Dynamic Microsimulation for Population Projections in Developing Countries: A Portable Application with a Country Case Study (DYNAMIS-POP-MRT)

Martin Spielauer, Olivier Dupriez

Abstract

Population projections are key for policy making and planning. Unlike the traditional cohort component projection method—a macro approach— microsimulation models can produce projections disaggregated by a broad variety of individual characteristics. They allow to model realistic life-courses and their diversity, and support the modeling of interactions between people. By developing the DYNAMIS-POP-MRT model for Mauritania, we aimed to demonstrate the feasibility and potential benefits of population microsimulation in the context of developing countries based on typically available data-sets. We start with a microsimulation implementation of DemProj, a macro model widely used in developing countries. We then extend the model to incorporate more detailed characteristics and behaviors. Extensions include a fertility module allowing realistic projections of fertility by level of education of the women. We also add a child mortality module that incorporates factors like mother's age and education. It's the inter-generational transmission of education is modeled. First union formation is also introduced as a key determinant of the timing of first births. For most behaviors, users of the model can choose between a base version resembling existing macro projections and a refined version. Results of the refined models can be aligned to the base versions thus re-producing existing macro projections for aggregate outcomes, while producing realistic life-courses and additional characteristics, and allowing further disaggregation of results. A simple study on the demographic effects of educational improvements in Mauritania is provided as an application example. The model is portable to other counties and freely available with extensive documentation and software.

Keywords: microsimulation, developing countries, population projections, disaggregation

Acknowledgments and disclaimer

The project, led by the World Bank Development Data Group, was funded by the World Bank Knowledge for Change Program (KCP III, grant number TF0A1095) and by the International Household Survey Network Trust Fund from the United Kingdom Department for International Development (DFID, grant number TF011722 executed by the World Bank Development Data Group).

The project aimed to demonstrate how new approaches—synthetic data generation and dynamic microsimulation—can take advantage and integrate data from multiple sources—in this case population censuses and household surveys—to inform development policies and research on development issues.

This paper draws heavily on the extensive model documentation describing the microsimulation population projection model DYNAMIS-POP applied to Mauritania (DYNAMIS-POP-MRT). The model will be expanded and adapted to other countries. A more advanced specialized education module is being developed and will be integrated in the model. Solutions to exploit geo-spatial data will be developed and integrated in the model, allowing fine geographic disaggregation of the model's input and output. Ethnic affiliation is added as an additional parameter. And other specialized modules will be developed. This first application to Mauritania is thus the first of what we envision to become a collection of models to be developed and published as public goods under the generic name DYNAMIS (Dynamic Micro-Simulation). The application examples given in this paper draw on the results of a pilot implementation of the model conducted with staff from the Centre Mauritaniend'Analyse de Politiques (CMAP) in May 2017.The results presented in this paper are the output of a pilot research project, and are not to be considered as final or official results.

The project was initiated and led by Olivier Dupriez, Lead Statistician at the World Bank Development Data Group. Martin Spielauer, expert in microsimulation and consultant for the World Bank, is the main author of the model. Staff from the Centre Mauritaniend'Analyse de Politiques (CMAP) and from the Office National de la Statistique de Mauritanie participated in the country adaptation of the model, in the provision and analysis of input data, and in the analysis of the model output.

The model was developed using MODGEN (a freeware developed and published by Statistics Canada) with Microsoft Visual Studio Community 2017.

1. Introduction

Population projections are key for policy making and planning. Changes in the size and composition of the population are key determinants of the demand for goods and services, from basic food and education, to energy and housing. Population projections, based on multiple scenarios, help governments and decision makers make informed decisions.

Most countries and international organizations, like the World Bank and the United Nations, produce population projections using the cohort-component method, a macro approach limited to a very small number of characteristics. This provides projections by age and sex at the national level, often dis-aggregated by urban/rural areas; large countries sometimes provide projections at a sub-national level. Given the high importance of education on human capital and its influence as the "single most important variable besides age and sex" on demographic behaviors (Lutz et.al. 1999), population projections that include education later became available for most countries, but that extension defined the technical limit of this approach.

A more advanced and flexible approach consists of dynamic microsimulation models. In this approach, the life course of a large sample of individuals is simulated. This approach is more complex, but it has major advantages: it produces detailed projections of a broad variety of individual characteristics, models realistic life-courses and their diversity, and supports the modeling of interactions between people. The idea is to start with a micro-data population and make it evolve over time, unlike the cohort-component approach, which starts with a simple distribution of a population by age and sex but cannot track individuals over time. The idea is not new (van Imhoff and Post, 1998), but it has become feasible and affordable only recently with the availability of free programming technologies and improvements in data availability and accessibility. Still, applications so far were limited to the developed world. The most prominent example is Demosim, the microsimulation population projection model developed and used by Statistics Canada to project the Canadian population by visible minority group (Caron-Malenfant et.al. 2010) and Aboriginal identity (Morency et.al. 2015), to project its labor force (Martel et.al. 2011), and to study the effect of educational improvements on the future size and composition of the Aboriginal labor force (Spielauer 2014). The aim of the DYNAMIS-POP-MRT model is to demonstrate the feasibility and potential benefits of population microsimulation in the context of developing countries, using data-sets available for a wide set of countries around the globe.

Population microsimulation can be seen both as a complement to macro-projections and as their replacement. Microsimulations can replicate macro-projection models, but allow for model extensions beyond the technical limitations of cohort-component models. This point is demonstrated by DYNAMIS-POP which was developed in two phases. The first phase reproduces the macro model DemProj (Stover & Kirmeyer 2001), widely used in developing countries. In our example, we use Mauritanian data for its parameterization. In the second phase, the model is extended to incorporate more detailed characteristics and behaviors. Again we based the model on Mauritanian data sources (the population census and a demographic survey), for which equivalent data are available throughout the world. Extensions include the fertility module, which allows a realistic projection of family sizes and fertility by education of the women. We also added a child mortality module, incorporating factors such as the mother's education and age. Education is modeled to include its inter-generational transmission; first union formation is introduced as key determinant of the timing of first births. For most behaviors, model users can choose between a base version resembling existing macro projections and a refined version. Results of the refined models can be aligned to the base versions thus re-producing existing macro projections for aggregate outcomes, while producing realistic life-courses and additional characteristics.

The model is implemented in Modgen, a freely available software package developed and maintained at Statistics Canada. We use an anonymized sample of the 2013 Population Census dataset as baseline population. Behavioral modules are based on an analysis of microdata from the Population Census and from the Multiple Indicators Cluster Survey (MICS) 2015.The model has an intuitive graphical user interface and runs on a standard personal computer. The code and statistical analysis scripts (Stata) are openly accessible and can be used for microsimulation model development and implementation.

The growing interest in disaggregated data for development planning and monitoring (including monitoring the Sustainable Development Goals)justify a push for more disaggregation in projections and simulations. Microsimulation also allows a more explicit incorporation of theory and policy levers in projections. For example, it allows modeling of inter-generational dynamics and analysis of downstream effects on demographic change of education policy interventions.

This paper is organized in five sections. Following this introduction, Section 2 provides some context and rationale for using dynamic microsimulation for population projections in developing countries. In Section 3 we describe the DYNAMIS-POP-MRT model implemented for Mauritania. Section 4 gives some concrete application examples of the model use in Mauritania. The model is extendable as well as portable to other counties. Some recent developments are discussed in Section 5. The Appendix provides links to extensive documentation and software downloads.

2. Rationale and Context

Microsimulation in the context of socioeconomic applications can be perceived as an experiment with a virtual society of thousands—or millions—of individuals. Central to dynamic microsimulation is the explicit modeling of the time dimension, following people and their families or households over time. Dynamic modeling lends itself naturally to the modeling of policies with a longitudinal component, e.g., educational investments, especially in the context of general rapid social, economic, and demographic change that make it difficult to assess the contribution of individual policies to overall trends without tracking and comparing the lives of individuals who form a society. Rapid social and demographic change are typical for developing countries. This makes microsimulation a valuable tool to complement traditional evaluations of effects of development programs.

Microsimulation is attractive both from a theoretical and a practical point of view. It supports research embedded into modern paradigms, such as the life-course perspective, while simultaneously providing a tool for what-if analysis of high policy relevance. Typical application areas include tax-benefit analysis, analysis of pension system adequacy and sustainability, and health insurance. Using microsimulation in demographic projections is a recent development. Although this has been discussed in literature for almost two decades (e.g., Imhoff & Post 1998), large-scale implementations are recent. Statistics Canada was the first statistical office to produce official population projections using microsimulation in 2004. Called Demosim, this model is implemented using the free microsimulation programming technology Modgen, developed and published by Statistics Canada. Variants of Demosim are currently being developed for several European countries and Australia (Marois et.al. 2017).

The strengths of microsimulation do not come without limitations which affect the applicability of this approach. From a methodological point of view, the central limitation of microsimulation is that the degree of model detail does not go hand-in-hand with overall prediction power. Providing more detailed models, something at which microsimulation excels, does not necessarily mean the models are "better." The ability to produce distributions comes at the price of losing predictive power in projecting means, and the ability to make very accurate statements in the short run does not necessarily lead to models that are useful for long-term projections. An analogy is weather forecasts: detailed models for the weather tomorrow, on a geographical scale, will not be of use for the projection of global climate changes over the next centuries. This also applies to socioeconomic models. The longer the time horizon and the more important the mean, the more the focus should be directed to the main driving forces and a solid theoretical foundation of these mechanisms. The reason for this can be found in what is called randomness, caused by accumulated errors and biases of variable values (for a discussion of randomness in microsimulation, see Imhoff & Post 1998). The fundamental problem lies in the trade-off between the additional randomness introduced by additional variables and misspecification errors caused by models that are too simplified. This means that the large number of variables that models can include, which is the feature that makes microsimulation especially attractive, at some point comes at the price of randomness and a decrease in prediction power that occurs as the number of variables increases.

There are two ways of dealing with this trade-off. The first is to keep models simple. The second is to combine the strengths of different modeling approaches. Not surprisingly, in many large-scale microsimulation models, some outcomes are aligned or calibrated towards aggregated numbers or projections obtained by external means. The effort to keep models simple often leads to macro models bypassing microsimulation as a modeling strategy, the choice often justified with the higher development costs of microsimulation. This choice ignores the fact that microsimulation can often reproduce the results of macro models if needed (and at comparable costs), while also allowing for step-wise refinements and removal of simplifying assumptions inherent to macro models. This is the development strategy we followed building DYNAMIS-POP. For most behaviors, model users can choose between a base version resembling existing macro projections and a refined version. In addition, results of the refined models can be aligned to the base versions thus reproducing existing macro projections for aggregate outcomes, while producing realistic life-courses and additional characteristics. This makes DYNAMIS-POP a flexible analysis and projection tool.

Another often-stated drawback of microsimulation is that such models have high data demands, and high costs of acquisition and compilation of data. It can be noted, however, that such costs are not explicit costs associated with the microsimulation itself, but represent the price to be paid for research in general, and informed policy making in particular. Recent advances in data availability and standardization could turn this argument around: microsimulation can make available data more policy relevant, as it complements traditional data analysis and combines such analysis with a what-if projection tool. In the case of population projections, required data are readily available for many countries and the model can be generic as the input and output (i.e., requirements and purpose) of the model are very much the same across countries.

3. The DYNAMIS-POP-MRT Model

DYNAMIS-POP is a customizable modular microsimulation application for population projections. The model is parameterized for Mauritania, but can be easily ported to other countries. It is implemented using Statistics Canada's freely available Modgen microsimulation language. It has a user-friendly graphical user interface with a help function for both the user interface and the model and its modules, parameters etc. Users can easily change parameters and create and save new scenarios. It also has rich table output, which can be exported to Excel. The model can generate projected micro-data files, for user-defined variables and points in time. It runs on a standard PC under Windows. The execution time of a model run depends on processor speed, population sample size, time horizon of the simulation, and the user's choice of model selection and alignment routines. A typical model run, starting from an initial population sample of 250,000 persons and a 100-year time horizon, is approximately 10 minutes. This time can increase substantially for some alignment options.

The model is highly modular and contains 11 main modules. Together, they create a fully functional model application, but modules can be adapted and extended, and new modules can be added. The generic DYNAMIS-POP model can thus be used as the core of a modeling platform.

DYNAMIS-POP is documented in detail. Also its code and statistical analysis files are available to model builders, who can use the step-by-step instructions as a textbook and toolbox for microsimulation model development, adaptation, and implementation. All resources are available from the project website at http://ihsn.org/projects/dynamis-pop

3.1 Starting population

The model starts from a baseline population file—a standard comma-separated variables (CSV) text file containing nine variables. Records can be weighted and the file length does not have to correspond to the true population size nor the size of the simulated population sample, which are parameters. According to these parameters, when the file is larger than the simulated starting sample, the model automatically samples from the starting population file. If the file is smaller than the chosen starting sample, the model replicates observations. All model output is automatically scaled to the total population size regardless of the chosen sample size for the simulation. Choosing larger samples will reduce Monte Carlo variability at the expense of additional time requirements to run the model.

The starting population file was generated from a 13% sample of the Mauritanian 2013 census (~450.000 records); a anonymized starting population file of 100.000 persons is publicly available. The variable LASTBIR (time of last birth, not available in the census dataset) was imputed from MICS 2015 data.

```
WEIGHT        person's sample weight          e.g., 8.275
BIRTH         time of birth                   e.g., 1966.532
SEX           sex                             0 Female, 1 Male
POR           province of residence           0 .. 12
EDUC          primary education               0 never entered, 1 dropout, 2 completed
POB           province of birth               0 .. 13, (13 for is abroad)
UNION         time of first union (marriage)  e.g. 1988.234
PARITY        number of children ever born    0, 1, ..
LASTBIR       time of last birth              e.g., 2012.34
```

3.2 Fertility

Model users can choose between two fertility modules. The base one corresponds to a typical macro model; the refined one models fertility separately by parity and includes a more detailed list of variables than just age. When using the refined module, the user can choose to align results to the macro model for producing the same aggregated outcomes.

Like in the DemProj model, the base version is parameterized by projected age distributions of births and total fertility rates. Model users can easily change the scenario of the projected total fertility rates (TFR) without having to change the age profile of fertility. Internally, the model automatically calculates fertility rates by age and period. The base fertility module is used in two alternative ways: as the model to be used to implement fertility, and as the benchmark model. In the latter case, it is used to produce the number of births to which the more detailed refined fertility model can be aligned.

The limitation of the base model is that it ignores important fertility differences, e.g., by number and timing of previous births. Because of this limitation, the model would only produce the right number of children but no realistic female life-courses, even if the future age-specific fertility was known.

The Refined Fertility Module models fertility by age, parity, educational attainment, and union status. First births are parameterized with separate age-specific fertility tables by education and union status. Higher-order births are modeled by proportional hazard regression models estimated separately by birth order. Models contain a baseline risk profile by duration since the previous birth and the relative risks for age group, education, and union status. Additionally, users can create scenarios for future trends.

3.3 Mortality

As for fertility, we provide two versions of the mortality model: a base version that resembles a typical macro model, and a refined version. The refined model focuses on infant mortality by mother's characteristics. When using the refined model, the user can align results to the macro model to produce the same aggregated outcomes in the number of deaths for an initial year.

In the base mortality module mortality is modeled by age and sex. Parameters are a mortality table (for age patterns) and projected period life expectancy. Within the application, the life table is scaled automatically for each year to meet the targeted life expectancy by calendar year and sex. If no national tables are available, standard mortality tables can be used.

An optional child mortality module focuses on mortality of children age 0 through 4. In addition to baseline mortality risks by age and sex, relative risks byage and level of education of mothers are used. When switched on, this module “overwrites” the base mortality model for children up to their fifth birthday. While this will typically alter the overall life expectancy, the user can align aggregate outcomes for an initial year; in this case, life expectancy is the same as in the initial year and future differences can be attributed to the changing composition of age and education of mothers. Additionally, the user can set specific time trends by age for child mortality, different from the overall trend (which is calculated automatically to meet the life expectancy parameter).

3.4. Migration

The model includes modules for internal migration, emigration, and immigration. All modules currently follow typical macro approaches. Internal migration is based on age-specific origin-destination matrices. The user can choose between a base version (by age group and sex) and a refined version, which adds education as a model dimension affecting the probability to leave a province. For easier scenario creation, probabilities to leave a province and distribution of destination provinces by age group and origin are parameterized separately.

Immigration is modeled specifying the total number of future immigrants by sex, their age distribution by sex, and the distribution of the province of destination by age group and sex. For immigrants, only their sex, place of residence, and age is initially known in the simulation. Other characteristics, including parity, education, union status, and time of last birth, are sampled from the foreign-born population of the same known characteristics, or the total resident population if no donors are found.

The emigration module is driven by age-specific emigration rates by province, age and sex. It is assumed that emigration pattern stay constant over time.

3.5 Primary Education

The modeling of education currently focuses entirely on primary education. Education is a key variable for first union formation, fertility, child mortality, and migration. At its core, the model is parameterized by probabilities to graduate from primary school by year of birth, province of birth, and sex, respectively. In addition, the model can introduce inter-generational transmission of education by specifying proportional factors (odds ratios) by mother's education. If this option is chosen, the aggregated educational outcome is automatically calibrated for a chosen year of birth to the overall probabilities. In such scenarios, the future dynamics are entirely driven by the changing educational composition of mothers. These model capabilities allow educational change to be separated into changes stemming from inter-generational dynamics, from inter-provincial migration, and from overall trends.

The layout of parameters was chosen to be as intuitive and generic as possible while allowing alternative ways to derive the parameters. For Mauritania, we used a proportional model that can distinguish general trends from inter-provincial differences, which were found very persistent. This allows for both intuitive and alternative scenarios, e.g., persistent inter-provincial differences versus convergence scenarios.

3.6 First Union Formation

Changes in the age at first union formation is one of the key mechanisms behind fertility changes and many developing societies currently experience a rapid increase in that age, partly resulting from educational expansion. The module for first union formation implements this process in two alternative ways: first by using a parametric Coale & McNeil model, and second by union formation risks by age and education. The Coale & McNeil model uses a very intuitive parameterization with three parameters by year of birth and education: (1) the earliest age of union formation observed, (2) the average age at first union formation, and (3) the proportion of persons ever entering a union. Alternatively, model users can specify age-specific rates of union formation. Such rates can also be derived by simulation from the Coale & McNeil model, which produces corresponding table output of the internally calculated rates. This may be useful for policy scenarios that set minimum marriage ages. First union formation is modeled only for females.

3.7 Data Requirements

Microsimulation is typically associated with high data requirements. This view stems from the predominant use of microsimulation for modeling highly complex systems, e.g., the operations of social insurance systems in the context of social and demographic change. In contrast to such models, which must depict individual life-courses in great detail and include educational choices, employment, earnings, family dynamics, savings, health, and retirement decisions, the data requirements for population projection models are modest. For most countries, the required data are readily available.

Required data for population projections are life tables for modeling mortality, origin-destination matrices for modeling migration, and age-specific fertility rates. Here, microsimulation does not differ from cohort-component models.But the power of microsimulation becomes obvious when adding variables and detail. In our application, these variables are internal migrations, primary education, first union formation, and parity—characteristics that allow for a more detailed modeling of the core demographic events. In addition to age, fertility can now account for education, partnership status, and the number and timing of previous births, resulting in the creation of realistic individual life-courses. Concerning mortality, we add a specialized module for child mortality by mother's characteristics (age and education), and migration now adds education as explanatory variable. As a consequence, the microsimulation model uses more of the characteristics contained in typical census data, such as education, and will usually require additional information from surveys.

In a nutshell, the microsimulation population projection presented in this report consists of 11 components and their related data requirements, which collectively add up to 16 variables. Of these components, five are based on the same data as a typical macro projection model:

- **Mortality** is based on a life table and projected changes of life expectancy. Parameters for mortality are usually available from published data. If no reliable national life table of mortality risks by single year of age is available, a standard life table for the world region can be used. The life table is used for age differences in mortality and automatically calibrated (i.e., re-scaled) to meet the second parameter of the model, life expectancy. Future life expectancy is scenario based.

- **Internal migration** is based on origin-destination matrices by age group and sex. In an optional model extension, education level is added as another dimension. Data are typically available from a population census. For easier parameterization and scenario building, the information is split into two tables: (1) the probability to migrate by age group, sex, and province, and (2) the destination of migrants by province of origin, sex, and age group. In the Mauritanian example, the probabilities to leave a province are modeled using logistic regression assuming the same age patterns in each province. Depending on sample size, these models can be replaced by cross-tabs when running the analysis on the full census. The transition matrices are produced by cross-tabs. In an optional extension, the same model can be parameterized by education group. The education variable is typically available from census data.
- **Immigration** requires projected total numbers and age distributions by province of destination and sex, which can be based on recent numbers, observed in the census, and by applying a time trend.
- **Emigration:** The emigration module orients itself on typical macro population projection models requiring a single parameter table, namely emigration rates by age group, sex, and province. In the Mauritanian case, emigration was modeled from census information collected on household members who left the country in the past 12 months; there were no other data sources available.
- **Fertility** is based on age-specific fertility rates and projected trends. Current rates can be tabulated from the census. The future age distribution and total fertility rate are scenario based.

The remaining six components require additional information. In the Mauritanian case, three of these models can be estimated from census data alone:

- **First union formation** uses a parametric (Coale & McNeil) model fitted from census data. The model is parameterized by the earliest age of union formation, average age, and proportion of females eventually entering a union. The parameters are by year of birth and education. Based on the observed proportion of women who have ever entered a union and the age distribution of those observed union formations, curves are fitted that allow projections into the future.
- At its core, the **primary education** module is based on census data, i.e., the highest level of schooling attended and the highest diploma obtained. An optional extension accounts for mother's education. The required proportional factor was estimated from the MICS 2015 survey.
- The **refined version of internal migration** uses education as an additional dimension.

The remaining components require information from household survey data. In the Mauritanian case, all additional information is available in the MICS 2015 data set:

- In the **refined version of fertility** module, higher-order births were estimated from MICS data, which contain retrospective birth histories that can be used to estimate past time trends.
- The optional **infant and child mortality** module requires retrospective birth histories and information of deaths as collected from mothers in the MICS survey. Besides an age baseline of child mortality, the model uses relative risks by mother's characteristics, namely age and education. Estimating these risk factors requires survey data, like the MICS in the Mauritanian case (in other countries, the Demographic and Health Survey – DHS – would be used). If not available, relative risks can be borrowed from comparable countries or based on research literature.

MODULE	POPULATION	Weight	Date Birth	Sex	Province of Residence	Province of Birth	Primary Education	Time First Union	Number Children born	Births past 12 months	Time last Birth	Province 12 months ago	Mother's education	Birth Child 1	…	Birth Child 15	Sex Child 1	…	Sex Child 15	Death Child 1	…	Death Child 15	Deceased past 12 months	Same as in macro models
		REQUIRED FOR CORE ANALYSIS											OPTIONAL CORE ANALYSIS											
		STARTING POPULATION												CHILD MORTALITY									MORTALITY	
Starting Population	residents	x	x	x	x	x	x	x	x	x	x													
Fertility Model A	female residents	x	x							x														x
Fertility Model B	female residents	x	x		x		x	x	x	x	x			o	o	o								
Internal Migration Model A	residents	x	x	x	x							x												x
Internal Migration Model B	residents	x	x	x	x		x					x												
Immigration	residents	x	x	x	x							x												x
Emigration	residents and recent emigrants	x	x	x	x							x												x
Primary Education	residents	x	x	x		x	x						o											
First Union Formation	female residents	x	x				x	x																
Child Mortality	female residents	x	x		x		x							x	x	x	x	x	x	x	x	x		
Mortality	residents and recent deceased	x	x	x	o																		x	x
Typical Data Sources	Census	x	x	x	x	x	x	x	x	x		x												x
	Surveys	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		
	Adminisrative, standard life tables		x	x	x																		x	x

Figure 3-1: Components, Variables, and Data Sources of the Microsimulation Application (x required; o optional)

3.10 Running the Model

The model's graphical user interface can edit parameter tables, create and save new scenarios, run the model, and view table output. Table results can be exported to Excel individually or collectively to an Excel Workbook. Parameters are organized in tables that can be accessed by clicking on the table name in the navigation area of the application (the left side of the interface). For easier navigation, the list of tables is grouped by topic. Clicking on a table name opens the table on the right side of the interface. Tables can be edited directly by the user. The application offers scenario control by saving all simulation results with all parameters. Users can create new scenarios by editing parameters and saving the scenario under a new name.

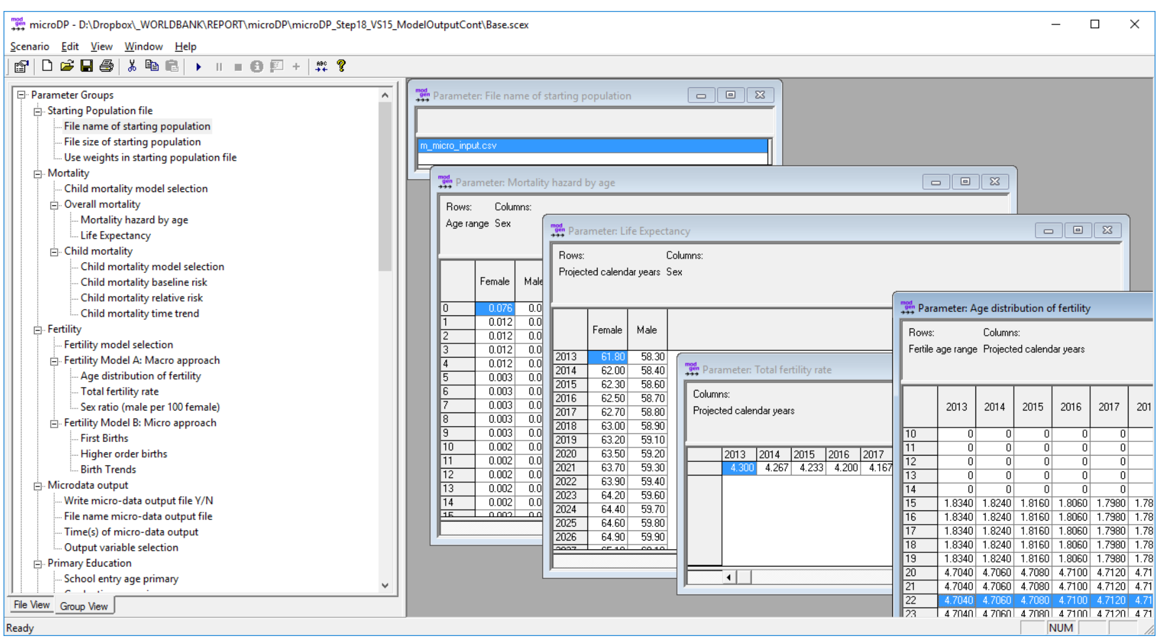

Figure 3.2: The User Interface, Parameter Tables. This screenshot shows the selection of parameter tables essential to the

base version of the model: the file name of the starting population, a mortality table, life expectancy, age-specific fertility, and total fertility rates.

The user interface is fully documented within the application. The menu offers users access to a detailed hyperlinked help option, which covers relevant aspects, such as editing parameters, creating scenarios, options for running the model, and viewing and exporting model results. Like the user interface, the model is also fully documented within the application. Users can access documentation from the help menu, including descriptions of the modules, parameters, model actors, and all table output.

In addition to model parameters, the user also controls some scenario settings. Most importantly, users can choose the time horizon of the simulation and number of replications simulated. When running more than one replicate, all model results are automatically calculated as averages over the replicates and distributional information (e.g., the coefficient of variation) is automatically available for each output table cell. This allows users to assess Monte Carlo variation in results.

The model produces two types of output: a collection of tables and micro-data files for selected moments in time. All tables are updated when running a simulation. Tables can have any number of dimensions. For example, population numbers can be displayed by age, year, sex, and province. The user controls how tables are displayed, e.g., a table by age group and year for selected province and sex, or by age group and province for a selected year and sex. Like parameter tables, output tables can be opened by clicking on the name in the navigation pane list. For easier navigation, output tables are grouped by topic.

4. Application Example: The demographic effects of educational improvements in Mauritania

The following simulation examples stem from the pilot implementation of the DYNAMIS-POP model with the Centre Mauritaniend'Analyse de Politiques (CMAP). Mauritania recently adopted a development strategy - the "Stratégie de Croissance Accélérée et de Prospérité Partagée" (SCAPP) - which includes universal access to primary education by 2030 as one of its policy goals.The DYNAMIS-POP-MRT model was applied (as a pilot research project) to assess the demographic impact of such an educational improvement. Two scenarios were compared, which differ only in the assumptions of enrollment and graduation from primary education. While the Base Scenario assumes a continuation of recent trends in primary school enrollment and graduation, Scenario 2 increments these rates linearly in order to reach the target of universal primary graduation by 2030. The demographic projections of the base scenario (births, deaths) closely resemble official population estimates, the lower fertility in the alternative scenario stems from later union formation as well as lower fertility of higher educated women. The lower mortality in the alternative scenario stems entirely from lower child mortality resulting from lower early teenage pregnancies and higher education of mothers. As in official population estimates, no international migration is modeled. Internal migration is modeled by origin-destination rates by age, sex, and education as observed today. This assumption leads to massive inflow of population to the capital Nouakchott.

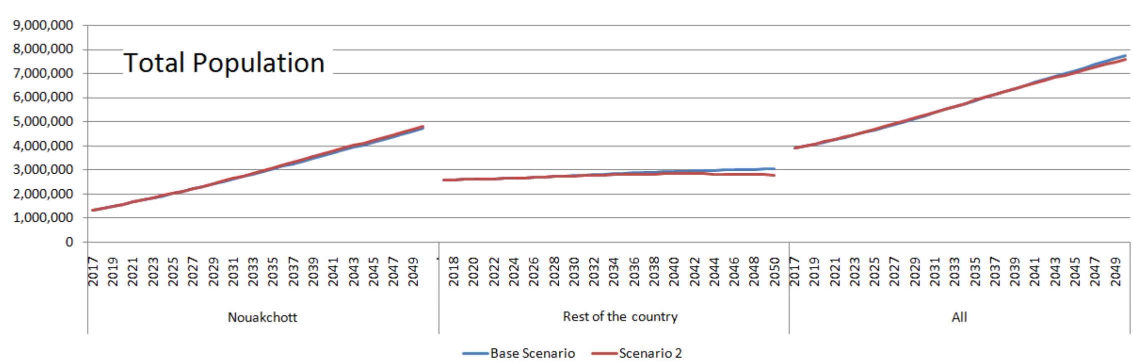

Figure 4.1. Projected total population

Figure 4.1. shows the rather small overall effect of the simulated educational expansion on the population size until 2050. As higher education increases migration to the capital, the effect is only visible in the rest of the country while the population size of Nouakchott is unaffected. When looking at the projected age pyramids (Figure 4.2), the effect of education improvements on migration becomes very visible, as the working age population below 40 is thinned out outside Nouakchott in the high education scenario. Also the number of very young cohorts is noticeable smaller due to the smaller population in reproductive age (and lower fertility of higher educated women)

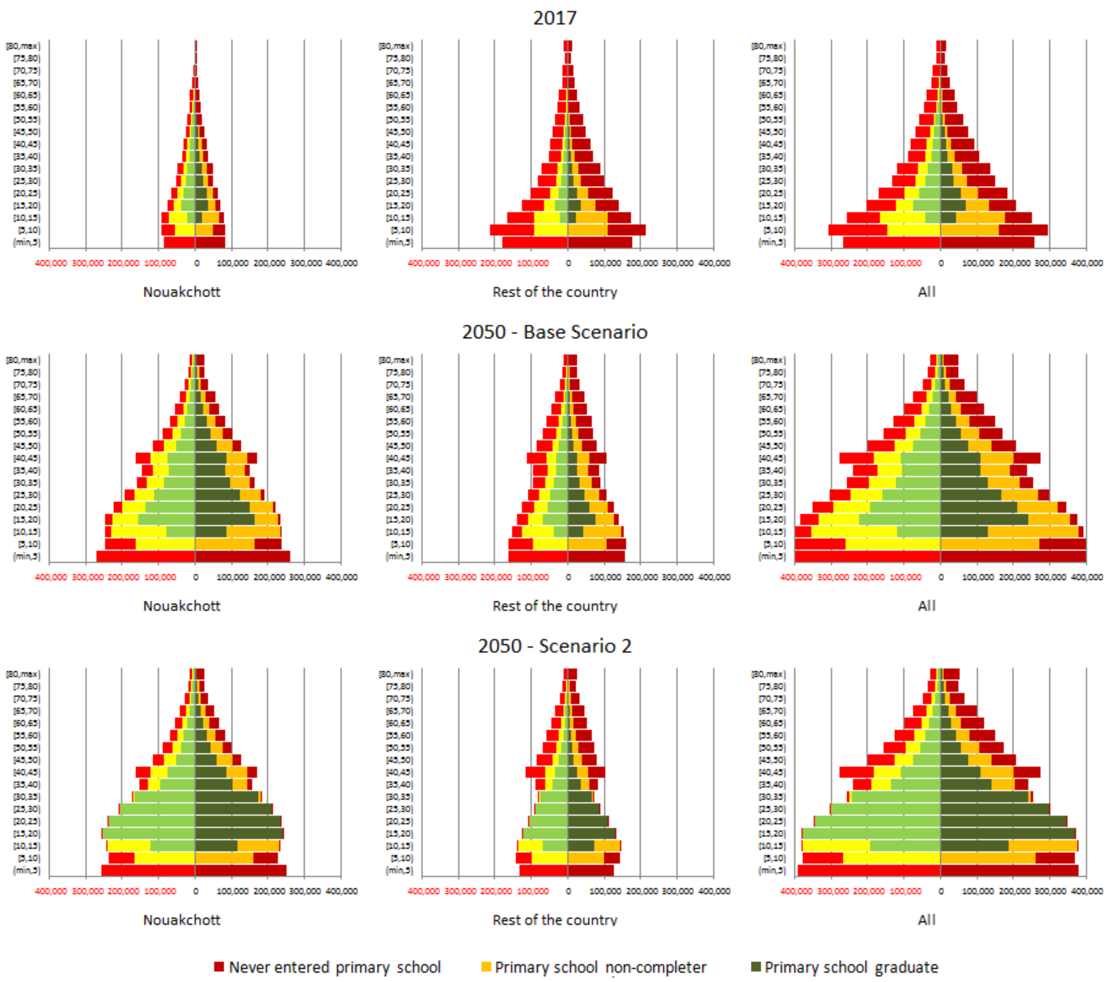

Figure 4.2. Population by age and sex 2017 and 2050

By 2050, the number of births is around 12% lower in the higher education scenario (Figure 4.3). Again, this effect is most noticeable outside the capital. Overall, the higher education leads to falling numbers of births around 2030 which is immediately after the education reform has reached its full effect.

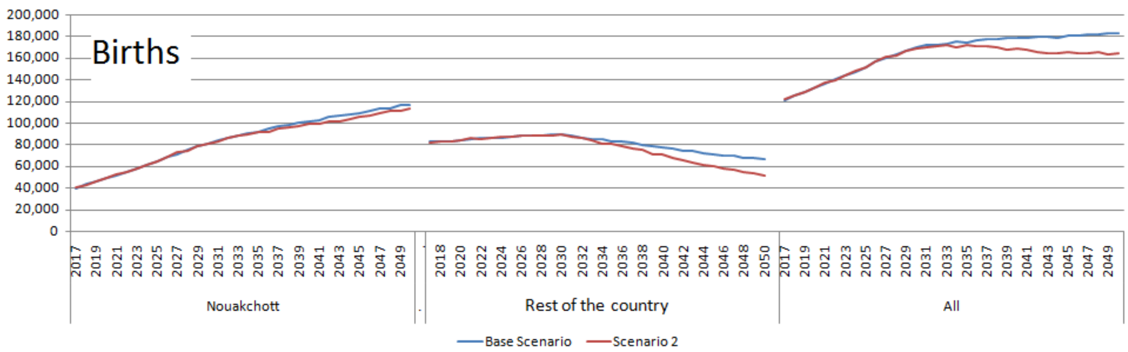

Figure 4.3. Projected births

The lower number of births also translates in a lower number of projected child deaths. Decomposing the effects reveals, that around 20% of the reduction of child deaths is additional to the lower number of births and can be attributed to less early teenage pregnancies and the higher education of mothers.

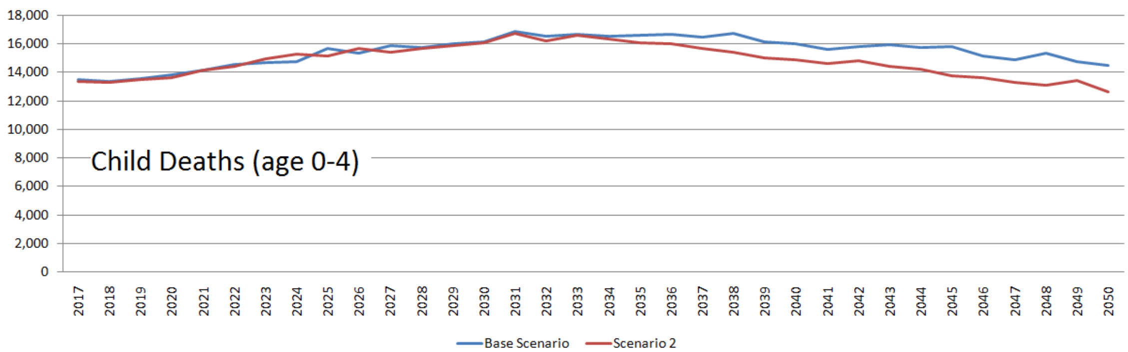

Figure 4.4. Projected child deaths

Figure 4.5. depicts the effect of the education reform on the education composition of the total active population (age 16-59). As older cohorts with lower education are replaced only gradually by the higher educated youth, the process to universal primary education is a slow process taking various decades. Due to the higher mobility of higher educated people, the process is faster in Nouakchott.

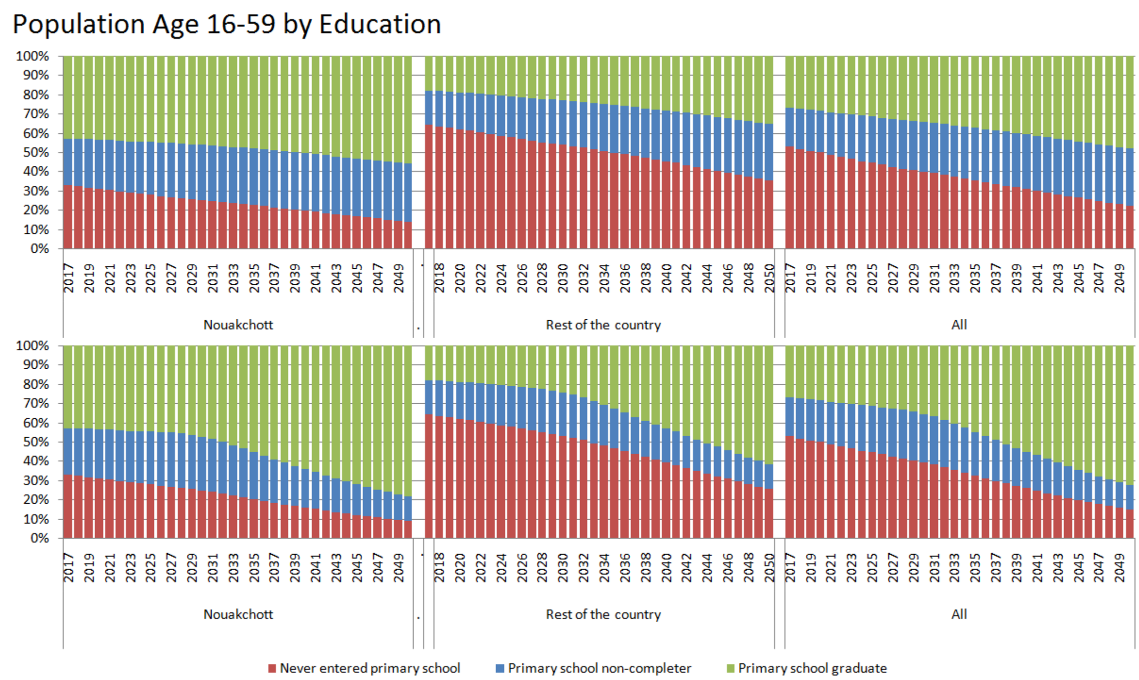

Figure 4.5. Projected population age 16-59 by education. On top: Base Scenario, bottom: Scenario 2

5. Outlook and Discussion

We presented in this report some output of a simple pilot application of dynamic microsimulation to population projections. The main purpose of this exercise was to illustrate the relevance and potential of the approach in the context of a developing country.

The project has confirmed the feasibility of the approach, including in relatively data deprived countries. A population census, and a demographic household survey such as a Demographic and Health Survey (DHS) or Multiple Indicators Cluster Survey (MICS), will provide the necessary input data. DHS and MICS surveys have been implemented in a large number of developing countries, and most datasets are made publicly available. (The MICS program is led by UNICEF -- http://mics.unicef.org/; the DHS program is sponsored by USAID -- http://dhsprogram.com/)

Particular attention was paid to maximize the replicability and adaptability of the model with the expectation to extend this pilot work undertaken in collaboration with CMAP to other countries. Currently the model is ported to Nepal. Also, the model itself, designed in a highly modular manner, will be upgraded and extended. On-going and planned activities include the development of a refined education projection module, integration of an ethnic (or other population grouping) variable for further disaggregation of parameters and results, and a focus on monitoring or informing selected indicators of the Sustainable Development Goals (SDGs). Already available are model extensions for tracking school attendance by grade including scenario support for educational goals and resource planning (e.g. number of required classrooms and teachers on a sub-national level). Other available extensions concern the introduction of additional geographical aggregation levels.

As each step of model refinements and extensions moves the microsimulation projections beyond the reign of comparable and available macro population projections, added attention has to be given to model validation. This is recognized in the ongoing work for Nepal; e.g. available data for both 2001 and 2011 allow for retrospective simulations (starting 2001).

In order to further improve accessibility of all modeling resources we also ported all statistical analysis and parameter generation from Stata to the open source software R. As meanwhile the openM++ open source implementation of the Modgen language became available (https://ompp.sourceforge.io/), an openM++ version of the application is planned. OpenM++ is platform independent (e.g. allowing to develop and run models in Linux) and supports running applications directly within R allowing for integration of the microsimulation model with data analysis and the post-processing of simulation results.

References

Caron-Malenfant, E.; Lebel, A. & Martel, L. (2010), **Projections of the Diversity of the Canadian Population 2006 to 2031**, Statistics Canada Catalogue no. 91-551-X. http://www.statcan.gc.ca/pub/91-551-x/91-551-x2010001-eng.pdf

Lutz, W., J.W. Vaupel, and D.A. Ahlburg, Eds. (1999) **Frontiers of Population Forecasting**. A Supplement to Vol. 24, 1998, Population and Development Review. New York: The Population Council.

Marois G, Sabourin P, & Belanger A (2017). A Microsimulation Approach for Modelling the Future Human Capital of EU28 Member Countries. IIASA Working Paper. IIASA, Laxenburg, Austria: WP-17-017

Martel, L., É. Caron Malenfant, J.-D. Morency, A. Lebel, A. Bélanger, N. Bastien (2011) **Projected trends to 2031 for the Canadian labour force**, Statistics Canada Catalogue no. 11-010-X, vol. 24, no. 8 http://www.statcan.gc.ca/pub/11-010-x/2011008/part-partie3-eng.htm

Morency, J-D, É. Caron-Malenfant, S. Coulombe, S. Langlois (2015) **Projections of the Aboriginal Population and Households in Canada, 2011 to 2036** - Statistics Canada Catalogue no. 91-552-X http://www.statcan.gc.ca/pub/91-552-x/91-552-x2015001-eng.pdf

Spielauer, M. (2014), **The relation between education and labour force participation of Aboriginal peoples: A simulation analysis using the Demosim population projection model**, Canadian Studies in Population 41(1-2), 144--164.https://ejournals.library.ualberta.ca/index.php/csp/article/viewFile/22095/16427

Stover, J., S. Kirmeyer (2001), **DemProj Version 4 - A Computer Program for Making Population Projections**, The POLICY Project Spectrum http://data.unaids.org/topics/epidemiology/manuals/demproj_manual_v_4_en.pdf

Van Imhoff, E. & W. Post (1998), **Microsimulation methods for population projection**, Population 10(1), 97--136. http://www.persee.fr/docAsPDF/pop_0032-4663_1998_hos_10_1_6824.pdf

Spielauer, M. (2011), **What is social science microsimulation?**, Social Science Computer Review 29(1), 9--20. http://ssc.sagepub.com/content/29/1/9.abstract

Willekens, F. (2001), **Theoretical and Technical Orientations Toward Longitudinal Research in the Social Sciences**, Canadian Studies in Population 28(2), 189-217 http://www.canpopsoc.ca/CanPopSoc/assets/File/publications/journal/CSPv28n2p189.pdf

Appendix: Online documentations and resources to run or reproduce the model

Documentation

DYNAMIS-POP is documented in detail. Also its code and statistical analysis files are available to model builders, who can use the given step-by-step instructions as a textbook and toolbox for microsimulation model development and implementation. All resources are available from the project website at http://ihsn.org/projects/dynamis-pop

The full technical documentation is available at:

- http://documentation.ihsn.org/dynamis/ [HTML]
- http://ihsn.org/sites/default/files/resources/dynamis.pdf [PDF]

Software for Model Users

Like all model resources, DYNAMIS-POP-MRT can be downloaded from the project website: http://ihsn.org/projects/dynamis-pop

DYNAMIS-POP-MRT is a Modgen application. To run Modgen applications on a PC, the Modgen Prerequisites have to be installed. This program is free and can be downloaded from the Statistics Canada Website at:http://www.statcan.gc.ca/eng/microsimulation/modgen/modgen. The Modgen Prerequisites version used for the population projection model is Version 12.1.

Software for Model Developers

The developer's version of DYNAMIS-POP-MRT as well as a complete suite of all model steps 1-19 are available from the project website: http://ihsn.org/projects/dynamis-pop

The model was implemented using **Modgen**, a generic microsimulation programming language developed and maintained at Statistics Canada. Modgen is free and can be downloaded from: http://www.statcan.gc.ca/eng/microsimulation/modgen/modgen. This site also provides training and information resources. The Version used is Modgen Version 12.1.0.13

Modgen 12.1. requires **Microsoft Visual Studio 17**. The free community version can be downloaded from the Microsoft site: https://www.visualstudio.com/downloads/. This software must be installed before Modgen is installed.

Data Analysis

All data analysis for deriving the model parameters was performed in Stata. The full set of Stata do-files is available here for download. Most Stata output was copy-pasted into Excel sheets for deriving and formatting the final parameters as used in the parameter tables of the model. These excel-sheets are available for download as well.