A stochastic multi-regional model for Italian population projections

Francesco Billari (francesco.billari@sociology.ox.ac.uk), University of Oxford - Department of Sociology, United Kingdom

Gianni Corsetti (giacorsetti@istat.it), Istat - Socio-demographic and Environmental Statistics Directorate, Italy

Rebecca Graziani (rebecca.graziani@unibocconi.it), Bocconi University - Department of Policy Analysis and Public Management, Italy

Marco Marsili (marsili@istat.it), Istat - Socio-demographic and Environmental Statistics Directorate, Italy

Eugenio Melilli (eugenio.melilli@unibocconi.it), Bocconi University - Department of Decision Sciences, Italy

Extended abstract

Introduction

In this work we show a methodology that will be implemented by Istat to produce stochastic forecasts of the Italian population for the period 2014-2065 at regional level.

In recent years stochastic (or probabilistic) population forecasting has, finally, received a great attention by researchers. In the literature on stochastic population forecasting, three main approaches have been developed (Keilman et al., 2002). The first approach is based on time series models. The second approach is based on the extrapolation of empirical errors, with observed errors from historical forecasts used in the assessment of uncertainty in forecasts (e.g., Stoto, 1983). Finally, the third approach referred to as random scenario defines the probabilistic distribution of each vital rate on the basis of expert opinions.

In this study we follow the latest approach, applying the method proposed by Billari et al. (2012), where the full probability distribution of forecasts is specified on the basis of expert opinions on future developments of the main components of the demographic change.

The main aim of this work is to exploit the results on national forecasts, obtained in the context of the above mentioned expert-based approach, as a reference to those relating to regional forecasts. As foreseen in the National Statistical Plan, Istat has the responsibility to produce, more or less regularly over time, the official regional population projections for Italy. Closely linked to this need, our reference model is therefore a multiregional one, where the twenty-one Italian regional populations are projected simultaneously and where, coherently, the number of internal in-migrants must always be equal to the number of internal out-migrants.

Stochastic method and expert opinions at national level

The conditional elicitation procedure makes it possible to elicit from experts information on the future marginal behaviour of a single indicator in terms of expected value and variability, but also on the across time correlation of each indicator and on the correlation between any two indicators at a given year or across time.

We have designed a questionnaire on the future national trends (Billari et al., 2013) according to such elicitation procedure. The questionnaire was submitted online to a representative sample of Italian demographers on Spring 2013.

In particular we have derived the joint forecast distribution of the pair *Male and Female Life Expectancies at Birth (MLE, FLE)*. The forecast distribution of *Total Fertility Rate (TFR)*, *Number of immigrants (IMM) , Number of emigrants (EMI)* and *Mean Age at Childbearing (MAC)* are derived separately.

We allow for correlation between sexes for mortality, as we aim at exploring the future evolution of the gender gap. Last, to reduce the complexity of the problem, some simplifying assumptions are made. In particular, we assume EMI, IMM, EMI and MAC being independent on the all components.

The method works by deriving the joint forecasting distribution of the indicators at 2030 and 2065, while the complete distribution of the process over the forecasting interval (from 2014 to 2065) is obtained by interpolation.

The forecasting distributions of the indicators at the two time points are assumed to be multivariate Gaussian, and the parameters are specified on the basis of the information elicited from experts.

The expert opinions, synthesized with an arithmetic mean, produce the average values we have assumed for our stochastic distributions of the indicators.

The variance of each indicator and the covariance among indicators are obtained from the conditional questions, summarized through the average values provided by the experts.

Calculation of the variances and covariance is more complex. In presence of a single indicator (EMI, IMM, EMI and MAC) variances and covariance of the bivariate random variables (i.e. TFR2030, TFR2065) are obtained by resorting to the rules of the standard bivariate normal variable.

In the case of a couple of indicators, as (MLE-FLE) for instance, variance and covariance of the multivariate normal random variable (MLE2030,FLE2030,MLE2065,FLE2065) are obtained by assuming conditional independence between MLE2065 and FLE2030 given the bivariate random variable (MLE2030,FLE2065).

Multi-regional model and stochastic regional forecasts

The stochastic methodology described above is a way to obtain results only at national level. Our aim is instead the production of stochastic forecasts for each Italian region, while it is rather not feasible to involve experts in an elicitation procedure for each of the twenty-one regions.

For these reasons we have developed a procedure which combines national stochastic expert-based forecasts with provisional non-stochastic regional forecasts based on observed data and on our experience of official projections-maker. Then, the combination of the two allows for stochastic forecasts at the regional level, as follows.

Focusing our interest on the mentioned demographic indicators, their future trends are derived by making assumption on a future convergence of regional projections (2014-based) to the expert-based Italian forecasts.

For each summary indicator we derive a regional forecast through time series analysis. We use the latest data available in Istat: 1952-2012 for age-specific fertility rates, 1974-2012 for life tables and 2008-2012 micro data for migration.

The convergence scenario of the Italian regions means that geographical differences in terms of demographic behaviour are fade out in the long run, where the year of full convergence is intended to be a year far beyond the time-horizon of the forecasts.

Following this approach, regional projections will converge to each of 2,000 national simulations drawn out from the stochastic distributions of the indicators at national level, specified on the basis of the elicitation procedure. In this way, for each regional summary indicator (each following a convergence path compatible to its characteristics) we draw 2,000 samples from its derived forecast distribution so to obtain, in a simulation based approach, the forecasts of the population sizes (by age and sex) for the twenty-one Italian regions.

Following a multiregional perspective we cannot ignore internal migration flows and also for this component we adopt a probabilistic approach. The main assumption states that future propensity of interregional migration, expressed by an O/D matrix of migration probabilities by age and sex, will be constant along the forecast period and equal to the observed one in the recent years (2008-2012). The use of migration probability from one region to another also ensures internal consistency of the multiregional model. That means, at each age and sex the sum of the destinations in the region j is equal to the sum of the origins from all other regions (i≠j) with destination j and, at the same time, the sum of the origins in the region j is equal to the sum of the destinations to all other regions (i≠j) having origin j.

Operatively we define six main geographical areas – by applying a log-linear model – formed by regions that in the recent past have shown a homogeneous migratory behaviour. This choice is due to small frequencies of interregional transfers that can be observed at age classes with low risk of migration and/or involving small sized regions.

The six clusters are:

- North-west: Piemonte, Valle d'Aosta, Lombardia.
- North-east: Trentino-Alto Adige, Veneto, Friuli-Venezia Giulia.
- North Apennines: Liguria, Emilia Romagna, Toscana.
- Central Adriatic: Umbria, Marche, Abruzzo, Molise.
- Lazio.
- South and Islands: Campania, Puglia, Basilicata, Calabria, Sicilia, Sardegna.

Then we use *kernel regression* to smooth age-schedules of migration probability from each geographical area to another by sex and single year.

Let ${}^{s}\mu_{x}^{hk}$ be the probability of an individual at age *x* and sex *s* to migrate from area *h* to area *k* (also in case of h=k), we assume that it is a gaussian random variable with:

- mean is equal to 2008-2012 mean of smoothed curves;
- variance is equal to 2008-2012 variance of smoothed curves.

Finally, the regional stochastic probabilities to migrate from each region to another are calculated by applying scale factors in order to consider the variability in terms of demographic size of the Italian regions.

Bibliography

Alho J.M. and B.D. Spencer. 1997. The practical specification of the expected error of population forecasts. Journal of Official Statistics. 13(3): 203-225.

Alho, J.M. and B.D. Spencer. 2005. Statistical demography and forecasting. New York: Springer.

Bernard A. and M. Bell. 2012. A Comparison of Internal Migration Age Profile Smoothing Methods. Working Paper 2012/01. Queensland Centre for Population Research, The University of Queensland.

Billari, F.C., R. Graziani and E. Melilli. 2012. Stochastic population forecasts based on conditional expert opinions. Journal of the Royal Statistical Society. Series A. 175(2): 491-511.

Billari, F.C., G. Corsetti, R. Graziani, M. Marsili and E. Melilli. 2013. Towards stochastic forecasts of the Italian population: an experiment with conditional expert elicitations. Eurostat-Unece, Work session on demographic projections, Rome, 29-31 October 2013.

Booth H. 2006. Demographic forecasting: 1980 to 2005 in review. International Journal of Forecasting. 22: 547–581.

CBS. 2011. Key figures of the population forecasts 2010-2060. Statline, Centraal Bureau voor der statistiek, http://statline.cbs.nl/statweb/.

Corsetti G. and Marsili M. 2012. A stochastic population projection from the perspective of a national statistical office. European Population Conference 2012. Stockholm, 13-16 June 2012, EAPS. http://epc2012.princeton.edu/papers/120635.

Eurostat. 2011. The greying of the baby boomers - A century-long view of ageing in European populations. Statistics in focus, n. 23/2011.

Graziani R. and Keilman, N. 2011. The sensitivity of the Scaled Model of Error with respect to the choice of the correlation parameters: A simulation study. Working Paper 37. Carlo F. Dondena Centre for Research on Social Dynamics, Bocconi University, Milan.

Istat. 2011. Il futuro demografico del paese - Previsioni regionali della popolazione residente al 2065. Statistiche Report, www.istat.it, 28 dec 2011. Rome.

Keilman, N., D.Q. Pham and A. Hetland. 2002. Why population forecasts should be probabilistic - illustrated by the case of Norway. Demographic Research. 6(15): 409-454.

Lee R. 1998. Probabilistic Approaches to Population Forecasting. Population and Development Review 24. Issue Supplement: Frontiers of Population Forecasting: 156-190.

Lee R.D. and L.R. Carter. 1992. Modeling and forecasting U.S. Mortality. Journal of the American Statistical Association, September. 87/419.

Lutz W., W.C. Sanderson and S. Scherbov. 1998. Expert-Based Probabilistic Population Projections. Population and Development Review. 24: 139-155.

Rogers A. and L. Castro. 1981. Model migration schedules. International Institute for Applied System Analysis. Laxenberg. Austria. RR-8 1-30.

Rogers, A. 1985. Regional Population Projection Models. Beverly Hills. CA: Sage.

Rowan S. and Wright E. 2010. Developing stochastic population forecasts for the United Kingdom: Progress report and plans for future work. Eurostat, Work session on demographic projections, Lisbon, 28-30 April 2010, Methodologies and Working papers, European Commission.

Shaw C. 2008. The National Population Projections Expert Advisory Group: results from a questionnaire about future trends in fertility, mortality and migration. Population trends n.134, Winter 2008, Office for national statistics.

Schmertmann C.P. 2003. A system of model fertility schedules with graphically intuitive parameters. Demographic Research. 9(5): 81-110.

Stoto, M. A. 1983. The accuracy of population projections. J. Am. Statist. Ass. 78: 13–20.

Tuljapurkar S., R.D. Lee and Q. Li. 2004. Random scenario forecast versus stochastic forecasts. International Statistical Review. 72: 185–199.