or above; zero otherwise	At time of the interview
11	Adults aged 30 or above in household	Binary	One if the household contains at least one (male or female) member aged 30 years or above; zero otherwise	At time of the interview
12	Men aged 15-29 with primary education or above	Binary	One if the household contains at least one male member aged 15-29 with completed primary education or above; zero otherwise	At time of the interview
13	Women aged 15-29 with primary education or above	Binary	One if the household contains at least one female member aged 15-29 with completed primary education or above; zero otherwise	At time of the interview
14	Men aged 15-29 with 1 year of secondary education or above	Binary	One if the household contains at least one male member aged 15-29 with one or more complete years of secondary education; zero otherwise	At time of the interview
15	Women aged 15-29 with 1 year of secondary education or above	Binary	One if the household contains at least one female member aged 15-29 with one or more complete years of secondary education; zero otherwise	At time of the interview
16	Dwelling has finished ceiling	Binary	One if the dwelling ceiling is made of concrete or tiles; zero otherwise	At time of the interview
17	Dwelling has finished floors	Binary	One if the dwelling floors are made of cement, brick, ceramics, polished wood, or boards; zero otherwise	At time of the interview
18	Dwelling connected to electricity	Binary	One if the dwelling is connected to an electric supply network; zero otherwise	At time of the interview
19	Household has agricultural land	Binary	One if the household works any agricultural land; zero otherwise	During the last 12 months
20	Household total agricultural land is above 2mz	Binary	One if the household works agricultural land of 1000 square meters or above; zero otherwise	During the last 12 months
21	Produces a subsistence crop	Binary	One if the household cultivates maize or beans; zero otherwise	During the last 12 months
22	Produces a high-value crop	Binary	One if the household cultivates coffee, cardamom, a horticultural crop, or other crops apart from maize and beans; zero otherwise	During the last 12 months
23	Household owns livestock	Binary	One if the household owns a livestock; zero otherwise	At time of the interview
24	Household owns a cellphone	Binary	One if the household owns a cellphone; zero otherwise	At time of the interview

25	Household owns a television	Binary	One if the household owns a television; zero otherwise	At time of the interview
26	Household owns a radio	Binary	One if the household owns a radio; zero otherwise	At time of the interview
27	Household owns a fridge	Binary	One if the household owns a fridge; zero otherwise	At time of the interview
28	Household owns a vehicle	Binary	One if the household owns a vehicle; zero otherwise	At time of the interview
29	Household owns livestock	Binary	One if the household owns livestock; zero otherwise	At time of the interview
30	Household purchased durable goods during last 12 months	Binary	One if the household made any expenditures on new durable goods during the last 12 months; zero otherwise	During the last 12 months
31	Household purchased furniture during last 12 months	Binary	One if the household made any expenditures on new furniture during the last 12 months; zero otherwise	During the last 12 months
32	Household purchased appliances during last 12 months	Binary	One if the household made any expenditures on new appliances during the last 12 months; zero otherwise	During the last 12 months
33	Household made home improvement during last 12 months	Binary	One if the household made any expenditures on construction, expansion, or improvement of their dwelling during the last 12 months; zero otherwise	During the last 12 months
34	Household had a savings account during the las 12 months	Binary	One if the household held a savings account during the last 12 months; zero otherwise	During the last 12 months
35	Household received a credit during the las 12 months	Binary	One if someone in the household received a credit during the last 12 months; zero otherwise	During the last 12 months
36	Household is beneficiary of Bono 10,000	Binary	One if household is beneficiary of social program Bono 10,000; zero otherwise	During the last 12 months
37	Household is beneficiary of Bono for schooling purposes	Binary	One if household is beneficiary of social program Bono for schooling purposes; zero otherwise	During the last 12 months
38	Household affected by a mild drought	Binary	One if the household has been affected by a mild drought (SPI<-1) during the past agricultural season (May to August); zero otherwise	During the last 3 years
39	Household affected by a moderate drought	Binary	One if the household has been affected by a moderate drought (SPI<1.5) during the past agricultural season (May to August); zero otherwise	During the last 3 years
40	Household has a migrant abroad	Binary	One if the household received remittances during the last 12 months; zero otherwise	During the last 12 months
41	Community-level rate of migrants abroad	Fraction [0-1]	Share of households in a household's community that received remittances during the last 12 months	During the last 12 months
42	At least one member works on dependent job	Binary	One at least one household member worked in dependent job during the last 12 months; zero otherwise	During the last 12 months
43	At least one member works on independent job	Binary	One at least one household member worked in independent job during the last 12 months: zero otherwise	During the last 12 months
44	At least one member works on independent job: sales	Binary	One at least one household member worked in independent sales job during the last 12 months; zero otherwise	During the last 12 months
45	At least one member works on independent job: agriculture	Binary	One at least one household member worked in independent agricultural job during the last 12 months; zero otherwise	During the last 12 months
46	At least one member works on independent job: manufacture	Binary	One at least one household member worked in independent manufacture job during the last 12 months; zero otherwise	During the last 12 months
47	At least one member works on independent job: professional	Binary	One at least one household member worked in independent professional job during the last 12 months; zero otherwise	During the last 12 months
48	Independent workers to dependents ratio	Ratio	Number of dependent household members (under 14 and over 65 years of age) divided by the number of members that worked on independent jobs in the last 12 months	During the last 12 months
49	Dependent workers to dependents ratio	Ratio	Number of dependent household members (under 14 and over 65 years of age) divided by the number of members that worked on dependent jobs in the last 12	During the last 12 months
50	Active working members to independent workers ratio	Ratio	Number of household members of working age (from 15 to 64 years old) divided by the number of members that worked on independent jobs in the last 12 months	During the last 12 months

51	Remoteness	Numeric	Distance from the village where the household is located to the nearest big market or population area	Static
52	Food security - hunger	Binary	One if in the past 30 days, someone in the household when to bed hungry because of lack of food; zero otherwise	During the last 30 days
53	Group participation – type 1	Binary	One if a member of the household participates in an agricultural cooperative, producers' association, patronage, or union organization; zero otherwise	At time of the interview
54	Group participation – type 2	Binary	One if a member of the household participates in the town council, water management council, political organization, women's group, or religious organization; zero otherwise	At time of the interview
55	Group participation – type 3	Binary	One if a member of the household participates in a credit cooperative (*caja rural*); zero otherwise	At time of the interview
56	Household type	Numeric	Gendered household type, 1 if just female head, 2 if just male head, 3 if both male and female primary decision makers	At time of the interview

Note: SPI = Standard Precipitation Index. A household is considered to have suffered a drought event if for any of the months that comprise the last *primera* agricultural season prior to the moment of the survey, the SPI was lower than -1 (in the case of the mild drought variable) or lower than -1.5 (in the case of the moderate drought variable). The monthly SPI was estimated using 20 years of historic precipitation data (2000-2020), at the municipality level, from **the Climate Hazards Group InfraRed *Precipitation* with Station data (*CHIRPS*)**.

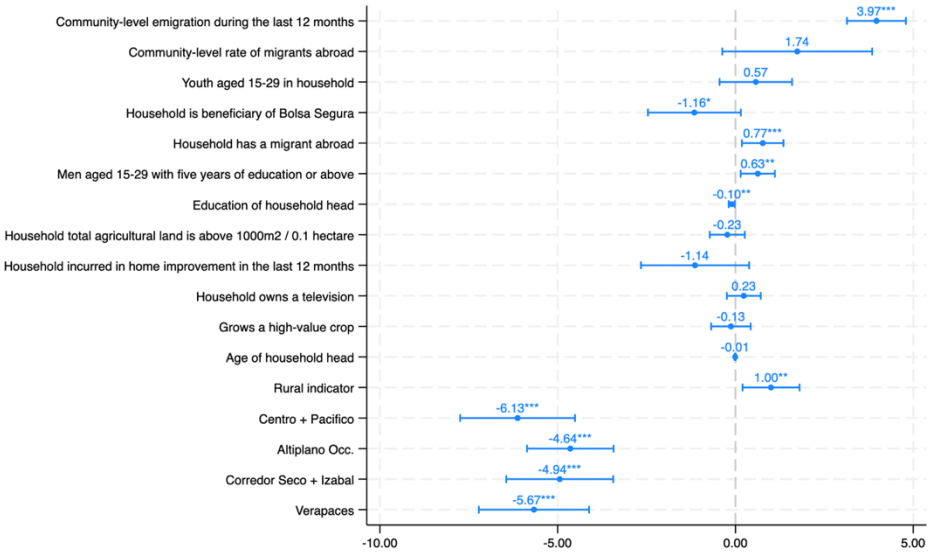
Appendix Table A.6. Illustrating the MPI-Honduras for a hypothetical household

	Question	Answer	Instructions	Score
a	Approximately how many households live in your community?	**40** households		
b	In how many of these households has someone migrated out of the country in the last 12 months?	**7** households	Calculate: (b / a) x 34	**5.95**
c	Does any member of your household currently receive remittances?	No **Yes**	0 9	**9**
d	How many members does your household have in total?	**5** people	Calculate: d x 1	**5**
e	Of all the MEN between the ages of 15 and 29 in this household, has anyone completed the last grade of primary education or higher?	No **Yes**	0 5	**5**
f	Of all the WOMEN between the ages of 15 and 29 in this household, has any completed the first grade of secondary education or higher?	**No** Yes	0 1	**0**
g	Is the head of your household male?	No **Yes**	6 0	**0**
h	Does any member of the household own land (with or without official title)?	**No** Yes	0 3	**0**
i	Does the household own any type of livestock such as cows, bulls, chickens/hens, pigs, or goats/sheep?	**No** Yes	2 0	**2**
j	Is the roof of the house made of tile or cement/concrete?	No **Yes**	0 3	**3**
k	Does any household member own a car or motorcycle?	**No** Yes	0 5	**0**
l	Do you have a television in your home?	No **Yes**	0 5	**5**
m	In the last agricultural season (May to September) was there a month in which it rained considerably less than usual?	**No** Yes	0 11	**0**
n	Is your home located in a rural area?	No **Yes**	0 1	**1**
o	In which department is your home located?	Copán Intibucá **La Paz** Lempira Ocotepeque Santa Bárbara	38 33 30 36 39 32	**30**

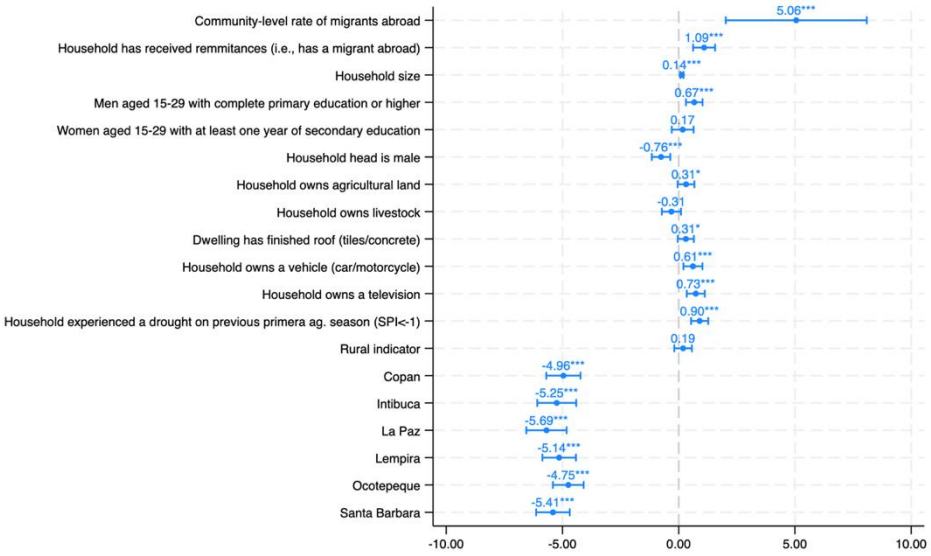
Below 50	**Low propensity**	Between 50 and 70	**Medium propensity**	Above 70	**High propensity**	**Final score**	**65.95**

Appendix Figure A.1. Estimated logit coefficients for the final MPI variables

Panel A. Guatemala



Panel B. Honduras



Note: This figure plots the logit coefficients for the final variables selected by the MPI calibration process in both Guatemala (panel A) and Honduras (panel B). The bars indicate 95% confidence bounds based on standard errors clustered at the household level. The estimation sample comprises 2,713 observations in Guatemala and 5,262 observations in Honduras. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Appendix B. Re-scaling of the final logistic model to obtain MPI scores

Overall objective: We want to find a linear transformation of the scalar $X\beta'$ such that the MPI score threshold below which a household is considered to have a low propensity to migrate is 50 and the MPI score threshold above which a household is considered to have a high propensity to migrate is 70. This choice of thresholds is arbitrary, but it allows for consistency across MPI products. The linear transformation we seek is represented by $MPI = (c + m\, X\beta')$, where c is a constant and m a multiplier.

To begin with, we have the logit formula, underlying the logistic regression used for calibrating the model:

$$Prob(Mig = 1) = p = \frac{e^{X\beta\prime}}{1 + e^{X\beta\prime}}$$

Resolving for $X\beta'$:

$$e^{X\beta\prime} = p + pe^{X\beta\prime} \quad \rightarrow \quad (1-p)e^{X\beta'} = p \quad \rightarrow \quad X\beta' = ln(\frac{p}{1-p})$$

The vector β is estimated from the final resulting model including all MPI variables. $X\beta'$ is the product of each model coefficient and the value of each variable, summed across all variables together with the constant:

$$X\beta' = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots$$

We use $p = 0.01$ and $p = 0.1$ as the threshold probabilities below and above which we consider households to be of low, medium, or high propensity to migrate. Consider $X_{0.01}$ and $X_{0.1}$ to be the vectors of explanatory variables X that, combined with the estimated β, result in the probability from the logit formula above being, respectively, 0.01 and 0.1.

Now, we can construct a system of equations at both thresholds and resolve for c and m:

$$c + m\, X_{0.01}\beta' = 50$$

$$c + m\, X_{0.1}\beta' = 70$$

Since $X_{0.01}\beta = ln\left(\frac{0.01}{1-0.01}\right) = ln\,(\frac{1}{99})$, we can resolve the first equation for c as follows:

$$c = 50 - m\, ln\,(\frac{1}{99})$$

And replace both c and $X_{0.1}\beta$ in the second equation, resulting in:

$$50 - \; m\; \ln\left(\frac{1}{99}\right) + m \ln\left(\frac{1}{9}\right) = 70$$

Resolving for m:

$$m\left[\ln\left(\frac{1}{9}\right) - \ln\left(\frac{1}{99}\right)\right] = 20$$

$$m = \frac{20}{\ln\left(\frac{99}{9}\right)} \cong 8.341$$

And thus

$$c \cong 88.326$$

One additional consideration is that, for clarity purposes, it is preferrable for binary variables to either positively contribute a few points or not to contribute at all to the final MPI score. When the logit coefficient for a binary variable is negative, though, this implies that when that variable takes a value of 1, the total MPI score should decrease. To avoid this, we can do a small modification to negative dummy coefficients, as follows. Suppose that both X_1 and X_2 are dummy variables, and that β_1 is negative while β_2 is positive. The regular expression for the MPI score is:

$$MPI = c + m\,[\beta_0 + X_1\beta_1 + X_2\beta_2 + \cdots]$$

Since $\beta_1 < 0$, we can rewrite the expression for the MPI score as:

$$MPI = c + m\,[\beta_0 - X_1|\beta_1| + X_2\beta_2 + \cdots]$$

Adding and substracting $|\beta_1|$:

$$MPI = c + m\,[\beta_0 - X_1|\beta_1| + |\beta_1| - |\beta_1| + X_2\beta_2 + \cdots]$$

$$MPI = c + m\,[\beta_0 + (1 - X_1)|\beta_1| + X_2\beta_2 - |\beta_1| \ldots]$$

$$MPI = [c - m|\beta_1|] + m\,[\beta_0 + (1 - X_1)|\beta_1| + X_2\beta_2 \ldots]$$

Then, with a negative coefficient ($\beta_1 < 0$), the expression above states that we can take their absolute value $|\beta_1|$ along with a modification in the X_1 dummy to $(1 - X_1)$, so that when $X_1 = 1$ then $(1 - X_1)|\beta_1| = 0$. In other words, we have reached an expression in which the partial MPI score contributed by this variable will be either null or positive.

Generalizing the latter expression indicates that we can take the absolute value of all negative coefficients and change the respective dummy variables to $(1 - X_i)|\beta_i|$ as long as we also subtract each term $-\,m|\beta_i|$ for all negative coefficients i from the constant c, resulting in:

$$c - \sum_{i \in NegDum} m|\beta_i|$$

Finally, with all elements in place, we can re-scale the β coefficients from the final logit model to those used for the MPI, where:

$$\beta_j^{MPI} \quad \rightarrow \quad m\beta_j$$

$$\beta_0^{MPI} \quad \rightarrow \quad m\beta_0 + c - \sum_{i \in NegDum} m|\beta_i|$$