

Stochastic population forecast: a supra-Bayesian approach to combine experts opinions and observed past forecast errors

Rebecca Graziani and Eugenio Melilli¹

Abstract

We suggest a method for developing stochastic population forecasts based on a combination of experts opinions and observed past forecast errors. The supra-Bayesian approach (Lindley, 1983) is used. The experts opinions on a given vital are treated by the researcher as observed data. The researcher specifies a likelihood, as the distribution of the experts evaluations given the rate value, parametrized on the basis of the observed past forecast errors of the experts and expresses his prior opinion on the vital rate by assigning a prior distribution to it. Therefore a posterior distribution for the rate can be obtained, in which the researchers prior opinion is updated on the ground of the evaluations expressed by the experts. Such posterior distribution is used to describe the future probabilistic behavior of the vital rates so to derive probabilistic population forecasts in the framework of the traditional cohort component model.

1 Introduction

Population forecasts are strongly requested both by public and private institutions, as main ingredients for long-range planning. Traditionally official national and international agencies derive population projections in a deterministic way: in general three deterministic scenarios are specified, low, medium and high scenarios, based on combinations of assumptions on vital rates and separate forecasts are derived by applying the cohort-component method. In this way, uncertainty is not incorporated so that the expected accuracy of the forecasts cannot be assessed: prediction intervals for any population size or index of interest cannot be computed. Yet the high-low scenario interval is generally portrayed as containing likely future population sizes. 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 builds on time series models and it derives stochastic population forecasts by estimating the parameters of population dynamics (e.g., vital rates) on the basis of past data. The second approach is based on the extrapolation of empirical errors, with observed errors from historical forecasts used in the assessment of uncertainty

¹Rebecca Graziani, Department of Decision Sciences and Carlo F. Dondena Center for Research on Social Dynamics, Bocconi University, Milan, Italy; email rebecca.graziani@unibocconi.it. Eugenio Melilli, Department of Decision Sciences Bocconi University, Milan, Italy; email eugenio.melilli@unibocconi.it.

in forecasts (e.g., Stoto, 1983). In particular Alho and Spencer (1997) proposed in this framework the so-called Scaled Model of Error, which was used for deriving stochastic population forecasts within the Uncertainty Population of Europe, (UPE) project. Roughly speaking the Scaled Model of Error defines a generic age-specific rate (in logarithm scale) as the sum of a point forecast and a gaussian error term, with variance and correlation across age and time estimated on the basis of the past forecast errors. Finally, the third approach referred to as random scenario defines the probabilistic distribution of each vital rate on the basis of expert opinions. In Lutz et al. (1998), the forecast of a vital rate 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 in the forecasting interval $[0, T]$ the vital rate forecast is obtained by interpolation from the starting known and final random rate. In Billari et al. (2010) the full probability distribution of forecasts is specified by expert opinions on future developments, elicited conditional on the realization of high, central, low scenarios, in such a way to allow for not perfect correlation across time.

The method, we suggest in this paper, draws on both the extrapolation of empirical errors and the random scenario approaches. The forecasts are based on a combination of expert opinions and past forecast errors. Such combination is the result of a formal approach known as supra-Bayesian introduced by Lindley in 1983. The experts opinions are treated as data to be used for updating a prior opinion on the distribution law of each vital rate. The derived posterior is then used as the rate forecast distribution. The method is described in detail in the next section.

2 The Proposal.

Let R be the quantity to be projected over the time period $[0, T]$; R can be an overall rate such as the total fertility rate or the male or female life expectancy, but it can also be an absolute quantity such as the sex-specific net migration flow. For convenience, we will refer to R as a rate. Denote by R_0 the (known) value of R at the starting time $t = 0$. In order to define the random process $\{R_t, t \in [0, T]\}$, i.e. to determine the joint distribution of all values of R in the interval $[0, T]$, we proceed as follows:

- we define the distribution of the random variable R_T ;
- by linear interpolation of R_0 and R_T , we define the whole distribution of the process.

To perform the first step, we resort to the supra-Bayesian approach, introduced by Lindley in 1983 and used, among others, by Winkler (1981) and Gelfand et al (1995) to model and combine experts opinions; Roback and Givens (2001) apply it in the framework of deterministic simulation models. Let x_1, \dots, x_k , be the evaluations on the rate R_T provided by k experts, usually expressed in terms of central scenarios.

The idea is to treat such evaluations as data provided by the experts. As in any inferential problem, the analyst is, then, asked to specify the joint distribution $f(x_1, \dots, x_k | \theta)$ indexed by a vector of parameters θ , of which one component is the rate R_T . Moreover in a Bayesian approach the analyst assigns a prior distribution to θ expressing his prior beliefs and knowledge on it. The Bayes theorem makes it possible to derive the posterior distribution of $\pi(\theta | x_1, \dots, x_k)$ and the marginal posterior distribution law of R_T can then be used as the required distribution of the forecasts of the rate R at time T .

In the following, for the sake of simplicity we consider $k = 2$ experts providing therefore two central scenarios for the rate at time T , x_1, x_2 . We assume that the observed sample vector (x_1, x_2) is the realization of a bivariate normal distribution having vector mean $\mu^T = (R_T, R_T)$ and covariance matrix $\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$. With such specification of the mean vector, the analyst states that he expects the experts to be unbiased in their evaluations, excluding a systematic underestimation or overestimation. A prior distribution has then to be assigned to R_T and Σ . It is reasonable to assume that R_T and Σ are independent and to choose a flat prior for R_T . Indeed if the analyst has prior information on the future value of the rate of interest, he can convey it into the analysis by considering such information as provided by an additional expert. As for Σ , if series of past forecasts of the experts along with the past observed values of the rate are available, then the analyst might decide to set the variances σ_1^2 and σ_2^2 equal to the observed mean squared errors of the forecasts of the first and the second expert respectively (s_1^2 and s_2^2); the correlation can be fixed at the correlation r of the forecast errors of the two experts. This is equivalent to assume a prior distribution degenerate on s_1^2, s_2^2 and r . Of course the same values can result from the estimation of more complex models fitted to the series of past forecast errors (as in Scaled Model of Error). Under such assignment of prior distribution, the posterior distribution $\pi(R_T | x_1, x_2)$ of R_T is derived:

$$\pi(R_T | x_1, x_2) = \mathcal{N} \left(x_1 \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} + x_2 \frac{\sigma_1^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}, \frac{\sigma_1^2\sigma_2^2(1 - \rho^2)}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \right).$$

The posterior mean of the rate is a linear combination of the experts' evaluations. The weights of the combination are functions of the covariance matrix Σ . The mixture gives more weight to the value given by the expert who has shown more accuracy in its past evaluations, that is a smaller mean squared forecast error. The combination is not convex, that is one of the weights can be negative and therefore the other greater than one. If for instance $\sigma_1^2 < \sigma_2^2$, then the weight associated to x_2 is negative, whenever $\rho > \frac{\sigma_1}{\sigma_2}$. If the correlation between the forecast errors of the two experts is positive and high and the forecasts of expert 2 have shown a greater variability than the forecasts of expert 1, then the mixture will give a negative weight to the evaluation of the expert who has shown to be less accurate (and greater than one to the other evaluation). Indeed in presence of a high positive correlation between the past forecast

errors, the errors of the two experts evaluations are expected to have the same sign (both experts are expected to underestimate or both to overestimate the future value of the rate) and the difference between the evaluations, that is $x_1 - x_2$, informs on the sign of the error. If negative, both experts are expected to overestimate the value rate and the linear combination shrinks the evaluation of expert 1 towards the future unknown value of R_T . The distribution law of R_T in this way is derived from both the experts evaluations and the observed forecast errors of the same experts, so that the suggested method can be seen as an attempt to combine the extrapolation of errors method and the random scenario approach.

Other assignments of prior distributions on Σ are of course possible. An inverted Wishart distribution, centered at Σ_0 with δ_0 degrees of freedom can be used, where Σ_0 can be defined on the basis of the past forecast errors, if available, or on information from any source. In this case the posterior distribution of R_T turns out to be a Student-t distribution, with mean having the same expression as in the previous case, with σ_2^1, σ_2^2 and ρ being the corresponding elements of Σ_0 .

3 An application

In this Section we apply the method previously described to obtain the forecast distribution at 2030 of the following summary indicators: Total Fertility Rate, Male and Female Life Expectancies at Birth. Indeed, age-specific fertility and mortality rates can be derived by resorting to specific models and then used as inputs for the cohort-component method of population forecast.

We consider two experts: the Italian National Statistical Office (ISTAT) and an expert whose forecasts are based on the so-called naive forecast method (see Keifitz, 1981). As for the second expert, forecasts over a time span of 20 years are obtained as follows: the Total Fertility Rate is held constant at the value observed at the beginning of the forecast period, while forecasts of the Male and Female Life Expectancies are derived by linear extrapolation, on the basis of series of past data. The 2030 forecasts of the summary indicators are derived on the same way. As for ISTAT, we use as forecast of each summary indicator, the central scenario at 2030 of the latest projections released by the Statistical Institute with starting year 2010. Moreover the responsible of the office uncharged of the Italian population projections provided us with series of past forecast errors. Table 1 shows for each indicator, the forecasts, x_1 and x_2 of the two experts along with their past forecast mean squared errors and the correlation between their past forecast errors to be used as σ_1^2, σ_2^2 and ρ respectively.

The mean and variance of the 2030 forecast distribution of each summary indicator is shown in Table 2. As we can observe from Table 1, for the Total Fertility Rate, σ_1^2 , the mean squared forecast error of ISTAT, is smaller than σ_2^2 , the mean squared forecast error of the naive expert. Moreover the ratio $\frac{\sigma_1}{\sigma_2}$ is smaller than ρ . Therefore, in the determination of the posterior mean of Total Fertility Rate, ISTAT

forecast has a weight greater than 1 (1.095) and the naive forecast a weight smaller than 0 (-0.095), this leading to a posterior mean (as shown in Table 2) equal to 1.58. For both Male and Female Life Expectancies at Birth, the posterior mean arises as a convex combination of the evaluations of the two experts. In all three cases the posterior variances are smaller than the mean squared errors of the past forecasts of the two experts.

Table 1: Evaluations for indicators, past forecast mean squared errors and correlation of the errors. TFR is the Total Fertility Rate, EM and EF are respectively Male and Female Life Expectancy at Birth.

Indicator	Istat		Naive expert		
	x_1	σ_1^2	x_2	σ_2^2	ρ
TFR	1.57	0.010	1.42	0.276	0.50
EM	82.20	0.480	83.02	0.248	0.65
EF	87.50	0.545	87.35	0.253	0.66

Table 2: Posterior mean and posterior variance of the 2030 forecast distribution of the Total Fertility Rate, TFR , the Male and Female Life Expectancy at Birth, EM and EF respectively.

Indicator	posterior mean	posterior variance
TFR	1.58	0.009
EM	82.96	0.239
EF	87.35	0.252

References

- Alho, J.M. and Spencer, B.D. (1997) “The practical specification of the expected error of population forecasts”, *Journal of Official Statistics*, **13**, 204–225.
- Billari, F.C., Graziani R. and Melilli, E. (2010) “Stochastic population forecasts based on conditional expert opinions”, Working Paper 33. Milan: Carlo F. Dondena Centre for Research on Social Dynamics, Bocconi University.
- Gelfand, A. E., Mallick, B. K. and Dey, D. K. (1995) “Modeling Expert Opinion Arising as Partial Probabilistic Specification”, *Journal of the American Statistical Association*, **90**, 430, 598–604.
- Keilman, N., Pham, D.Q. and Hetland, A. (2002) “Why population forecasts should be probabilistic - illustrated by the case of Norway”, *Demographic Research*, **6**, 15, 409–454.

- Keyfitz, N. (1981) “The limits of Population Forecasting”, *Population and Development Review*, **7**, 4, 579–593.
- Lindley, D. (1983) “Reconciliation of Probability Distributions”, *Operations Research*, **31**, 5, 866–880.
- Lutz, W., Sanderson, W.C. and Scherbov, S. (1998) “Expert-Based Probabilistic Population Projections”, *Population and Development Review*, **24**, 139–155.
- Roback, P.J and 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.