

An agent-based decision model of migration,
embedded in the life course –
Model description in ODD+D format

Anna Klabunde *
Sabine Zinn †
Matthias Leuchter ‡
Frans Willekens §

June 10, 2015

Abstract

This report contains the model description for a prototype model of migration decision making which is based on the theory of planned behavior. It makes use of empirically estimated demographic transition rates and is thus, to our knowledge, the first decision model of migration which is embedded in the life course. Moreover, it is, other than most agent-based models, a continuous time model which makes heavy use of survival theory and competing risks. The model description follows the ODD+D (“Overview, Design Concepts, Details + Decisions”) protocol as suggested in [5].

Keywords. Discrete event simulation, agent-based model, migration, theory of planned behaviour.

*Max Planck Institute for Demographic Research, Germany. *Email:* klabunde@demogr.mpg.de

†Leibniz Institute for Educational Trajectories, Germany.

‡Max Planck Institute for Demographic Research, Germany

§Max Planck Institute for Demographic Research, Germany.

1 Overview

1.1 Purpose

1.1.1 What is the purpose of the study?

The purpose of the study is to develop a prototype for a new generation of models which combine microsimulation of demographic processes with decision-making and interaction in an agent-based fashion. The ultimate goal is to build a model which can forecast timing and location of migration flows, in different time periods and in reaction to external events such as policy change, war or climate change.

The proposed model is essentially a combination of two different model types: A demographic projection model describing life course dynamics, and an agent-based model for describing unobservable decision processes related to migration. The timing of all life events other than migration follows stochastic processes, whereas the migration decision is the outcome of a decision process.

For modelling the migration decision, it is important to find the set of relevant factors influencing a person's decision which is just large enough to render the model useful. Increasing the number of parameters beyond this minimum is dangerous as it becomes difficult to grasp the parameter space in its entirety. If the parameter space is small enough we will be able to understand the model's behavior at different points in the parameter space by applying a systematic sensitivity analysis.

This very first version will focus on the migration decision only. Neither the location choice nor return or circular migration will be modelled at this point. Nevertheless we will outline possible alternative ways of modelling as well as desirable extensions throughout this model description.

1.1.2 For whom is the model designed?

The model is designed, in this version, for other scholars. Future versions might be adapted so that they become useful for policy makers, too.

1.2 Entities, state variables and scales

1.2.1 What kind of entities are in the model?

The model includes individuals who are organized in households in the home country and the host country. Individuals can be connected through network links.

Extensions

- A future version of the model might include a labor market; in this case one would have to include firms.
- Alternatively, the model could include a policy maker as an actor and one could thus endogenize policy decisions.
- One could model higher level entities such as neighborhoods.
- Agents in the host country or in several countries of origin could be modelled.

1.2.2 By what attributes (i.e. state variables and parameters) are these entities characterised?

Agents have the following attributes:

- gender (binary: m,f)
- marital status (categorical; 4 levels: never married, married, divorced, widowed)
- identification number of spouse
- number and identification numbers of children (discrete numeric: $[0,20]$; eg. 2)
- income (continuous numeric, in monetary units; rounded to five decimal places)
- location (binary: host country, home country)
- household membership: Identification number of household of which the agent is a member
- migration stage (categorical: no intention, intention, planning, preparation, migration)
- time in current stage (continuous numeric, in years $[0, \dots 110.00000]$; rounded to five decimal places, e.g. 4.56132 years)

Households have the following attributes:

- capital (continuous numeric: $[0, \dots \infty)$; rounded to two decimal places, e.g. 520.90)

- number of household members (discrete numeric: $[0, 1, \dots, 20]$; eg. 2)
- identification numbers of household members

Extensions

- Here, we do not distinguish between different labor market status. This choice was made because of the situation in Senegal, the origin country in our data set, where labor market transitions are usually not clear-cut and underemployment, self-employment, and agricultural work are very common.
- Future model versions might include the exact location. In this version we refrain from modelling the location decision.
- The modeller might be interested in self-selection of migrants according to educational level, which is likely to be the case in reality. Thus, this feature will be included in the next model version. Other dimensions of heterogeneity might be ethnicity, religion or health.
- Additionally, we might want to study the implications of inequality on migration and vice versa. Therefore, agents might be specified to differ in their initial wealth.

1.2.3 What are the exogenous factors/drivers of the model?

There is a set of exogenous stochastic events influencing the agents' life course. These are modelled using the R microsimulation package Micsim ([9]) and fed into the model. The duration until the occurrence of demographic events such as marriage, childbirth or death is drawn from empirically estimated waiting time distributions. The distribution functions of waiting times are fully determined by transition rates. These might depend on age, marital status and number of children. These processes are:

- Death: All agents are always exposed to death. Mortality rates are assumed to depend on age and gender.
- Marriage: The unmarried population is at risk of marriage from the age of 16 until the age of 65; marriage rates are assumed to depend on age only and follow the Coale-McNeil model.
- Dissolution of marriage: The married population is at risk of dissolution of marriage.

- Child birth: women from the age of 16 until the age of 45 are at risk of child birth; fertility rates depend on age, marital status and time elapsed since last birth
- Income: Income accrues to agents from the age of 16 to their death. Wages depend on location (home country or host country) and are drawn from an empirically determined distribution.

Furthermore, policy is an exogenous factor to the model, which is assumed not to change in this model version. Policy determines how easy it is for a migrant to actually cross the border when he attempts to do so.

Extensions

- Dissolution of marriage can be made to depend on age and/or duration of marriage.
- Income can be made a function of seniority and/or increased each year to model inflation.

1.2.4 How is space included in the model?

In this version of the model, space is only included in the form of two regions, one representing the home country and the other representing the host country. In the home country, households are arranged on a grid so that social closeness can be easily modelled by means of spatial distance. Network neighbors are the members of those households who are in the Moore neighborhood of an agent's household in the very first day of the agent-based simulation. Even if one of those network neighbors migrates he retains the same network links. Whenever an agent moves - be it to another household, e.g. through marriage, or to the host country - he forms additional network links to the members of his new household and to the individuals in the neighboring households. Thus, spatial distance is used to express social distance, not actual physical distance.

Extensions

The model could be extended to include more than one host country. Alternatively or additionally, the exact location decision of the migrant can be modelled and studied. For this purpose, the modeller has to decide on the level of detail of the spatial representation, minding that increased spatial complexity is likely to decrease tractability.

1.2.5 What are the temporal and spatial resolutions and extents of the model?

- Space: A grid size of $200 * 100$ is chosen to accommodate a maximum of 10,000 households. The grid is divided in half to represent two countries. All members of one household are allocated on the same square of the rectangular grid.
- Time: Time is modelled on a continuous time scale. The simulation horizon can be of any desired length, thus, overlapping generations can be modelled.

1.3 Process overview and scheduling

A demographic microsimulation is run before the agent-based simulation. The microsimulation determines for each individual all demographic states that an individual will experience throughout his life and the amount of time the individual will spend in each state, except for the migration decision, which will be modelled in the agent-based model. All possible states and transitions are described in section 3.4.1. Thus, within the agent-based model, agents experience demographic events deterministically. The timing of events is measured down to the level of milliseconds, thus there is a very clear order in which agents act or are affected by transitions. The items in brackets denote submodels which are described in more detail in section 3.4.1. The following events happen in the order they are scheduled:

- Births ('Birth')
- Deaths
- Marriage ('Marriage')
- Divorce ('Household dissolution')
- Retirement at age 65
- Update of intention to migrate ('Attitude', 'Social norms', 'PBC')
- Drop out of migration decision process ('Exit')
- Progress to next decision stage ('Progress')
- Migration

At the end of each day: Agents consume and save ('Consumption'). At the end of each month: Agents receive income ('Income').

2 Design Concepts

2.1 Theoretical and empirical background

2.1.1 Which general concepts, theories or hypotheses are underlying the model's design at the system level or at the levels of the submodels (apart from the decision model)?

The decision model on migration is the core of the model. All other demographic events are determined by the underlying microsimulation model. At the systems level, networks ties serve as transmission mechanism for three different things: information, financial and other types of help (social capital), and social norms through observed behavior of others.

What is the link to complexity and the purpose of the model?

Migration decisions cannot be seen as independent from one another. This has been shown in many different studies. The reason for this is that potential migrants receive information and help from previous migrants and their behavior serves as an example for the feasibility and desirability of migration. Thus, migration moves are correlated. Any policy or other exogenous effect on the migration system will never only concern those directly affected, but also many others indirectly. These second-round effects are often not foreseeable by analytical or statistical analysis only.

2.1.2 On what assumptions is/are the agents' decision model(s) based?

The decision model is based on the Theory of Planned Behavior (TPB) [1]. The basic assumption is that an agent makes a decision in at least two stages: First, he/she determines his/her attitude, social norms and perceived behavioral control associated with a certain behavior. Then, he/she forms a behavioral intention. Whether an intention translates into action is determined by the actual level of control over an action. For details see e.g. [2] and fig. 1. In our model, we additionally assume that between the formation of an intention and the eventual action, i.e. migration, a planning phase and a preparation phase occur. We assume that forming an intention takes time, as well as proceeding from intention to planning, from planning to preparation, and from preparation to action. Agents can exit the decision process leading to eventual migration at any phase and any point in time when the basis of their positive evaluation of migration – i.e. attitudes,

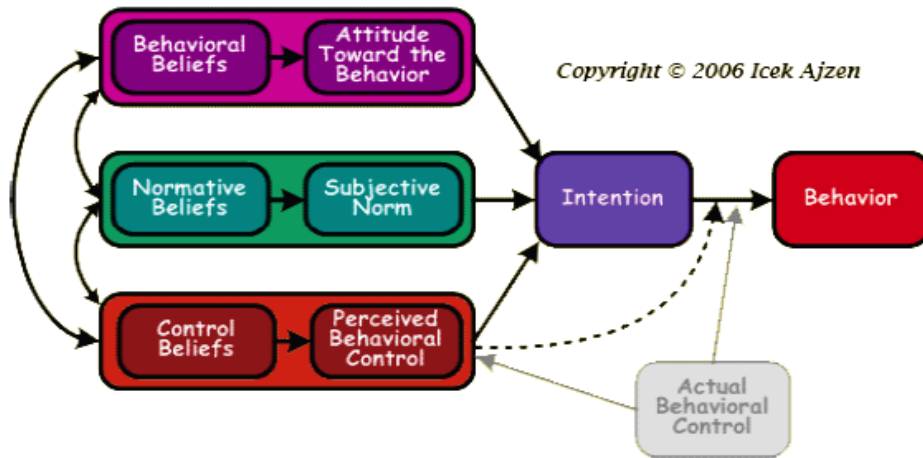


Figure 1 – Schematic representation of the TPB, taken from <http://people.umass.edu/aizen/tpb.diag.html>

social norms or perceived behavioral control – changes due to changes in factors such as policy, demographic events, or the behavior of others. The time an agent spends in a phase is stochastic. It is determined by an exponential distribution. Its parameter is calibrated taking into consideration the average time between migration attempt(s) and eventual migration in the MAFE data.

2.1.3 Why is/are certain decision model(s) chosen?

There are several reasons for choosing the Theory of Planned Behavior as decision model. The first reason is that it is an established theory from Social Psychology. Second, it has been shown to be an empirically relevant predictor of behavior (see [3] for a meta-analysis of empirical findings related to the TPB). Third, the TPB offers a behavioral heuristic which is apt for deliberate decisions that involve high levels of uncertainty, such as the migration decision. We know that the migration decision has far-reaching consequences, many of which are unforeseeable due to the interconnectedness of migration decisions, the stochastic behavior of the labor markets in both host and home country, and unpredictable future policy change. Fourth, often a fairly long time passes between the formation of a positive migration attitude and the actual migration move, and frequently the migration does not occur in the end despite of a positive attitude. In contrast to other

decision models, e.g. utility maximization, the TPB makes it possible to incorporate attrition during the decision process, which is ultimately an explanation for the relatively low observed flows of migration despite positive migration attitudes ([7]). Fifth, the TPB allows the inclusion of many different background factors evidently influencing the migration decision, without losing rigour: Factors with a clearly defined interpretation can be included in the model, such as attitudes, norms and beliefs.

2.1.4 If the model/submodel (e.g. the decision model) is based on empirical data, where do the data come from?

The data will come from the MAFE data set. The MAFE project is coordinated by INED (C. Beauchemin) in partnership with the Université catholique de Louvain (B. Schoumaker), Maastricht University (V. Mazzucato), the Université Cheikh Anta Diop (P. Sakho), the Université de Kinshasa (J. Mangalu), the University of Ghana (P. Quartey), the Universitat Pompeu Fabra (P. Baizan), the Consejo Superior de Investigaciones Cientificas (A. Gonzalez-Ferrer), the Forum Internazionale ed Europeo di Ricerche sull'immigrazione (E. Castagnone), and the University of Sussex (R. Black). The MAFE project has received funding from the European Community's Seventh Framework Programme under grant agreement 217206. The MAFE-Senegal survey was conducted with the financial support of INED, the Agence Nationale de la Recherche (France), the Région Ile de France and the FSP programme 'International Migrations, territorial reorganizations and development of the countries of the South'. For more details, see: <http://mafeproject.site.ined.fr/>

2.1.5 At which level of aggregation is the data available?

MAFE data is available at the household and the individual level.

2.2 Individual Decision-Making

2.2.1 What are the subjects and objects of the decision-making?

The subjects of decision-making are the model agents, which represent potential and actual migrants. The object of decision-making is whether or not to migrate from Senegal to another country. The MAFE data set comprises information on individuals migrating to a wide range of different countries. In our model, we abstract from the location choice and model a single artificial destination country.

On which level of aggregation is decision-making modelled?

Decisions are made at the individual level.

Are multiple levels of decision making included?

Agents are organized in households which pool resources. Household members have an impact on one another's migration attitude. Moreover, agents are influenced by network neighbors. The decision itself, however, is made at the individual level.

2.2.2 What is the basic rationality behind agent decision-making in the model? Do agents pursue an explicit objective or have other success criteria?

Agents pursue happiness, which is influenced by a high expected level of income and by being close to family. Thus, high income and being close to family are factors that influence migration attitude and constitute possible reasons for migration in this model. The goal of increasing one's happiness is implicit in the decision rule through the choice of relevant, salient outcomes of migration which are included in the computation of the agent's attitude. Since return is not an option in this model version, agents do not evaluate the success of their choice. Once an agent has migrated he remains in the host country.

2.2.3 How do agents make their decisions?

Agents who are not currently considering migration re-evaluate their situation after an exponentially distributed waiting time has passed. Agents who are currently at any stage of the decision process leading to migration re-evaluate their situation before moving to the next stage. The evaluation takes place following the TPB. That is, agents update their migration attitude, their current social norms and their perceived behavioral control (see section 3.4.1 for details), which jointly determine the migration intention. The migration intention is transformed into a value between 0 and 1 using a logit framework (see section 3.4.1 for details). The alternative to migration is remaining in the current state, whose evaluation by the agent is normalized to 0. Thus, if an agent's intention > 0 , he has a propensity to start or to continue considering migration. A uniformly distributed random number between zero and one decides whether the agent stays in the current decision phase, or if he wasn't considering migration yet, he enters the first stage.

Next, the rate of moving to the next decision stage (e.g. planning, or preparation) is determined using an exponential survival model. The rate of the model is composed of a baseline rate and an individual component, so that individuals with higher intentions have shorter expected waiting times (see section 3.4.1 for details). Once the time comes that the agent is scheduled to move to the next decision stage, he re-evaluates his situation as described above, and so forth. Possible transitions between different decision stages are shown in fig. 2.

2.2.4 Do the agents adapt their behaviour to changing endogenous and exogenous state variables? And if yes, how?

The agents adapt to several endogenous as well as exogenous factors. This happens through updating of beliefs, either of the valuation of potential outcomes of migration or of their subjective probabilities of occurrence if the agent migrates. Valuations change when children are born, marriages form or dissolve, or when the agent's household composition changes, e.g. through a household member's migration. Subjective probabilities are affected by the experience of others that the agent learns about and by the help offered by others.

In this model, the subjective probability of a potential outcome to occur is defined as the proportion of agents in their network for whom a particular event occurred. Thus, for determining the subjective probability that the income will increase through migration, the agent computes the proportion of migrant network neighbors whose income after migration was larger than his own current income. If the agent has not yet observed any migrants he assumes a probability of .5, which is then updated as soon as new information becomes available.

The subjective probability related to perceived behavioral control is updated in a similar manner: The importance of visa regulation and border enforcement is affected by policy, whereas the subjective probability to be able to successfully manage the border crossing is affected by the experience of others. Here, the agent uses the proportion of successful migration attempts he knows about from his network neighbors as proxy for the probability to overcome the difficulties of the border crossing.

The migration cost is assumed to decrease by a fixed amount if someone has network neighbors in the host country. The rationale behind this is that other migrants often provide information and direct support regarding housing and job search in the host country. This lowers the cost of migration, part of which is foregone income through extended periods of unemployment

in the host country. The subjective probability to control the migration cost is affected by the household's financial situation, which is a function of household composition and joint income of the household members. The agents consider the proportion of days, of all days spent in the current decision stage, on which they would have been able to afford the migration cost and use this as probability to be able to afford the migration cost in the period in which they eventually migrate.

Social norms are updated as well; they are just the proportion of network neighbors who have ever migrated. These beliefs affect behavior in a probabilistic way (see section 3.4.1).

Extensions

- The migration cost could be made to depend linearly on the number of migrant network neighbors. Alternatively, the migration cost could be a decreasing but convex function of number of migrant network neighbors.
- A more sophisticated model would include labor market policies in the host and home countries, which would then affect the subjective probability of a beneficial financial outcome of a migration event.
- The migration cost might be affected by policy.

2.2.5 Do social norms or cultural values play a role in the decision-making process?

Social norms play a role; they are directly part of the Theory of Planned Behavior and influence migration intention.

2.2.6 Do spatial aspects play a role in the decision process?

The model contains two regions, a home and a host country. Agents' network neighbors are also distinguished according to the country in which they live (home or host country). Only agents who have migrated can provide information on their migration success and only they can help potential migrants achieve a lower migration cost.

Extensions

Future versions will include a location decision. This will also assign more importance to the exact location of network neighbors.

2.2.7 Do temporal aspects play a role in the decision process?

There are waiting times between transitions to each new stage (intention formation, planning, preparation, migration). These waiting times follow an exponential distribution. The demographic waiting times from the microsimulation model influence the decision indirectly by determining the timing of marriage, birth, etc, which influence migration intention. Furthermore, the empirical demographic rates can depend on age.

2.2.8 To which extent and how is uncertainty included in the agents' decision rules?

We use a discrete choice logit framework to account for two kinds of model uncertainty: Neither can agents predict the future perfectly, nor can the modeller include all factors which potentially influence a person in reality. The valuation can thus be considered a random variable. Thus, even if intention is positive, an agent proceeds to the next stage only with a certain probability. There is a clear analogy to utility maximization: As shown by [4], the logit formulation describes observed behavior if the underlying decision model is myopic stochastic maximization of a utility function (see also [6]). The waiting time between stages is also stochastic.

2.3 Learning

2.3.1 Is individual learning included in the decision process?

There is an update of subjective probabilities regarding the occurrence of different outcomes and control factors, but this is not learning in the strict sense of learning about the true state of the world and how best to behave in it. Different subjective probabilities may reflect actual changes in exogenous or endogenous factors. However, changes in subjective probabilities are not always improvements in accuracy towards the true probabilities for two reasons. First, within the model framework, agents might be observing outliers and thus deteriorate in their accuracy. Second, at a meta level, the model might be misspecified or there might be measurement error in the data. Thus, the true probabilities are uncertain and it is unclear whether the subjective approximation performed by the agents does actually improve their predictions.

2.3.2 Is collective learning implemented in the model?

No.

2.4 Individual Sensing

2.4.1 What endogenous and exogenous state variables are individuals assumed to sense and consider in their decisions?

Individuals sense all of their own state variables as well as the current policy situation. However, they do not know if and when demographic events or changes in employment status will occur. They only know about them once they have occurred. They also sense several of the other agents' state variables (see below).

Is the sensing process erroneous?

No.

2.4.2 What state variables of which other individuals can an individual perceive?

They know several of their current network neighbors' state variables (migrant or not, income, family status) as well as aggregate statistics on past migration behavior of all previous network neighbors (how many increased their income compared to the potential migrant's current income).

Is the sensing process erroneous?

No.

2.4.3 What is the spatial scale of sensing?

Network neighbors are family members and those individuals who are in an agent's Moore neighborhood at birth, as well as those individuals who live in the same household at some point during the simulation and those that live in neighboring households. Whenever an agent moves to a new household he gains additional network neighbors. This also happens when other individuals move to the agent's household, since links are always bidirectional. Only network neighbors' state variables can be sensed.

2.4.4 Are the mechanisms by which agents obtain information modelled explicitly, or are individuals simply assumed to know these variables?

They are assumed to know these variables.

2.4.5 Are the costs for cognition and the costs for gathering information explicitly included in the model?

No.

2.5 Individual Prediction

2.5.1 Which data do the agents use to predict future conditions?

They use current information and assume it is a good proxy for future conditions.

2.5.2 What internal models are agents assumed to use to estimate future conditions or consequences of their decisions?

Consequences of their decision to migrate are estimated by observing previous migrants' income.

2.5.3 Might agents be erroneous in the prediction process, and how is it implemented?

Yes, because the income in the host country is a random draw from a distribution (see section 3.4.1), agents might be wrong about whether or not their income will increase through migration.

2.6 Interaction

2.6.1 Are interactions among agents and entities assumed as direct or indirect?

They are direct, also they mainly take the form of information transfer which is not modelled explicitly. Apart from information, social capital is transferred as well, which manifests itself in a lower migration cost if someone knows previous migrants. Within households, agents share financial means.

2.6.2 On what do the interactions depend?

They depend on whether agents live in the same household and on whether they are network neighbors.

2.6.3 If the interactions involve communication, how are such communications represented?

Communication is indirect through transfer of state variable values.

2.6.4 If a coordination network exists, how does it affect the agent behaviour? Is the structure of the network imposed or emergent?

At the beginning of an agent's life, he forms links to his parents and older siblings (no matter where they live), to all other members of the household (e.g. unmarried aunts and uncles or grandparents if the mother is not married) and to all members of the households which are located in the individual's household's Moore neighborhood¹. Those network links remain throughout life, regardless of location in future periods. Additionally, new links are formed when agents move to a new household, e.g. through marriage. Agents then include all members of the new household and all members of neighboring households in their network. Links are bidirectional. Thus, agents are also automatically linked to other agents moving into their household. Agents exchange information only with direct network neighbors. They are indirectly influenced by their neighbors' neighbors as well, because their decisions depend on the information that they have received. Moreover, network neighbors in the host country help newly arrived migrants to find jobs. This is modelled as a decrease in the migration cost for those people who have network neighbors in the host country.

Alternatives

Of course there are a lot of possibilities for network formation; here a very simple one is chosen. Maybe the most straightforward extension would be homophily, so that agents who are similar in terms of skill and/or family status and/or employment status are more likely to connect. Another option would be to cut and form new network ties at regular intervals. Finally, it is possible to have link strength increase and decrease with spatial proximity and to cut links when they have become too weak. This is particularly interesting if one wants to model return and circular migration. The remaining link formation mechanisms summarized in [6] (transitive closure, hierarchy, popularity effects) could also be tried.

¹For the initial setup see section 3.2.

2.7 Collectives

2.7.1 Do the individuals form or belong to aggregations that affect and are affected by the individuals?

Yes, they belong to households. Financial means are shared within households. New households form when agents get married and dissolve when they get divorced.

Are these aggregations imposed by the modeller or do they emerge during the simulation?

The very first setting is imposed by the modeller, but then they emerge during the simulation.

2.7.2 How are collectives represented?

Agents are assigned to households. Households themselves are entities as well, but they do not act. Capital is registered at the household level because financial means are shared within households.

2.8 Heterogeneity

2.8.1 Are the agents heterogeneous?

Yes.

If yes, which state variables and/or processes differ between the agents?

Agents differ in the following state variables:

- gender
- marital status
- identification number of spouse
- number and identification numbers of children
- income
- location (host or home country)
- household membership

- migration stage
- time in current stage

At the time of birth, agents differ only in gender, location and household membership. Some processes occur only during certain ages or depend on age in probability of occurrence (see section 1.3). The valuation and subjective probabilities for attitude formation and perceived behavioral control depend on state variables. The decision process itself does not differ between agents.

2.8.2 Are the agents heterogenous in their decision making? If yes, which decision models or decision objects differ between the agents?

They are only heterogeneous in their valuations and subjective probabilities, but the decision model itself is the same for everyone.

2.9 Stochasticity

2.9.1 What processes (including initialisation) are modelled by assuming they are random or partly random?

The following processes include random components:

- Birth: The agent's gender is random
- Location on the grid
- Waiting times until demographic transitions
- Waiting time in migration decision stages
- Random number draw to determine if someone drops out of the decision process

2.10 Observation

2.10.1 What data are collected from the ABM for testing, understanding and analysing it, and how and when are they collected?

This has not been decided.

2.10.2 What key results, outputs or characteristics of the model are emerging from the individuals? (Emergence)

This cannot be answered yet.

3 Details

3.1 Implementation Details

3.1.1 How has the model been implemented?

The model will be implemented in NetLogo. The stochastic demographic processes will be modelled in R.

3.1.2 Is the model accessible, and if so where?

Not yet.

3.2 Initialisation

3.2.1 What is the initial state of the model world, i.e. at time $t=0$ of a simulation run?

The initial values of the state variables have not been determined yet. They will be modelled according to an actual population, based on MAFE data. Initial simulated agents will all have a real-life counterpart in the MAFE data. After initialisation, of course, the life courses of real and simulated individuals will most likely diverge. The initial proportions of people in each state will be taken from data. Married individuals in the data set are also married in the model setup.

Some proportion of the initial population will be in the host country already when the simulation starts. Every host country household – that is, all of its members – will be connected through a network link with one random household in the home country, and thus all of its members. Initial income for all individuals will be drawn from a distribution (see below).

3.2.2 Is the initialisation always the same, or is it allowed to vary among simulations?

It will be allowed to vary among simulations to test the impact of different parameter settings. Some calibration procedure yet to be determined might be applied.

3.2.3 Are the initial values chosen arbitrarily or based on data?

They will eventually all be based on data.

3.3 Input Data

3.3.1 Does the model use input from external sources such as data files or other models to represent processes that change over time?

Yes, it uses input files for all the demographic processes using the MicSim package in R. This is achieved through the ‘r’-extension in Netlogo.

3.4 Submodels

3.4.1 What, in detail, are the submodels that represent the processes listed in ‘Process overview and scheduling’?

Demographic events

In a first step, waiting time distributions between demographic transitions are estimated from the MAFE data. These waiting times serve as input for a stochastic microsimulation, where the life courses of a population are simulated. Over time, agents can be exposed to different events such as, for example, getting divorced, having a child, or dying. These events are called competing risks. The event that happens next to a person in the simulation is that event which has the shortest waiting time. For the estimation of waiting times from the MAFE data, death is an absorbing state. The migration transition will not be part of the microsimulation, but the estimated rates will be used for validating the decision model in the ABM. The migration transition will be replaced by a decision process, modeled in an agent-based fashion. The microsimulation is run before the agent-based model is initiated. Thus, most demographic transitions, as well as the time they will die, have already been scheduled for all individuals when the agent-based simulation starts.

The following table 2 depicts all possible states and transitions between states for the home country population which are pre-determined and thus do not depend on decision-making and interaction within the agent-based model. For the sake of brevity, the maximum number of children included in the table is 3, however, more children are possible in the simulation. The upper bound of number of children that can be born to a person is set to 13, the maximum number of children observed in the MAFE data set.

Table 1 – Possible pre-determined transitions of a female agent. NM = never married, M = married, D = divorce, W = widowed. Numbers denote numbers of children. De = Death.

Nr.	State	Possible new states
1	(NM, 0)	(M, 0), (NM, 1), De
2	(M,0)	(W, 0), (D, 0), (M, 1), De
3	(NM,1)	(M, 1), (NM, 0), (NM, 2), De
4	(W, 0)	(M, 0), (W, 1), De
5	(D, 0)	(M, 0), (D, 1), De
6	(M, 1)	(W, 1), (D, 1), (M, 0), (M, 2), De
7	(NM, 2)	(M, 2), (NM, 1), (NM, 3), De
8	(W, 1)	(M, 1), (W, 0), (W, 2), De
9	(D, 1)	(M, 1), (D, 0), (D, 2), De
10	(M, 2),	(W, 2), (M, 1), (D, 2), (M, 3), De
11	(NM, 3)	(M, 3), (M, 2), (NM, 3+), De
12	(W, 2)	(M, 2), (W, 1), (W, 3), De
13	(D, 2)	(M, 2), (D, 1), (D, 3), De
14	(M, 3)	(W, 3), (D, 3), (M, 2), (M, 3+), De
15	(W, 3)	(M, 3), (W, 2), (W, 3+), De
16	(D, 3)	(M, 3), (D, 2), (D, 3+), De

Marriage, fertility, mortality, and divorce rates are assumed to depend on age and gender only. This will be refined in later versions of the model if possible. The first and most natural candidate for a factor on which fertility rates should depend is the number and the age of children already born.

In the ABM, agents experience demographic transitions which were computed in the microsimulation.

There are a few exceptions to the above-mentioned procedure, which arise from the fact that we link individuals in family relations. In the ABM, men become fathers deterministically when their wives experience childbirth events. Divorce is similar: We assume that divorce is triggered by women, thus men get divorced deterministically.

Widowhood is deterministic: Agents transition to widowhood when their spouse dies. The transition to widowhood is not included in the microsimulation, only death is. Children can die, too. In the microsimulation, this implies a transition for the mother, in that the number of her children decreases by one. Dead agents are removed from the agent-based model.

The microsimulation model does not include a mate matching procedure. That is, life-courses of paired people are modelled independently of each other. This is in contrast to the ABM where paired agents are part of a household. Therefore, re-marriage after divorce or widowhood is not part of the microsimulation for men, because it is not possible to pre-determine if they will be divorced or widowed and thus available at a given point in time, because this depends on the stochastic transitions that the wife experiences. Thus, re-marriage rates for newly single men are determined in the ABM (see below).

In our data set, polygamous relationships are frequent: In 2008, 33.9 % of married women and 17.3% of married men in the data set lived in polygamous marriages. If we restricted our analysis to monogamous relationships, the proportion of unmarried agents would be much larger in the simulation than in reality. Therefore, we have to include polygamy as an option in the model. Men can thus experience several marriage events in a row, or, in other words, they can get married again without having been divorced. We distinguish three different transition rates for men, namely to first, second and subsequent marriages.

Thus, while women in the microsimulation experience the whole range of possible transitions in table 2, except widowhood for the reasons outlined above, men experience only first and subsequent marriages and death. Everything else is determined in the agent-based model.

After a migration event, a new remaining life course is simulated for a migrant, employing migrant-specific transition rates to life events.

Birth

When a woman experiences a childbirth event, the man she is married to is assumed to be the father. For the newborn, a new agent is created and endowed with characteristics as determined by the microsimulation; in this version only a randomly assigned gender. Also this new agent has an identity in the microsimulation and thus a pre-determined life course apart from the migration decision. At birth, new agents form network links to their own family members (parents and siblings), whether or not they live in the same household. In addition, they form links to all other members of their household (e.g. unmarried aunts or uncles or grandparents, if the mother is not married) as well as to all members in all households in the Moore neighborhood. These links remain throughout life.

Marriage

6 months before a marriage event is scheduled for a person² he joins a marriage market; a queue like described in [8]. Every individual carries a stamp with their scheduled marriage date. Beforehand, a logit model which predicts the probability to be married to a certain individuals with given characteristics has been estimated on the MAFE data. Like in [8], covariates are age of the man, age difference (in categories), marriage history of the partner and number of children with previous partners³. Like suggested by [8], because we only have information on couples who did actually get married, actual individuals from the data set are paired randomly to create observations for pairings that did not result in marriage. Thus, the outcome variable is 1 if in reality a marriage was observed, and 0 otherwise. Proceeding in this way we obtain a compatibility measure between 0 and 1 for each other individual in the simulation.

Every time a new individual enters the market, the following mate-matching routine is run. The individual with the earliest marriage time is chosen. This individual makes their partner choice. All individuals in the marriage market are checked whose marriage time is within one year of the agent's marriage time (from 6 months before the agent's marriage time to 6 months after). Of those, the individual selects the partner whose compatibility score is highest. Now, the compatibility score is interpreted as a probability again and a random draw determines if the individuals do get

²or, if the marriage date becomes available at shorter notice, like in the case of divorced male agents or widowed agents, as soon as possible

³Once different educational levels are introduced, the educational level will also be a covariate.

married. If they do, they are paired with each other and leave the marriage market. There are some restrictions to mating: Relationships involving incest or extreme age differences are excluded, and people can only marry those currently residing in the same country. In order to make sure that individuals who are unable to find a partner do not block the marriage market, they give up if there is no one in the marriage market who matches their criteria and there are at least 20 people in the marriage market.

Newly married agents form a new household together. If any of the two partners is living with his children from a previous union, they move in with the new couple. If it is not the first wife for the male agent, the new wife moves in with the man's existing household. Agents form network links to members of their new household.

Household dissolution

If a woman experiences a divorce event, a divorce occurs. For the ex-husband a new life course is scheduled which replaces the previous one. For this, the microsimulation is called from within the agent-based model.

If a couple divorces, the household dissolves. Both former partners form new households. If the couple has any children under the age of 16 they stay with the mother⁴. If one of several wives divorces a man, she moves out of the household and forms a new household.

Agents join the marriage market again in the way described above if a re-marriage event is scheduled for them.

If an agent is widowed, he computes a new life course as well to replace the pre-determined life course.

Income

At the end of each month, agents receive their income. All incomes are added to household capital. The two countries have different average income levels. When an agent turns 16 and after migration an income is drawn from an empirically determined income distribution that differs by country⁵.

⁴This is arbitrary and could be replaced by the assumption to be assigned to one of the parents randomly or to the mother with a certain probability determined empirically.

⁵This means we assume there is no inflation. Inflation can be easily introduced, as well as some seniority component to the income.

Consumption

At the beginning of a month it is computed how much capital a household can save. Adults need to consume m_1 and children under the age of 16 need to consume m_2 . Then, every day the capital $c_{h,t}$ of household h is updated as follows:

$$c_{h,t} = c_{h,t-1} - A_{h,t} \frac{m_1}{30} - K_{h,t} \frac{m_2}{30} \quad (1)$$

where $A_{h,t}$ is the number of adults in the household and $K_{h,t}$ is the number of children. The remaining capital is saved.

In case the household does not have enough capital to feed everyone throughout the month, i.e. $c_{h,t} < A_{h,t}m_1 + K_{h,t}m_2$, every household member computes a new life course for themselves as described earlier, in which the mortality rate is increased by a fixed proportion.

Retirement

Once the agents reach the age of 65 they retire. They stop earning income. If both spouses have retired or if the retiree is living by himself, the agent or the agent and his spouse join the household of that of their children who lives in the wealthiest household. If the retired couple or individual does not have any children, it joins the household of the network neighbor with the wealthiest household. If the agent/couple does not have any network neighbors, they die.

Attitude

Every agent in the home country updates his migration attitude when he is scheduled to progress to a new stage in the migration decision process (see fig. 2). The same is true for social norms and perceived behavioral control.

The migration attitude of agent i at time t is defined as:

$$A_{i,t} = ew_{i,t}bw_{i,t} + ef_{i,t}bf_{i,t} \quad (2)$$

where ew is the evaluation of higher income in the host country, bw is the subjective probability to achieve higher income, ef is the evaluation of family reunification, bf is the subjective probability to achieve family reunification.

The functional form of this is arbitrary and assumes that higher income and family reunification are perfect substitutes with respect to attitude formation. One could think of modelling attitude - in analogy to utility func-

tions - as a Cobb-Douglas function so that agents have a preference for equal contributions of higher income and family reunification.

Other possible outcomes of migration to potentially include are risk hedging, better living conditions, saving for retirement or housing, and the formation of human capital.

In this simple version, bf is assumed to be 1, because once migrated the agent will reunite with his family with probability 1. The other components of attitude formation are computed as follows:

$$ew_{i,t} = a_1c_{h,t} + a_2w_{i,t} + a_3(A_{h,t} + K_{h,t}) \quad (3)$$

where $c_{h,t}$ is the capital of the household of agent i at time t , $w_{i,t}$ is the agent's income at time t , $A_{h,t}$ is the number of adults in the household of i at time t , $K_{h,t}$ is the number of children and a_1 , a_2 and a_3 are weighing parameters. In this version parameters do not differ between individuals, but can be allowed to do so easily.

Next,

$$ef_{i,t} = a_4M_{i,t} \quad (4)$$

where $M_{i,t}$ is the number of previous family members (parents, grandparents, current spouses, children, grandchildren, siblings) who have migrated and a_4 is a fixed weighting parameter. Again, the functional forms are arbitrary.

The subjective probability to achieve higher income with migration $bw_{i,t}$ is computed as follows. For each of the agent's network neighbors in the host country the income after migration is compared to the agent's income since they turned 16. The subjective probability $bw_{i,t}$ is then the proportion of the agent's network neighbors in the host country who have had higher incomes than the agent's income. The total number of person-day observations of similar network neighbors in the host country and the number of those observations which represented higher incomes are recorded so that $bw_{i,t}$ can be updated at any given day. If no migrants have been observed yet, the agent assumes a $bw_{i,t} = .5$, which is then updated as soon as new information is available.

Social norms

Social norms $SN_{i,t}$ are updated at the same points in time as attitudes; they are just the proportion of network neighbors who have ever migrated.

PBC

Perceived behavioral control, which is also updated at the same points in time as attitude, is

$$PBC_{i,t} = -(pb_t cb_{i,t} + pc_{i,t} cc_{i,t}) \quad (5)$$

where pb_t is the importance of border enforcement at time t , $cb_{i,t}$ is the subjective probability that border enforcement will influence the migration attempt of agent i at time t , $pc_{i,t}$ is the importance of the migration cost to agent i at time t , and $cc_{i,t}$ is the subjective probability that the migration cost precludes the agent's migration.

The elements of PBC are defined as follows. The level of border enforcement⁶ pb_t is exogenously given and assumed to be known to the agents, e.g. through the media. The subjective probability that the border enforcement will hinder the agent's migration, $cb_{i,t}$, is the proportion of failed migration attempts of the agent's network neighbors. The migration cost is assumed to depend on whether someone knows a migrant. It is thus $pc_{i,t} = pc$ if someone does not know a migrant, and $pc_{i,t} = pc - dc$ if someone does know another migrant. The last element to be defined is the subjective probability to not be able to afford the migration cost, $cc_{i,t}$. The total amount of funds that an agent has available to pay the migration cost is the household's capital. The agent keeps track of $c_{i,t}$ and computes the average monthly capital of his current household in its current composition. The proportion of months in which $c_{i,t} < pc_{i,t}$, i.e., the proportion of months in which the agent would not have been able to afford the migration cost, is $cc_{i,t}$.

Exit

Now, a migration intention of agent i at time t can be computed from the attitude, the social norms and the perceived behavioral control:

$$I_{i,t} = a_5 A_{i,t} + a_6 SN_{i,t} + a_7 PBC_{i,t} \quad (6)$$

where a_5 , a_6 and a_7 are weighting parameters. While attitude and social norms are positive values, PBC is negative, so that the agent's intention $I_{i,t}$ might range from minus infinity to infinity. If intention is negative, the agent drops out of the decision process. Otherwise, the intention value is transformed to a probability $prob$ to move to the next decision stage:

⁶This very broadly incorporates all kinds of policy measures which make migration more difficult.

$$prob_{i,t} = \frac{1}{1 + exp(-I_{i,t})} \quad (7)$$

The agent makes a random draw in the interval $[0, 1]$. If the value drawn is smaller than $prob_{i,t}$, the agent moves to the next decision stage (i.e., intention, planning, or preparation). Otherwise, the agent exits his current stage and moves back to the beginning of the decision making process.

Progress

Agents enter the first stage of the migration decision making process, the ‘No intention’ stage, when they turn 14. Once an agent enters a new stage, the rate of moving to the next stage is determined using an exponential survival model with the intention as the only idiosyncratic parameter. The probability density function of the exponential distribution is

$$p(t; \lambda_{i,t}) = \lambda_{i,t} e^{-t\lambda_{i,t}} \quad (8)$$

so that the expected waiting time is $1/\lambda_{i,t}$. The idiosyncratic parameter $\lambda_{i,t}$ is determined as:

$$\lambda_{i,t}(\rho, I_{i,t}) = \rho e^{a_8 * I_{i,t}} \quad (9)$$

where ρ is the baseline rate for an intention value of 0, $I_{i,t}$ is the intention computed as in equ. 6, and a_8 is a weighing parameter.

Entering a new stage in the migration decision process and experiencing a new demographic event are competing risks. If a demographic event happens to an individual before the waiting time to the next decision stage is over, a new intention value is computed right away after the demographic event has happened in order to compute a new waiting time for the next step in the migration decision as shown in eq. (9). The reason for this is that the demographic event can change the evaluation of the attractiveness and feasibility of migration dramatically. Thus, the waiting time has to be adjusted accordingly.

Once the time comes that the agent is scheduled to move to the next decision stage, he re-evaluates his situation as described in equations (6) and (7), and so forth.

Migration

When an agent has passed the assigned waiting time in the preparation stage, he computes his intention one more time and performs another random draw. If this random number is smaller than $prob_{i,t}$, and if the agent’s

household capital exceeds the migration cost, the agent finally attempts migration. Otherwise, the agent moves to the very first stage of the decision process. Once an agent has decided to migrate, actual border control pb_t comes into play. An agent's probability to successfully migrate $probm_{i,t}$ is determined as

$$probm_{i,t} = \frac{1}{1 + \exp(-\frac{1}{pb_t})}, pb_t > 0. \quad (10)$$

A random draw determines if the agent manages to cross the border. If the migration attempt is successful, the agent changes his location to the host country. If the agent is unmarried or the spouse is still in the home country, the migrant still officially belongs to his former home country household, despite living abroad. This means he still joins in the means sharing within households. His contribution to the household income are remittances. If the agent is married and the spouse migrated before him, both spouses form a new household in the host country. Thus, they share income within their new household and remittances to the home country stop. If the agent is married, the agent's under-age children stay with the spouse in the home country. If the agent is single or the spouse has migrated before him, the agent's children accompany him to the host country. If the migration attempt is unsuccessful the agent remains in the home country and moves back to the first decision stage (No-intention stage).

3.4.2 What are the model parameters, their dimensions and reference values?

This can only be answered later.

3.4.3 How were the submodels designed or chosen, and how were they parameterised and then tested?

The submodels were chosen with the idea in mind to have stochastic demographic processes and a conscious migration decision. The parameterization and testing will follow later.

References

- [1] I. Ajzen. The theory of planned behavior. *Organizational behavior and human decision processes*, 50(2):179–211, 1991.

- [2] I. Ajzen and J. Klobas. Fertility intentions: An approach based on the theory of planned behavior. *Demographic Research*, 29:203 – 232, 2013.
- [3] C. J. Armitage and M. Conner. Efficacy of the theory of planned behaviour: A meta-analytic review. *British Journal of Social Psychology*, 40(4).
- [4] D. McFadden. Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, pages 105–142, 1973.
- [5] B. Müller, F. Bohn, G. Dreßler, J. Groeneveld, C. Klassert, R. Martin, M. Schlüter, J. Schulze, H. Weise, and N. Schwarz. Describing human decisions in agent-based models ODD + D, an extension of the ODD protocol. *Environmental Modelling & Software*, 48(0):37 – 48, 2013.
- [6] T. A. Snijders and C. E. Steglich. Representing micro–macro linkages by actor-based dynamic network models. *Sociological Methods & Research*, 2013.
- [7] F. Willekens. An agent-based simulation model (abs) of migration based on the theory of planned behaviour. Technical report, Paper presented at the workshop ‘Recent Developments and Future Directions in Agent-Based Modelling in Population Studies, Leuven, 18-19 Sept 2014, 2014.
- [8] S. Zinn. A mate-matching algorithm for continuous-time microsimulation models. *International Journal of Microsimulation*, 5(1):31–51, 2012.
- [9] S. Zinn. *Package MicSim*. Published on CRAN, 2014.

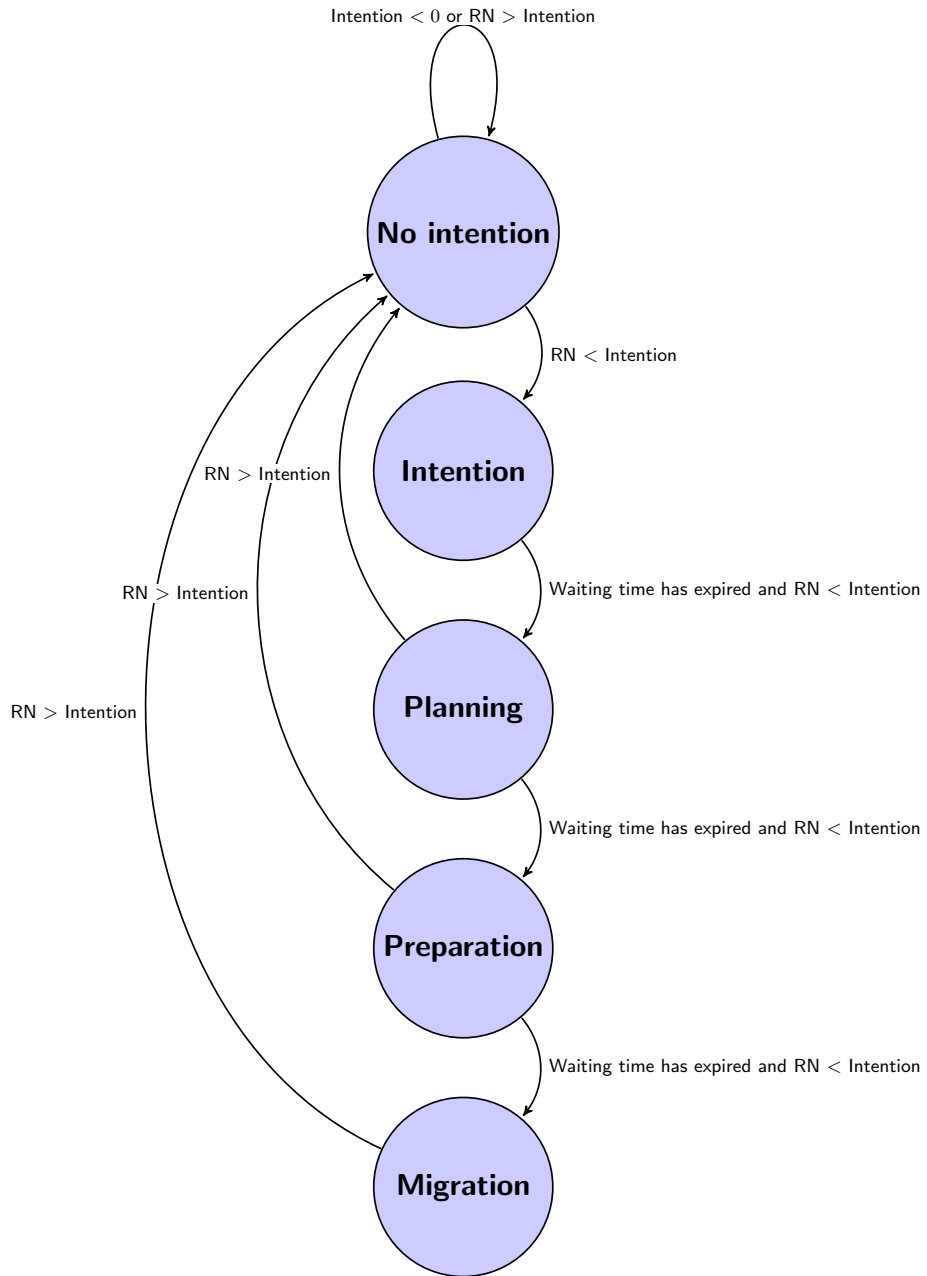


Figure 2 – Possible transitions between different states. RN = random number, uniformly distributed between zero and one.