National Workshop

on

Subnational Population Projections using Census Data

Beijing, China

17-18 January 2013

Table of contents

Session I: Projecting subnational population: Introduction

I.1.   Objective

The Session introduces the participants to main approaches to subnational population projections. Participants are familiarized with the advantages and disadvantages of various approaches, and with guidelines for choosing an appropriate method.

I.2.   Main approaches to subnational projections: Top-down or bottom-up

Subnational population projections are increasingly in demand. They are an important information source for regional and local policy makers. Subnational population projections are also important for countries that are large and have a diverse population. In such circumstances, national population projections may not effectively and accurately reflect the demographic settings in its administrative or regional entities.

There are two basic approaches to create sub-national projections:

- Top-down approach: a projection for the entire country is performed, and the results are then disaggregated (downscaling).
- Bottom-up approach: projections for all sub-national entities are created separately and then aggregated to obtain the results for the entire country. A variation to this approach – multi-state or multi-regional projection model – is much more demanding, as it performs simultaneously the projections for all sub-national populations.

Each of these two basic approaches has advantages and disadvantages. Choosing the right approach depends on data availability, time horizon, expected users, etc.

In addition to the above two main approaches, combinations of different approaches/methods have been used. India has produced population projections for the country and its states using the cohort component projections for all large states (with more than 10 million Inhabitants). For smaller states or locale, a variant of the Growth Differential Methods has been used, and for very small entities, extrapolations with appropriate mathematical formula were applied.

I.3.   Elements of subnational projections: Demographic components revisited

The preparation of subnational projections in general does not require other input information than the projections of national populations. However, one has to be aware of the following challenges.

Because subnational populations are smaller than the national one, data may show more irregularities and fluctuations. This affects all elements of demographic components: births or fertility, deaths or mortality and migration. If a population is small, its age composition is also subject to random fluctuations.

International migration for national populations is usually integrated into the projection as net migration, which is the balance of immigration and emigration. If migration is of relatively small magnitude, international migration has often been ignored (e.g. assumed to be zero). The same treatment is almost always not possible for subnational projections, since the magnitude of internal migration is in most cases larger than international migration, and therefore has significant demographic impact that cannot be ignored.

In sum, while national and subnational projection work with the same types of demographic data, subnational projections need significantly more data (when prepared for all subnational entities together), and the data may be subject to random fluctuations or limited availability.

I.4.   Projecting urban and rural populations

Projections for urban and rural populations are an important part of national projection exercises as they illustrate the demographic dynamics of two main socioeconomic population groups. In principal, all projections methods are applicable, for instance the cohort-component projection method or appropriate ratio methods. The United Nations has developed and used for many years a robust and parsimonious method that uses the Urban-Rural Growth Differential (URGD) as driving force (United Nations, 1974). The URGD method follows a logistic growth pattern and is therefore adequate for modeling urbanization as a diffusion process. Other countries also have used this method (see Registrar General, 2006).

The United Nations methods are documented in United Nations (1974), United Nations (1980) and United Nations (2012)

I.5.   Resources

- United Nations (1974). Manual VIII - Methods for Projections of Urban and Rural Population
- United Nations (1980). Patterns of urban and rural population growth.
- United Nations (2012). World Urbanization Prospects, The 2011 Revision.
- Registrar General (2006). Population Projections for India and States 2001-2026.
- George, M. V., Smith, S. K., Swanson, D. A., & Tayman, J. (2004). Population Projections.

Session II: The migration component of sub-national projections

II.1. Objective

The Session presented particular methods that can be used in obtaining subnational migration estimates. Data from consecutive censuses are used, in particular from the questions on residence five years ago.

II.2. Obtaining migration data from a national census

Subnational projections are not very different from projections for whole countries. They have, however, additional requirements to obtain estimates and prepare assumptions for the migration component. Sub-national population projections need to incorporate, in addition to the international migration element, migration within the country between the sub-national entities (provinces, regions, urban and rural areas, etc.).

One crucial source of information about subnational migration flows is population census. The Session covers five parts: (a) the cohort-component method to be used; (b) obtaining internal migration data from population censuses; (c) methods to be used to obtain inter-provincial migration data; (d) methods for obtaining urban-rural migration data; and (e) comparing results by using different estimation methods.

Emphasis of the Session is to use the place of usual residence 5 years ago information obtained from the 2000 and 2010 censuses to obtain inter-provincial and urban/rural migration data by age and sex.

II.3. Resources

Session III: Preparing subnational projections using a top-down approach

III.1. Objective

In Session III, participants will be introduced to a robust and simple method of generating subnational population projections that is appropriate when detailed data are lacking or for a 'back-of-the-envelope' calculation.

III.2. The ratio method and its variations

Ratio methods express the population or population growth of a smaller area as a proportion of the population or the population growth of a larger area that is the parent of the smaller area.

The ratio method has several variations. All of the different implementation is common in that the area containing the population to be projected is part of a larger area (parent area) for which a projection already exists. Ratio methods are often used where areas exist in a hierarchical structure.

The ratio method projects the population of a particular subnational entity/area by making assumptions about the relation between the subnational entity and the larger area it belongs to.

- **Constant-Share Method**
  The constant share method holds the smaller area's share of the parent population constant at a certain point in time, usually at the base year. Data requirement are light: population figures for the smaller areas are required for one point in time only, and for the parent area a population projection. Like all variations of the ratio method, the constant-share method is usually applied to total populations, sometimes disaggregated by sex.
- **Shift-Share Method**
  The shift-share method allows for changes in population shares of the smaller areas over time. The trends of changing shares may be determined by inspecting past trends or by formulating appropriate assumptions. The shift-share method does not guarantee plausible results, so care must be taken and constraints introduced to avoid extremely large subpopulations or even negative population figures.
- **Share-of-Growth Method**
  Instead of population shares as in two preceding methods, the share-of-growth method is based upon population growth rates for the subpopulations. The method assumes that the observed population growth of the smaller areas will be constant over the projection period. The share-of-growth method experiences problems if there are growth rates with opposite signs for the subpopulations.

III.3. Advantages and shortcomings

The ratio method, and especially the constant-share variant, has the advantage of needing only minimal information for the task of projecting subnational populations.

The ratio methods are shortcut methods, and therefore cannot yield all the information that more complex methods, like the cohort-component methods, can generate. They also have some specific shortcomings. Problems arise especially in cases where some subnational entities have negative growth, while the other entities continue to grow, or when the size of the subnational entities is very diverse, ranging from very small to large ones.

III.4. Resources

- Smith, S. K., Tayman, J.,Swanson, D. A. (2001, pp. 176-180).
- Rowland (2003), pp.456-458.
- George et al. (2004), pp. 570-571.

Session IV. Preparing subnational projections using a bottom-up approach

IV.1. Objective

The Session discusses the approaches for the preparation of subnational population projection using a bottom-up approach. Participants are introduced to two variations of the cohort-component method that are suitable for subnational projections. Advantages and data requirements for each method are presented.

IV.2. National projections with subnational components: An integrated approach

A national projection with subnational components is usually performed by a central authority of the country, often in collaboration with regional or subnational institutions. One approach could be the simultaneous preparation of all subnational projections using the standard cohort-component method and software. In a second step, the national projection is then obtained by aggregating all subnational projections. For the latter, the projection software needs to have special functionality.

Another, more comprehensive and elegant, approach is to employ a multi-state or multi-regional projection framework. Here the subnational populations and their demographic characteristics enter the model together. The model has special – and demanding – requirements on internal migration. For each subnational area, migration needs to be fully specified in terms of immigration by sending subnational area and for emigration by destination area. Altogether, this forms a gross migration matrix with as many rows and columns as there are subnational regions. If international migration needs to be included, an additional virtual region could be added representing the outside world.

Both approaches produce projections that are consistent and transparent, that is the country's total population is identical to the sum of all sub-populations. In addition, the two approaches generate a rich set of demographic indicators.

IV.3. Stand-alone subnational projections using the cohort-component method

There is a need for standalone population projections that benefit for the advantages of the cohort-component method while without the heavy administrative and organizational overhead of coordinated national population projections with subnational components. Such need arises from emerging and ad-hoc policy and planning requirements that need population projection data as its basis.

IV.4. Resources

- Smith, S. K., Tayman, J.,Swanson, D. A. (2001, pp. 176-180).
- Rowland (2003, pp.456-458).
- George et al. (2004, pp. 570-571).

Resources

Bibliography

Arriaga (1994a). Population Analysis with Microcomputers, Volume I: Presentation of Techniques, Bureau of the Census.

Arriaga (1994b). Population Analysis with Microcomputers, Volume II: Software and Documentation, Bureau of the Census.
URL: http://www.census.gov/population/international/software/pas/pasdocs.html

Bryan, T. (2004). Population Estimates. In J. S. Siegel and D. A. Swanson (Eds.), *The methods and materials of demography* (2nd ed., pp. 523–560). Elsevier Academic Press.

George, M. V., Smith, S. K., Swanson, D. A., and Tayman, J. (2004). Population Projections. In J. S. Siegel and D. A. Swanson (Eds.), *The methods and materials of demography* (2nd ed., pp. 561–602). Elsevier Academic Press.

Hobbs, F.B. (2004). Age and Sex Composition. In J. S. Siegel and D. A. Swanson (Eds.), *The methods and materials of demography* (2nd ed., pp. 125–173). Elsevier Academic Press.

ILO (2011). ILO Estimates and Projections of the Economically Active Population 1990-2020 (Sixth Edition). Geneva

O'Neill, B. C., Balk, D., Brickman, M., and Ezra, M. (2001). A Guide to Global Population Projections. *Demographic Research*, *4*, 203–288. doi:10.4054/DemRes.2001.4.8

Popoff, C., and Judson, D. H. (2004). Some Methods of Estimation for Statistically Underdeveloped Areas. In J. S. Siegel and D. A. Swanson (Eds.), *The methods and materials of demography* (2nd ed. Ppp. 603–641). Elsevier Academic Press.

Preston, S. H., Heuveline, P., and Guillot, M. (2001a). Population Projection. In *Demography: Measuring and Modeling Population Processes.* (pp. 117–137). Oxford and Malden, MA: Blackwell Publishers.

Preston, S. H., Heuveline, P., and Guillot, M. (2001b). Modelling Age Patterns of Vital Events. In *Demography: Measuring and Modeling Population Processes.* (pp. 191–210). Oxford and Malden, MA: Blackwell Publishers.

Registrar General (2006). *Population Projections for India and States 2001-2026*.

Rowland, D.T. (2003). *Demographic Methods and Concepts.* Oxford University Press, New York

Smith, S. K., Tayman, J.,Swanson, D. A. (2001). *State and local population projections. Methodology and Analysis*. Kluwer Academic/Plenum Publishers, New York.

United Nations (1952). Manuals on Methods of Estimating Population. *Manual I: Methods of Estimating Total Population for Current Dates*. Sales No. E.52.XIII.5.

United Nations (1956a). Manuals on Methods of Estimating Population. *Manual II: Methods of Appraisal of Quality of Basic Data for Population Estimates*. Sales No. E.56. XIII.2.

United Nations (1956b). Manuals on Methods of Estimating Population. *Manual III: Methods for Population Projections by Sex and Age*. Sales No. E.56.XIII.3.

United Nations (1980). *Patterns of urban and rural population growth*. POPULATION STUDIES, No. 68, Sales No. E.79.XIII.9

United Nations (1983). *Manuals X. Indirect Techniques for Demographic Estimation*. Sales No. E.83.XIII.2.

United Nations (1989a). *Projection methods for integrating population variables into development planning*. Module One: Conceptual issues and methods for preparing demographic projections. ST/ESA/SER.R/90.

United Nations (2008). *Principles and Recommendations for Population and Housing Censuses. Revision 2*, Statistical Papers, No. 67/Rev. 2, Sales No. E.07.XVII.8.

United Nations (2010). *World Population Prospects: The 2006 Revision*, Vol. III, Analytical Report, New York
   URL: http://www.un.org/esa/population/publications/WPP2006RevVol_III/WPP2006RevVol_III_final.pdf

United Nations (2011). World Population Prospects: The 2010 Revision, Highlights and Advance Tables.
   URL: http://esa.un.org/unpd/wpp/Documentation/pdf/WPP2010_Highlights.pdf

United Nations (2012). *World Urbanization Prospects, The 2011 Revision. Methodology*. New York.
   URL: http://www.census.gov/population/international/software/pas/pasdocs.html

United Nations Economic Social Commission for Asia and the Pacific (2011). 2011 ESCAP Population Data Sheet. Bangkok.
   URL: http://www.unescap.org/publications/detail.asp?id=1465Session

Willekens F, Rogers A, 1978, Spatial population analysis: methods and computer programs, RR· 78-18, International Institute for Applied Systems Analysis, Laxenburg, Austria.

World Population Prospects: The 2010 Revision (Online)
URL: http://esa.un.org/unpd/wpp/index.htm

Software

Spectrum v. 4.51
Spectrum can be downloaded from: http://www.futuresinstitute.org/spectrum.aspx

MORTPAK v. 4.3
MORTPAK can be obtained from the United Nations Population Division

United States Census Bureau Population Analysis System (PASEX).

PASEX consists of 45 spreadsheets for population analysis. The PASEX documentation is contained in the second volume of the manual Population Analysis with Microcomputers.(Arriaga, 1994b).
It can be downloaded from: http://www.census.gov/population/international/software/pas/pasdocs.html.
For a brief description, see Annex A.

United Nations Models in Excel

The United Nations Tools in Excel are implementations of various models used by the United Nations Population Division for projecting demographic indicators. For a brief description, see Annex B.

Annex A: Population Analysis System (PASEX)[1]

PASEX (Population Analysis Software in Excel) currently consists of 45 spreadsheets. The PASEX documentation is contained in Volume II of the manual Population Analysis with Microcomputers.

A brief explanation of the calculation performed by each of the spreadsheets is presented here, grouped by main demographic topics.

Age Structure

There are 11 spreadsheets for the analysis of the age structure of the population. They are, in alphabetical order:

ADJAGE Adjusts any population total by sex to a given age structure.

AGEINT Interpolates between two population age structures. The user can specify if the interpolation is to be linear or exponential.

AGESEX Analyzes the age reporting in a population age and sex distribution. Calculates age and sex ratio indices and the United Nations age/sex accuracy index.

AGESMTH Smoothes a population age distribution using several methods.

BASEPOP Prepares the age and sex distribution of a population for making a projection, using a strong smoothing process.

BPSTRNG Prepare the age and sex distribution of a population for making a projection, using a strong smoothing process.

GRPOP-YB Makes a graph using population age structures for two or three dates, by year of birth of each cohort.

MOVEPOP Moves the population age distribution pertaining to a specific date to another date.

OPAG Distributes open-ended age groups (younger than 80+) into 5-year age groups up to 80 years and over.

PYRAMID Makes an age pyramid by sex, with absolute numbers and percentages of the population data.

SINGAGE Calculates the Whipple, Myers, and Bachi indices of age heaping based on enumerated population by single years of age.

Mortality

There are 15 spreadsheets for analyzing the mortality of a population. Some spreadsheets were prepared for populations with acceptable data, while others assume that only limited data are available. Although the set of mortality spreadsheets is the largest, we did not include those techniques developed and published by the United Nations in the MORTPAK package. The analysis of mortality may require the use of both the U.S. Census Bureau spreadsheets (PASEX) and the MORTPAK package. The PASEX spreadsheets related to mortality are:

ADJMX Adjusts a pattern of mortality rates by age for both sexes combined or for each sex (age-specific death rates or central death rates from an empirical or existing life table) for obtaining a user-specified number of deaths in a population.

BTHSRV Estimates infant mortality rates based on information on the number of children born during a year prior to the census and the number still alive at the census date.

E0LGST Fits a logistic function to values of life expectancies at birth for each sex simultaneously, given two or more observed values of life expectancies and the two asymptotes of the logistic.

GRBAL Uses the technique developed by Brass for estimating the completeness of reporting of deaths over 5 years of age in relation to a population (Brass, 1975).

---

[1] URL: http://www.census.gov/population/international/software/pas/pasdocs.html

INTPLTM and INTPLTF Interpolate male and female life tables, respectively, between the values of two given sets of "pivotal" life tables.

LOGITQX and LOGITLX These two spreadsheets are used for smoothing functions of a life table, using logits of the $q(x)$ and $l(x)$ functions of a life table.

LTMXQXAD Constructs a life table from age-specific death rates or from the probabilities of dying between two specific ages.

LTNTH Selects a Coale-Demeny model life table, region North, that will reproduce a given crude death rate pertaining to a given population age structure.

LTPOPDTH Constructs and smooths a life table for both sexes or one sex at a time, using population and death data.

LTSTH Is the same as LTNTH, but using region South of the Coale-Demeny model life tables.

LTWST Is the same as LTNTH, but using region West of the Coale-Demeny model life tables.

PREBEN Estimates the level of mortality for ages 5 years and above during an intercensal period (Preston-Bennett, 1983).

PRECOA Uses the technique developed by Preston and Coale for evaluating an available age distribution of deaths in relation to the population.

There are 13 spreadsheets for analyzing the fertility of a population. Most spreadsheets are for populations whose registration of births is not complete.

ADJASFR Adjusts a given pattern of age-specific fertility rates to reproduce a desired total number of births.

ARFE-2 and ARFE-3 Use a technique developed by Arriaga for estimating fertility rates based on information on the average number of children ever born and a pattern of fertility.

ASFRPATT Provides age-specific fertility rates pertaining to a given total fertility rate.

CBR-GFR Calculates the crude birth rate and the general fertility rate based on a desired total fertility rate.

CBR-TFR Estimates the crude birth rate and the total fertility rate, based on the general fertility rate.

PFRATIO Uses the P/F ratio technique, originally developed by Brass, for adjusting reported age-specific fertility rates to the "actual" level of fertility.

RELEFERT Uses Rele's technique for estimating the gross reproduction rate of a population for one or two 5-year periods prior to the census date.

REL-GMPZ Uses the technique developed by Brass and associates "for the evaluation and adjustment of fertility estimates obtained from retrospective reports of birth histories or features of birth histories."

REVCBR Calculates crude birth rates during two or three 5-year periods prior to the census date, based on the age structure of the population.

TFR-GFR Estimates the total fertility rate and the general fertility rate based on a crude birth rate.

TFRLGST Fits a logistic function to two or more values of total fertility rates, and interpolates and extrapolates. It requires asymptotic values.

TFRSINE Fits a sine function to two values of total fertility rates, and interpolates and extrapolates.

Internal Migration

There is one spreadsheet that estimates intercensal migration between two areas.

CSRMIG Provides intercensal net migration estimates between two areas.

Urbanization and Distribution of the Population

There is one spreadsheet providing several indices for analyzing the urbanization process in a population.

URBINDEX Calculates several urbanization and population distribution indices.

Other Spreadsheets

There are four additional spreadsheets that can be used for several purposes; they are:

CTBL32 Uses a technique for adjusting information in a table to a set of desired marginal totals (totals of rows and/or columns).

FITLGSTC Fits a logistic function to three (or a multiple of three) observed equidistant values of any index. It does not need asymptotic values.

LOGISTIC Fits a logistic function to values of any index given two or more observed values of the index and the two asymptotes of the logistic.

SP Constructs a stable population based on life tables by sex and age-specific fertility rates (or the intrinsic growth rate.

Annex B: United Nations Projection Models in Excel

The United Nations Projection Tools in Excel are implementations of various models used by the United Nations Population Division for projecting demographic indicators. A brief explanation of the calculation performed by each of the spreadsheets is presented here, grouped by main demographic topics.

Fertility

There are 2 spreadsheets for the projection of fertility, one for the level of fertility, expressed as Total Fertility, and one for age patterns of fertility, expressed as proportionate age-specific fertility rates:

- **UNPD_FerModel_v1** Implements the United Nations Model of Fertility Decline. The model consists of a double logistic function and three different parameter sets for three different pathways of fertility decline. For a description, see http://www.un.org/esa/population/publications/WPP2006RevVol_III/WPP2006RevVol_III_final.pdf

- **UNPD_ASFR** The spreadsheet interpolated between an observed, user-supplied base pattern of fertility and 15 different target age patterns of fertility. The target age patterns were obtained from typical observed empirical patterns and are indexed by the Mean Age at Childbearing (MAC).

Mortality

There is one spreadsheet for projecting the level of mortality, expressed as life expectancy at birth for males and females. For the projection of the age pattern of mortality, the United Nations MORTPAK software package can be used. The Excel spreadsheet related to mortality is:

- **UNPD_MorModel_v.1** The United Nations model of increasing life expectancy at birth is based on a double-logistic function and five different parameter sets for five different pathways of mortality decline (Increase in life expectancy).

International Migration

There is one spreadsheet that can be used for the calculation of model migration schedules, especially created for net-migration.

- **UNPD_Migration Age Patterns** The United Nations model age patterns of net migration are based on observed empirical age patterns and allow for the distribution of total net migration figures into age patterns typical for family migration, female labor migration and male labor migration.