Stochastic Population Forecasting
Based on Combinations of Expert Evaluations
Within the Bayesian Paradigm

**Francesco C. Billari · Rebecca Graziani ·
Eugenio Melilli**

Abstract This article suggests a procedure to derive stochastic population forecasts adopting an expert-based approach. As in previous work by Billari et al. (2012), experts are required to provide evaluations, in the form of conditional and unconditional scenarios, on summary indicators of the demographic components determining the population evolution: that is, fertility, mortality, and migration. Here, two main purposes are pursued. First, the demographic components are allowed to have some kind of dependence. Second, as a result of the existence of a body of shared information, possible correlations among experts are taken into account. In both cases, the dependence structure is not imposed by the researcher but rather is indirectly derived through the scenarios elicited from the experts. To address these issues, the method is based on a mixture model, within the so-called Supra-Bayesian approach, according to which expert evaluations are treated as data. The derived posterior distribution for the demographic indicators of interest is used as forecasting distribution, and a Markov chain Monte Carlo algorithm is designed to approximate this posterior. This article provides the questionnaire designed by the authors to collect expert opinions. Finally, an application to the forecast of the Italian population from 2010 to 2065 is proposed.

F. C. Billari
Department of Sociology and Nuffield College, University of Oxford, Oxford, UK
e-mail: francesco.billari@nuffield.ox.ac.uk

R. Graziani (✉)
Department of Policy Analysis and Public Management and Carlo F. Dondena Center for Research on
Social Dynamics, Bocconi University, Milan, Italy
e-mail: rebecca.graziani@unibocconi.it

E. Melilli
Department of Decision Sciences, Bocconi University, Milan, Italy
e-mail: eugenio.melilli@unibocconi.it

Introduction

Population forecasts are in strong demand by both public and private institutions as crucial ingredients for long-range planning. Official national and international agencies have traditionally derived population projections in a deterministic way, specifying multiple scenarios (almost always three) based on combinations of assumptions on the future behavior of demographic components. These scenarios—usually referred to as low, medium, and high scenarios—yield separate forecasts of the population by age and sex after the cohort-component method is applied. Indeed, virtually all population forecasts are based on this method so that the forecast of the population reduces to the forecasts of the three main components of the demographic change: fertility, mortality, and migration. When a deterministic approach is adopted, uncertainty is not explicitly incorporated in the model, and thus 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 the likely future values of population levels and/or other demographic indicators.

More recently, stochastic (or probabilistic) population forecasting has received growing attention by researchers and, to a lesser extent, by forecasting agencies (see Land 1986). Three main approaches to stochastic population forecasting can be roughly distinguished in the literature (Keilman et al. 2002). The first approach adopts standard procedures in time series analysis: for each indicator, a model is fitted to past series of data, and forecasts are obtained by extrapolation. Forecasting models are developed both within the classic (frequentist) and the Bayesian approach. The pioneering demographic forecasting model that uses times series analysis in a classical framework is the one attributed to Lee and Carter (1992), which was originally proposed to forecast mortality and was later modified to address fertility forecasting (see Lee 1993; Lee and Tuljapurkar 1994). Several extensions, generalizations, and modifications of Lee-Carter approach have been proposed more recently (Booth et al. 2002; Cairns et al. 2006, 2011; Hyndman and Booth 2008; Hyndman and Ullah 2007; Hyndman et al. 2013). Booth and Tickle (2008) reviewed methods based on generalizations of the Lee-Carter procedure. Some of these generalizations are based on ideas from functional data analysis. Within this approach, Hyndman and Ullah (2007) provided robust forecasts of age-specific mortality and fertility rates, while Hyndman et al. (2013) discussed coherent mortality forecasts for more subpopulations, using functional principal components models of simple functions of the rates. Booth et al. (2006) and Shang et al. (2011) discussed interesting comparisons and evaluations of some of these methods for mortality forecasting. Also, Bayesian hierarchical time series models have been proposed to derive fertility (Alkema et al. 2011) and mortality (Raftery et al. 2013) forecasts. As a sign that these approaches are entering the mainstream of demographic forecasting, the 2012 Revision of United Nations World Population Prospects (United Nations, Department

of Economic and Social Affairs, Population Division 2013) complements traditional deterministic scenarios with stochastic forecasts that are obtained by applying Bayesian time series models (see also Raftery et al. 2012, 2013, 2014).

The second main way to address stochastic population forecasting 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 (Alders et al. 2007).

Here, we follow the third approach, also known as "random scenario." In this approach, the forecast distribution of demographic components is derived by starting from suitably elicited expert opinions. Usually, a summary indicator for each of the three basic components of the population evolution is considered during the elicitation stage. Indicators have to be particularly meaningful to experts: that is, they have to be related to the measures they use to describe and study a demographic phenomenon. 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 a 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 through interpolation between the starting known and the final random values. In Billari et al. (2012), the probability distribution of each indicator over the interval $[t_0, T]$ is specified using expert opinions at time T on the indicator, conditional on its value at an intermediate point. More precisely, after having fixed a point t_1 in (t_0, T), the expert is required to elicit the mean and a specific upper quantile for each indicator at time t_1; in a second step, the mean and a specific upper quantile of the same indicator at time T, given its value at time t_1. Assuming that the pair of values of the indicator at times t_1 and T follows a bivariate normal distribution and interpolating the intermediate values, the whole distribution of each indicator over the interval $[t_0, T]$ is determined. In this way, correlations across time for each indicator are obtained indirectly through marginal and conditional evaluations. Finally, the joint probability distribution for all indicators is obtained assuming independence among indicators. In this approach, one can rely for elicitations on a single expert or on several experts; in the latter case, inputs for the projections can be obtained by simply averaging all the elicitations.

In this article, we suggest a new method, through which we derive expert-based stochastic population forecasts that explicitly take into account the relationships both between demographic components and between experts. We build on the conditional expert-based approach of Billari et al. (2012). Our contribution addresses two main issues.

First, we define for each expert a joint distribution for the demographic indicators that allows for both across-time correlation for each indicator and for the dependence between different indicators at the same time and across time. Whether it is essential to allow for dependence between different demographic components is questionable. In many contexts, it might be better to ignore possible weak correlations than to try to estimate them (as in the case of multivariate time series forecasting compared with independent univariate time series). For these reasons, it is common practice in population forecasting to assume independence among fertility, mortality, and migration. Nevertheless, we believe that the most sensible approach is to be neutral in some

respect—that is, to allow for dependence without imposing it. In our proposal, the presence, degree, and type of relationship among demographic components stem from the evaluations of the experts on their joint development through time. The independence among demographic components is therefore not excluded because it might be the result of expert opinions. As we will see in the application of our method to the forecast of the Italian population, correlations among demographic components are indeed quite low.

Second, we propose 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 (albeit dated) review on this topic, see Genest and Zidek (1986). Generally speaking, pooling expert opinions means to merge many individuals' probability distributions on unknown objects into a single collective assignment; in our context, these objects are the demographic indicators of interest. Classical pooling methods proceed by averaging expert opinions; for instance, the linear rule derives the collective assignment through the (possibly weighted) average. Similarly, one can define geometric or logarithmic pooling rules. McConway (1981) discussed some theoretical features and properties of this class of combination methods. As observed by Genest and Zidek (1986) and Dietrich (2010), pooling methods based on averaging are affected by the lack of a normative basis for choosing weights: generally, no specific indications can be given concerning the choice and interpretation of weights. If the same experts have provided forecasts before and the accuracy of those forecasts can be evaluated against actual values, then the forecast variance associated with each expert can be computed. Hence, in this case, it is possible to weight experts' evaluations through their inverse forecast variances. This could be a feasible strategy in the case of past forecasts released by official agencies—considering a forecasting agency as a single expert. An alternative approach could be to weight experts on the basis of evaluations of their actual expertise in the field, measured by means of their publication rate or in the production of projections within official agencies.

An expert's opinion is the expression of the person's beliefs. Nevertheless, it can be seen as based on the experience that the individual has had with the issue at hand. Experts who have undergone similar training and who have shared the same knowledge environment will typically share quite a large amount of information. The actual opinions of experts may therefore differ for two main reasons. First, experts may not share exactly the same information. Second, they may not interpret shared information in the same way. Linear and classical pooling methods do not explicitly consider this difference. In this regard, Dietrich (2010) distinguished the information aggregation problem from the opinion pooling one; he suggested a method for combining opinions within a Bayesian approach using all information spread asymmetrically over the individuals. In this article, we address the same issue, but we make use of a different approach: the Supra-Bayesian method of pooling, introduced by Morris (1974) and then developed by others, such as French (1980, 1981), Winkler (1981), Lindley (1983, 1985), and Gelfand et al. (1995). Roback and Givens (2001) applied the Supra-Bayesian approach

to deterministic simulation models. By assuming that expert opinions are data, the Supra-Bayesian approach makes it possible to combine expert opinions on unknown quantities within the formal framework provided by Bayesian inference. The analyst is therefore asked to specify a likelihood function, to be parameterized in terms of the unknown objects, and a prior distribution for the parameters. The posterior distribution, obtained by applying Bayes' theorem, updates the analyst's prior opinion on the basis of the evaluations provided by the experts and can then be used as a collective distribution for the unknown quantities of interest. This approach takes into account and exploits the variability of the expert evaluations. In this sense, the higher the number of experts, the more informative the procedure is.

Through the choice of the likelihood function, the Supra-Bayesian approach makes it possible to model different kinds of dependence structure. A standard choice (see, e.g., Lindley 1983, 1985) is a multivariate normal distribution. When opinions are elicited on several objects and at different time points (as in our case), the choice of a multivariate normal distribution requires the specification of a large number of parameters: marginal means and variances, and correlations between objects and between experts. Albert et al. (2012) suggested a hierarchical random-effects model as a more parsimonious approach, in the sense of the number of parameters to be specified, especially when the number of experts is large. Here, we suggest implicitly deriving the dependence structure of the expert evaluations by using a mixture model. We assume that experts can be grouped into a given number of clusters according to information they share. In its simplest version, the model assumes that the number J of clusters is fixed by the analyst, but a quite straightforward generalization can treat J as a random parameter. In any case, we let expert evaluations determine 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 centers of the clusters' distributions are then assumed to be independently generated from the same distribution, centered on the unknown vector of future values of the indicators. In this way, we can account for the heterogeneity of the expert evaluations that results from their holding different pieces of information. At the same time, we achieve the goal of allowing for dependence between experts without explicitly imposing it.

A fundamental step in any expert-based forecasting procedure is the actual elicitation of expert opinions. To this aim, we designed a questionnaire in collaboration with researchers of the Italian National Statistical Institute (ISTAT), an official forecasting agency; we then submitted the questionnaire online to a panel of Italian demographers. The questionnaire is an enhanced version of the elicitation procedure introduced by Billari et al. (2012).

The Proposed Approach

Following common practice in demographic forecasting, we derive the forecast of the population by age and sex through a cohort-component model (Cannan 1895). The

inputs are age schedules of fertility and mortality, the distribution of migrants by age, and a starting population by age and sex. Age schedules and distributions are, in turn, derived by a series of summary indicators. In what follows, we view summary indicators as the lens through which experts derive information on population. From among the many available options, we use the following indicators: total fertility rate (complemented by a timing indicator), male and female life expectancies at birth, and male and female number of immigrants and emigrants.

As anticipated, our method derives the joint forecast distribution of all summary indicators on the basis of the evaluations provided by several experts, and it allows for both dependence between indicators (across time for a single indicator, and at the same time and across time between any two indicators) and dependence between expert opinions. Nevertheless, to reduce the dimensionality of the problem, we make some assumptions on the dependence structure across components. We consider two pairs of indicators, one made up by the total fertility rate and the total number of immigrants and the second made up by the male and female life expectancies at birth; this way, the possible association within each of the two pairs is taken in account. Conversely, the two pairs are assumed mutually independent, and both pairs are independent on the total number of emigrants. In fact, dependence is quite obvious for some pairs of demographic components—for instance, for male and female life expectancies at birth. In other cases, dependence is more questionable. We allow for a correlation between fertility and immigration because the literature shows that with fertility decline, the rising importance of migrants for childbearing becomes an important issue (Sobotka 2008). In this context, the assimilation of immigrants to the fertility patterns of their countries of destination (see, e.g., Parrado 2011; Parrado and Morgan 2008) will lead to very weak correlations. Recent evidence has documented the existence of lagged correlations between migration and population reproduction in contemporary world population trends (Billari and Dalla Zuanna 2013). Finally, the total numbers of immigrants and emigrants are then split by sex on the basis of a deterministic rule. Age schedules for mortality, fertility, and migration are then derived by means of standard nonstochastic models (which we discuss in more detail later).

Expert evaluations are elicited according to the conditional procedure suggested by Billari et al. (2012). For all indicators, experts are asked to provide both marginal and conditional, central and high, scenarios. In this way, information on the means, the variability, and the across-time and across-indicators correlations can be derived.

We describe how our method works referring to two generic indicators, R_1 and R_2, to be forecasted over the interval $[t_0, T]$. We split the forecast interval into two subintervals by considering an inner point t_1. We begin by deriving the distribution of the vector $\mathbf{R} = (R_{11}, R_{12}, R_{21}, R_{22})$, where R_{ij} is the random variable associated with the value of R_i at time t_j; $i = 1, 2$ and $j = 1, 2$, where $t_2 = T$. The joint distribution of the two indicators over the entire forecast interval can then be obtained resorting to interpolation techniques. Several methods, from the simplest linear interpolation to more complex methods based on splines or on wavelets, can be chosen on the basis of both computational issues and/or demographic considerations. The selected method clearly may affect the final forecasts. A minimum requirement, here, is that the variance of each rate increases with time. This can be achieved by choosing a linear interpolation even if it is known that such choice could underestimate the uncertainty at intermediate points.

Consider K experts, and denote by $\mathbf{x}_i$ 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 earlier, we use the Supra-Bayesian approach to derive the joint forecast distribution of the indicators. Accordingly, $x_1, x_2, \ldots, x_K$ are treated as data. Under a Bayesian framework, the analyst has to specify the likelihood $f(x_1, \ldots, x_K \mid R_{11}, R_{12}, R_{21}, R_{22})$—that is, the joint distribution of the scenarios parameterized by the unknown future values of the indicators, and a prior on such parameters. For the likelihood, following Lindley (1983), a possible choice is a Gaussian distribution centered on $\mathbf{R}$, so as to assume that experts are not systematically biased. Indeed, the choice of the Gaussian distribution can be primarily motivated by mathematical convenience because under such an assumption, the computation of the posterior distribution of the vector of the indicators is greatly simplified. Nevertheless, the construction of a likelihood of this kind is cumbersome because of the number of terms of the covariance matrix to be specified. The covariance matrix is a square matrix of dimension $4K$, made up by the covariance matrices of the evaluations on the components of $\mathbf{R}$ for each expert and by the covariances between the evaluations provided by any two experts on any two components of $\mathbf{R}$. The elements of the first set of covariances (the covariance matrices for each expert) can be specified on the basis of the elicitations obtained through a questionnaire. Simplifying assumptions can be introduced to reduce the number of the remaining entries. However, the square matrix obtained in this way needs to satisfy the constraints of a covariance matrix. Moreover, with this choice of the likelihood, the analyst is asked to fix the correlations between the evaluations of any two experts—correlations that are one of the foci of the current work.

To overcome the difficulties implied by a Gaussian distribution, we suggest using a mixture model. As we shall show, in this way, we achieve a greater parsimony—and, at the same time, we let the expert evaluations shape the structure of the dependence among future values of the demographic components, thus reducing the number of specifications by the analyst. The proposed model assumes the existence of J groups of experts. The number J is fixed by the analyst. However, the model can be modified by taking a random J—that is, with the number of groups endowed with a prior distribution. The mixture turns out to be a useful and efficient technical device that allows us to reduce the dimensionality of the problem while accounting for dependence of expert evaluations. For each expert, group membership is determined on the basis of the elicited scenarios. Indeed, the assignment of each expert to a specific group might be rather difficult and somewhat arbitrary. This is why we prefer a mixture model to a hierarchical random-effect model, as used by Albert et al. (2012). Within each group j, the members' opinions are sampled from a multivariate Gaussian distribution, centered on 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 a multivariate Gaussian distribution centered on the vector $\mathbf{R}$ of the unknown future values of the indicators. In this way, we explicitly account for the 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. For each j, the covariance matrix of the vector μ_j is set equal to Σ_j divided by a fixed constant k_0. The matrices $\Sigma_1, \ldots, \Sigma_J$ are then sampled from an inverse-Wishart distribution with center Σ_0 and n_0 degrees of freedom. Finally, the weights $p_1, \ldots, p_J$ of the mixture are distributed according to a Dirichlet distribution with parameters $\alpha_1, \ldots, \alpha_J$.

Schematically, the hierarchical multivariate mixture model we are assuming can be described as follows:

$$x_i \big| \mu_1, \ldots, \mu_J, \Sigma_1, \ldots, \Sigma_J, p_1, \ldots, p_J \ \text{iid} \sim \sum_{j=1}^{J} p_j N_4 \left(\mu_j, \Sigma_j \right) \quad i = 1, \ldots, K$$

$$\mu_j \big| \Sigma_j \ \text{ind} \sim N_4 \left(R, \frac{1}{k_0} \Sigma_j \right) \quad j = 1, \ldots, J$$

$$\Sigma_j \ \text{iid} \sim IW \left(\Sigma_0, n_0 \right) \quad j = 1, \ldots, J$$

$$p_1, \ldots, p_J \sim Dir \left(\alpha_1, \ldots, \alpha_J \right)$$

$$R \sim N_4 \left(\mu_R, \Sigma_R \right).$$

In the preceding description, $N_q(\mu, \Sigma)$ denotes the q-variate normal distribution with a mean of μ and a covariance matrix Σ; $IW(\Sigma, n)$ is the inverse-Wishart distribution with scale matrix Σ and n degrees of freedom; and $Dir(\alpha_1, \ldots, \alpha_J)$ is the Dirichlet distribution with parameters $(\alpha_1, \ldots, \alpha_J)$. It is worth emphasizing that under the suggested hierarchical model, the analyst is asked to specify $10J + 4$ parameters; however, in the case of a multivariate normal likelihood, the parameters to be specified are $10 + (K(K-1)/2)$. In this sense, if the number of experts is large enough—that is, if it is greater than four—we have a gain in parsimony.

Notice that for computational convenience, conjugate prior distributions are assigned to the parameters involved in the model. The analyst must fix the following hyperparameters: $k_0, \Sigma_0, n_0, \alpha_1, \ldots, \alpha_J, \mu_R, \Sigma_R$. The general idea is to specify these hyperparameters to produce noninformative (i.e., highly spread) prior distributions such that the resulting posteriors are mainly determined by data (in this case, by expert evaluations). Moreover, assuming (as usually done within the Supra-Bayesian approach) that the analyst is not an expert, the center μ_R of the distribution of $\mathbf{R}$, representing the prior guess on the future behavior of the rates, can be specified relying on the projections given by some official agencies.

The number k_0 is the ratio of the within-group variance to the variance of the group means (i.e., a measure of the between groups variance). The choice of a small value for k_0 implies a prior confidence in the existence of the groups. The value of n_0 affects the spread of the prior on the groups' covariances: the smaller n_0, the higher such a spread will be. The smallest admissible value is the order of the covariance matrix Σ_j—in this case, $n_0 = 4$.

The matrix Σ_0 is the center of the prior on the covariance matrices Σ_js of the groups' distributions. A reasonable choice is to specify it on the basis of the values elicited from the experts. Indeed, for each expert, we can work out a covariance matrix based on her/his marginal and conditional scenarios (as discussed in the next section) and relying on standard results of a multivariate Gaussian distribution (see Billari et al. (2012: appendix) for detailed computations). We suggest setting Σ_0 equal to the arithmetic mean of such matrices, multiplied by a given constant in order to increase the elicited variances of the indicators and therefore to take into account the fact that experts often tend to underestimate the variability of their forecasts.

For $\alpha_1, \ldots, \alpha_J$, according to the properties of the Dirichlet distribution, the lower the value of sum of the α_j, the higher the variability will be. Moreover, for each j, α_j describes the prior probability for an expert to belong to the jth group. Hence, a standard choice representing prior ignorance can be to set $\alpha_j = 1 / J$ for each $j = 1, 2, \ldots, J$.

As for Σ_R, according to our noninformative setting, high values should be chosen for the variances. The covariances could be fixed equal to 0 in order to ensure *a priori* independence among the indicators.

The posterior distribution of the summary indicators at the future time points, derived according to the Bayesian paradigm, is then used as the forecast distribution. Such a distribution cannot be derived in closed form, but it can be approximated by means of a Gibbs sampler worked out on the basis of well-known results in the literature on mixture models (see, e.g., Lavine and West 1992).

Implementation

The implementation of our proposal requires both expert opinions as inputs and an algorithm to perform the necessary computations to derive the forecast distribution. In this section, we describe the questionnaire used to elicit the expert opinions and the algorithm developed for the implementation of the method.

Elicitation of Expert Opinions Through a Questionnaire

To elicit expert opinions, we designed a questionnaire in collaboration with researchers of the Population Division of the Italian National Statistical Institute (ISTAT), the official forecasting agency in Italy. The elicited evaluations are then used to derive the forecast of the Italian population during 2010–2065, the interval used by ISTAT in the latest release of Italian deterministic population projections. Following the procedure described in the previous section, the forecasting interval is divided into two subintervals, with 2030 as the intermediate point. Experts are asked to provide scenarios of the indicators of interest at the two time points of 2030 and 2065, with a mean scenario and a high scenario.

The questionnaire is divided in two sections. In the first section, experts are asked to discuss the drivers that might affect the development of the fertility, mortality, and migration. This information is qualitative, and it can help the analyst in the specification of the priors on the model parameters as well as set the stage for quantitative elicitations. In the second section, on which we focus here, the quantitative elicitations of experts are elicited using the conditional procedure developed in Billari et al. (2012). For the forecasting distribution of a single indicator (e.g., the total number of emigrants), each expert is asked for the mean of one indicator given the (real or hypothetical) value of the same indicator at the previous time. In the case of the joint forecast distribution of a pair of indicators, such as total fertility rate and total number of immigrants, each expert is asked for the mean of one indicator conditional on the values taken by the same indicator and the other indicator at the same time or at the previous time. In the same way, conditional quantiles are elicited, focusing on

the high scenario: experts are instructed to consider a high scenario as one with a .1 probability of being exceeded by the future actual value. After the conditional means and quantile are given, and assuming a normal distribution, it is possible to derive each expert's evaluations on both the central scenarios and on the marginal variability and on the required correlations.

Questions in the second section are divided in four sections: fertility and immigration, male and female life expectancies, emigration, and mean maternal age at birth. The evaluation on mean maternal age at birth is used in the derivation of age schedules of fertility. Therefore, the first two parts aim at eliciting opinions on a pair of indicators and have the same structure. At the beginning, the expert is asked to provide central scenarios of the two indicators at 2030 and 2065 and also a high scenario of one of them at 2030. The subsequent questions elicit conditional central and high scenarios. Typical queries are as follows: "Assuming that the total number of immigrants is 100,000 in 2030, provide a central scenario for the total fertility rate in 2030," or "Assuming that the total number of immigrants is 100,000 in 2030 and that the total fertility rate in 2013 is 1.5, provide a central scenario for the total fertility rate in 2065." Parts 3 and 4 of the second section of the questionnaire concern single indicators; thus, central and high scenarios are elicited conditioned on the values taken by the same indicator at the previous time points.

The questionnaire was sent to 30 Italian experts (professional academic demographers); 16 demographers answered the questions on mortality and emigration, and 14 answered the questions on fertility and immigration. As an illustration, Table 1

Table 1 Some results of the elicitation procedure

	Total Number of Immigrants (in thousands) and TFR				Male and Female Life Expectancies at Birth		
	Expert 1	Expert 2	Expert 3		Expert 1	Expert 2	Expert 3
$c(IM_{2030})$	250	200	350	$c(EM_{2030})$	81	81	83
$c(IM_{2065})$	250	150	500	$c(EM_{2065})$	84	83	87
$c(TFR_{2030})$	1.6	1.5	1.5	$c(EF_{2030})$	87	86	88
$c(TFR_{2065})$	1.6	1.56	1.7	$c(EF_{2065})$	90	87	89
$\sigma(IM_{2030})$	100.43	60.30	39.27	$\sigma(EM_{2030})$	0.78	3.10	1.48
$\sigma(IM_{2065})$	120.30	86.20	−87.41	$\sigma(EM_{2065})$	1.10	3.49	1.64
$\sigma(TFR_{2030})$	0.14	0.13	0.11	$\sigma(EF_{2030})$	1.10	3.13	1.68
$\sigma(TFR_{2065})$	0.41	0.16	0.23	$\sigma(EF_{2065})$	1.91	3.16	7.23
$\rho(IM_{2030}IM_{2065})$	0.97	0.75	0.45	$\rho(EM_{2030}EM_{2065})$	0.71	0.89	0.45
$\rho(TFR_{2030}\,TFR_{2065})$	0.53	0.03	−0.53	$\rho(EF_{2030}EF_{2065})$	0.29	0.95	0.04
$\rho(IM_{2030}TFR_{2030})$	0.55	0.03	0.32	$\rho(EM_{2030}EF_{2030})$	0.72	0.98	0.92
$\rho(IM_{2065}TFR_{2065})$	0.03	0.12	0.22	$\rho(EM_{2065}EF_{2065})$	0.57	0.66	0.99
$\rho(IM_{2030}TFR_{2065})$	0.12	0.05	−0.15	$\rho(EM_{2030}EF_{2065})$	0.42	0.95	0.05
$\rho(IM_{2065}TFR_{2030})$	0.07	0.17	−0.76	$\rho(EM_{2065}EF_{2030})$	0.51	0.89	0.04

displays, for selected experts, central scenarios (in the table denoted by $c(\cdot)$) along with standard deviations and correlations of selected indicators, as derived from the answers of the questionnaire.

Figures 1 and 2 depict the expert central scenarios for immigration and total fertility rate, and for male and female life expectancies at birth, respectively. As shown in Table 1 and Figs. 1 and 2, expert evaluations are noticeably variable. The method that we suggest actually embodies such variability in a formal framework.

The Algorithm

Here we describe the algorithm designed for the procedure, which has been implemented in a computer program using MATLAB, to obtain stochastic population forecasts using our method through simulation. The male and female populations are separately forecasted and grouped into age intervals. The forecasting period is divided into time intervals that have the same length as the age intervals. For each forecasting interval, the algorithm follows the cohort-component method: that is, (1) starting from an initial population, the population for each age group is projected forward by applying age- and sex-specific survival rates derived from life expectancy; (2) the number of births for each subgroup over the time interval and the number of births still alive at the beginning of the next interval are computed using fertility and mortality rates; and (3) net migration flows are added, and the number of immigrants and births from immigrants still alive at the beginning of the next time interval are projected forward.

The algorithm requires the following inputs:

- The starting population by age and sex
- The length of the time intervals dividing the forecasting period (equal to the length of the age intervals)
- The (conditional) scenarios provided by the experts on all summary indicators
- The values of the prior parameters
- The number of groups assumed in the mixture model
- The number N of samples to draw for each indicator along with the number of draws to discard so to ensure the convergence of the Markov chain Monte Carlo (MCMC) procedure
- The confidence level $1 - \gamma$ of the predictive intervals

The algorithm works through four steps. At the first step, means, standard deviations and correlations (across time and between indicators) of the evaluations of each expert are worked out.

At the second step, N samples drawn for each indicator by means of a Gibbs sampler are stored. On the basis of standard results on Bayesian finite mixture models, because of the choice of the priors, the full conditionals are worked out in a closed form. For each indicator, the values at each time point of the forecasting interval are obtained by choosing suitable interpolation techniques. Samples of net migration flows are computed and split by sex. Therefore, at this step, we obtain for each summary indicator N draws from the joint distribution of the values at each of the points in the forecasting interval.
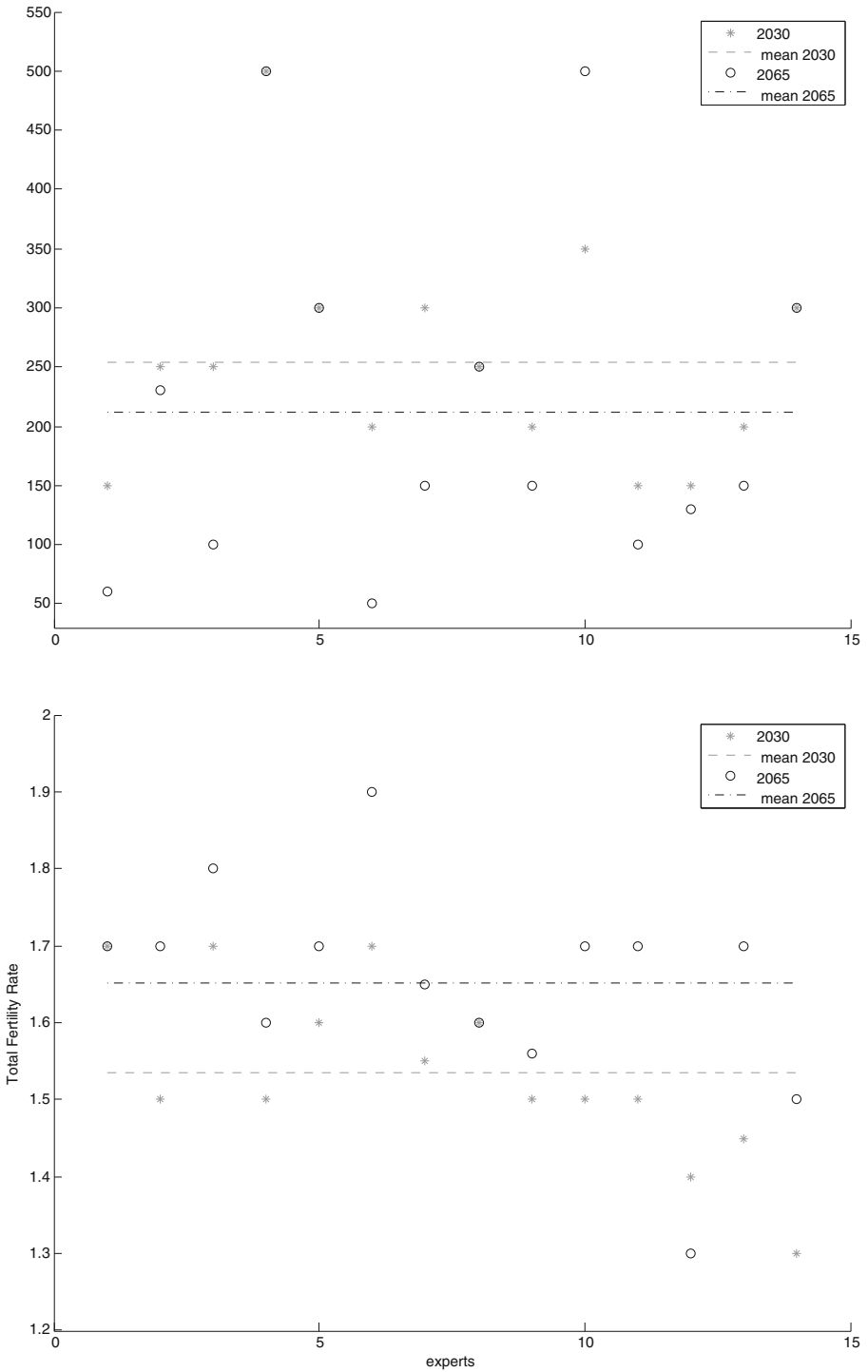
Fig. 1 Experts' central scenarios on number of immigrants (in thousands) and total fertility rate at 2030 and 2065
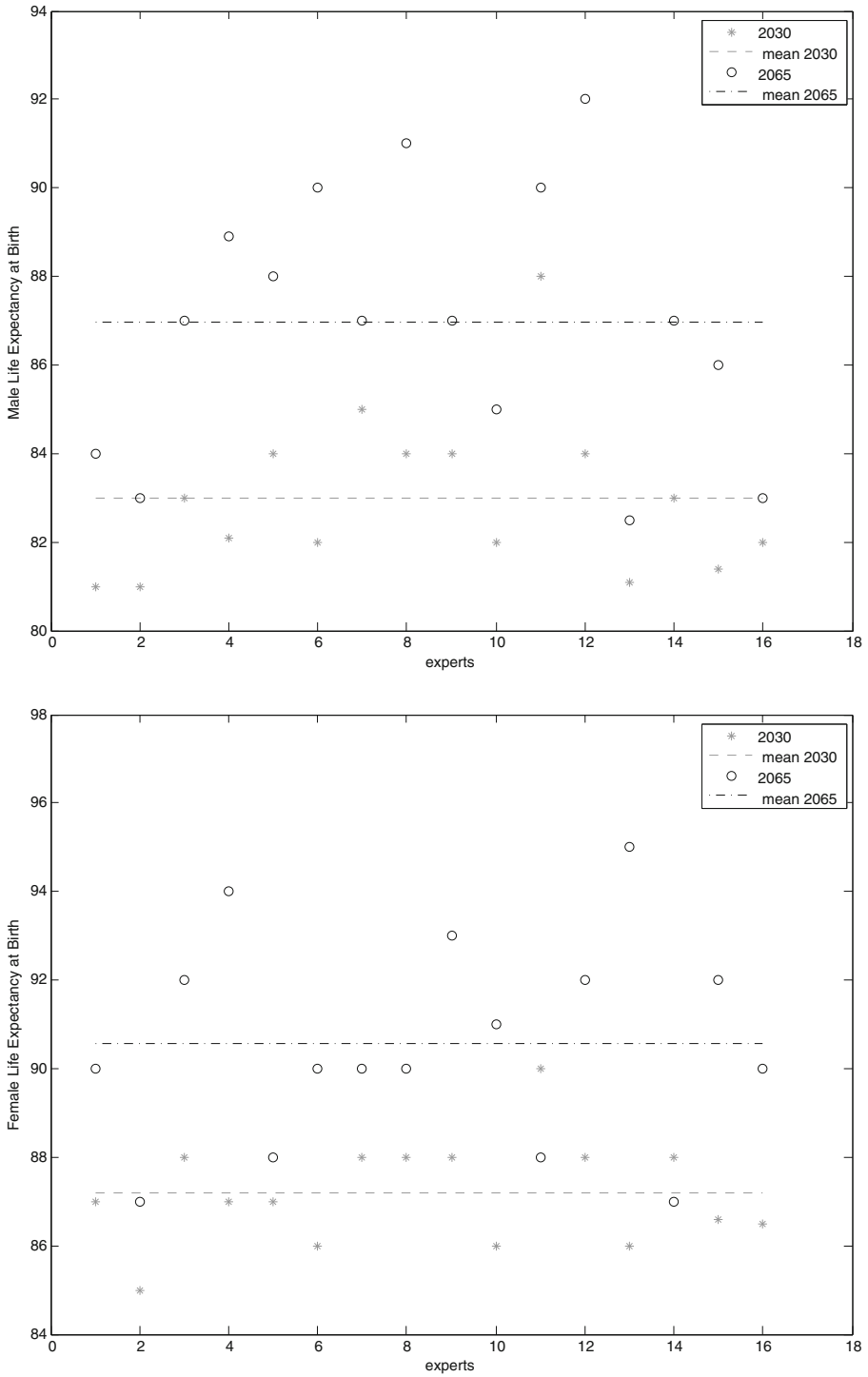
Fig. 2 Experts' central scenarios on male and female life expectancies at birth at 2030 and 2065

At the third step, the whole set of age- and time-specific quantities is derived by starting from the summary indicators. In this implementation, we choose extremely simple ways to obtain smooth age-specific quantities. In particular, N matrices of male and N matrices of female age- and time-specific mortality rates are obtained from the corresponding life expectancies at birth on the basis of the extended model life tables provided by the United Nations. N matrices of age- and time-specific fertility rates are derived from the vectors of total fertility rates and the vectors of mean maternal ages at birth, using a rescaled normal model, as in Billari et al. (2012). For migration, N matrices of male and N matrices of female age-specific net migration flows are derived from the corresponding vectors of total net flows, applying a rescaled gamma model, as in Billari et al. (2012). This is a simplifying assumption that assumes the absence of pre-school, retirement, and post-retirement peaks in the age profile of migrations, with the only peak being related to labor migration. More general models of migration for the derivation of the age schedules could also be implemented.

At the fourth and last step, the matrices of age-specific mortality rates by sex and the matrices of age-specific fertility rates along with the matrices of the migration flows by age and sex are used as inputs of the cohort-component method to derive N matrices of the total population by age and time for each sex. Such matrices can be considered as draws from the corresponding distribution of the male and female populations so that forecasts and prediction intervals can be derived from them. For instance, the forecast of the female population size of a fixed age and for a fixed forecasting time is given by the arithmetic mean of the N corresponding draws of the female population size for the considered age and forecasting time. An estimate of the $(1 - \gamma)\%$ prediction interval is given by the interval determined by the $(\gamma / 2)N$th and $(1 - \gamma / 2)N$th sample quantiles. Similarly, stochastic forecasts for summary indicators related to the age structure of the population (e.g., dependency ratios) can be obtained.

An Application: Bayesian Expert-Based Stochastic Forecasts of the Italian Population

We now illustrate the results of the stochastic forecast of the Italian population from 2010 to 2065, obtained by applying our proposed method. The actual results are based on $N = 10,000$ samples of 20,000 drawn from the joint distribution of the demographic indicators at the two time points of 2030 and 2065. We assessed the convergence of the MCMC procedure by visual inspection and, more formally, by Raftery and Lewis's diagnostic tests and other tests implemented in the R package *coda* (Plummer et al. 2006; Raftery and Lewis 1992). The prior parameters are specified as follows. The prior mean μ_R of the vector of the indicators is set equal to the vector of the central scenarios provided by ISTAT. For Σ_R, the covariances are all fixed equal to 0, while the prior variances are specified by inflating the values derived from high/low ISTAT scenarios. We chose this strategy to specify a diffuse prior. Table 2 shows the prior means and standard deviations for each indicator. As discussed earlier, Σ_0 is given by the arithmetic mean, computed over all experts, of the covariance matrices of their elicitations, multiplied by a suitable positive constant in order to inflate it. The spread parameters k_0 and n_0 are chosen to ensure a high variability in the prior on the groups' means

Table 2 Prior means and standard deviations and posterior forecasts with 85 % forecast intervals for each indicator, number of immigrants and of emigrants (in thousands)

Indicator	Prior Mean		Prior SD		Forecasts (85 % forecast intervals)				
	2030	2065	2030	2065	2010	2030		2065	
Total Fertility Rate	1.5	1.5	0.5	0.5	1.42	1.53	(1.36 1.70)	1.63	(1.36 1.89)
Mean Maternal Age at Birth	31.8	32.0	1	1	31.4	31.8	(31.3 32.3)	31.8	(30.69 32.91)
Male Life Expectancy	82.8	86.6	3	3	79.5	82.93	(80.36 85.58)	86.89	(83.70 90.13)
Female Life Expectancy	87.7	91.5	3	3	84.6	87.21	(85.01 89.44)	91.02	(86.89 95.04)
Number of Immigrants	321	304	100	100	408.66	321.19	(297.68 346.58)	314.35	(275.98 348.29)
Number of Emigrants	101	128	30	30	83.81	101.34	(89.01 115.83)	101.09	(90.08 111.50)

and covariance matrices. More specifically, k_0 is set equal to 1, and n_0 is set equal to the dimension of vector **R**. Finally, we set $\alpha_j = 1 / J$, thus ensuring a diffuse prior assignment on the groups' probabilities.

The mixture model is fitted for several choices of the number J of mixture components. For each of the resulting models, the Bayesian information criterion (BIC) is computed. For comparison, we also consider the model obtained by assuming that the expert evaluations are sampled from the same multivariate Gaussian distribution, with mean μ and covariance matrix Σ (nonmixture model). A multivariate Gaussian distribution centered on **R** and with covariance matrix Σ / k_0 is assigned to μ. The same prior assignment used for the mixture model is specified for **R**; an inverse-Wishart prior distribution, with parameters Σ_0 and n_0, is assigned to Σ. The values of n_0 and k_0 are chosen as for the mixture model. Table 3 shows the forecasts—that is, the posteriors' means and standard deviations, and correlations for the total number of immigrants (in thousands) and total fertility rate as obtained from the mixture model for $J = 2, 3, 4$, and 5, and from the nonmixture model. As shown in the table, the posterior values obtained by letting J vary do not differ significantly, thus proving the robustness of the model with respect to the choice of the number of components. The forecasts under the nonmixture model are characterized by higher values for immigration and by an overall higher variability. The correlations between the two indicators are, in all cases, quite low. As for the model fit, the mixture with two components scores the smallest BIC value among all the models considered. Thus, the mixture with two components provides the best fit of the expert evaluations. This result supports the choice of the mixture model because it improves on the nonmixture model in terms of goodness of fit. The choice of the smallest mixture model is not surprising given the number of observations (i.e., the number of experts). In particular, the two groups of experts identified by the mixture model differ mainly in terms of correlations, as shown in Table 4. Similar results for model fit are obtained for mortality and emigration.

Our preferred specification, which we now show and discuss, is therefore a forecast based on the mixture model with two components. Table 2 shows the forecasts along with the 85 % prediction intervals for all the demographic indicators at 2030 and 2065. Figure 3 depicts the forecast of the Italian population and the corresponding 85 %

Table 3 Posterior means, standard deviations and correlations of total number of immigrants (IM, in thousands) and total fertility rate (TFR)

	Mixture Model				Nonmixture 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
$\sigma(IM_{2030})$	65.20	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

prediction interval. Table 5 shows the forecasts of the total population and of the elderly dependency ratio (the ratio of the population older than 65 to the population aged 14–65, as a percentage) at 2030 and 2065, along with their 85 % prediction intervals.

Table 4 Posterior means, standard deviations, and correlations of total number of immigrants (IM) and total fertility rate (TFR) in the two subgroups of experts

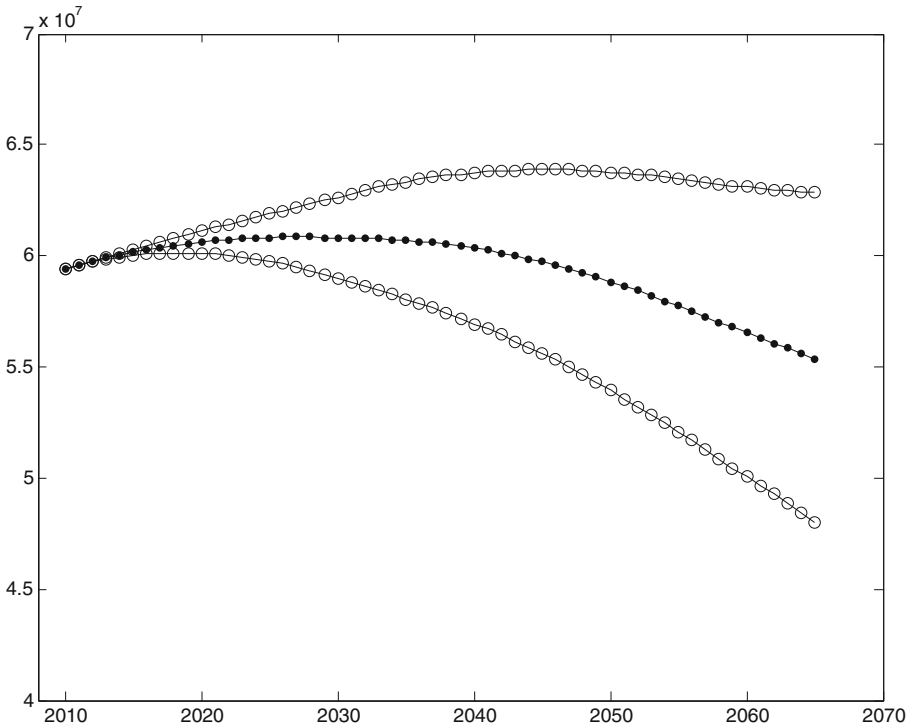|	Group 1	Group 2
$E(IM_{2030})$	270.45	269.51
$E(IM_{2065})$	243.00	237.81
$E(TFR_{2030})$	1.53	1.53
$E(TFR_{2065})$	1.64	1.63
$\sigma(IM_{2030})$	69.70	52.00
$\sigma(IM_{2065})$	90.00	79.50
$\sigma(TFR_{2030})$	0.64	0.84
$\sigma(TFR_{2065})$	0.52	0.49
$\rho(IM2_{2030}IM_{2065})$	0.59	0.39
$\rho(TFR_{2030}TFR_{2065})$	0.48	0.17
$\rho(IM_{2030}TFR_{2030})$	0.67	0.21
$\rho(IM_{2065}TFR_{2065})$	−0.39	0.34
$\rho(IM_{2030}TFR_{2065})$	−0.32	0.41
$\rho(IM_{2065}TFR_{2030})$	0.47	−0.32

Fig. 3 Forecasts of the Italian population from 2010 to 2065 with 85 % forecast intervals

Our forecasts of the Italian population show a likely growth between 2010 and 2030. Although the starting population is 60.34 million, the central forecast for 2030 is 61.65 million; the lower and upper bounds of the prediction interval are, respectively, 59.83 and 63.48 million. Between 2030 and 2065, the population is forecasted to be more likely decrease than to increase. The central forecast is 55.88 million, but as expected, there is more uncertainty, with a prediction interval from 48.50 to 63.34 million. Thus, a further increase is not ruled out.

As results of the forecast, we provide age- and sex-specific population figures; forecasts of any indicator related to the age structure of the population can be derived. In particular, the forecasts of the elderly dependency ratio show that there is virtually no uncertainty that the Italian population will continue to age during the whole interval. Our method forecasts a rise of elderly

Table 5 Stochastic forecast of the Italian population and of the elderly dependency ratio with 85 % prediction intervals

	2010	2030		2065	
Total Population	60.340	61.651	(59.828 63.478)	55.876	(48.504 63.342)
Elderly Dependency Ratio	32.20	45.87	(43.52 48.31)	68.58	(59.11 79.14)

dependency ratio from 32.2 % in 2010 to 45.9 % in 2030 (with prediction interval ranging from 43.5 to 48.31) and to 68.6 % in 2065 (with prediction interval ranging from 59.1 to 79.1).

Comparison With Other Approaches

We can assess the performance of our preferred forecast (i.e., the mixture model) by comparing its results against those obtained under different approaches.

First, what is the impact of correlation between indicators? To assess this, we applied the mixture model with two components, dropping the assumption of correlation between indicators and therefore assuming that the three components of demographic change—fertility, mortality, and migration—are independent. The resulting forecasts did not show significant differences in terms of point predictions and associated variability. This was indeed expected because of the low correlations derived from the expert evaluations as shown in Table 3.

Second, we derived an alternative expert-based stochastic forecast by applying the method proposed by Billari et al. (2012), using a simple linear combination of expert opinions. As displayed in Table 6, compared with the mixture model, the linear combination shows a higher variability in forecasts, with uncertainty regarding the 2065 population ranging from 42.42 to 63.51 million.

Third, we compared the mixture model with official deterministic scenarios by ISTAT. The comparison is shown in Table 6. Mixture expert-based predictions are, on average, lower than ISTAT scenarios, with the central scenario of ISTAT not contained within the 85 % forecast interval in 2030. Expert evaluations on future trends of demographic components seem to depart from the assumptions behind ISTAT forecasts. Very recently, ISTAT has undertaken a reconstruction of the Italian population on the basis of the data made available by the 2011 census. The revised total population size at 2012 is 59.40 million, but it is not clear whether further revisions will be made. Both our forecasts and ISTAT projections, based on the same nonrevised starting population at 2010, would then overpredict the total population in 2012. However, our predictions show a smaller forecast error if looking at the central forecast.

Table 6 Forecasts of the Italian population, with 85 % intervals

		2030		2065	
Stochastic Forecast[a]					
Mixture model		61.651	(59.828 63.478)	55.876	(48.504 63.342)
Linear combination		61.152	(58.583 63.733)	52.936	(42.416 63.511)
Forecast[b]					
Mixture model	60.340	61.651	(59.828 63.478)	55.876	(48.504 63.342)
ISTAT scenarios	60.340	63.482	(61.675 65.205)	61.305	(53.390 69.125)

[a] Comparison with stochastic forecasts based on linear pooling of expert opinions.

[b] Comparison with deterministic forecasts.

Conclusions and Discussion

In this article, we suggest a method for deriving stochastic population forecasts on the basis of a combination of expert evaluations. Expert opinions are elicited through of a specially designed questionnaire in which experts are asked to provide (conditional) scenarios on summary indicators of the components of the demographic change. In this way, opinions are elicited not only on the central scenarios but also on the marginal variability and on the correlations across time and between indicators. Moreover, our method allows for the dependence between expert opinions, combining opinions in the framework provided by the Supra-Bayesian approach. A mixture model is used to allow for the dependence between experts, under the assumptions that experts belong to a limited number of groups to which they are assigned according to their actual opinions. The forecast distribution is derived as a posterior distribution, and a Gibbs sampling algorithm is designed to approximate this posterior. The approach is implemented through a specially written MATLAB program. We apply the method, using several Italian demographers as experts, and provide stochastic forecasts of the Italian population from 2010 to 2065. These forecasts are also compared with the scenarios provided by the official forecasting agency and with forecasts obtained under alternative approaches based on the same expert elicitations.

Our proposal lies within the expert-based random scenario approach (Lutz et al. 1998). There is debate among demographers regarding the feasibility and reliability of forecasts based on expert opinions (see, e.g., Alho and Spencer 1990; Booth 2006; Lee 1999), and some issues that emerge in this debate are worth discussing here. One of the problems is that experts may be highly influenced by recent trends or idiosyncrasies. Moreover, experts tend to be overconfident in their opinions, potentially generating forecasts that have a lower variability than would be desirable. Finally, expert-based forecasts cannot generally be validated with data, contrary to (for instance) time series models. Although we recognize the broad validity of these critiques, we can defend our expert-based random scenario proposal from different perspectives. First, and very simply, when past data are too limited or even unavailable, expert-based forecasting may be the only option. This applies also to phenomena that emerge suddenly, such as low fertility or switching migration regimes. If it is possible to learn from other cases (i.e., countries) that are leading trends (Alkema et al. 2011; Raftery et al. 2012), some kind of expert opinion for the leading countries is essential. More importantly, and from a conceptual point of view, experts are de facto involved in all population forecasts, regardless of whether these forecasts are explicitly defined as expert-based—and not only because the forecaster unavoidably uses her/his own expertise and is sometimes the only expert. In actual practice, differences exist on the stage of the forecasting procedure at which the involvement of experts takes place, in the way they are asked to contribute to the forecast, and in the formalization of their involvement. When forecasts are based on time series models, expert opinions are used in the choice of the statistical model; when a Bayesian approach is followed, expert opinions are also involved in the specification of the prior distribution of the parameters. In the approach based on the extrapolation of empirical errors, experts (usually official agencies) are involved in providing central trajectories for the stochastic projections and, in some cases, for the *a posteriori* evaluation of the forecasts. The uniqueness of the random-scenario approach, in the version proposed by our method, is that it uses experts as the sole source for the

derivation of the forecasts. This strategy does not imply at all that information on past demographic trends or on other countries is neglected. Indeed, if experts are truly experts (as the ones consulted in a routine way by forecasting agencies), then in their (marginal or conditional) evaluations, they take into account this information because they have knowledge of it. Additional research is undoubtedly needed on how to better gain knowledge from experts in human population forecasting, including, in particular, efforts to limit their overconfidence. This issue has been tackled in other areas of forecasting, in fields as diverse as politics, economics and finance epidemiology, and population ecology (e.g., Speirs-Bridge et al. 2010; Tetlock 2005; Tyszka and Zielonka 2002).

To conclude, it is important to emphasize some specific limitations of our modeling approach. In the current implementation of our method, we did not explicitly consider the uncertainty in the initial age and sex distribution of the population, which is particularly problematic given some inconsistencies between the census-based and the register-based population in Italy. Nevertheless, this uncertainty could, in principle, be handled using the same expert-based probabilistic approach we propose: that is, experts could be called to express opinions on the baseline population as well. Moreover, we did not explicitly consider the uncertainty that arises by moving from summary indicators of fertility, mortality, and migration to the full age schedules. Finally, a drawback thus far unavoidable in expert-based approaches to stochastic population forecasting is that they must generally be performed by resorting to summary demographic indicators, even if models built using age-specific rates could, in some cases, be a better alternative. Further research will need to be targeted to these limitations.

References

Albert, I., Donnet, S., Guihenneuc-Jouyaux, C., Low-Choy, S., Mengersen, K., & Rousseau, J. (2012). Combining expert opinions in prior elicitation. *Bayesian Statistics, 7,* 503–532.

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.

Alho, J. M., & Spencer, B. D. (1990). Error models for official mortality forecasts. *Journal of the American Statistical Association, 85,* 609–616.

Alho, J. M., & Spencer, B. D. (1997). The practical specification of the expected error of population forecasts. *Journal of Official Statistics, 13,* 204–225.

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., & Dalla Zuanna, E. (2013). Cohort replacement and homeostasis in world population, 1950–2100. *Population and Development Review, 39,* 563–585.

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

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

Booth, H., Hyndman, R. J., Tickle, L., & De Jong, P. (2006). Lee-Carter mortality forecasting: A multi-country comparison of variants and extensions. *Demographic Research, 15*(article 9), 289–310. doi:10.4054/DemRes.2006.15.9

Booth, H., Smith, L., & Maindonald, J. (2002). Applying Lee-Carter under conditions of variable mortality decline. *Population Studies, 56,* 325–336.

Booth, H., & Tickle, L. (2008). Mortality modelling and forecasting: A review of methods. *Annals of Actuarial Science, 3,* 3–43.

Cairns, A. J. G., Blake, D., & Dowd, K. (2006). A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance, 73,* 687–718.

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., & Khalaf-Allah, M. (2011). Mortality density forecasts: An analysis of six stochastic mortality models. *Insurance: Mathematics and Economics, 48,* 355–367.

Cannan, E. (1895). The probability of a cessation of the growth of population in England and Wales during the next century. *Economic Journal, 20,* 505–515.

Dietrich, F. (2010). Bayesian group belief. *Social Choice and Welfare, 35,* 595–626.

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

French, S. (1981). Consensus of opinion. *European Journal of Operational Research, 7,* 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,* 598–604.

Genest, C., & Zidek, J. V. (1986). Combining probability distributions: A critique and an annotated bibliography. *Statistical Science, 1,* 114–135.

Hyndman, R. J., & Booth, H. (2008). Stochastic population forecasts using functional data models for mortality, fertility and migration. *International Journal of Forecasting, 24,* 323–342.

Hyndman, R. J., Booth, H., & Yasmeen, F. (2013). Coherent mortality forecasting: The product- ratio method with functional time series models. *Demography, 50,* 261–283.

Hyndman, R. J., & Ullah, S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics and Data Analysis, 51,* 4942–4956.

Keilman, N., Pham, D. Q., & Hetland, A. (2002). Why population forecasts should be probabilistic— Illustrated by the case of Norway. *Demographic Research, 6*(article 15), 409–454. doi:10.4054/DemRes.2002.6.15

Land, K. C. (1986). Methods for national population forecasts: A review. *Journal of the American Statistical Association, 81,* 888–901.

Lavine, M., & West, M. (1992). A Bayesian method for classification and discrimination. *The Canadian Journal of Statistics, 20,* 451–461.

Lee, R. D. (1993). Modeling and forecasting the time series of US fertility: Age distribution, range and ultimate level. *International Journal of Forecasting, 87,* 187–202.

Lee, R. D. (1999). Probabilistic approaches to population forecasting. *Population and Development Review, 24*(Suppl.), 156–190.

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

Lee, R. D., & Tuljapurkar, S. (1994). Stochastic population forecasts for the United States: Beyond high, medium, and low. *Journal of the American Statistical Association, 89,* 1175–1189.

Lindley, D. (1983). Reconciliation of probability distributions. *Operations Research, 31,* 866–880.

Lindley, D. (1985). Reconciliation of discrete probability distributions. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, & A. F. M. Smith (Eds.), *Bayesian Statistics 2* (pp. 375–390). Amsterdam, The Netherlands: North-Holland.

Lutz, W., Sanderson, W. C., & Scherbov, S. (1998). Expert-based probabilistic population projections. *Population and Development Review, 24,* 139–155.

McConway, K. J. (1981). Marginalization and linear opinion pools. *Journal of the American Statistical Association, 76,* 410–414.

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

Parrado, E. (2011). How high is Hispanic/Mexican fertility in the United States? Immigration and tempo considerations. *Demography, 48,* 1059–1080.

Parrado, E., & Morgan, S. P. (2008). Intergenerational fertility patterns among Hispanic women: New evidence of immigrant assimilation. *Demography, 45,* 651–671.

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News, 6,* 7–11.

Raftery, A. E., Alkema, L., & Gerland, P. (2014). Bayesian population projections for the United Nations. *Statistical Science, 29,* 58–68.

Raftery, A. E., Chunn, J. L., Gerland, P., & Sevcikova, H. (2013). Bayesian probabilistic projections of life expectancy for all countries. *Demography, 50,* 777–801.

Raftery, A. E., & Lewis, S. (1992). One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo. *Statistical Science, 7,* 493–497.

Raftery, A. E., Li, N., Sevciková, H., Gerland, P., & Heilig, G. K. (2012). Bayesian probabilistic population projections for all countries. *Proceedings of the National Academy of Science, 109,* 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,* 447–476.

Shang, H. L., Booth, H., & Hyndman, R. J. (2011). Point and interval forecasts of mortality rates and life expectancy: A comparison of ten principal component methods. *Demographic Research, 25*(article 5), 173–214. doi:10.4054/DemRes.2011.25.5

Sobotka, T. (2008). The rising importance of migrants for childbearing in Europe. *Demographic Research, 19*(article 9), 225–248. doi:10.4054/DemRes.2008.19.9

Speirs-Bridge, A., Fidler, F., McBride, M., Flander, L., Cumming, G., & Burgman, M. (2010). Reducing overconfidence in the interval judgments of experts. *Risk Analysis, 30,* 512–523.

Stoto, M. A. (1983). The accuracy of population projections. *Journal of the American Statistical Association, 78,* 13–20.

Tetlock, P. E. (2005). *Expert political judgment: How good is it? How can we know?* Princeton, NJ: Princeton University Press.

Tyszka, T., & Zielonka, P. (2002). Expert judgments: Financial analysts versus weather forecasters. *Journal of Psychology and Financial Markets, 3,* 152–160.

United Nations, Department of Economic and Social Affairs, Population Division. (2013). *World population prospects, the 2012 revision.* New York, NY: United Nations.

Winkler, R. L. (1981). Combining probability distribution from dependent information sources. *Management Science, 27,* 479–488.