New set of population projections by age, sex, and educational attainment for 171 countries of the world: methods and challenges

Samir K.C. and Erich Striessnig

Introduction

The reconstruction and projection of the population by age, sex, and educational attainment for 120 countries of the world were first published in 2007 and 2010, respectively (Lutz et al. 2007; K.C. et al. 2010). The data have been used by researchers and planners, for example, to analyze the age-dimension of the relationship between human capital and economic growth (Lutz, Cuaresma, and Sanderson 2008; Chappuis and Walmsley 2011; Eberstadt 2012), to study the impacts of natural disasters (Cavallo et al. 2010) as well as the determinants of vulnerability to natural disasters (Striessnig, Lutz, and Patt 2013), demographic and health related issues (KC and Lentzner 2010; Prettner, Bloom, and Strulik 2012), predicting armed conflict (Herge et al. 2011, unpublished paper). Detail discussion about including education as an important variable in addition to age and sex for measuring demographic heterogeneity were published recently (Lutz and KC 2010; Lutz and KC, 2011).

In the past, the basic assumptions regarding the future of fertility, mortality, and migration were mostly taken over from the World Population Prospects (UNPD, 2008). The education projections by four levels, namely never been to school, some primary, secondary, and tertiary, were derived from fitting a single global trend (by age and sex). Education differentials in fertility and mortality were mostly based on the literature as well as own analysis of a few censuses (IPUMS) and surveys. These differentials were assumed to be constant for the entire period of projection without making use of further available expertise. Population distribution by age, sex, and educational attainment were compiled from various sources. The initial distribution along with the assumptions, the distribution of population by age, sex, and education was projected up to the 2050 for 120 countries of the world.

The aim of this contribution is to present for the first time a new set of basic assumptions regarding the future of fertility, mortality, and migration currently under development at the newly established Wittgenstein Centre for Global Human Capital and Demography (WiC) for a total of 195 countries of the World. In contrary to previous assumption-gathering exercises, the new set of assumptions is derived from argument-based expert opinions. These WiC assumptions will be the basic demographic inputs for the new population projections by age, sex, and education that we will present in this paper. The number of education categories is increased from four to six to allow for more detailed disaggregation and the number of countries is expanded from 120 to 171 countries, for which the data were available, covering more than 97% of the World's population. Initial distribution of population by age, sex, and education are prepared using most recent censuses (IPUMs) or surveys and the education variables were recoded to match the ISCED definition. Education differentials in fertility, mortality, and migration are mostly based on the estimation using census (IPUMS) and survey data as well as available literature. Various methods of dealing with the differentials are fine-tuned and some additional complexities have been introduced (e.g. allowing child mortality to depend on the education of the mother). Finally, the education projection is improved by allowing, in addition to the global, country as well regional trends to influence future attainment levels.

Our aim, therefore, is to present the methodological advancement and challenges in introducing education as an additional dimension in the evolution of population distributions. The ultimate goal of this paper is to serve as a step-by-step manual to conduct population projections by age, sex, and educational attainment which will be a tool highly regarded by users who want to do their own multi-state projections. At the same time, this paper will eventually serve as a background paper for a population projection software package in R in the coming future.

Data and Methods

Over the last two years, six teams of researchers at the WiC have been working to prepare a new set of assumptions regarding the future evolution of fertility, mortality, and migration in 195 countries with a population of more than 100,000. The assumptions do primarily rely on the argument-based opinions of experts from all around the world collected through a web survey (Bilal et al.) as well as from six smaller groups of top experts who validated the experts' opinions. These opinions, together with the past country-, region- and global-experience, were then used by the different teams of researchers at the WiC to come up with a final set of assumptions for TFR, sex-specific life expectancies, and migration flows for 195 countries of the world between 2010-2100. This is equivalent to the UN's Medium scenario in terms of "middle of the road" or "business as usual" scenario.

Education differentials in fertility, mortality, and migration are estimated from various censuses (IPUMS) and surveys (e.g. DHS). In many cases the differentials are also derived and compiled from the existing literature. Country-specific assumptions for the future differentials were prepared based on the analysis of the differentials observed in the past, e.g., for countries with a TFR above 1.8, current fertility differentials are assumed to converge to a standard set of differentials by the time TFR reaches below 1.8 or by 2030, whichever occurs later.

One of the most time consuming tasks was the preparation of the distribution of the initial population by age, sex, and education. A team of three researchers prepared a final set of distributions by scanning all possible data sources. In case there were multiple sources of data available, the decision on which was the best was based on the careful analysis of the data quality. Country-specific education categories were harmonized according to the UNESCO's International Standard Classification of Education (ISCED-97) and a final set of data was prepared for 171 countries with six levels of educational attainment namely: no education (E1), incomplete primary (E2), completed primary (E3), completed lower secondary (E4), completed upper secondary (E5), and completed post-secondary education (E6).

Given this data, the task is now to prepare a model that can project the population by age, sex, and educational attainment. What follows is a brief summary of challenges that are worth mentioning (main points).

Projecting population by age and sex

As a first step, population is projected by age and sex (and without the additional dimension of education) using the rates of fertility and life expectancies as implied by the experts' judgments. The age-specific fertility rates and life tables corresponding to these judgments are derived using the age profiles from the UN Medium scenario.

Projecting population by age, sex, and educational attainment

As a second step, the level of educational attainment is added to the base year population distribution by age and sex. The projections are done at a five yearly steps starting in 2000, 2005 or 2010, depending on the time of the census or survey from which the initial education distribution by age and sex is derived. Of course the year of the census or the survey is not necessarily a year ending with 0 or 5. Therefore, we have to first project the age-sex-specific population distribution forward (year ending with 3 or 4) or backward to get our base-year distribution.

Education-specific life tables

At each five-yearly step, the population distribution by age, sex, and educational attainment for the adult population (aged 15 and over) is projected forward using education-specific life tables. These are derived from the "overall" life table and the known education-differentials in life expectancy at age 15 as provided by the literature. Gompertz' relational model is used to determine the age-schedule for each education group by minimizing the difference between the total number of age-sex-specific deaths and the sum of deaths by age and sex from the individual education categories. It is assumed that the survival ratio during a five-yearly period depends on the education status at the beginning of the interval.

Mortality of children under 15

As an additional sophistication, we derive the mortality of children aged under15 by their mother's education. The education-specific differentials are taken from a review of the literature and also from own estimates using surveys. Once again, these are used to minimize the difference between overall age-specific deaths under 15 and the aggregation of age-sex-specific deaths by education category.

Education scenarios

Once the population is projected to the end of a five yearly period applying the previously derived education-specific survival ratios, education transitions are implemented to match a pre-defined set of education profiles corresponding to the medium population scenario for the 15-34 age-group. It is assumed that there are no education transitions between the six education states after age 34.

Education-specific ASFRs

Using the overall ASFRs for the total population and the education-specific relative ratios of TFR, education-specific ASFRs are derived by applying Brass' relational logit model. Similar to the education-specific life tables, the difference between the overall age-specific births and the sum of age specific births from all education categories is minimized.

Education-specific migration flows

Finally, we implement education-specific rates of migration. While applying given set of rates of inflows and outflows, at the global level often the inflows and outflows by education within a age-sex categories is not zero due to the complexity in applying the full scale multistate model at the global level., adjustments are made to ensure that globally net migration by age, sex, and educational attainment is zero.

Different scenarios

While we are currently in the process of finalizing the "medium" projection scenario, additional "high" and "low" projection scenarios will be included in the final paper together with detailed scenario definitions.

Results

Our preliminary results show that the Global population will peak at a level of 9.45 billion around the year 2070 and will then start to decline. This result is different from the UN's projected medium scenario, where population will keep on rising till the end of the century beyond 10 billion. The difference in the projected number at the global level is mainly due to different sets of assumptions on fertility.

In terms of educational attainment, according to the education trend scenario, the population in 2100 will be much more educated than in 2010. Figure 1 shows the total population for 171 countries by six levels of education for population aged 15 and above along with the total population for children under 15 years of age. Population will increase by about 2 billion in the next 50 years. However, the proportion of adults with at least high school completed will increase significantly.

In Figure 2 we show the composition of the population by age, sex, and educational attainment for two very different demographic histories: the one of Kenya and the one of Japan. While in Kenya, experts expect the population to more than double over the next five decades, Japan is going to lose one fourth of its population compared to today, such that by 2060 the two countries will have reached at nearly the same population size. Yet, in talking about what these two countries have in common, demographically speaking, we have thus already come to an end. While Kenya, on the one hand, will still have a rather young population with its majority concentrated at prime working ages, we see Japan's population represented almost by a pyramid turned upside down. This picture reflects tremendous aging and an associated increase in the conventional old-age dependency ratio that has often been taken to imply worsening socioeconomic prospects, if not economic collapse, for the future. But one might also point at the stunning increases in education and assume that in a likely future of high tech industries, fewer and much better educated young people is just what the Japanese labour market needs. In more scientific terms, the question refers to the trade-offs between age-structure and education-structure with regard to their contributions to current and future wellbeing. This highlights yet another important aspect behind the necessity of population projections by age, sex, and education.

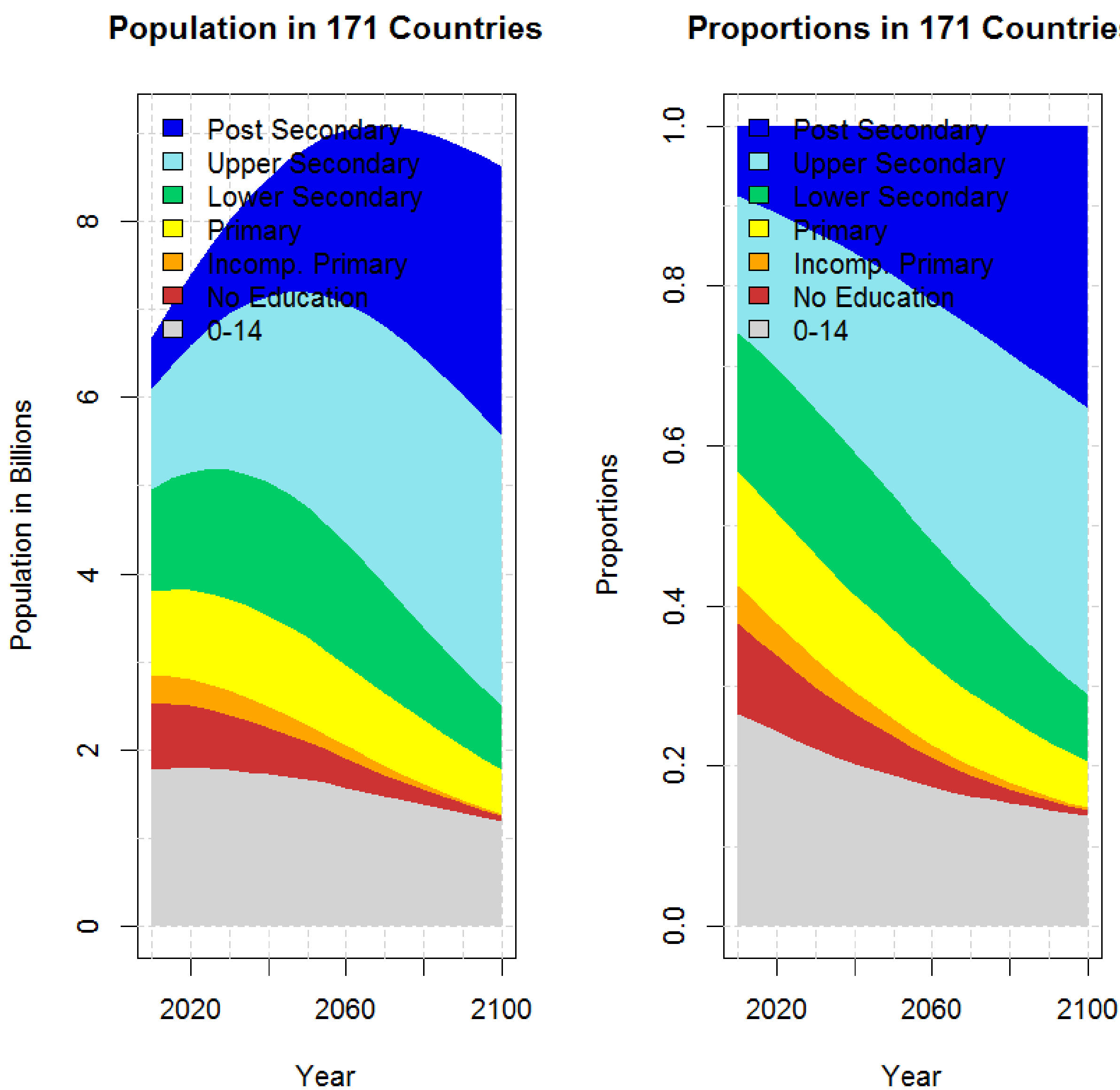

Figure1. Global Population Projection by Educational Attainment in 171 Countries (more than 97% of the World Population) under Medium Scenario

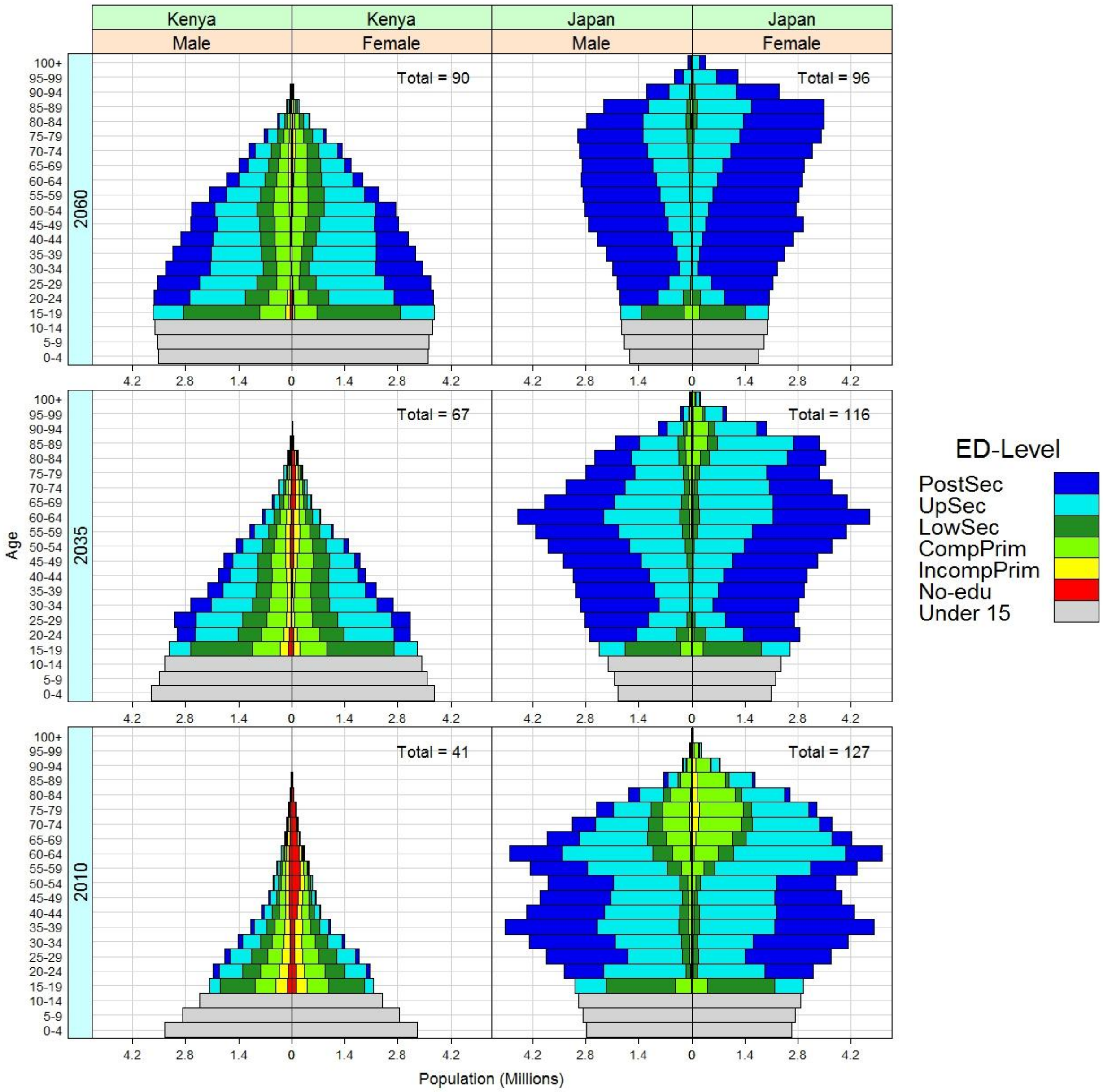

Figure 2. Population by Age, Sex, and Education, Kenya and Japan, 2010, 2035, 2060. Source: Own calculations. (*Note: "Total = xx" on the right hand side of each pyramid represents total population in the country in millions*)

Discussion and Conclusion

This paper will present a step-by-step methodology developed in producing a set of population projection by age, sex, and educational attainment for almost all countries of the world. The details regarding the assumptions and the results are not presented and discussed in this extended abstract at this stage because they are in the process of finalization. This is an ongoing work that started two years ago involving more than 15 researchers and will be completed by the end of the year 2012. We will present all the details in the final paper.

Usefulness of the data and methodology

Since the Vienna University of Economics and Business, one of the pillar institutions of the WiC, also hosts an eminent platform for the development of “The Comprehensive R Archive Network” (CRAN), the “Wic-Model” is entirely programmed in R (R Development Core Team 2008). The plan behind this is to later use CRAN to make the model publicly available in the form of a population projection software package. Eventually, it should be possible for anyone in the interested community to do multi-state population projections for a range of different scenario options at the cost of a few mouse clicks.

References

Cavallo, Eduardo A., Sebastian Galiani, Ilan Noy, and Juan Pantano. 2010. “Catastrophic Natural Disasters and Economic Growth.” SSRN eLibrary (April 28). http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1597507.

Chappuis, Thomas and Terrie L. Walmsley. 2011. “Projections for World CGE Model Baselines”. GTAP Research Memorandum No. 22. https://www.gtap.agecon.purdue.edu/resources/download/5625.pdf

Eberstadt, Nicholas N. 2012. “Looking towards 2030: A new world coming into focus”. Economic Affairs. 32 (1): 17-25.

Hegre, Havard, Joakim Karlsen, Havard Mokleiv Nygard, Havard Strand, and Henrik Urdal. 2011. “Predicting Armed Conflict, 2010-2050”. http://folk.uio.no/hahegre/Papers/PredictionISQ_Final.pdf

K.C., Samir, Bilal Barakat, Anne Goujon, Vegard Skirbekk, and Wolfgang Lutz. 2010. “Projection of Populations by Level of Educational Attainment, Age, and Sex for 120 Countries for 2005-2050.” Demographic Research 22 (March): 383–472. doi:10.4054/DemRes.2010.22.15.

KC, Samir, and H. Lentzner. 2010. “The Effect of Education on Adult Mortality and Disability: A Global Perspective.” Vienna Yearbook of Population Research 2010 8: 201–235. doi:10.1553/populationyearbook2010s201.

Lutz, Wolfgang, Jesus Crespo Cuaresma, and Warren Sanderson. 2008. “The Demography of Educational Attainment and Economic Growth.” Science 319 (5866): 1047–1048.

Lutz, Wolfgang, Anne Goujon, Samir K.C., and Warren Sanderson. 2007. “Reconstruction of Population by Age, Sex and Level of Educational Attainment of 120 Countries for 1970-2000.” Vienna Yearbook of Population Research 2007: 193–235.

Lutz, Wolfgang and Samir K.C. “Global Human Capital: Integrating Education and Population”. 2011. Science: 333 (6042), 587-592.

Lutz, Wolfgang and Samir K.C. “Dimensions of global population projections: what do we know about future population trends and structures?” 2010. Phil. Trans. R. Soc. B. 365 (1554) 2779-2791.

Prettner, Klaus, David E. Bloom, and Holger Strulik. 2012. Declining Fertility and Economic Well-being: Do Education and Health Ride to the Rescue? PGDA Working Paper. Program on the Global Demography of Aging. http://ideas.repec.org/p/gdm/wpaper/8412.html.

R Development Core Team. 2008. “R: A Language and Environment for Statistical Computing.” R Foundation Statistical Computing.

Striessnig, Erich, Wolfgang Lutz, and Tony Patt. 2013. “Effects of Educational Attainment on Climate Risk Vulnerability.” Ecology and Society. 18(1): 16.