A stochastic population projection from the perspective of a national statistical office

Gianni Corsetti, Marco Marsili

Istat

1. Introduction

Istat has a longstanding tradition in the regular production of population projections. Since the mid-80's Istat has been in charge of official deterministic projections in Italy. This approach includes plausible variants based on different assumptions regarding the future evolution of each demographic factor, in the more general framework of the cohort-component model.

Latest (2011-based) official projections have been developed on a set of assumptions about future – until 2065 - levels of fertility, mortality and migration, according to the so called *scenario* approach: a *main* variant assumed as best performance of the future demographic trend, integrated with two variants, namely the *high* and *low* scenario, having the task of defining, albeit in a deterministic approach, the level of the future uncertainty. Furthermore, projections have been produced separately for each region (NUTS2 level); consequently data for Italy as a whole is an outcome resulting from the sum of 21 regional forecasts. With regard to regional assumptions we adopted a *convergence scenario* for each demographic component. That means all regions reach the same value in a given year in the future that is beyond the time horizon of our forecasts.

In the last years some national statistical offices have started to produce population forecasts in a stochastic approach. The main goal of probabilistic population projections is to obtain prediction intervals of demographic variables and thus to measure projection uncertainty. With variant projections, on the other hand, the user has no idea how likely they are, so he has to trust that the experts have provided them with scenarios representing the "most likely" variant and its plausible borders (Abel et al., 2010).

Stochastic forecasts therefore have the advantage of providing to user the level of likelihood that a particular future population value will occur given a set of assumptions about the underlying probability distributions.

In recent years several methods of stochastic forecasts have developed. They can be grouped under three widely recognised approaches, each one of them giving the probability distributions for fertility, mortality and migration:

- probabilistic projections based on analysis of past forecast errors;
- probabilistic projections based on expert elicitations;
- probabilistic projections based on time series analysis.

In this paper we expose the first attempts to produce stochastic population projections for Italy in addition to the official deterministic population projections currently released by Istat.

We have implemented two different methods: a forecast based on the *Conditional Expert Opinions* (Billari et al., 2012) and a second one on the *Scaled Model of Error* (Alho and Spencer, 1997). The first method, falling within the so called *random scenario* approach, is built on the use of expert opinions in the definition of conditional probability distributions for the selected demographic indicators. The second one is based on the extrapolation of empirical errors, where assessment of uncertainty is modelled according to the analysis of past projections errors.

We start from hypotheses made under official projections produced by Istat. Both the methods use the official deterministic assumptions as input.

With regard to expert-based method, the expert opinions were then replaced by the official scenarios (main and high); while for the scaled model of error the input consists of age-specific rates for fertility and mortality and the absolute values by age for international migration.

In summary, deterministic projections were taken into account as starting inputs of two stochastic forecasting methods, but their inclusion in the data processing was different.

In the next section we describe operational choices and procedures adopted for the processing of latest deterministic Istat projections (Istat, 2011).

In the third section we expose the stochastic methods and their implementation on the basis of the input at our disposal.

Finally we explore the results arising from the two stochastic methods and we make a comparison between the stochastic and deterministic approaches.

2. Official deterministic projections: data, methodology and assumptions

2.1 Data and model

The calculations are developed on the basis of the traditional cohort-component model, processing demographic events for each calendar year by region, gender and single cohort. The assumptions were created by implementing for each component of the population change well-known standard projection models on latest time series available at Istat: 1952-2008 for age-specific fertility rates, 1974-2008 for life tables, 2005-2009 micro data for internal migration and outmigration. Time series were then completed until the year 2010, using provisional data on the total intensity of each demographic component and making some estimates about the requested breakdown. The base year population is the one observed on Jan 1st 2011.

Given that a convergence scenario among the Italian regions has been hypothesized the territorial differences in term of demographic behaviour are expected to fade out in the long run. The operational choice in order to ensure convergence varies for each of the three demographic components. These are described below together with the assumptions used to derive future values of the main demographic synthetic indicators in association with their own age profiles.

2.2 Fertility assumptions

We performed some time-series analyses in order to extrapolate forecasts of the total fertility rate (TFR). We chose a LogisticARIMA(1,1,0) to model its future evolution for Italy as a whole and for each single region separately. The main convergence scenario finally provides that, from 2011 to 2130 (convergence year), each region converges linearly towards the national context. Alternative scenarios were made on the basis of confidence intervals of the regional estimates and repeating the same procedure of the main scenario.

The outcome of the above procedure is a TFR that at national level increases in 2011-2065 from 1.42 children per woman to 1.61 according to the main scenario. A stronger recovery of the TFR is expected under the high scenario, 1.83 children per woman, while it remains almost stable in the low scenario, with a levelling off at 1.38 in the long run.

The fertility age schedules have been modelled using a system of quadratic splines (QS model) developed by Schmertmann (2003). The Schmertmann model describes the shape of the age fertility rates (ASFR) using only three parameters:

- the starting age of fertility α;
- the age P at which fertility reaches its peak level;
- the youngest age H above P at which fertility falls to half of its peak level.

The QS model fits five quadratic polynomials to ASFR schedules. The resulting shape function is continuous with the first derivate also continuous. Thanks to appropriate mathematical restrictions the shape function is uniquely determined by the index ages [α, P, H].

Fitting the past age patterns for each region and for Italy as a whole from 1952 to 2010 we define future trends of three parameters α, P, H until 2065 as follows:

- α is modelled as a AR(1) and it's assumed to be the same for each variant
- P and H are modelled as a LogisticARIMA(2,1,0).
- Regional estimates of A, P and H converge by 2130 to the values obtained for Italy as a whole.

Table 1 shows the assumptions on TFR and Schmertmann's parameters for Italy by variant.

Table 1 - Schmertmann's parameters and TFR by variant, Italy, 2011-2065

Year	α	P			H			TFR		
		low	main	high	low	main	high	low	main	High
2011	11.96	32.37	32.52	32.65	37.54	37.67	37.80	1.40	1.42	1.44
2020	12.39	32.24	32.92	33.49	37.46	37.91	38.32	1.38	1.46	1.53
2030	12.68	32.01	33.17	34.11	37.26	37.95	38.58	1.37	1.49	1.60
2040	12.85	31.71	33.36	34.58	37.10	37.98	38.77	1.37	1.53	1.67
2050	12.95	31.44	33.50	34.93	36.94	38.00	38.91	1.37	1.56	1.74
2060	13.04	31.20	33.62	35.19	36.79	38.00	39.01	1.37	1.60	1.80
2065	13.07	31.10	33.67	35.31	36.73	38.00	39.05	1.38	1.61	1.83

2.3 Mortality assumptions

In order to derive future level and age pattern of mortality the standard Lee-Carter model was performed. The model describes the shape of the log-mortality using the following three parameters: k_t, a_x and b_x. The first is a general mortality index varying over time while a_x and b_x are age-depending parameters. The three parameters are linked each other by almost precise relationships, so that it's quite simple to derive them by fitting the model to the time series 1974-2000 of mortality rates. The predicted values of mortality can then be found by projecting into the future only the parameter k_t. Because of the linearity of k_t at national level, we modelled it with a random walk with drift for the period 2011-2065 in order to obtain the main scenario. Alternative scenarios are generated from confidence intervals of k_t obtained by using the standard error of k_t derived from the input data.

Regional assumptions are then obtained, for each parameter and for each scenario, by converging regional predictions to the national one by 2165.

Looking at the more important results, life expectancy at birth should increase, especially as regards men, though not at the same pace as the one registered in the past 30 years. In particular, in the main scenario life expectancy at birth for men rises from 79.5 to 86.6 years (+7.1) and for women, from 84.6 to 91.5 (+6,9).

Table 2 – Life expectancy at birth and at 65 years by sex and variant, Italy 2011-2065

Year	Men						Women					
	At birth			At 65 years			At birth			At 65 years		
	low	main	high	low	main	high	low	main	high	low	main	High
2011	79.2	79.5	79.8	18.3	18.4	18.6	84.3	84.6	84.9	21.8	22.0	22.2
2020	80.2	81.2	82.1	18.9	19.5	20.2	85.1	86.2	87.2	22.4	23.2	24.0
2030	81.4	82.8	84.1	19.7	20.7	21.6	86.2	87.7	89.2	23.2	24.5	25.7
2040	82.5	84.2	85.7	20.5	21.7	22.8	87.1	89.1	90.8	24.0	25.5	27.0
2050	83.4	85.3	87.0	21.1	22.5	23.8	87.9	90.2	92.2	24.6	26.5	28.2
2060	84.1	86.2	88.1	21.6	23.2	24.7	88.6	91.1	93.4	25.1	27.3	29.2
2065	84.4	86.6	88.6	21.8	23.5	25.1	88.8	91.5	93.8	25.4	27.6	29.7

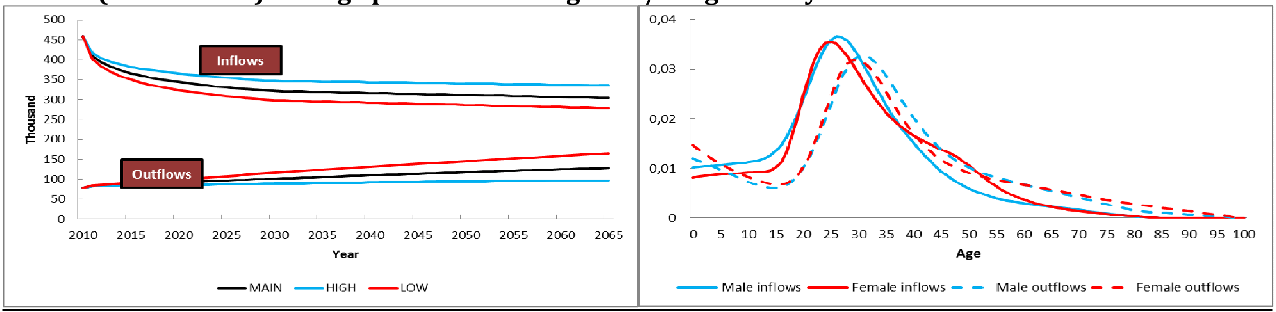
2.4 Migration assumptions

In recent years immigration flows have become more pronounced in Italy, particularly because of the growing number of arrivals from Eastern Europe and southern areas of the world. As a result, international migration has now become a separate and crucial component in the projection as it determines important changes both regarding sex and age structure of the population and the ethnic and cultural composition of the country itself. Accordingly, in order to capture the more recent trends, we focus our analysis just on the last six years, namely 2005-2010. This limit, combined with the difficulty of making predictions on international migration, does not recommend the use of an approach based on the analysis of time series, to such an extent we preferred to choose a simple model of projection.

Our solution is that in the first year of projection (2011) the total inflows and outflows are, respectively, the average value calculated over the last five observed years. Then, in accordance with the overall framework of convergence we assume for fertility and mortality as well, we assume that inflows and outflows should converge to the same level in the long run. The above procedure was applied for each region and it gives the results for Italy as a whole shown in Figure 1. What differentiates the three variants is the year in which convergence is supposed to verify: 2130 in the main, 2095 in the low and 2165 in the high scenario. The way the outflows converge in the long run is linear, while it's almost linear for inflows. Actually, for the first years of projection we introduced a decreasing factor of reduction to immigration, in order to avoid a too strong impact on the overall dynamic of the population.

Once the total level of outmigration is obtained the distribution by age is derived fitting a Rogers-Castro model to the 2005-2009 micro data (Rogers and Castro, 1981). The model parameters thus estimated are kept constant throughout the forecast period (Figure 1).

Figure 1 – International migration: Expected values of immigrants and emigrants by calendar year and variant (in thousand) and age profiles of immigrants/emigrants by sex



Although not representing a subject of specific analysis in this paper, it is worth remembering the important role played by internal migration for countries as Italy when dealing with regional projections. Its contribution in determining developments for regional population still remains significant, above all when compared to the weaker roles played by natural components, namely fertility and mortality. Over the last five years an average of about 1,4 million people changed the place of residence for internal destinations in Italy, 25% of which due to interregional movements the remaining 75% due to intraregional ones.

In our regional projections internal migration is processed by building a multi-regional matrix of probabilities by region of origin and destination, sex and age. Such a matrix, applied to the population at risk to perform an interregional migration, provides for each year of projection a coherent number of immigrants/emigrants to/from each region. In specific terms, the projection migration matrix of probabilities is defined fitting a Rogers-Castro model on 2005-2009 micro data for each couple of regions of origin/destination. Alternative scenarios are then built in a deterministic approach by increasing or decreasing the level of internal migration on the basis of specific push-pull factors among the Italian regions (for instance, changing by 5% each year the propensity to move towards Northern regions).

3. <u>Implementation of two stochastic models at national level</u>

Before describing the details of the procedures we used for implementing our stochastic approach, we should make some considerations about the reasons that helped us in the choice of the methods.

The *Scaled Model of Error* has been widely used in international studies for the last decade. It represents a point of reference for scholars, like us, who for the first time aim at producing probabilistic forecasts. It should also be emphasized its ease of use, thanks to the availability of PEP software, the parameters for its application and an extensive and depth bibliography on the subject.

The second method, the *Conditional expert opinion model*, is rather new, although the theoretical assumptions upon which it relies are less recent. Nevertheless, in our opinion this method is easy to understand and implement and it allows a discrete flexibility with regard of data and analysis needs.

Even at the risk of repeating ourselves, we emphasize once again the improper way we have used this method: we, actually, consider ourselves as experts and our deterministic projections as responses of the experts. However, we have the willing to use it more appropriately in the future. In fact, in the framework of a national research project (PRIN) coordinated by Bocconi University of Milano we are actively participating to the preparation of a questionnaire that will be given to national experts (mainly demographers). The goal will be to obtain elicitations from them on the future evolution of main demographic indicators.

Finally, we point out that even with regard to *Scaled Model of Error* our official forecasts provide the inputs needed to run the software PEP.

3.1 Expert-based method

The method is based on the elicitation of a set of parameters that allow to describe the future evolution of each demographic component. The method proceeds through a series of subsequent expert-based conditional evaluations on (summaries of) demographic indicators, given the values of the rates at some previous time points (Billari, Graziani and Melilli, 2010).

Each indicator (R) is required to be predicted in two time-points: an intermediate year (t_1) and the end of the projection horizon (t_2). In our study we consider t_0=2011, t_1=2040, t_2=2065, so generating two subintervals, 2011-2040 and 2040-2065.

It is also assumed that the vector (R_{t1}, R_{t2}) is distributed as a bivariate normal. The next step is to obtain the values of demographic parameters for each forecast year by interpolation with linear or quadratic functions, the choice between them depending on the best fit to the observed past trends.
Finally, age-specific rates are derived from synthetic indicators through the application of demographic models mentioned below.

We chose to summarize demographic components through a series of indicators:

a. Total fertility rate (TFR).
   We obtain the time-series 2011-2065 trough linear interpolation from 2011 to 2040 and from 2040 to 2065 thanks to the elicitations of the experts. In order to derive ages-specific rates we make use of the methodology proposed by Schmertmann (2003). The parameters to be taken into account are then the age *P* at which fertility reaches its peak level and the youngest age *H* above *P* at which fertility falls to half of its peak level.
b. Life expectancy at birth by sex (EM, EF).
   Interpolation of life expectancy at birth is obtained through a quadratic function. The age-specific distribution is derived from a Lee-Carter model, the parameters a_x and b_x deriving from the assumptions of the deterministic model adopted by Istat for Italy. Then k_t is obtained ex-post by constraint with the assumed levels of life expectancy.
c. Migration by sex (IMM, IMF, EMM, EMF)
   As concerns emigration we interpolated linearly from 2011 to 2040 and from 2040 to 2065 the expert's elicitations. Immigration values for each calendar year are obtained by a quadratic function. Both for emigration and immigration age specific values are obtained processing a Castro-Rogers model; the model parameters emerge from deterministic projections.

We assume the indicators being mutually independent. According to the standard stochastic methodology for each indicator R we focus our interest on the joint Gaussian distribution of R_{2040} and R_{2065}, where:

- $\boldsymbol{\mu_1}$ is the main scenario for the indicator R at the time point t_1;
- $\mathbf{q_1}$ is the high scenario (e.g. the quantile of order *q*) for the indicator R at the time point t_1;
- $\boldsymbol{\mu_2 = E(R_{2065} | R_{2040} = \mu_1)}$ is the main scenario for the indicator R at the time point t_2;
- $\mathbf{q_2 = E(R_{2065} | R_{2040} = q_1)}$ is the conditional main scenario at time t_2 given that, at time t_1, the indicator assumes the value of the high scenario fixed in the previous step (q_1);

As described above we consider the assumptions produced under the deterministic Istat projections as expert opinions providing the necessary input for implementing the expert-based method. Table 2 shows the latest Istat assumptions under the scenarios main and high. In fact, for implementing purposes it suffices to consider the elicitations provided for the main variant and for one of the two possible alternatives.

Once collected all the necessary input we get the conditions to define the stochastic process for each demographic indicator. Finally 1,000 samples are drawn from the corresponding bivariate distributions and a cohort-component model has been processed for each of them. Table 3 shows means, variances and correlation coefficients where, q is the order of quantile of the normal random variable R_{t1}.

Table 2 - Istat assumptions on fertility, life expectancy at birth and international migration

	2010	2040	2065
μ(TFR)	1.41	1.53	1.61
q(TFR)		1.67	1.83
μ(P)	32.57	33.36	33.67
q(P)		34.58	35.31
μ(H)	37.75	37.98	38.00
q(H)		38.77	39.05
μ(EM)	79.20	84.20	86.60
q(EM)		85.70	88.60
μ(EF)	84.40	89.10	91.50
q(EF)		90.80	93.80
μ(IMM)	199,880	146,048	141,286
q(IMM)		158,460	155,487
μ(IMF)	231,895	170,163	162,568
q(IMF)		184,891	179,850
μ(EMM)	39,738	55,898	64,204
q(EMM)		47,930	50,427
μ(EMF)	33,630	53,824	63,917
q(EMF)		43,771	46,861

Table 3 - Means, variances, correlations obtained from ISTAT scenarios assumptions (*q=0.9*)

	μ_{2040}	μ_{2065}	σ^2_{2040}	σ^2_{2065}	P
TFR	1.53	1.61	0.01	0.04	0.844
P	33.36	33.67	0.90	2.39	0.803
H	37.98	38.00	0.38	0.98	0.800
EM	84.20	86.60	1.37	3.58	0.800
EF	89.10	91.50	1.76	4.69	0.804
IMM	146,048	141,286	93,814,101	200,969,177	0.753
IMF	170,163	162,568	132,083,454	291,920,602	0.761
EMM	55,898	64,204	38,652,243	147,778,082	0.866
EMF	53,824	63,917	61,529,835	228,400,817	0.861

3.2 Scaled model of error

Probabilistic simulations have been produced by implementing the Scaled Model of Error (Alho and Spencer, 1997) through the use of software PEP: (http://joyx.joensuu.fi/~ek/pep/userpep.htm). The computer program PEP was used to produce the forecast for 18 European countries in the Uncertain Population of Europe Program (UPE) project (Alders et al., 2007).

This model treats the age-specific rates for fertility and mortality and the age-specific numbers of net migration as statistical distributions. The model also requires the specification of correlation of the error for each demographic component across age and over time, and correlation between male and female mortality.

In detail, the logarithm of a generic age-specific rate, *log R(j,t),* is modelled as follows:

$$\log R(j,t)=\log \hat{R}(j,t)+X(j,t), \quad j=1\ldots J,\ t=1\ldots T$$

where $R\hat{}(j,T)$ is the value of the age-specific rate (or the absolute number of net-migrants for migrations) coming from our deterministic projection and representing the expected value of a statistical distribution, j is the notation for age, t is the notation for time, and X(j,t) is the distance (error) between the true value of the age-specific rate and our input value. This last component is decomposed through a sum of forecast errors:

$$X(j,t) = \varepsilon(j,1) + \ldots + \varepsilon(j,t).$$

The model assumes that the error increments are of the form:

$$\varepsilon(j,t) = S(j,t)(\eta_j + \delta_{jt})$$

where (Graziani and Keilman, 2011):

- *S(j,t)* are deterministic scale terms;
- the variables η_j are only age dependent and they are assumed to have a Normal distribution with mean 0, variance k_j, and the correlation (η_i,η_j) having an AR(1) structure;
- the variables δ_{jt} are assumed to be uncorrelated across time, to have, for every time *t*, a Normal distribution with mean 0 and variance $1- k_j$, while the correlation (δ_i, δ_j) is treated as for the η_j terms;
- the variables η_j and δ_j are assumed to be uncorrelated.

Finally the assumptions on the model parameters – *k, S and correlations* - are the same as those used in the UPE project. The final results are ex-post aggregated after having launched 1,000 simulations.

4. Main results

A preliminary consideration is that the implementation of the two methods produces different outcomes. The use of Pep software provides, for each simulation, statistics on population and life expectancy by age, sex and calendar year. Therefore no information is available about demographic flows (deaths, births and migration). The expert based method, on the other hand, provides a full outcome, including each component of the population change. For this reason a full comparison making cannot be carried out between the outputs of the two methods, limiting our analysis just to the total population and to its age structure.

Table 4 displays the median values of the total population for 2012-2065 produced under different stochastic forecasts: the Scaled model of error (SME) and the Expert-based method (EBM) at three different values of *q*. Despite not being properly correct to make comparisons with the deterministic approach, latest Istat deterministic projection (main variant) is also reported in the table. From this latter point of view the SME seems to reproduce faithfully the values of the official forecast, whereas for the EBM the more we move away from the base year the more the distance increases, especially when elicitations are retained not to be very accurate.

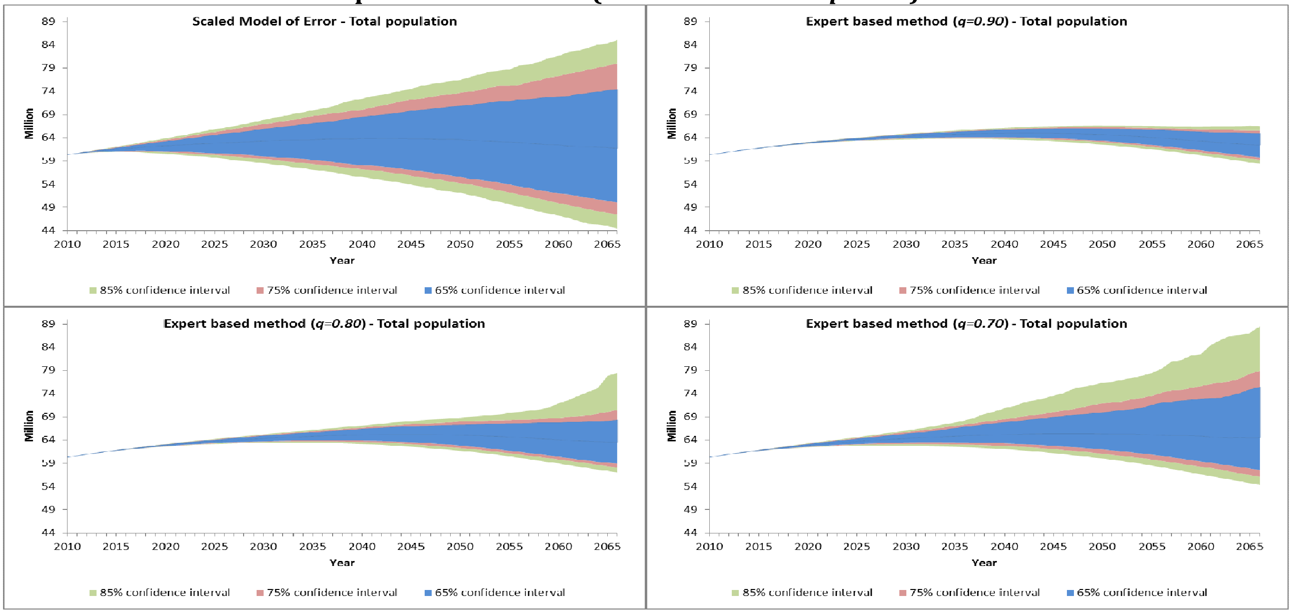
Table 4 – Total population: deterministic projections and probabilistic forecasts (median value) from Scaled model of error and Expert based method (different levels of *q value*)

Year	Deterministic projections (main variant)	Scaled model of error	Expert-based method (q=0.90)	Expert-based method (q=0.80)	Expert-based method (q=0.70)
2012	60,916,192	60,886,589	60,942,431	60,942,735	60,942,311
2020	62,497,034	62,290,647	62,883,995	62,892,989	62,878,771
2030	63,482,851	63,302,297	64,299,265	64,352,058	64,354,279
2040	63,889,453	63,846,000	64,951,855	65,145,122	65,193,374
2050	63,546,405	63,636,822	64,661,195	65,090,457	65,335,533
2060	62,169,504	62,468,252	63,294,853	64,082,011	64,891,271

Figure 2 shows the evolution of total population at different levels of confidence intervals. One can easily observe how EBM allows to define forecasts with a lower level of uncertainty. Nevertheless, we have to stress how the EBM could result quite sensitive to the level of accuracy (*q value*) we assign to the expert opinion. In quality of absolute beginners in the attainment of probabilistic forecasts, such a result represents a warning for us: as far as we can our goal is to find out a way for measuring uncertainty, but we see that doing more tests with different methods or selecting different options of the same method we are not in a condition to establish with certainty what is the "true" uncertainty.

After this necessary premise, our next comments regard only the projections made with the Stochastic Model of Error and the Expert Based Method with q=0.7 (hereinafter EBM0.7). Our intention is to have comparable levels of uncertainty, that is the point on which we focus our interest, since in our study these latter two alternatives produce closer results.

Figure 2 – Total population: forecasts and 85%, 75% and 65% Confidence Intervals (in million) from Scaled model of error and Expert based method (different levels of *q value*)



Figures 3-5 show the evolution of the population by main age-group. At a first look SME and EBM0.7 provide projections giving similar trend along time. This result is implicitly derived from the age composition of the population in the base-year and from the quite common assumptions on the demographic flows. Looking at uncertainty, it must be stressed how our forecasts seem to produce very low variability throughout the initial period of the forecast, particularly with regard to EBM0.7 where uncertainty practically does not exist, at least until the year 2020. After which, in the mid and long term EBM0.7 shows a lower variability as compared to the one obtained from SME. For instance, examining the confidence interval at 65% by 2065 from SME, we find a bandwidth of 8 million for the 0-14 age group, 15 million for the 15-64 age group, 8.6 million for the 65 and over age group. The corresponding values according to EBM0.7 are equal to 6.8, 8 and 4.8 million.

Figure 3 – Population aged 0-14: forecasts and 85%, 75% and 65% Confidence Intervals (in million) from Scaled model of error and Expert based method (*q=0.70*)

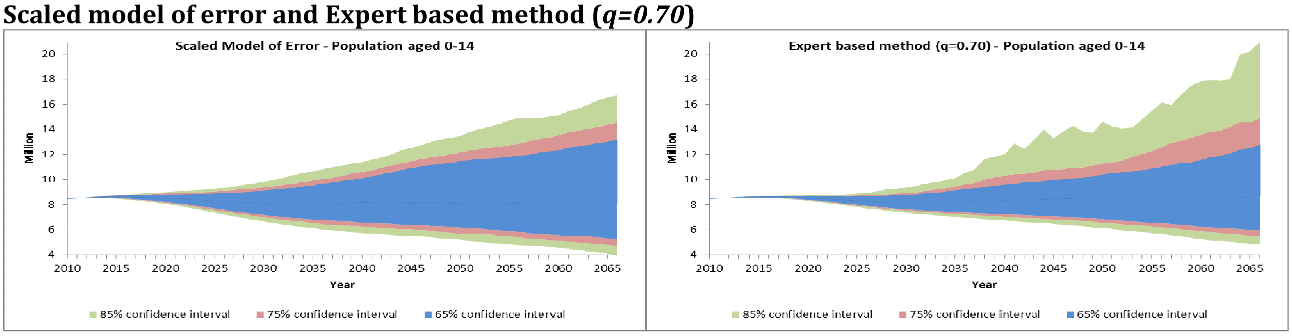


Figure 4 – Population aged 15-64: forecasts and 85%, 75% and 65% Confidence Intervals (in million) from Scaled model of error and Expert based method (*q=0.70*)

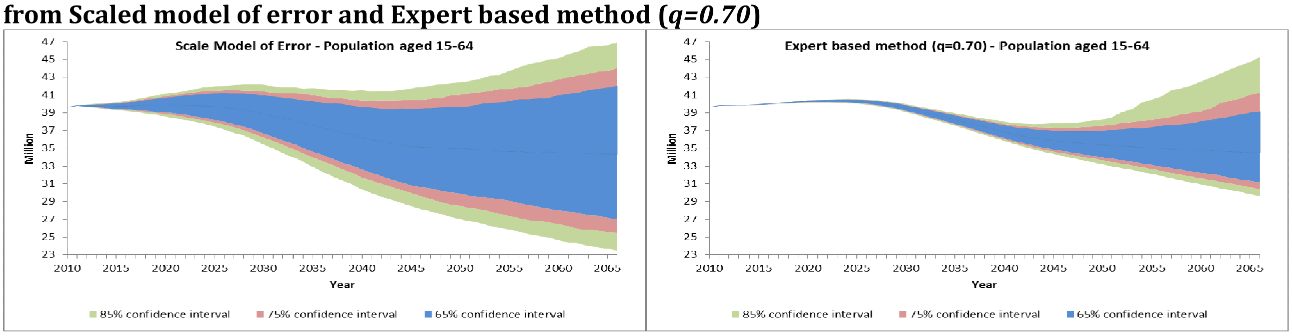
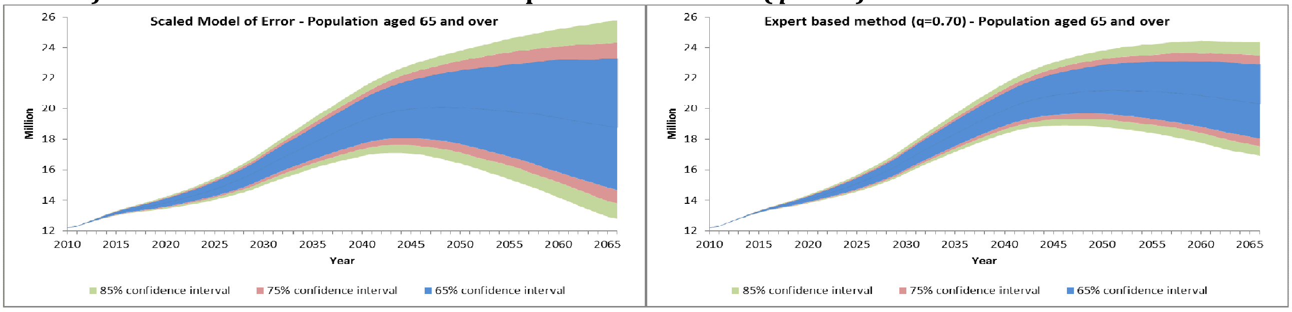


Figure 5 – Population aged 65 and over: forecasts and 85%, 75% and 65% Confidence Intervals (in million) from Scaled model of error and Expert based method (*q=0.70*)



With some rare exception, as in the case of the 0-14 age group at a 85% level of confidence interval, the lower variability highlighted by EBM0.7 is in relation to the way variances and correlations are obtained (Billari, Graziani and Melilli, 2010). In SME variance is estimated from past forecast errors. Then forecasts are obtained working directly on fertility and mortality age-specific rates, by adding them the shocks with variance and correlation across age and time, as estimated on the basis of the past forecast errors time series. In EBM approach, instead, a first step is to randomize the indicators (TFR, EM, EF, IMM, IMF) on the basis of expert opinions and a second one is to derive the age-specific rates from given demographic models. Another relevant cause is that expert opinion from national statistical offices could produce a smaller variance, compared to the variance used in the UPE project.

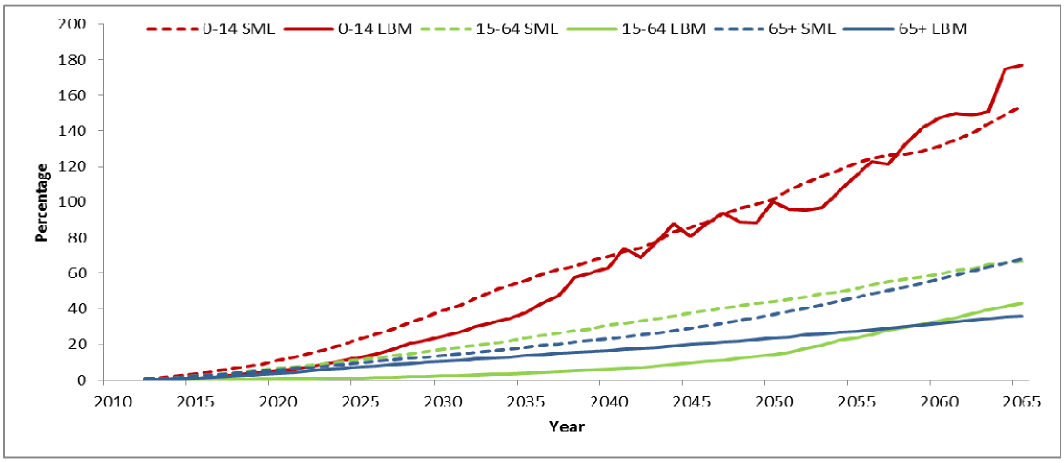
To better understand which component of the population – young, adult or elderly – is expected to present less or more uncertainty we consider the following indicator:

$$I_{a\_y} = 100 * \left(F_{U85\ a\_y} - F_{L85\ a\_y} \right) / F_{M\ a\_y},$$

where $F_{M\ a\_y}$ is the median value of a forecast in the year y for the age group a; $F_{U85\ a\_y}$ and $F_{L85\ a\_y}$ are the upper and lower bounds of the 85% confidence interval of the same forecast. In other words I_{a_y} is a standardised indicator, expressing the uncertainty in percentage points.

Figure 6 shows the percentage relative interval of uncertainty at 85% confidence level by age group for SME and EBM0.7. As it was easy to imagine the young population is characterized with a greater level of uncertainty. On the other hand, it's interesting to note that, *ceteris paribus*, adult and elderly present surprisingly the same trend of uncertainty, whichever method is taken into consideration.

Figure 6 – Percentage relative interval of uncertainty at 85% confidence level for each age group from two stochastic methods



Dealing with outcome produced by EBM0.7 in term of demographic flows (outcome not available for SME), it appears clearly how births are affected by more uncertainty than deaths (Figure 7). For what concerns uncertainty of migration there aren't substantial differences between inflows and outflows, although the former present a greater bandwidth in the long term (Figure 8).

Figure 7 – Births and deaths: forecasts and 85%, 75% and 65% Confidence Intervals (in thousand) from Expert based method (*q=0.70*)

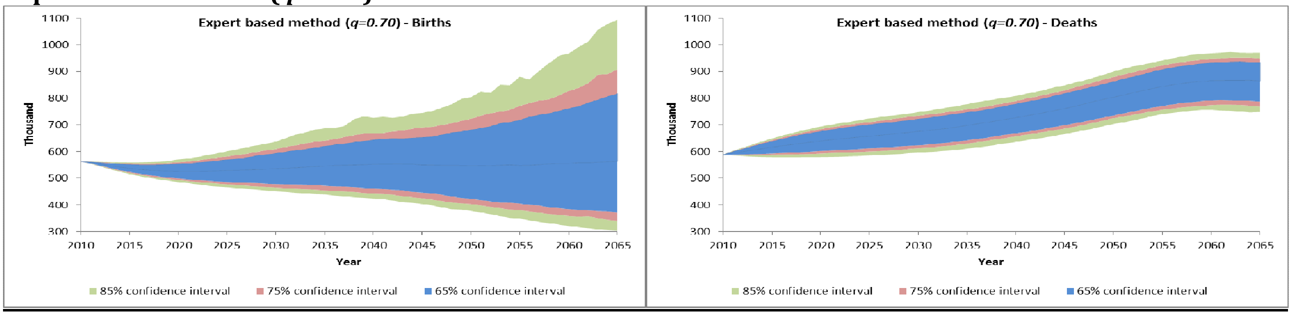


Figure 8 – Immigrants and emigrants: forecasts and 85%, 75% and 65% Confidence Intervals (in thousand) from Expert based method (*q=0.70*)

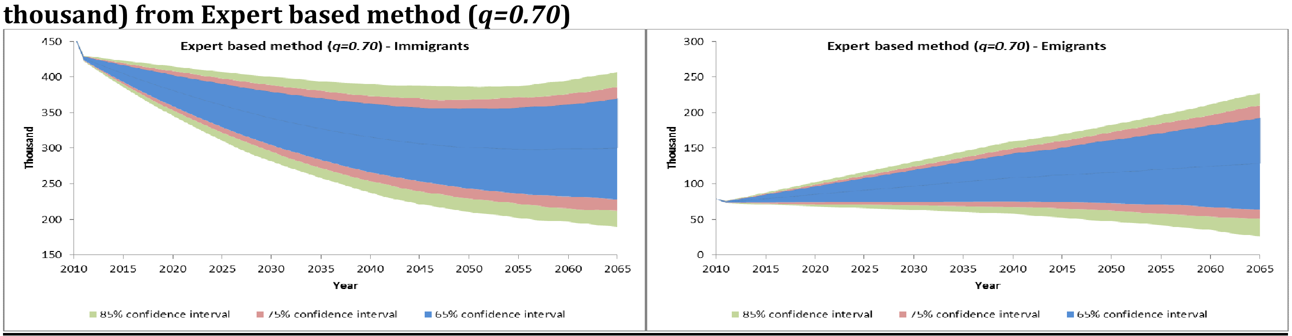


Figure 9 – Stochastic Population pyramid in 2065: forecasts and 85%, 75% and 65% Confidence Intervals (in thousand) from Scaled model of error and Expert based method (*q=0.70*)

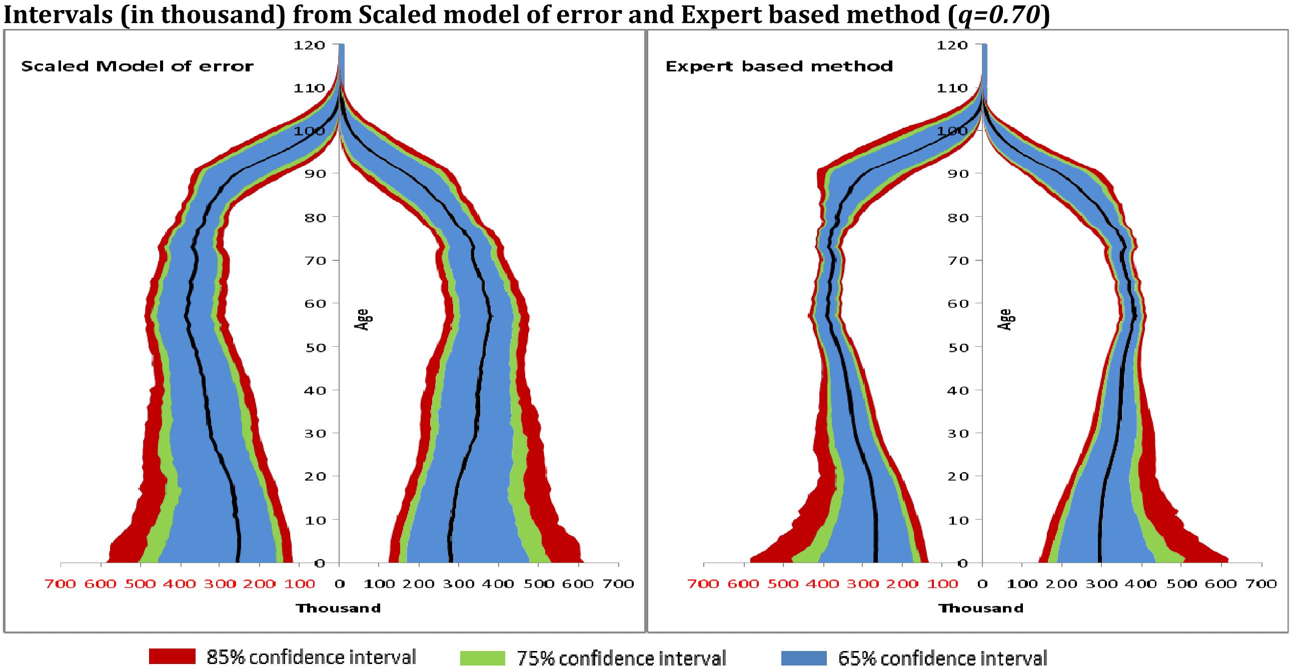


Figure 9 shows the stochastic population pyramid by 2065 for both methods, SME and EBM0.7. With this representation we can confirm what has emerged from previous analyzes, given that the role played by the different components of the population change on the forecast uncertainty can be immediately seen. Without any surprise, uncertainty due to mortality is lower than uncertainty associated with fertility. Considering the increasing population aging for Italy, most of deaths will occur among the adult and, even more, among the elderly populations. So, mortality will involve mostly people aged 55 and more, in other words people already born at the beginning of our forecast period. Migration will also play an important role, particularly among adult ages and, in correlation with fertility, a more pronounced impact at younger ones. Anyway, there's no doubt that fertility will play a major role in term of uncertainty, whichever method one can use, as it involves partially births from cohorts not yet born. From this latter point of view we stress that uncertainty released under EBM comes very close to uncertainty released under SME.

5. Conclusion

For making probabilistic forecasts, many data are required and several data assumptions need to be specified. Furthermore, despite the availability of increasingly powerful hardware and software more sophisticated than in the past, data processing can be really time consuming in order to achieve a sufficient number of simulations. A lot of details must be specified step by step, so that the subjective element, implied in the projection-maker, persists substantially even in the probabilistic approach, exactly as, and perhaps even to a greater extent, than in the deterministic approach.

We can say this having experienced some simulations on a national basis, but this issue is more pronounced when we come down to regional level. In this regard it should be noted that the NSI's are often responsible for the preparation of projections on a regional basis, just as Eurostat is responsible for preparing projections for each Member State. Despite not being on the scope of this paper, we consider highly important for our institutional responsibilities, as NSI, to facilitate any methodological effort in the direction of the stochastic approach under a multi-regional perspective. From that point of view the most challenging methodological issue is, according to us, the inclusion of internal migration under a stochastic perspective.

In this paper we present our first attempts to develop stochastic population projections for Italy. Bearing in mind our preeminent experience in the production of official deterministic projections for Italy, this new activity can then be placed at an experimental stage. We didn't choose a stochastic method *a priori*, but we decided to use a comparative approach between two alternatives: the Scaled Model of Error, one of the most known and used in the production of stochastic forecasts, and an expert-based method, developed by Bocconi University.

Results show that both methods provide median projections really close to each other, but this outcome is largely originated by the use of the same data input, namely last (2011-based) official deterministic projections by Istat. For the same reason our probabilistic forecasts do not differ much from the main variant of the official projections. Therefore the former can provide useful information on the accuracy of the latter.

We then focused on the uncertainty of the projections, which is the main objective when introducing stochastic approaches. Our results confirm what has already been noted from past studies conducted by leading scholars in this field: assessing the level of uncertainty is clearly the crucial question for demographers but the more options we have at disposal, the more challenging becomes guessing a suitable choice. Once uncertainty has been quantified, the aim should be not only to find a way for reducing it that is – needless to say – a really important matter. Some of the simulations we presented here, for instance, show that uncertainty can be very low or even almost entirely absent, particularly in the first years of forecasting. In conclusion, an approach born with the spirit to take seriously on board the problem of uncertainty in population projections can come to the paradox of producing too much accurate forecasts.

Acknowledgements

We are grateful to Rebecca Graziani from Bocconi University, for providing us with information on software PEP and its utilization.

References

1. Abel G.J., Bijak J. and Raymer J. (2010) A comparison of official population projections with Bayesian time series forecasts for England and Wales, Population Trends 141: 95-114.

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

3. Alho, J. M. and Nikander, T. (2004). Uncertain population of Europe—summary results from a stochastic forecast. http://www.stat.fi/tup/euupe/rp_reports_and_pub.html.

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

5. Alders M., Keilman, N. and Cruijsen, H. (2007) Assumptions for long-term stochastic population forecasts in 18 European countries. European Journal of Population, 23(1): 33-69.

6. Billari, F. C., Graziani, R. and Melilli, E. (2010) Stochastic population forecasts based on conditional expert opinions. Working Paper 33. Carlo F. Dondena Centre for Research on Social Dynamics, Bocconi University, Milan.

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

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

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

10. 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.

11. Keilman, N. and Pham D. Q. (2004) Empirical errors and predicted errors in fertility, mortality and migration forecasts in the European Economic Area. Discussion Paper 386 August 2004, Statistics Norway.

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

13. Lee R.D., Carter L.R. (1992) Modeling and forecasting U.S. Mortality, Journal of the American Statistical Association, September, vol. 87, n.419.

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

15. Rogers A., Castro L. (1981) Model migration schedules, International Institute for Applied System Analysis, Laxenberg, Austria, RR-8 1-30, November 1981.

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

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