Expert-based stochastic population forecasting: a bayesian approach to the combination of expert evaluations

Rebecca Graziani[1]

(with F. Billari[2] and E. Melilli[1])

[1] Bocconi University and Carlo F. Dondena Center for Research on Social Dynamics, Milano, Italy
[2] Nuffield College, University of Oxford, Oxford, United Kingdom

Joint Eurostat/UNECE
Work Session on Demographic Projections,
30th October 2013

Outline

- Expert-Based stochastic Population forecasting.
- Conditional elicitation procedure.
- Combination of expert opinions.
  - The Supra-Bayesian approach.
  - Mixture model.
- Forecast of the Italian population from 2010 to 2065.

Expert-based approach

- Our proposal lays within the Expert-Based approach to stochastic forecasting.
- The forecasting distribution of demographic components are worked out on the basis expert opinions.

Forecasts of summary indicators

The forecasts are obtained within the usual framework of the cohort-component method. We focus on the construction of the forecasting distribution of summary indicators of the three components of the demographic change:

- Total Fertility Rate;
- Male and Female Life Expectancies at Birth;
- Immigration and Emigration.

Simplifying assumptions

- We model jointly
  - Total Fertility Rate and Immigration
  - Male and Female Life Expectancies
- We assume that such two pairs are mutually independent and both independent on Emigration.
- The method derives the joint forecasting distribution of the indicators at two time points and the entire process is obtained by interpolation.
- Total Immigration and Emigration are then split by sex on the basis of deterministic rules.
- Age-schedules are derived for all indicators on the basis of deterministic models.

The aim

- ☞ To allow for dependence between indicators: across-time for a single indicator, at the same and across-time between any two indicators.
- ☞ To allow for dependence between expert opinions: on the same indicator at the same time and across time and on any two indicators at the same time and across time.

Conditional evaluations

- Expert evaluations are elicited by means of the conditional elicitation method introduced in our previous work "Stochastic Population Forecasts based on Conditional Expert Opinions" in JRSS A (2012).
- The method makes it possible to elicit evaluations
  - on the future values of the indicators;
  - on their expected variability;
  - on the across-time correlation of each indicator;
  - on the correlation (at the same time and across time) between indicators.

Combination of expert opinions

- ☞ The issue addressed 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.
- ❑ 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.

Traditional pooling methods

- Classical pooling methods proceed by averaging in some way expert opinions; linear rule defines the (possibly weighted) standard average as collective assignment. Similarly are defined geometric or logarithmic pooling rules.
- Pooling methods based on averaging suffer of a lack of a normative basis for choosing the weights.

Information aggregation vs opinion pooling

- Even if an expert's opinion is the expression of the person's beliefs, nevertheless it is usually based on the experience the individual has had with the problem at hand. Indeed, experts who have been trained in the same framework will typically share a quite large amount of information.
- So their current opinions may differ for two main reasons: first, they don't share exactly the same evidence (i.e., they have at least partially different information), second they do not interpret common facts in the same way.
- Linear and classical pooling do not explicitly consider this difference.

Supra-Bayesian approach

- We suggest to resort to the Supra-Bayesian approach introduced by Morris (1974) and then developed by many authors; French (1980, 1981), Winkler (1981), Lindley (1983, 1985), Rubin (1983), Gelfand et al (1995). Roback and Givens (2001).
- Such approach makes it possible to combine expert opinions on unknown quantities within the formal framework provided by the Bayesian approach to statistics.

Supra-Bayesian approach

- Expert opinions are treated as data.
- The analyst is therefore asked to choose
  - ☞ a likelihood function, to be parametrized in terms of the unknown quantities.
  - ☞ a prior distributions for the parameters.
- The posterior distribution, obtained by applying the Bayes theorem 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.

Two indicators, k experts

- Let consider the case of two indicators, R_1 and R_2 and let the forecasting interval $[t_0, T]$ be split into two sub-intervals, by considering an inner point t_1 in $[t_0, T]$.
- We show how to derive the joint forecast distribution of the vector of the values of the indicators R_1 and R_2 at time t_1 and t_2, that is $R = (R_{1t_1}, R_{1t_2}, R_{2t_1}, R_{2t_2})$, where $t_2 = T$. The entire joint distribution of the pair over $[t_0\ T]$ can be obtained by interpolation.
- Consider k experts.
- Let $(x_1, x_2, \ldots, x_k)$,where $x_i = (x_{i,1t_1}, x_{i,1t_2}, x_{i,2t_1}, x_{i,2t_2})$ the vector of central evaluations provided by expert i, on the pair of indicators at times t_1 and t_2.

Set-up of the model

- $(x_1, x_2, \ldots, x_k)$ are treated as data.
- the analyst has to specify the likelihood $f(x_1, \ldots, \ x_k | R_{1t_1}, R_{1t_2}, R_{2t_1}, R_{2t_2})$.
- the analyst has to assign a prior distribution to the vector $(R_{1t_1}, R_{1t_2}, R_{2t_1}, R_{2t_2})$ expressing his prior beliefs and knowledge on it.

☞ The Bayes theorem makes it possible to derive the posterior distribution $\pi(R_{1t_1}, R_{1t_2}, R_{2t_1}, R_{2t_2} | x_1, \ldots, x_k)$, that can therefore be used as the joint forecast distribution of the indicators R_1 and R_2 at t_1 and t_2.

Choice of Likelihood function

- One possible choice could be a Normal distribution with marginals centered at $R = (R_{1t_1}, R_{1t_2}, R_{2t_1}, R_{2t_2})$, and with a given covariance matrix.
- Indeed the choice of the Normal distribution can be primarily motivated by mathematical convenience since under such assumption the computation of the posterior distribution of the indicators vector is greatly simplified.
- On the other hand, the construction of a likelihood of this kind is cumbersome, due to the number of terms of the covariance matrix to be specified.

☞ With this choice, the analyst is asked to specify the correlations between the evaluations of any two experts, our current focus.

Our proposal

- ☞ We suggest to implicitly derive the dependence structure by resorting to a mixture model.
- We assume that experts can be grouped into a given number of clusters, according to the shared information.
- We fix the number of clusters, but we let expert evaluations determine the cluster membership for each expert.
- 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.

Our proposal

- The centers of the distributions for each cluster are assumed to be independently generated from a distribution, centered at the unknown vector of future values of the indicators.
- We are able to account for the heterogeneity of the expert evaluations due to their owning different pieces of information, while achieving the goal of allowing for correlation between experts, without explicitly fixing it.

The hierarchical model

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

Prior

- ☞ Diffuse but proper priors are assigned to the parameters, so to avoid computational issues.
- ❑ Σ_0 is the center of the prior distribution on the covariance of groups components. We suggest to specify it on the basis of the values elicited from the experts. We set it equal to the arithmetic mean of such matrices multiplied by a given constant so to increase the marginal variances of the indicators, as elicited from the experts.
- ❑ μ_R, as center of the prior assigned to vector R, represents a prior guess on the future values of the indicators. We can fix it by considering the forecasts of the indicators released by the national and international statistical agencies.
- ❑ k_0 and n_0 determine the spread of the prior on the groups means and covariances respectively. They can be chosen so to achieve a high spread.
- ❑ α_j' s are set equal to $\frac{1}{J}$, since a priori there is no knowledge favoring one component rather than another.

Posterior

☞ Under such assignment of priors, the posterior distribution of the indicators at the two time points can be approximated by means of a Gibbs Sampler, on the basis of well-known results concerning Bayesian mixture models.

Forecast of the Italian Population

- The hierarchical model is applied to derive forecasts of the Italian population from 2010 up to 2065, on the basis of the evaluations elicited from a group of Italian experts by means of a questionnaire.
- The model is fitted for several values of J. The forecasts are robust with respect to the choice of the number of clusters, so that the mixture model turns out to be a useful and efficient technical devise making it possible to reduce the dimensionality of the problem while accounting for the dependence of expert evaluations.

- The results are shown for the case of $J = 2$ clusters, the corresponding model having the smallest BIC value, also with respect to the non-mixture ($i.i.d$) case.

Total Fertility and Immigration forecasts

Table: Posterior means, standard deviations and correlations of Total Number of Immigrants (IM) and Total Fertility Rate (TFR)

	mixture model				non mixture model
	two groups	three groups	four groups	five groups	
$E(IM_{2030})$	288.98	282.91	281.74	280.63	299.40
$E(IM_{2065})$	270.53	264.91	254.77	256.53	285.89
$E(TFR_{2030})$	1.53	1.53	1.52	1.53	1.53
$E(TFR_{2065})$	1.63	1.62	1.62	1.63	1.63
$\sigma(IM_{2030})$	65.2	60.30	66.05	65.26	72.11
$\sigma(IM_{2065})$	85.90	90.51	86.96	85.89	99.27
$\sigma(TFR_{2030})$	0.10	0.12	0.11	0.10	0.16
$\sigma(TFR_{2065})$	0.16	0.18	0.23	0.16	0.20
$\rho(IM_{2030} IM_{2065})$	0.42	0.44	0.41	0.45	0.44
$\rho(TFR_{2030}\ TFR_{2065})$	0.38	0.35	0.27	0.29	0.57
$\rho(IM_{2030}\ TFR_{2030})$	0.20	0.12	0.10	0.20	0.11
$\rho(IM_{2065}\ TFR_{2065})$	-0.09	-0.05	-0.13	-0.07	-0.10
$\rho(IM_{2030}\ TFR_{2065})$	-0.07	-0.08	-0.12	-0.13	0.10
$\rho(IM_{2065}\ TFR_{2030})$	-0.09	-0.05	-0.06	0.04	-0.18
BIC	291.09	333.54	346.67	380.67	305.10

Forecasts Total Fertility and Immigration

Table: Forecasts with 85 % Forecast intervals

Demographic Indicator	2010	2030		2065	
Total Fertility Rate	1.42	1.53	(1.36 1.70)	1.63	(1.36 1.89)
Mean Maternal Age	31.4	31.8	(31.3 32.3)	31.8	(30.69 32.91)
Male Life Expectancy	79.5	82.93	(80.36 85.58)	86.89	(83.70 90.13)
Female Life Expectancy	84.6	87.21	(85.01 89.44)	91.02	(86.89 95.04)
Number of Immigrants	408.66	321.19	(297.68 346.58)	314.35	(275.98 348.29)
Number of Emigrants	83.81	101.34	(89.01 115.83)	101.09	(90.08 111.50)

Table: Italy: 2030 and 2065 Total Population: Istat Scenarios and Stochastic Forecasts with 85 % Forecast intervals (in millions).

	2010	2030		2065	
Istat Scenarios	60.626	63.482	(61.675 65.205)	61.305	(53.390 69.125)
Stochastic Forecast (BGM 2012)	60.626	62.890	(59.430 60.772)	56.852	(48.572 65.959)
Stochastic Forecast (mixture model)	60.626	61.651	(59.828 63.478)	55.876	(48.504 63.342)

Further work

- Questionnaire.
- Interpolation.
- Age-schedules.
- Incorporation of uncertainty of starting population.