Working Paper

DT/2018/5

Introducing Uncertainty on Fertility and Survival in the Spanish Population Projections: A Monte Carlo Approach

Abstract

In this paper we present a methodology to generate stochastic population projections combining the cohort-component method with Monte Carlo simulation of two of the main demographic inputs: fertility rates (by age) and survival probabilities (by age and gender). The Monte Carlo simulation is based on a parsimonious parameterization of the corresponding curves and a multivariate time series model that is used to simulate future scenarios. The paper also includes a procedure to condition the stochastic projections to a long-term benchmark using the information provided by a panel of European countries that share with Spain the main drivers of the dynamics of fertility and survival. This procedure aims to tame idiosyncratic effects that may arise when considering only the information related to just one country.

Written by Ainhoa Osés-Arranz[T] and Enrique M. Quilis[TT]
Revised by Ángel Estrada and Álvaro Pastor
Approved by Carlos Cuerpo

Key words: Fertility, Survival, Demography, Spanish Population, Population Projections, Monte Carlo, VAR Models

JEL: C53, J11, J13

[T] Spanish Independent Authority for Fiscal Responsibility. Current affiliation: Irish Fiscal Advisory Council.

[TT] Spanish Independent Authority for Fiscal Responsibility.

We thank I. Arroyo, C. Cuerpo, J.L. Escrivá, A. Estrada, J. Fernández-Huertas, R. Frutos, M.A. García, G. López-Molina, M. Morano and A. Pastor for stimulating discussions. The views expressed herein are those of the authors and not necessarily those of the Spanish Independent Authority for Fiscal Responsibility.

Index

1 Introduction

Population projections are a key ingredient in the formulation of fiscal policy, especially with regards to the evolution of government expenditure linked to the level and composition of population, chiefly pensions, education and health. Its impact on monetary policy is also gaining momentum due to its role in the determination of the natural interest rate (see Gagnon et al. (2016) for a detailed analysis).

In addition, dynamic macroeconomic modeling using the Overlapping Generation (OLG) framework requires population projections of considerable length for its proper numerical resolution (see Auerbach and Kotlikoff (1987) and Heer and Maussner (2009)).

In this paper we present a methodology to generate population projections combining the cohort-component method with Monte Carlo simulation of two of the main demographic inputs: fertility rates (by age) and survival probabilities (by age and gender). The Monte Carlo simulation is based on a parsimonious parameterization of the corresponding curves and a multivariate time series model (Vector of Autoregressions, VAR) that is used to simulate future scenarios. The model considers separately the dynamics of the national and the foreign population, allowing for a detailed analysis of the migration phenomenon.

The paper also includes a procedure to condition the stochastic projections to a long-term benchmark using the information provided by a panel of European countries that share with Spain the main drivers of the dynamics of fertility and survival. This procedure aims to tame idiosyncratic effects that may arise when considering only the information related to just one country.

The model can be used to generate both deterministic and probabilistic scenarios. In the latter case, the model allows for quantifying the reliability of the projections by means of the corresponding density forecasts, including an attribution of the uncertainty to the basic demographic inputs.

The structure of the paper is as follows. Sections 2 to 4 cover the theoretical framework followed along the paper, starting from the mathematical structure of the model: the cohort-component projection method. This method allows to project the population structure –by cohorts and gender– according to its main determinants: internal dynamics conditioned on fertility and survival, immigration and emigration. Total population is split by nationality, considering nationals and foreigners as two separate groups that interact only through the process of acquisition of the Spanish nationality. The third section is devoted to the implementation of the Monte Carlo projections, in order to gauge the underlying uncertainty related to the behavior of fertility rates and survival probabilities. The fourth section presents a procedure to condition the stochastic projections using the information provided by a panel of European countries that share with Spain the main drivers of the dynamics of fertility and survival. This conditioning procedure is made using a two-step approach that, apart from its ease of exposition and transparency, allows us to combine both sources of information in a numerically tractable way. The empirical results on fertility rates and survival probabilities are covered in section 5. Section 6 concludes.

2 General framework: projection using the cohort method

Population is projected through the cohort-component method –the most widely used framework by the statistical offices in the world–, with an additional novelty: it allows for nationality breakdowns. For a given nationality (in our case, for Spanish national residents and for Spanish residents without domestic nationality) the model allows to project population in a way that each cohort or age group evolves as a function of different inputs. In particular, demographic dynamics are determined by the surviving fraction of the group of individuals that was one year younger in the previous period, migratory flows and nationality-acquisition probabilities. Newborns are included in the model through the specific fertility rates of women at childbearing age during the preceding period.

The model offers a breakdown by age groups, gender and nationality, as shown in equations [2.1] and [2.2]. This differentiation is denoted by indices i, j and v respectively, where the age group i lies within the interval $[1, I]$ being $I = 101$. $j = \{1,2\}$ determines gender by men and women, and $v = \{n, f\}$ specifies whether the population group is national (i.e., nationality holder) or foreigner (i.e., resident without domestic nationality).

The population dynamics can be represented in matrix form through a time-dependent first-order Markov chain with an exogenous term related to the migratory flow (Luenberger 1979; Girosi and King 2008):

[2.1] $$N_t^v = \Phi\,(F_{t-1}^v, S_{t-1}^v)N_{t-1}^v + M_t^v - E_t^v \pm A(a_{t-1})N_{t-1}^f$$

Being:

- N_t^v: number of people of nationality v existing in period t.
- Φ: transition matrix that evolves over time, comprising specific fertility rates (F) by nationality, as well as survival rates (S) by nationality.
- M_t^v: number of immigrants of nationality v at time t.

- E_t^v: number of emigrants of nationality v at time t.

- A: transition matrix that evolves over time, comprising specific probabilities of nationality acquisition of foreigners (a). The sign of this term is positive when calculating the national population, whereas it becomes negative for foreign population. This is the case given the fact that acquiring nationality merely implies a transferal of inhabitants from the foreign group to the national one.

In all cases, the first I elements refer to men $(j = 1)$, whereas the following I elements apply to women $(j = 2)$. In order to take into account this breakdown by gender and age, the structure of the model is expanded in following equation:

[2.2]

$$
\begin{bmatrix} N^v_{1,1,t} \\ \vdots \\ \vdots \\ N^v_{I,1,t} \\ N^v_{1,2,t} \\ \vdots \\ \vdots \\ N^v_{I,2,t} \end{bmatrix}
=
\left[\begin{array}{cccc|cccc}
0 & \cdots & \cdots & 0 & F^v_{1,2,t-1} & \cdots & \cdots & F^v_{I,2,t-1} \\
S^v_{2,1,t-1} & & & \vdots & 0 & 0 & & 0 \\
\vdots & \ddots & & \vdots & \vdots & & \ddots & \vdots \\
0 & & S^v_{I,1,t-1} & 0 & 0 & \cdots & \cdots & 0 \\
\hline
0 & \cdots & \cdots & 0 & F^v_{1,2,t-1} & \cdots & \cdots & F^v_{I,2,t-1} \\
\vdots & \ddots & & \vdots & S^v_{2,2,t-1} & 0 & 0 & \vdots \\
\vdots & & \ddots & \vdots & \vdots & \ddots & 0 & \vdots \\
0 & \cdots & \cdots & 0 & 0 & \cdots & S^v_{I,2,t-1} & 0
\end{array}\right]
\times
\begin{bmatrix} N^v_{1,1,t-1} \\ \vdots \\ \vdots \\ N^v_{I,1,t-1} \\ N^v_{1,2,t-1} \\ \vdots \\ \vdots \\ N^v_{I,2,t-1} \end{bmatrix}
+
\begin{bmatrix} M^v_{1,1,t} \\ \vdots \\ \vdots \\ M^v_{I,1,t} \\ M^v_{1,2,t} \\ \vdots \\ \vdots \\ M^v_{I,2,t} \end{bmatrix}
-
\begin{bmatrix} E^v_{1,1,t} \\ \vdots \\ \vdots \\ E^v_{I,1,t} \\ E^v_{1,2,t} \\ \vdots \\ \vdots \\ E^v_{I,2,t} \end{bmatrix}
\pm
$$

$$
\left[\begin{array}{cccc|cccc}
a_{1,1,t-1} & \cdots & \cdots & 0 & 0 & \cdots & \cdots & 0 \\
\vdots & \ddots & & \vdots & \vdots & \ddots & & \vdots \\
\vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots \\
0 & \cdots & \cdots & a_{1,I,t-1} & 0 & \cdots & \cdots & 0 \\
\hline
0 & \cdots & \cdots & 0 & a_{1,2,t-1} & \cdots & \cdots & 0 \\
\vdots & \ddots & & \vdots & \vdots & \ddots & & \vdots \\
\vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots \\
0 & \cdots & \cdots & 0 & 0 & \cdots & \cdots & a_{I,2,t-1}
\end{array}\right]
\times
\begin{bmatrix} N^f_{1,1,t-1} \\ \vdots \\ \vdots \\ N^f_{I,1,t-1} \\ N^f_{1,2,t-1} \\ \vdots \\ \vdots \\ N^f_{I,2,t-1} \end{bmatrix}
$$

In the calculation of newborns, gender considerations are taken into account. The probability of a newborn being male is multiplied to the fertility rates on the part of the matrix referring to men, while their reciprocal lies on the women's side. In the model, fertility rates will be non-zero only for what is deemed as childbearing maternal age, which lies within the [15, 49] age range.[1]

After calculating the national and foreign population for a given time period, the total resident population is the addition of both:

[2.3] $$N_t = N_t^n + N_t^f$$

In sum, the cohort-component methodology allows for understanding the population dynamics in a detailed manner, namely by allowing for age, group, sex and nationality differentiations through the different components that affect the equation. Furthermore, this approach offers flexibility for carrying out deterministic and stochastic simulations based on the main fundamentals of demographic phenomena. Likewise, the model can also be modified in accordance with the available information; for instance, emigration can be regarded as a flow that is endogenously determined.

[1] Note that age groups 15 and 49 are somewhat special, because they comprise 15 years and less (15-) and 49 years and more (49+). The same holds for age group 101 because it comprises 101 years and more (101+).

3 Unconditional projections of fertility and survival

As we have seen, the main inputs of the model are the expected values of fertility rates, survival probabilities and migratory flows, apart from the initial population that acts as the starting condition for the population dynamics. In this way, the uncertainty around the central projection must be traced to the uncertainty around the expected values for the demographic determinants of the projection.

The quantification of this uncertainty is not an easy task, due to the complex interplay of economic, institutional and cultural factors that lie behind the determination of fertility rates, survival probabilities and migratory flows. We have followed a model-based approach that combines a parsimonious parameterization of the cross-section fertility and survival probabilities curves with multivariate time series models[2].

This approach, affine to the one used in finance to model the dynamics of the yield curve[3], provides a comprehensive and consistent approach to curve modeling, which is especially important since the cross-section (age) of the fertility curves and survival probabilities curve is a temporal dimension by itself. At the same time, the estimation of the corresponding curves generates a vector of time series that can be modeled and projected by means of a VAR model. Furthermore, the use of the Monte Carlo method for the VAR projection produces a set of stochastic scenarios consistent on the cross-section dimension and coherent with the observed dynamics of the curves.

In the remaining of this section we will present the technical details of the Monte Carlo projection of the fertility and survival curves.

3.1 Fertility rates

The starting point is a parametric functional representation of the historical fertility curves. Among several candidates, we have chosen a Gaussian function because it allows for a flexible, parsimonious and intuitive parameterization of the observed fertility rates:

---

[2] In particular, Vector of Autoregressions (VAR) models.

[3] See Diebold and Rudebusch (2013) for an in-depth exposition.

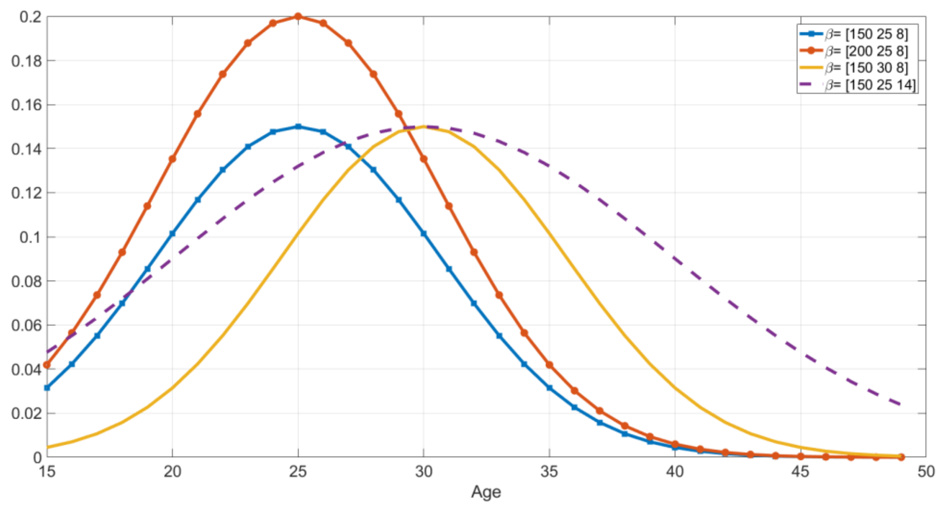
[3.1] $$f_{i,t} = \beta_{0,t} \exp[-\left(\frac{i-\beta_{1,t}}{\beta_{2,t}}\right)^2] + e_{i,t} \qquad i \in [i^-, i^+] \qquad t = 1, \dots, n$$

Being:

- $f_{i,t}$: number of child per woman at (fertile) age i in time t.
- i: age index, enclosed in the predetermined range $[i^-, i^+]$=$[15{,}49]$.
- $e_{i,t}$: error term that completes the decomposition of the observed fertility rates into a systematic component linked to age and an unsystematic term unrelated to age.
- $\beta_{0,t}$: scale parameter that measures the global intensity of the fertility.
- $\beta_{1,t}$: location parameter closely related to the average age at maternity.
- $\beta_{2,t}$: dispersion parameter that measures the concentration of the fertility distribution around the average age at maternity.

The role of each parameter can be seen in figure 1:

Figure 1: Gaussian parameterization of the fertility curve



The Monte Carlo projection proceeds in seven steps:

- 1: Estimate the parametric model [3.1] for each year.
- 2: Stack the corresponding parameters as a vector time series:

$$\beta_t = [\beta_{0,t} \ \beta_{1,t} \ \beta_{2,t}]' \qquad t = 1, \dots, n$$

- 3: Estimate a VAR model for the β series:

[3.2] $$\beta_t = c + \Psi_1\beta_{t-1} + \Psi_2\beta_{t-2} + \ldots + \Psi_p\beta_{t-p} + U_t \qquad t = p+1, \ldots, n$$

Where c is a vector of intercepts and Ψ_h, $h = 1, \ldots, p$, are 3x3 matrices. The term U_t represents a vector of zero-mean stochastic Gaussian shocks, which are serially uncorrelated and with constant variance-covariance (VCV) matrix:

$$U_t \sim iid\ N(0, \Sigma_u)$$

The estimation of the VAR model and the generation of the density forecasts by means of Monte Carlo simulation use the MATLAB functions written by Quilis (2016).

- 4: Make M draws of the shocks according to a multivariate Gaussian distribution for the projection interval:

$$u_t^h \sim N(0, S_u) \qquad t = n+1 \ldots n+T$$

Being S_u the Maximum Likelihood Estimate (MLE) of Σ_u, the VCV matrix of the shocks.

- 5: Plug in [3.2] each of the M draws of the shocks using β_n as initial condition. In this way, we derive the M possible paths for the three parameters that define the fertility curve.
- 6: For each Monte Carlo draw h, plug in the projected parameters in order to generate M alternative fertility curves for the time span $t = n+1 \ldots T$.
- 7: Compute the M paths for the population according to the cohort-component model [2.1] – [2.3].

3.2 Survival probabilities

The Monte Carlo projection of the survival probabilities closely resembles the procedure used for fertility rates. First, we have chosen a potential function to summarize the cross-section behavior of the survival probabilities[4] for each time period.

[4] We have dropped the gender index to keep the notation to a minimum, but all the calculations are performed separately for men and women.

We have chosen this functional form because it provides a very good statistical fit and its parameters have a direct demographic meaning, as we will see below.

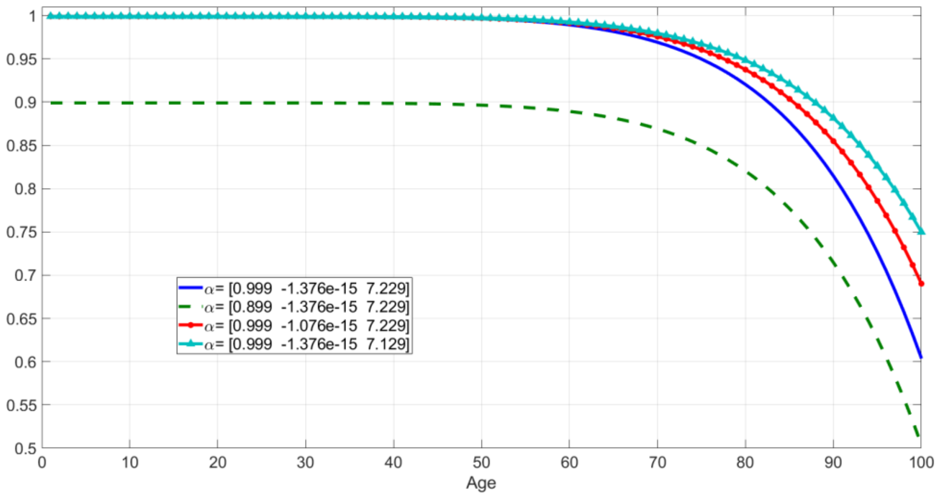
[3.3] $$s_{i,t} = \alpha_{0,t} + \alpha_{1,t}\, i^{\alpha_{2,t}} + e_{i,t} \qquad t = 1, \dots, n$$

Being:

- $s_{i,t}$: probability of moving from age i at $t-1$ to age $i+1$ at time t.
- i: age index, that covers the full range from 0 to 100+.
- $e_{i,t}$: an error term that completes the decomposition of the survival probabilities into a systematic component linked to age and an unsystematic term.
- $\alpha_{0,t}$: fixed term (global reference) related to the average survival rate but unrelated to age. In this way, the second term of [3.3] acts as a drag on this fixed component.
- $\alpha_{1,t}$: parameter that translates the (power law) influence of age on the survival probabilities in a linear fashion. This parameter has a negative sign, reflecting the reduction of survival probabilities as people grow older.
- $\alpha_{2,t}$: parameter that controls the influence of age on the survival probabilities according to a power law at time t. The greater $\alpha_{2,t}$, the greater the effect. Although both α_1 and α_2 determine the slope of the curve, the influence of $\alpha_1$can be related to the (linear) intensity of the age-decay effect while α_2 more influential than α_1 when determining the slope.

The interactions of the different parameters can be seen in the next figure.

Figure 2: Potential function parameterization of the survival curve



The Monte Carlo projection method operates in the same way as for the fertility rates, adapted to constrain the simulated probabilities in the [0,1] range.

4 Conditional projections of fertility and survival

The long-term nature of population projections suggests the use of the maximum amount of available information, including the historical evidence provided by different countries that: (i) show economic convergence among them; (ii) share a common cycle; and (iii) have similar political and social frameworks. Introducing information from a panel of countries helps to tame idiosyncratic effects that may arise when considering only the information related to just one country.

As exposed below, the fertility and survival patterns among European countries share important common features. In order to incorporate them in our modeling framework, we have conditioned the projections described in the previous section to a long-term benchmark derived from a factor analysis of a panel of European countries.

We have chosen the number of children per women and the life expectancy at birth as synthetic measures of fertility and survival, respectively. Considering a panel of *M* countries, we have represented their joint evolution through the following factor model (Mardia et al, 1979):

[4.1] $$\gamma_{j,t} = \lambda_{j,1} z_{1,t} + \lambda_{j,2} z_{2,t} + \cdots + \lambda_{j,r} z_{r,t} + e_{j,t} \qquad j = 1 \ldots M \quad t = 1 \ldots n$$

Being:

- $\gamma_{j,t}$: Number of children per women (or life expectancy at birth) of country *j* at time *t*.
- $z_{h,t}$: Common factor *h* at time *t*, where $h = 1..r$ being $r < M$.
- $\lambda_{j,h}$: Loading of factor *h* on country *j*.
- $e_{j,t}$: Idiosyncratic term of country *j* at time *t*.

The estimation of the factor model by Principal Components (PC) provides a way to recover the factors conditioned on the observed sample of the panel γ. This process is based on the score matrix that, after normalization to unity, generates a long-term benchmark for the synthetic demographic measures simply combining the long-term projections for the individual countries. This combination is based on the scores linked to the first factor because, according to the PC method, it explains the largest part of the variance of the observed variables.

This long-term benchmark provides a sort of attractor for the unconditional projections for the synthetic measures. The force of attraction exerted by the benchmark increases linearly with time, being close to zero in the short term and becoming completely binding in the long term. Technically, this process resembles the projection of a (simplified) Brownian bridge[5].

Since the fertility and survival curves are represented by means of three parameters, we have chosen one of them to be conditioned on the long-term benchmark. The selected parameter is the one that is most empirically related with the synthetic measure: the intensity parameter, both for fertility and survival.

[5] A Browning bridge is a stochastic process characterized by known and fixed initial and terminal conditions and non-stationary (random walk) dynamics (Huynh et al, 2008).

5 Empirical results

In this section we present the main results of the application of the two-step methodology explained in the previous sections to the Spanish and European demographic data. The results on fertility rates are firstly exposed, and are then followed by the findings on survival.

5.1 Fertility rates

The historical annual fertility curves used for estimation, which cover the period 1975-2015, are fitted through a Gaussian function as specified in [3.1]. As plotted in figure 3, the historical time series follow a marked down and rightward trend. The latter reflects the overall postponement of birth giving, where the peak age shifts from 27 to 33 years.

Figure 3: Fertility curves in Spain

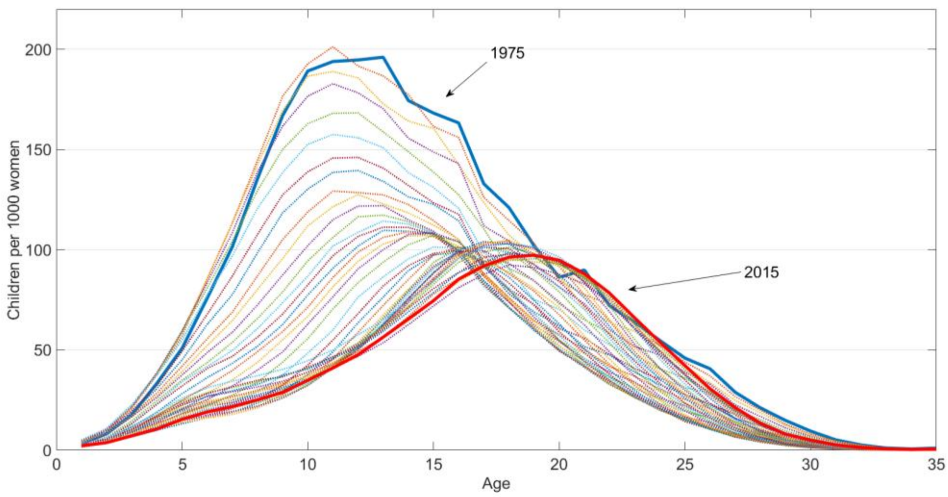
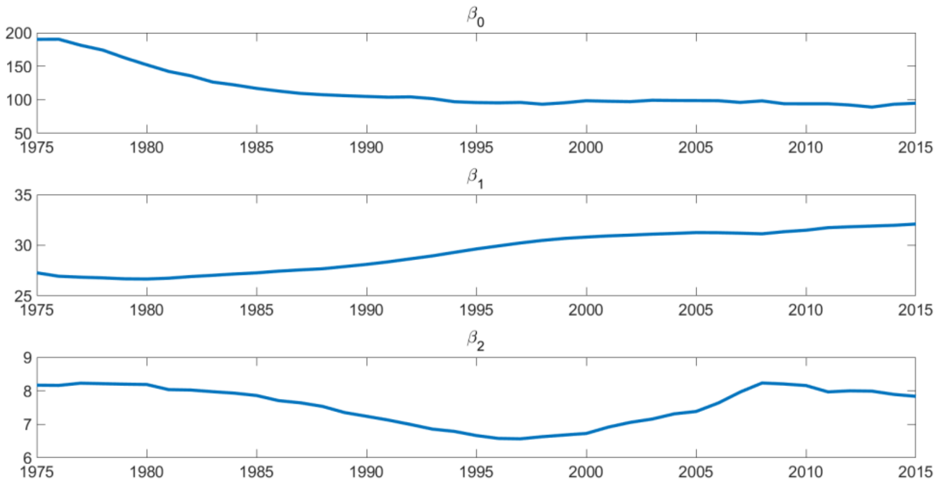


Figure 4 displays the evolution of each estimated parameter over the period 1975-2015. The analysis of these estimates provides a deeper insight into the trends on fertility over the last decades. The scale coefficient reflects the decreasing height of the curves, implying lower fertility rates being reached with time. In terms of the location parameter, its increasing tendency mirrors the rightward move of the fertility curves, where the average maternal age increases with time. The last coefficient shows that the fertility curves over the post-crisis period have experienced a dispersion similar to the initial years analyzed (roughly 1975-1985), compensating in part the downward pressure exerted by the other parameters on the average number of children per women.

Figure 4: Least squares estimation of fertility curves coefficients



Importantly, the application of a multivariate model is also motivated by the empirical correlation of the time series of the three estimated coefficients, as represented in the following matrix:

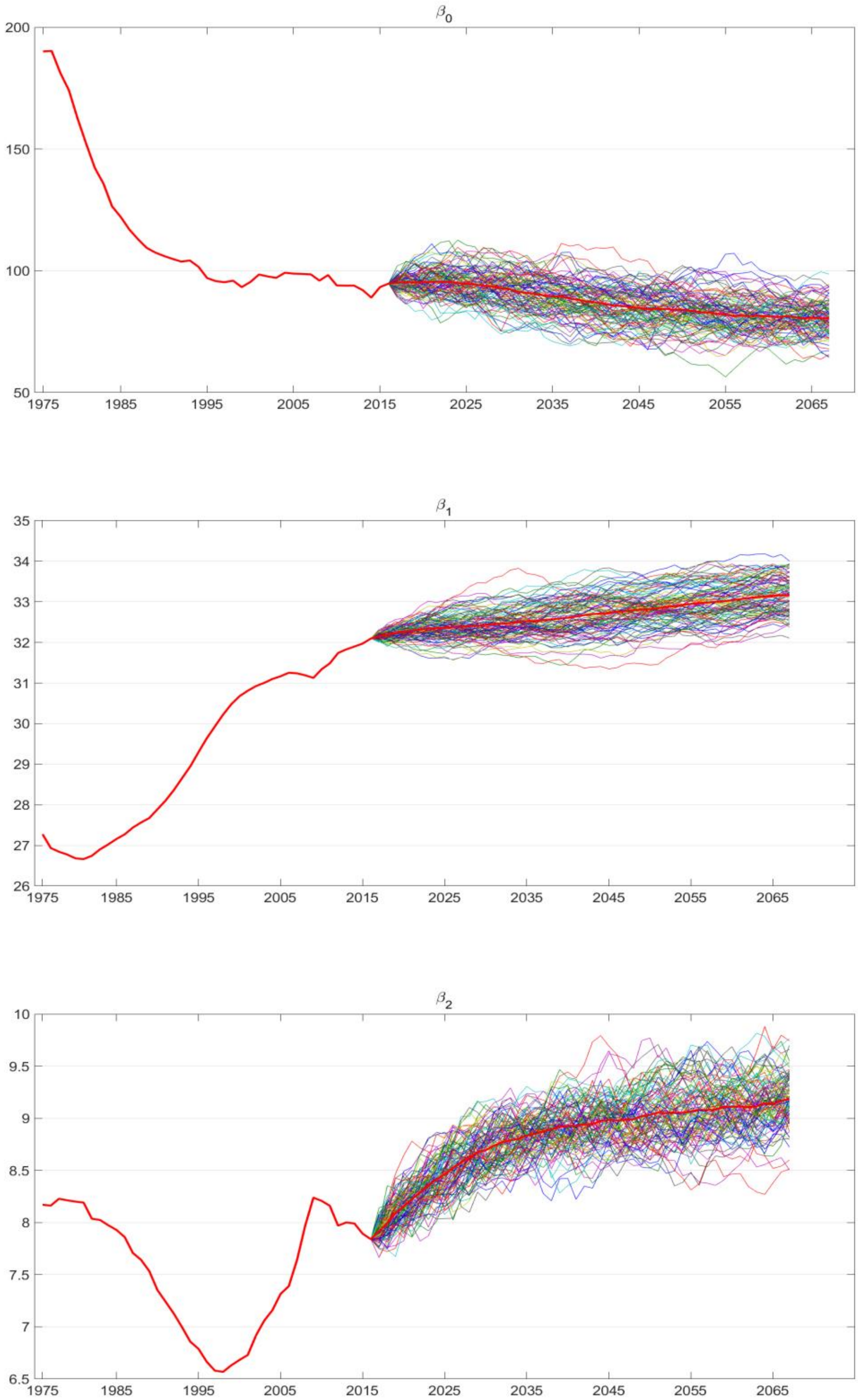
Table 1: Correlation matrix of the fertility curves coefficients

	β_0	β_1	β_2
β_0	1	0.75	0.53
β_1	0.75	1	0.26
β_2	0.53	0.26	1

The Monte Carlo Vector Autoregressive projections of each coefficient is reflected in figure 5[6].

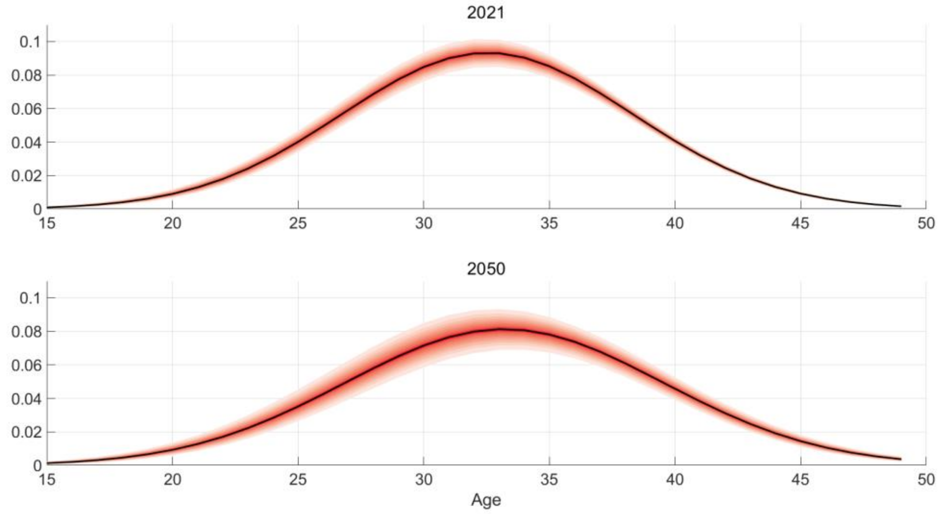
[6] For visualization reasons, only 10 (out of 1,000) Monte Carlo projections are displayed in the figure.

Figure 5: Monte Carlo VAR projections of fertility parameters



Plugging the stochastically projected parameters in the corresponding Gaussian function yields the forecasted density functions of the full set of fertility curves (Figure 6).

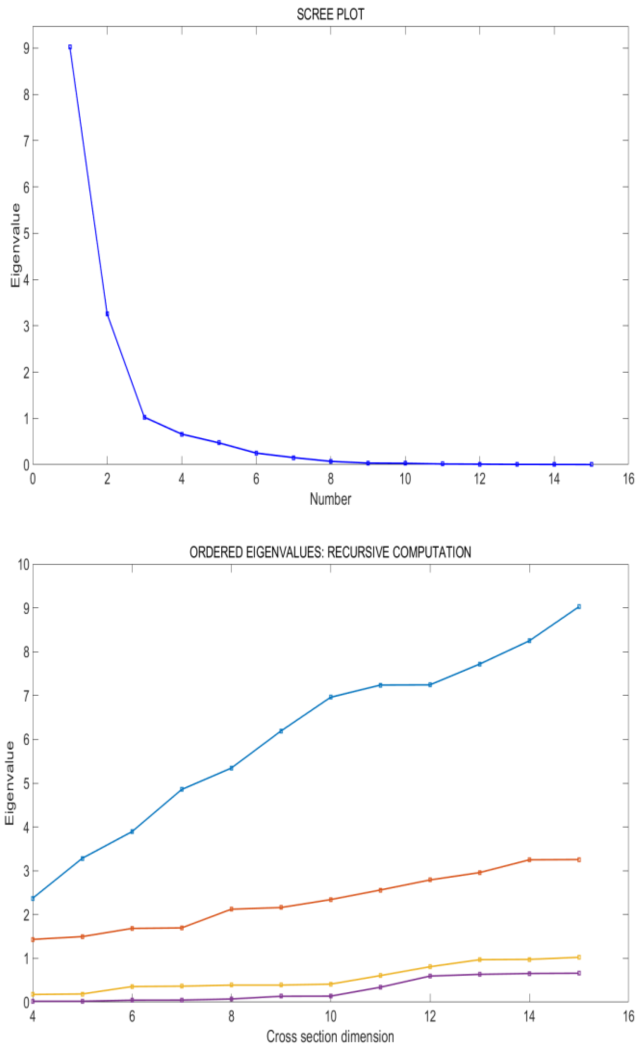
Figure 6: Monte Carlo projections of the fertility curves



Conditioned projections are generated through the estimation of a factor model to a panel of European countries. Using data provided by Eurostat, we have selected 15 countries for the time period 1971-2016. The selected countries are Belgium, Denmark, Ireland, Greece, Spain, France, Italy, Netherlands, Austria, Portugal, Finland, Sweden, UK, Norway and Switzerland. We have formed the panel by considering a sample of at least 10 countries capable of providing the longest possible series.

After standardization of the observed number of children per woman for all the selected countries, the scree plots of the eigenvalues of the correlation matrix suggest that two factors account for most of the common variance of the data (figure 7):

Figure 7: Scree plots of the European number of children per woman



The estimation of the bi-factor model by PC yields the following results:

Table 2: Factor model for the European number of children per woman

	Loadings		Communalities			Scores	
Country	1f	2f	1f	2f	All factors	1f	2f
Austria	0.95	-0.17	0.90	0.03	0.93	0.11	0.05
Belgium	0.79	0.51	0.63	0.26	0.89	0.12	0.27
Denmark	0.46	0.69	0.21	0.47	0.68	0.04	0.10
Finland	-0.48	0.60	0.23	0.36	0.59	0.00	0.05
France	0.85	0.32	0.72	0.10	0.82	0.07	0.12
Greece	0.80	-0.54	0.64	0.29	0.93	0.04	-0.19
Ireland	0.88	-0.46	0.77	0.21	0.98	0.19	-0.44
Italy	0.96	-0.20	0.92	0.04	0.96	0.18	0.05
Netherlands	0.72	0.58	0.52	0.33	0.85	0.09	0.20
Norway	0.77	0.52	0.60	0.27	0.86	0.10	0.21
Portugal	0.84	-0.48	0.71	0.23	0.94	0.06	-0.17
Spain	0.88	-0.43	0.77	0.18	0.96	0.10	-0.18
Sweden	0.20	0.59	0.04	0.35	0.40	0.01	0.04
Switzerland	0.89	0.08	0.80	0.01	0.81	0.05	0.06
UK	0.75	0.34	0.57	0.11	0.68	0.03	0.07
Median	0.80	0.48	0.64	0.23	0.86	0.07	0.05
MAD	0.25	0.14					

Note: Median and MAD (Median absolute distance) for the second factors are compute using the loadings in absolute value.

The estimation of the bi-factor model confirms the high degree of commonality among the selected European countries regarding fertility. The first factor explains most of their synchronized dynamics as well as its global non-stationary behavior. The second factor is basically linked to a differential pattern between southern countries and northern countries. The explanatory power of the second factor is much more limited than the one of the first factor (0.23 vs 0.64). With the corresponding caveats due to the limited sample and data frequency (annual), we can consider that the deviations of the individual synthetic fertility measures with respect to their common factor are stationary, although persistent.

Sweden and, especially, Finland[7] are the countries less affine to the common factors of the selected countries. Nevertheless, we have kept them in the model to retain a wide spectrum of demographic records and knowing that their scoring (i.e. its contribution to the estimation of the common factors) is consequently very low (see table 3).

Considering the Spanish case, it is noticeable its high linkage with both factors (i.e., with loadings above the corresponding medians). This commonality allows us to

[7] Finland is the only country with a negative loading on the first factor.

include international (European) information as a conditioning benchmark for the Spanish projections.

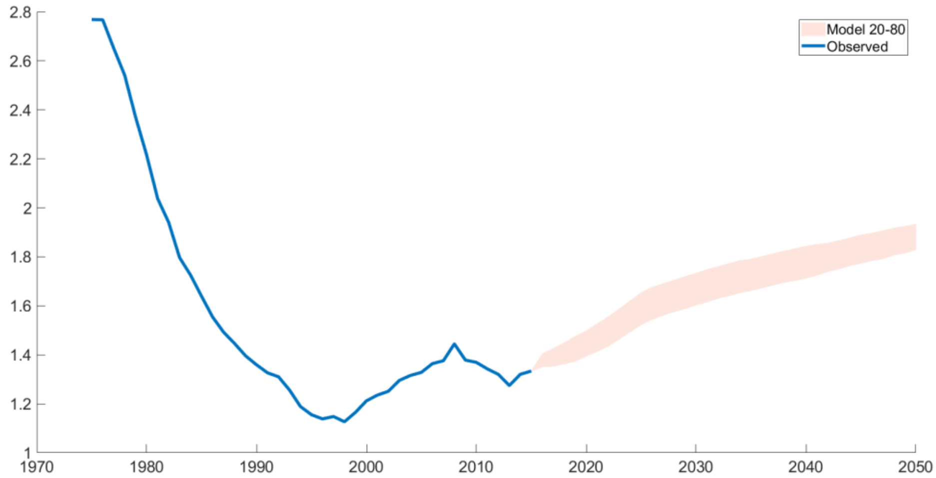
Considering the scores linked to the first factor and the long-term projections compiled by the United Nations we have estimated a long-term benchmark (2100) for the European countries of 1.83 children per woman. The following table 3 presents the detailed calculations:

Table 3: Number of children per woman. Long-term benchmark (2100)

Country	Scores	Weights	UN 2095-2100
Austria	0.11	0.09	1.79
Belgium	0.12	0.10	1.87
Denmark	0.04	0.03	1.86
Finland	0.00	0.00	1.83
France	0.07	0.06	1.94
Greece	0.04	0.03	1.76
Ireland	0.19	0.16	1.91
Italy	0.18	0.15	1.79
Netherlands	0.09	0.07	1.83
Norway	0.10	0.08	1.86
Portugal	0.06	0.05	1.76
Spain	0.10	0.09	1.72
Sweden	0.01	0.01	1.93
Switzerland	0.05	0.04	1.70
UK	0.03	0.03	1.86
Sum / Weighted average		1.00	1.83

The corresponding conditional stochastic projections, represented as density forecasts rather than as a point estimate, are depicted in figure 8.

Figure 8: Projected number of children per woman



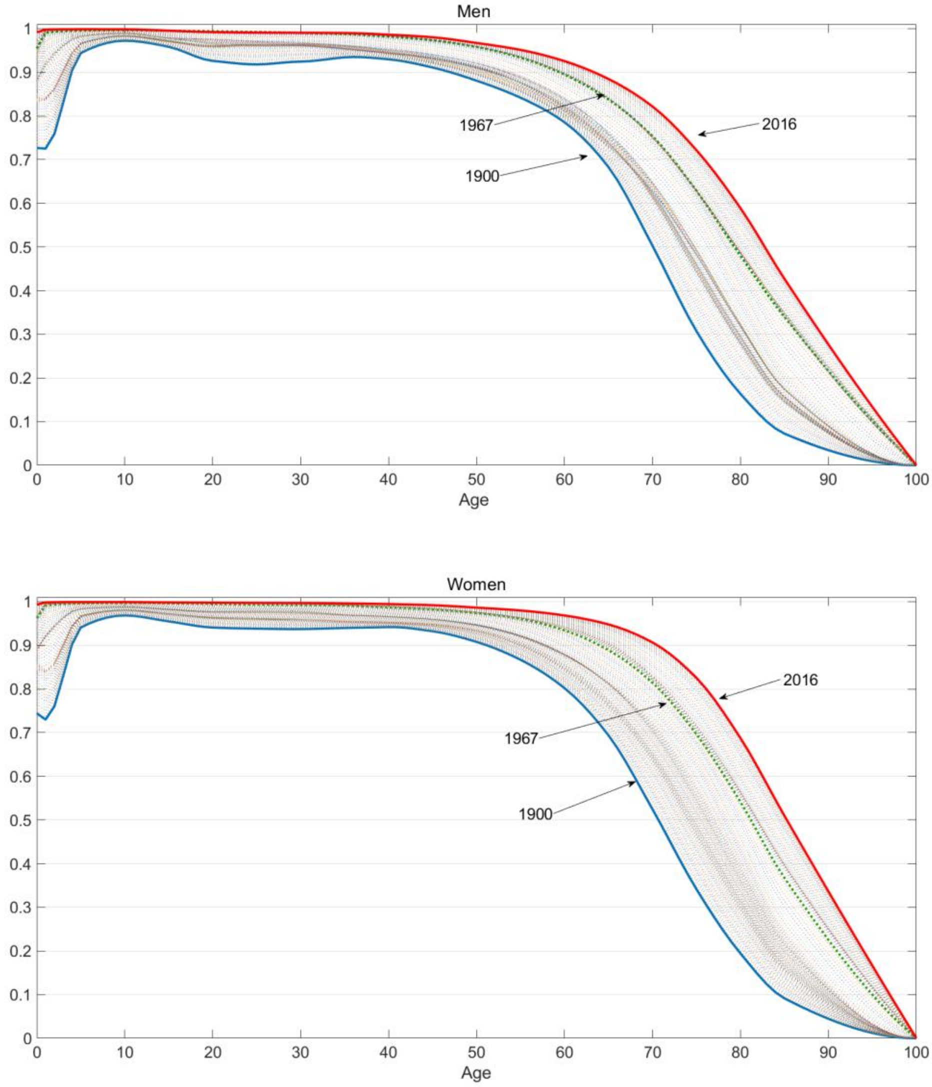
The projections eventually converge around a value of 1.80 although with a more contained rate of growth from 2030 onwards.

5.2 Survival probabilities

The survival curves –fitted through a power curve for each year and disaggregated by age and gender (see figure 9[8])–, are susceptible to a VAR modeling. This allows us to project, via Monte Carlo, the three parameters characterizing the curves and consistently generate the corresponding survival curves.

[8] Data for 1900-1974 is interpolated annually and by age groups, due to limitations in historical data published by the INE.

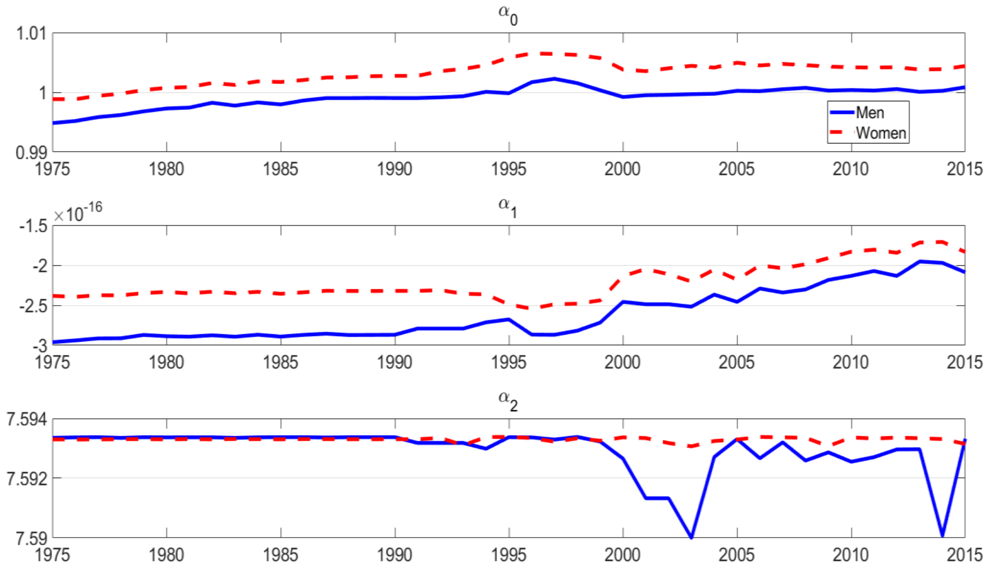
Figure 9: Survival probability curves: men and women



The methodological changes in the completion of mortality tables by the INE since 1991 preclude the availability of homogeneous time series of the parameters. For this reason, we have chosen the subsample 1991-2015 for fitting the survival curves.

The corresponding time series of the six parameters (three for men and three for women) are represented in figure 10. The parameters that quantify the intensity of the age decay (α_1) show an upward trend that explain most of the historical increase in life expectancy. The other parameters may be considered as stationary, especially the exponent (α_2). Another remarkable feature is the pairwise correlations between men and women: α_0=0.85, α_1=0.62[9] and α_2=-0.24.

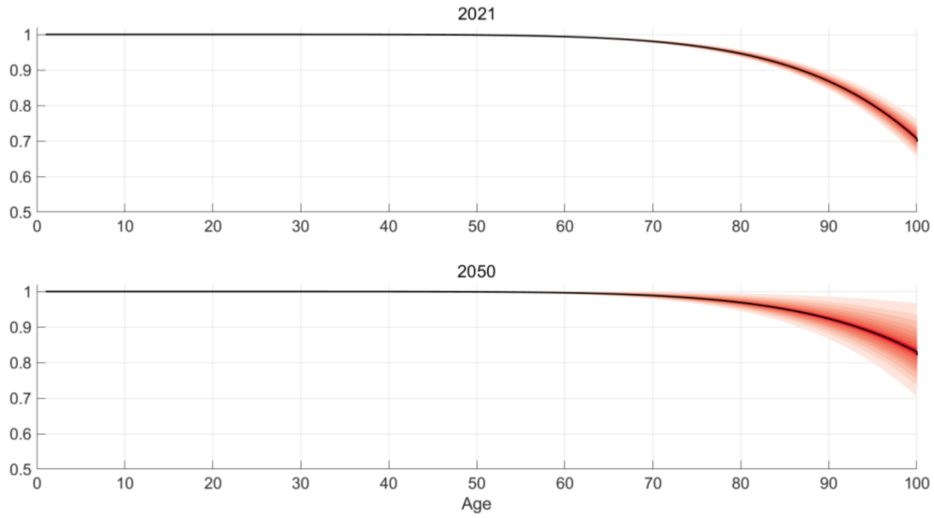
Figure 10: Survival parameters



Taking into account the stationary behavior of the global reference and the exponent, these have been projected around their mean and standard deviation as a bivariate white noise process. The intensity parameters (α_1) for men and women are projected using a VAR model according to the procedure previously outlined and applied for the fertility parameters. The Monte Carlo simulation yields the following density forecasts for the survival curves (figure 11):

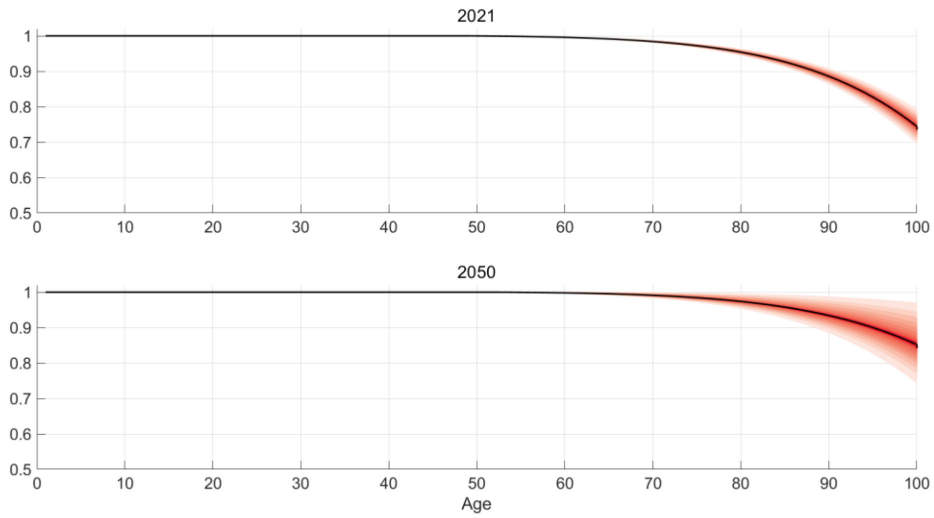
[9] This is 0.99 if the two outlying observations of the men parameter are corrected.

Figure 11a: Monte Carlo projections of the survival curves (men)



The corresponding density forecasts for women are depicted below:

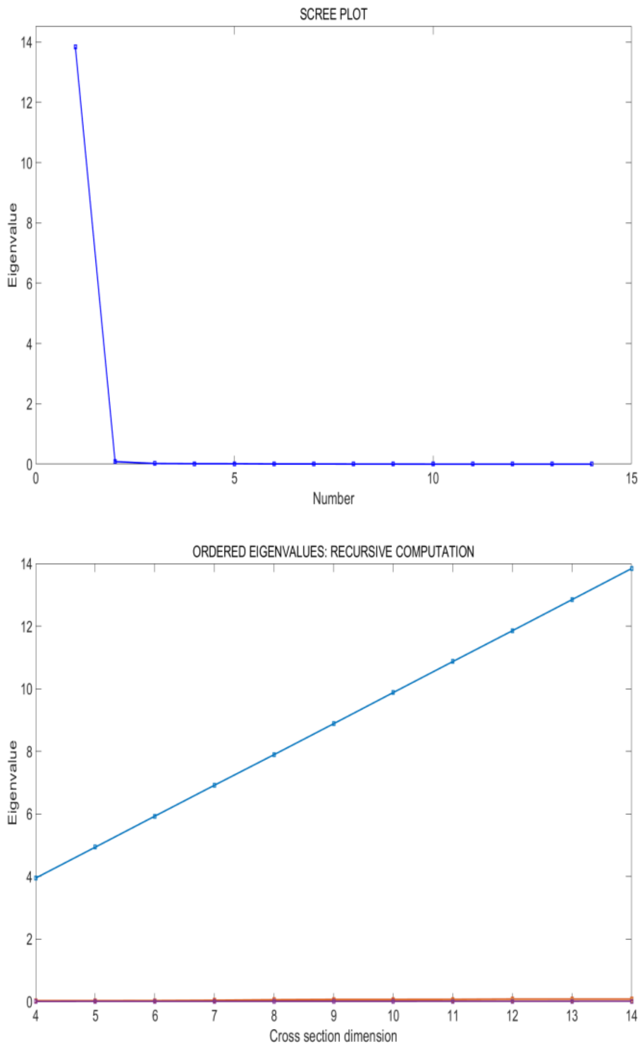
Figure 11b: Monte Carlo projections of the survival curves (women)



Again, conditioned projections are generated through the estimation of a factor model to a panel of European countries. Using data from Eurostat, we have selected 14 countries with data up to the mid-1980s. The selected countries are Belgium, Denmark, Ireland, Germany, Greece, Spain, Italy, Netherlands, Austria, Portugal, Finland, Sweden, Norway and Switzerland. The time period is 1985-2016. Yet again, we have formed the panel by considering a sample of at least 10 countries capable of providing the longest possible series.

After standardization of the observed life expectancies for all the analyzed countries, the scree plots of the eigenvalues of the correlation matrix (Figure 12) suggest that only one factor is needed to account for most of the common variance of the data:

Figure 12: Scree plots of the European life expectancy at birth



The estimation of the one-factor model by PC yields the following results:

Table 4: Factor model for the European life expectancy at birth

Country	Loadings	Communalities	Scores
Austria	0.9953	0.9906	0.0689
Belgium	0.9964	0.9928	0.0904
Denmark	0.9881	0.9764	0.0272
Finland	0.9969	0.9938	0.1051
Germany	0.9928	0.9857	0.0450
Greece	0.9947	0.9895	0.0616
Ireland	0.9937	0.9875	0.0516
Italy	0.9946	0.9893	0.0602
Netherlands	0.9865	0.9732	0.0239
Norway	0.9966	0.9932	0.0959
Portugal	0.9976	0.9951	0.1326
Spain	0.9954	0.9909	0.0707
Sweden	0.9937	0.9873	0.0510
Switzerland	0.9973	0.9946	0.1205
Median	0.9950	0.9901	0.0653
MAD	0.0024	0.0047	0.0265

The estimation of the one-factor model confirms the high degree of commonality among the selected European countries regarding survival (proxied as life expectancy), even higher than for fertility (proxied as the average number of children per woman). The common factor explains most of their non-stationary (upward-trended) behavior. Now the deviations of the individual synthetic survival measures with respect to their common factor are purely idiosyncratic (i.e. country-specific) and stationary.

Considering the Spanish case, it is noticeable its high linkage with the factors (i.e., with a loading slightly above the corresponding median). Again, this high commonality allows us to include international (European) information as a conditioning benchmark for the Spanish projections.

Considering the scores linked to the first factor and the long-term projections compiled by the United Nations, we have estimated a long-term benchmark (2100) for the European countries that involve life expectancy converging to 90.37 for men and 94.08 for women. The following table 5 presents the detailed calculations:

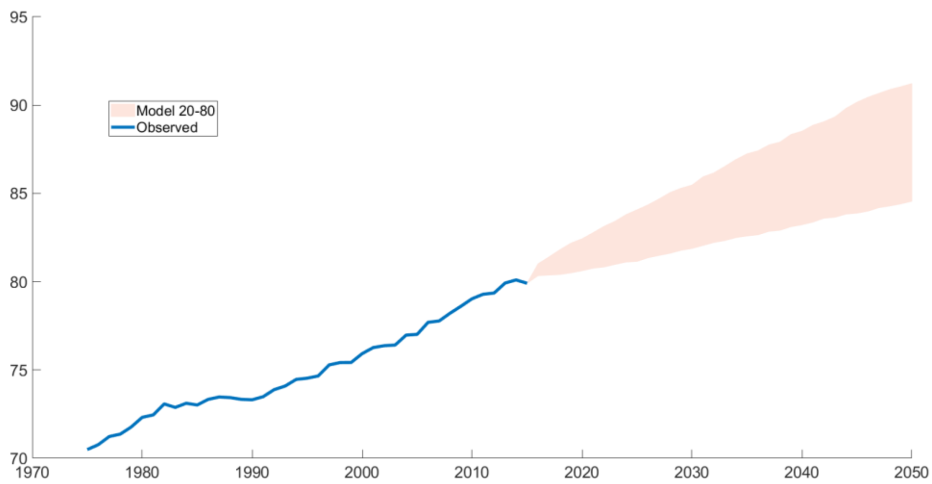
Table 5: Life expectancy at birth. Long-term benchmark (2100)

Country	Scores	Weights	UN 2095-2100 Men	Women	Both
Austria	0.07	0.07	90.33	93.97	92.09
Belgium	0.09	0.09	89.82	93.15	91.43
Denmark	0.03	0.03	89.81	92.03	90.88
Finland	0.11	0.10	89.51	94.02	91.69
Germany	0.04	0.04	89.92	93.16	91.49
Greece	0.06	0.06	89.87	93.63	91.69
Ireland	0.05	0.05	90.76	93.47	92.08
Italy	0.06	0.06	91.09	95.13	93.05
Netherlands	0.02	0.02	90.34	92.64	91.45
Norway	0.10	0.10	90.66	93.38	91.97
Portugal	0.13	0.13	89.78	94.54	92.09
Spain	0.07	0.07	90.57	95.96	93.18
Sweden	0.05	0.05	91.15	93.65	92.36
Switzerland	0.12	0.12	91.52	95.07	93.24
Sum / Weighted average		1.00	90.37	94.08	92.16

Note: Aggregation of men and women assumes a masculinity ratio at birth of 0.52.

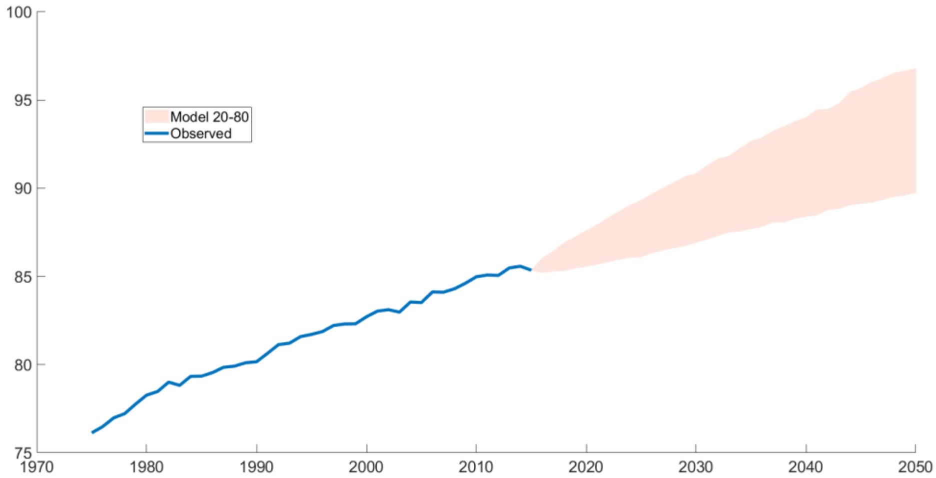
The corresponding life expectancy at birth conditional stochastic projections, represented as density forecasts, are depicted in the next figures.

Figure 13a: Life expectancy at birth (Men)



The corresponding projection for women:

Figure 13b: Life expectancy at birth (Women)



The projections eventually converge to a value close to 88 years for men and 93 for women in 2050. We should note that the uncertainty around those estimates is rather high, pointing to the difficulty of modeling and projecting survival.

6 Final remarks

After generating stochastic projections for the fertility curves and the survival probabilities curves – either conditional or unconditional –, we can plug them in the general cohort-component model to compile stochastic population projections. The remaining input of the model that for which projections are needed under this framework is namely migration flows. This can be introduced in the model using the results provided by Fernández-Huertas and López-Molina (2018), which are based on a multilateral gravity model to forecast migration flows up to 2100.

The modeling approach used to represent the joint dynamics of the curves considers them as a consistent object, a feature especially convenient since their domain (age) has a clear temporal dimension. This consistency allows us to represent the dynamics without losing information about the cross-section dimension.

Finally, the two-step approach provides a tractable way to incorporate the information provided by a cross-section of countries, avoiding most of the complexities linked to other procedures (Alkema et al., 2011; Raftery et al., 2012, 2013).

7 References

Alkema, L., Raftery, A.E, Gerland, P., Clark, S.J., Pelletier, F., Buettner, T. and Heillig, G.K. (2011). Probabilistic Projections of the Total Fertility Rate in All Countries. *Demography*. August; 48(3): 815–839.

Auerbach, A.J., and Kotlikoff, L.J. (1987). Dynamic Fiscal Policy. Cambridge University Press.

Diebold, F. and Rudebusch, G. (2013). Yield Curve Modeling and Forecasting. Princeton University Press.

Fernández-Huertas, J. and López-Molina, G. (2018). Predicting Spanish Emigration and Immigration, AIReF, Working Paper.

Gagnon, E., Johannsen, B.K., and Lopez-Salido, D. (2016). Understanding the New Normal: The Role of Demographics. Discussion Series 2016-080. Board of Governors of the Federal Reserve System.

Girosi, F. and King, G. (2008). Demographic Forecasting: Methods, Models, and Empirics. Princeton University Press.

Heer, B. and Maussner, A. (2009). Dynamic General Equilibrium Modeling. Springer.

Huynh, H.T., Lai, V.S., and Soumaré, I (2008). Stochastic Simulation and Applications in Finance with MATLAB programs, John Wiley and Sons.

Luenberger, D.G. (1979). Introduction to Dynamic Systems: Theory, Models and Applications. John Wiley and Sons.

Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979). Multivariate Analysis. Academic Press.

Quilis, E.M. (2016). BayVAR: Bayesian Vector Autoregressive (BVAR) Modeling. MATLAB Central File Exchange.

Raftery, A.E., Li, N., Ševčíková, H., Gerland, P. and Heilig, G.K. (2012). Bayesian Probabilistic Population Projections for All Countries, *Proceedings of the National Academy of Sciences of the United States of America*, 109: 13915–13921.

Raftery, A.E., Chunn, J.E., Gerland, P. and Ševčíková, H. (2013). Bayesian Probabilistic Projections of Life Expectancy for All Countries, *Demography*, 50: 777-801.