WP 12.3
16 October 2013

UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

STATISTICAL OFFICE OF THE EUROPEAN UNION (EUROSTAT)

Joint Eurostat/UNECE Work Session on Demographic Projections
organised in cooperation with Istat
(29-31 October 2013, Rome, Italy)

Item 12 – Bayesian approaches

Expert-Based Stochastic Population Forecasting: A Bayesian Approach to the Combination of the Elicitations

Francesco Billari, University of Oxford
Rebecca Graziani, Bocconi University
Eugenio Melilli, Bocconi University

Abstract

This paper builds on Billari et al. (2012) and suggests a method that derives expert-based stochastic population forecasts in such a way to account for relationships both between demographic components and between experts. Indeed, two main issues are addressed. First, we aim at defining for each expert a joint distribution for the demographic indicators that allows for both across-time correlation for a single indicator and for dependence between indicators at the same time and across-time. The second goal we pursue in this work is to find a suitable way to combine opinions elicited from several experts, to be used as the basis for the forecasts. We suggest to address this issue by resorting to the so-called Supra-Bayesian approach, introduced by Morris (1974) and then developed by many authors. Such approach makes it possible to combine expert opinions on unknown quantities within the formal framework provided by the Bayesian approach to statistics, by assuming that such opinions are data. The analyst is therefore asked to specify on one side a likelihood function, to be parametrized in terms of the unknown objects, and on the other side a prior distribution for the parameters. The posterior distribution, obtained by applying the Bayes theorem and updating the analyst prior opinions on the basis of the evaluations provided by the experts, can then be used as a collective distribution for the unknown quantities of interest. As for the likelihood, we suggest to resort to a mixture model approach. We assume that experts can be grouped into a given number of clusters, according to the shared information. The number of clusters is fixed by the analyst, but we let expert evaluations determine the cluster memberships. Finally, an application to the forecast of the Italian Population from 2010 up to 2065 is proposed.

Key Words: Random scenarios; Conditional elicitation; Supra-Bayesian; Mixture model; Markov Chain Monte Carlo.

1 Introduction

Population forecasts are strongly requested by both public and private institutions as crucial ingredients for long-range planning. Traditionally, official national and international agencies derive population projections in a deterministic way, specifying scenarios based on combinations of assumptions on the future behavior of demographic components. Uncertainty is not explicitly incorporated so that the expected accuracy of the forecasts cannot be assessed: prediction intervals for any indicator depending on the future population structure cannot be computed. Nevertheless, it is common practice to consider high-low scenario intervals as containing likely future values of the indicators of interest.

Recently, stochastic (or probabilistic) population forecasting is finally receiving a growing attention by researchers and, to a lesser extent, by official agencies. Three main approaches to stochastic population forecasting can be roughly distinguished in the literature (Keilman et al. 2002). The first approach adopts standard procedures of time series theory: for each indicator a model is fitted to past series of data and forecasts are obtained resorting to usual extrapolation techniques. Models are developed both within the frequentist and the Bayesian approach. The best-known approach using times series in a classical framework is the one due to Lee and Carter (1992) Recently Bayesian hierarchical time series models have been proposed for deriving fertility (Alkema et al. 2011) and mortality (Raftery et al. 2012) forecasts; United Nations Population Division provides stochastic forecasts obtained by applying these models. The second approach is based on the extrapolation of empirical errors, with observed errors from historical forecasts used in the assessment of uncertainty; see, e.g., Stoto (1983). Alho and Spencer (1997) proposed in this framework the so-called Scaled Model of Error, which was used to derive stochastic population forecasts within the UPE (Uncertainty Population of Europe) project; see Alders et al. (2007). Finally, in the third approach, known as random scenario, the forecast distribution of the demographic components is derived on the basis of suitably elicited expert opinions. In Lutz et al. (1998) the forecast of a demographic indicator at a given future time T is assumed to be the realization of a random variable having Gaussian distribution with parameters specified on the basis of expert opinions. For each time t of the forecasting interval $[t_0, T]$ the forecast is obtained by interpolation of the starting known and the final random values. In Billari et al. (2012) the joint probability distribution of the indicators in the interval $[t_0, T]$ is specified using expert opinions at time T conditional on their values at an intermediate point; in this way correlations across time are obtained indirectly through conditional evaluations. This work builds on the last cited paper by by Billari et al. (2012). We suggest a method that derives expert based stochastic population forecasts in such a way to account for relationships both between demographic compo- nents and between experts. Indeed, two main issues are addressed. First, we aim at defining for each expert a joint distribution for the demographic indicators that allows for both across-time correlation for a single indicator and for dependence between indicators at the same time and across-time. There is a certain debate among researchers about the advisability and/or necessity of allowing for dependence between different components. We believe that the most sensible approach is to be neutral in some respect, that is to allow for dependence without imposing it. We think that the presence, degree and type of relationship among demographic components should stem from the evaluations of the experts on their joint development through time.

The second goal we pursue in this work is to find a suitable way to combine opinions elicited from several experts, to be used as the basis for the forecasts. A wide literature is available on the problem of the aggregation of expert opinions. For a very good and comprehensive, even if not very recent, review on this topic see Genest and Zidek (1986). Generally speaking, pooling opinions means to merge many individuals' probability distributions on unknown objects into a single collective assignment; in our context these objects are the chosen demographic indicators. Classical pooling methods proceed by averaging in some way expert opinions; for instance, the linear rule defines the (possibly weighted) standard average as collective assignment.

In this paper, we suggest to combine expert opinions resorting to the so-called Supra-Bayesian method of pooling, introduced by Morris (1974) and then developed by many authors; we cite French (1980, 1981), Winkler (1981), Lindley (1983, 1985), Gelfand et al (1995). Roback and Givens (2001) apply it in the framework of deterministic simulation models. Within such approach expert opinions on unknown quantities are combined in the formal framework provided by the Bayesian approach to statistics, by assuming that opinions are data. The analyst is therefore asked to specify a likelihood function, to be parametrized in terms of the unknown objects, and a prior distribution for the parameters. The posterior distribution, obtained by applying the Bayes theorem and updating the analyst prior opinions on the basis of the evaluations provided by the experts, can then be used as a collective distribution for the unknown quantities of interest.

Through the choice of the likelihood, the Supra-Bayesian approach makes it possible to model different kinds of dependence structure. In this work, we suggest to implicitly derive the dependence structure of the expert evaluations by resorting to a mixture model approach. We assume that experts can be grouped into a given number of clusters. The number of clusters is fixed by the analyst, but we let expert evaluations determine the cluster memberships. We assume that within each cluster, expert evaluations are generated by the same distribution. This makes it possible to account for the variability of the evaluations of experts exposed to the same information. The centres of the clusters distributions are then assumed to be independently generated from the same distribution, centred at the unknown vector of future values of the indicators. In this way, we are able to account for the heterogeneity of the expert evaluations due to their owning different pieces of information. Moreover we achieve the goal of allowing for dependence between experts, without explicitly fixing it.

The paper is structured as follows. In Section 2 the suggested method is discussed in detail. In Section 3 the method is applied for the derivation of stochastic forecasts of the Italian Population from 2010 to 2065.

2 The model

Following a common practice we work out population forecasts on the basis of the cohort-component model, so that we can focus on summary indicators of the three components of the demographic change: Total Fertility Rate for fertility, Male and Female Life Expectancies at Birth for mortality, Male and Female Number of Immigrants and Emigrants for migration. Our method derives the joint forecast distribution of all summary indicators on the basis of the evaluations provided by several experts; the aim, as emphasized in the Introduction, being to allow for both dependence between indicators (across-time for a single indicator, at the same time and across-time between any two indicators) and dependence between expert opinions. Nevertheless, to reduce the dimension of the problem, we make the following assumptions on the dependence structure. We model jointly Total Fertility Rate and Total Number of Immigrants, on one hand, and Male and Female Life Expectancies at Birth, on the other hand. We assume that such two pairs are mutually independent and both independent on the Total Number of Emigrants. The Total Numbers of Immigrants and Emigrants are then split by sex on the basis of a deterministic rule and age-schedules are derived for all indicators on the basis of deterministic models.

The evaluations are elicited according to the conditional procedure suggested by Billari et al. (2012). Experts are asked to provide central scenarios for the indicators of interest; moreover they are asked for conditional scenarios so that it is possible to gather information on the marginal variability and on the across-time correlation of each single indicator, and on the correlation between any two indicators at the same time and across-time.

We describe how our method works referring to two indicators R_1 and R_2 to be forecasted over the interval $[t_0, T]$. We split the forecast interval into two sub-intervals, by considering an inner point t_1 and we begin by deriving the distribution of the vector $R = (R_{11}, R_{12}, R_{21}, R_{22})$, where R_{ij} is the random variable associated with the value of R_i at time t_j, $i = 1, 2$ and $j = 1,2$ and $t_2 = T$. The joint distribution of the two indicators over the entire forecast interval can then be obtained resorting to interpolation techniques. Consider K experts, and denote by x_i the four dimensional vector of central scenarios provided by expert i, on the pair of indicators at times t_1 and t_2.

As mentioned in the Introduction, we resort to the Supra-Bayesian approach to derive the joint forecast distribution of the indicators. According to such approach, $x_1, x_2, \ldots, x_K$ are treated as data. Within a Bayesian approach, the analyst is asked to specify the likelihood $f(x_1, \ldots, x_K | R_{11}, R_{12}, R_{21}, R_{22})$, this being the joint distribution of the scenarios parametrized by the unknown future values of the indicators, and a prior on such parameters.

As for the likelihood we suggest to resort to a mixture model.

Our model assumes the existence of J groups of experts. The number J is assumed to be fixed by the analyst, but the model can be generalized by taking J as an additional parameter. The group membership is determined for each expert on the basis of the elicited scenarios. Indeed the assignment of each expert to a specific cluster might be rather difficult and turn out to be quite an arbitrary choice. Within each group j, the members opinions are sampled from the same multivariate Gaussian distribution, centered at μ_j and having covariance matrix Σ_j, with $j = 1, \ldots, J$. The group means $\mu_1, \ldots, \mu_J$ are independently distributed according to the same multivariate Gaussian distribution centered at the vector R of the unknown future values of the indicators. In this way we explicitly account for two possible sources of lack of agreement between the experts mentioned in the Introduction: experts can be exposed to different information or they can interpret the same shared body of information in different ways.

More precisely, we rely on the following hierarchical multivariate mixture model:

$$
\begin{aligned}
& x_i | \mu_1, \ldots, \mu_J, \Sigma_1, \ldots \Sigma_J, p_1, \ldots, p_J \text{ ind} < \textstyle\sum_{j=1}^{J} \quad p_j N_4(\mu_j, \Sigma_j) \quad i = 1, \ldots, K \\
& \mu_j | \Sigma_j \text{ ind } < N_4(R, \frac{1}{k_0} \Sigma_j) \quad j = 1, \ldots, J \\
& \Sigma_j \text{ iid} < IW(\Sigma_0, n_0) \quad j = 1, \ldots, J \\
& p_1, \ldots, p_J < Dir(\alpha_1, \ldots, \alpha_J) \\
& R < N_4(\mu_R, \Sigma_R)
\end{aligned}
$$

In a Bayesian approach, the analyst is asked to specify the prior distributions on the parameters of the likelihood. We resort for computational convenience to the usual assignment of conjugate proper priors and we specify their hyperparameters so to yield flat and therefore noninformative distributions. The hyperparameters to be chosen are $k_0, n_0, \Sigma_0, \alpha_1, \ldots, \alpha_J, \mu_R, \Sigma_R$. k_0 and n_0 impact on the spread of the prior distributions on the groups means and on the groups covariances, respectively. They are chosen as small as possible, in order to increase the priors variability. Σ_0 is the centre of the prior on the covariance matrices of the groups distributions. We specify it on the basis of the values elicited from the experts. Indeed, for each expert we can work out the covariance matrix of his evaluations. We suggest to set Σ_0 equal to the arithmetic mean of such matrices; this mean could be multiplied by a given constant so to increase the elicited variances of the indicators and therefore to take into account the fact that, often, experts tend to underestimate the variability of their forecasts. According to the properties of the Dirichlet distribution, the lower is the value of $\sum_{j=1}^{J} \alpha_j$ the higher is the variability. A standard choice is then to set $\alpha_j = \frac{1}{J}$. μ_R is the centre of the prior assigned to vector R; it represents a prior guess on the future values of the indicators and then can be specified using all available information. For instance, it can be fixed by considering the central scenarios provided by national and international statistical agencies. Finally, Σ_R is the covariance matrix of the prior distribution on R. We suggest to choose rather high variances so to specify a diffuse prior, thus limiting the impact of μ_R and to set the covariances equal to 0; this choice corresponds to an assumption of a priori independence of the indicators.

The posterior distribution is then the forecast distribution of the pair of indicators at the two considered time points. Such distribution cannot be derived in closed form, but it can be approximated by means of Gibbs sampler worked out on the basis of well-known results in the literature on mixture models (see, among others, Lavine and West 1992).

3 An application: Forecast of the Italian Population

Here we illustrate the results of the forecast of Italian Population from 2010 to 2065, obtained by applying the method. Expert evaluations were collected by means of a questionnaire, designed jointly with researchers of the Population Division of the Italian National Statistical Institute (ISTAT). The questions are posed according to the elicitation procedure introduced by Billari et al. (2012). The elicited evaluations are used for deriving the forecast of the Italian population over the interval [2010, 2065], this being the interval used by ISTAT in the latest release of Italian population projections. The results are based on $N = 10000$ samples out of 20000 drawn from the joint distribution of the considered demographic indicators at the two time points 2030 and 2065. The prior parameters are specified as described in the previous section. k_0 is set equal to 1 and n_0 to the dimension of the vector of indicators, so to increase the spread of the prior on the groups means and the groups covariance matrices; α is fixed equal to 1, μ_R is set equal to the vector of central scenarios provided by ISTAT; as for Σ_R, and the variances are derived from the high-low ISTAT scenarios while the covariances are all fixed equal to 0. Σ_0, as discussed, in Section 2 is given by the arithmetic mean computed over all experts of the covariance matrices of their elicitations. We provide the results assuming that experts can be grouped into $J = 4$ clusters. We ran the algorithm for several values of J, but the derived forecasts did not show significant differences, this proving the robustness of the method with respect to the choice of the number of clusters.

Table 1 shows (posterior) means and prediction intervals of the demographic indicators at 2030 and 2065 along with the scenarios (central, high and low) provided by ISTAT, for comparison purposes. Table 2 shows the 2030 and 2065 total population forecasts with their 85% prediction intervals, along with ISTAT scenarios. Population is forecasted to grow from 2010 to 2030: the starting population is 60.63 million at the beginning of 2010, the forecast for 2030 is 61.83 million, the lower bound and upper bound of the prediction interval being respectively 60.77 and 62.98 million. This could be compared to ISTAT non-probabilistic projections, for which the central scenario at 2030 results in a total population of 63.48 million; the low scenario is 61.68 million and the high scenario 65.20 million. From 2030 to 2065, the population is forecasted to decrease to 57.67 million, and as expected there is more uncertainty, with a prediction interval 54.16 and 61.01 million. Again, this could be compared to ISTAT projections, for which the central scenario at 2065 is 61.31 million; the low and high scenarios are 53.39 million and 69.13 million, respectively. Therefore our forecasts show a smaller variability with respect to the interval determined by the high-low ISTAT scenarios. Forecasts of any indicator related to the age-structure of the population can be derived. In particular, the forecasts of the elderly dependency ratio (population aged more than 65 over population aged between 14 and 65, as a percentage) show that is virtually no uncertainty on the fact that the Italian population will continue to age during the whole interval. As compared to 32.2 percent in 2010, the elderly dependency ratio is projected to rise to 46.2 percent in 2030 (with prediction interval ranging from 42.7 to 53.1 percent) and to 65.5 percent in 2065 (with prediction interval ranging from 57.8 to 72.9 percent).

Table 1: 2030 and 2065 Demographic Indicators: ISTAT Scenarios and Stochastic Forecasts with 85 % intervals

Demographic Indicator	2010	2030				2065			
		Istat		Mixture model		Istat		Mixture model	
Total Fertility Rate	1.42	1.49	(1.37 1.60)	1.509	(1.36 1.66)	1.61	(1.38 1.83)	1.55	(1.36 1.72)
Mean Maternal Age	31.4	31.8	(31.3 32.3)	31.8	(30.69 32.91)	32.6	(30.9 32.7)	32.2	(30.59 33.74)
Male Life Expectancy	79.5	82.8	(81.4 84.1)	82.89	(80.39 85.34)	86.6	(84.4 88.6)	86.69	(82.80 90.38)
Female Life Expectancy	84.6	87.7	(86.2 89.2)	87.53	(85.69 89.43)	91.5	(88.8 93.8)	91.24	(87.86 94.62)
Number of Immigrants	408.66	321.19	(297.68 346.58)	314.35	(275.98 348.29)	303.85	(278.02 335.34)	298.26	(261.72 330.03)
Number of Emigrants	83.81	101.34	(89.01 115.83)	101.09	(90.08 111.50)	97.29	(97.29 164.62)	122.54	(106.08 133.95)

Table 2: 2030 and 2065 Total Population: ISTAT Scenarios and Stochastic Forecasts with 85 % Intervals

	2010	2030		2065	
Istat Scenarios	60.626	63.482	(61.675 65.205)	61.305	(53.390 69.125)
Stochastic Forecast	60.626	61.825	(60.772 62.977)	57.672	(54.164 61.009)

4 Conclusions

In this paper we suggested a method for deriving stochastic population forecast on the basis of a combination of expert evaluations. Expert opinions are elicited by means of a special designed questionnaire; experts are asked to provide (conditional) scenarios on the summary indicators of the main components of the demographic change. In this way, elicit opinions not only on the central scenarios, but also on the marginal variability and the correlation across-time and across indicators. Moreover, our method allows for dependence between expert opinions, combining them in the framework provided by the Supra-Bayesian approach. The dependence is induced by the choice as likelihood of a mixture model. A Gibbs sampler is designed to approximate the forecast distribution of the summary indicators, being in this approach a posterior distribution. The program developed in MATLAB for the implementation of the suggested method is available on demand from the second author. Stochastic forecasts of the Italian population from 2010 up to 2065 are derived, as an application of the proposed method and comparisons with latest released ISTAT projections are provided. It is important to emphasize here some of the caveats also because they suggest lines for future research. We did not explicitly consider, in the current implementation of the method, the uncertainty in the initial age and sex distribution of the population, although it could be handled using the same expert-based probabilistic approach. Moreover, we did not explicitly consider the uncertainty deriving from the way summary indicators given by the experts are translated into age-specific rates or absolute numbers.

References

Alders, M., Keilman, N. & Cruijsen, H. (2007). Assumptions for long-term stochastic population forecasts in 18 European countries. European Journal of Population 23, 33–69.

Alkema, L., Raftery A. E., Gerland P., Clark S. J., Pelletier F., Buettner, T. & Heilig, G. K. (2011). Probabilistic Projections of the Total Fertility for all countries. Demography 48, 815–839.

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

French, S. (1980). Updating of belief in the light of someone else's opinion. Journal of the Royal Statistical Society A 143, 43–48.

French, S. (1981). Consensus of opinion. European Journal of Operational Research s7, 332–340.

Gelfand, A. E., Mallick, B. K. & Dey, D. K. (1995). Modeling Expert Opinion Arising as Partial Probabilistic Specification. Journal of the American Statistical Association 90 (430), 598–604.

Genest, C. & Zidek, J. V. (1986). Combining Probability Distributions: A Critique and an Annotated Bibliography. Statistical Science 1 (1), 114–135.

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

Lavine, M. & West, M. (1992). A Bayesian Method for Classification and Discrimination. The Canadian Journal of Statistics 20, 451–461.

Lee, R.D. & Carter, L.R. (1992). Modelling and forecasting U.S. mortality. Journal of the American Statistical Association 87, 659–671.

Lindley, D. (1983). Reconciliation of Probability Distributions. Operations Research 31 (5), 866–880.

Lindley, D. (1985). Reconciliation of Discrete Probability Distributions. Bayesian Statistics 2, eds. J. M. Bernardo et al., Amsterdam: North- Holland, 375–390.

Lutz, W., Sanderson, W.C. & Scherbov, S. (1998). Expert-Based Probabilistic Population Projections. Population and Development Review 24 s, 139–155.

Morris, P. A. (1974). Decision analysis expert use. Management Science 20, 1233–1241.

Raftery, A.E., Li. N., Ševčíková , H., Gerland, P. & Heilig, G.K. (2012). Bayesian probabilistic population projections for all countries. - Supporting information. Proceedings of the National Academy of Science, 109 (35), 1–7.

Roback, P.J & Givens, G.H. (2001). Supra-Bayesian pooling of priors linked by a deterministic simulation model. Communications in Statistics – Simulation and Computation 30, (3) 381, 447–476.

Stoto, M.A. (1983). The Accuracy of Population Projections. Journal of the American Statistical Association 78 (381), 13–20.

Winkler, R. L. (1981). Combining Probability Distribution from Dependent Information Sources. Management Science 27 (4), 479–488.