UNIVERSIDAD CARLOS III DE MADRID
working papers

Working Paper 08 - 45
Statistics and Econometrics Series 12
September 2008

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

A METHODOLOGY FOR POPULATION PROJECTIONS: AN APPLICATION TO SPAIN

Andrés M. Alonso[*], Daniel Peña[†], and Julio Rodríguez[‡]

Abstract

This paper looks at projections for the Spanish population by sex and age for the period of 2005 to 2050. These were carried out using forecasts for birth and mortality rates, and migration. These rates are calculated using two main sources of information. First, a multivariate time series model was applied for the series of variables from the 1970 to 2001 period. Second a model was estimated for life expectancy and for a synthetic fertility index. Both sources of information were combined to obtain the forecasts for the rates. Immigration rates are predicted by assuming three possible scenarios based on the maximum proportion that immigrants will represent in the Spanish population. With these variables a structure of ages and sex for the Spanish population is estimated using a cohort component model.

Keywords: Population projections, time series, factorial model, bootstrap.

JEL Classification: C32, C53 and J11

[*] Departamento de Estadística. Universidad Carlos III de Madrid, C/ Madrid 126, 28903 Getafe (Madrid), e-mail: andres.alonso@uc3m.es

[†] Departamento de Estadística. Universidad Carlos III de Madrid, C/ Madrid 126, 28903 Getafe (Madrid), e-mail: daniel.pena@uc3m.es

[‡] Departamento de Análisis Económico: Economía Cuantitativa. Universidad Autónoma de Madrid, Campus de Cantoblanco, 28049 Madrid, e-mail: jr.puerta@uam.es

1 Introduction.

Studies on population forecasts are a high priority for national statistical organizations (in Spain, the National Statistical Institute (NSI)) or international organizations such as the United Nations or the World Bank. Population projections differ by geographic area, time horizons and type of use. Thus, for example, in regions, communities or autonomous states, forecast horizons to the nearest decade are usually used (see, IEA (1995) and IECM (2004)) whereas in projections on a national level the horizons are extended to several decades (see García-Ferrer and del Hoyo (1991), USCB (2000), INE (2001, 2004) and Hyndman and Booth (2008)). On the other hand, the demand for these types of long term projections is quite diverse: human resource planning; energy resource planning; health and pension system planning. This last element in particular has been widely studied in the Spain (see Herce et al (1995, 1996), Diez (2000), Jimeno (2002) and Ahn et al (2005) among others), mainly in relation to the hypothesis that the decrease in fertility and increase in life expectancy will lead to a reduction in the active population which must support the cost of Social Security benefits.

One aspect which has been less studied is that of population forecasting in different educational cycles. One study which focused on this area was carried out by the Statistical Institute of Andalusia (Spain) for the period of 1998 - 2016 (see INE (2000)). Recently, Alonso-Meseguer and Sosvilla-Rivero (2004) looked at the gross enrolment rate and carried out projections until the year 2050 on registration and educational system expenses under different scenarios of migration flow and the integration of immigrant populations. These scenarios are based on different hypotheses of the volume of future migratory flow which were done by the National Statistical Institute in its revision of Spanish population forecasts, using the 2001 Census. Alonso et al (2007) provided the demand projections for Spain's official educational stages in period 2005 - 2050.

The population forecasts can be obtained using a wide range of procedures and models: the cohort component method; aggregate time series analysis; micro simulation; structural models and functional data modelling (see Booth (2006) for a review on stochastic population forecasting). In this work a cohort component method and time series analysis were combined as proposed by Lee and Tuljapurkar (1994), and shown by Keilman et al (2002) and Wilson and Bell (2004) for the United States, Norway and Australia respectively. Specifically, we use a dynamic factor model similar to that proposed by Lee and Carter (1992) and Lee and Tuljapurkar (1994), but with the introduction of restrictions in the common factor in order to model the mortality and fertility rates by age groups. The component method requires the establishment of future paths for the three basic components of the population changes: mortality, fertility and migration. The prediction of these components is based on the sieve bootstrap procedure proposed by Alonso et al (2002, 2004). A major advantage of this combined approach is that we obtain distribution functions of the forecasts for mortality and fertility rates for each single age and sex. Regarding net migration rates, we center our attention in immigration and we propose an analysis of immigration using different growth models. Using these models gives us future growth of the immigrant population if we make assumptions regarding the upper bounds in the percentage of immigrants out of the total population.

The rest of this paper is divided into four sections. Section 2 presents the projection methodology of the populations by single-year age and sex. In Section 3 this methodology is applied to the Spanish data for period 1970 - 2001. Also forecasts for mortality, fertility and migration rates are generated from 2005 to 2050. Finally, Section 4 describes the results of the population projections and compares them to those published by the Spanish National Statistical Institute (INE, 2004).

2 General Forecasting Methodology

2.1 Introduction

The cohort component method is widely used by official organisms for population projections. This method is based on dividing initial population numbers into cohorts defined by age and sex, and on updating each new population period for each age and sex group keeping in mind the components of the demographic change: fertility, mortality and migration. The following system of equations defines the cohort component method:

$$\begin{aligned} P_{0,t}^{(s)} &= B_t^{(s)} - D_{0,t}^{(s)} + M_{0,t}^{(s)}, \\ P_{e,t}^{(s)} &= P_{e-1,t-1}^{(s)} - D_{e,t}^{(s)} + M_{e,t}^{(s)}, \text{ with } e \in \{1, 2, \ldots, e_{max}\}, \end{aligned} \tag{1}$$

where P denotes the population at time t, B, D and M denote births, deaths and net migration in the period $(t-1, t)$, respectively; e denotes age and s denotes sex and takes the values F, M, i.e. female and male. This system is completed with the following equations for the components of fertility, mortality and net migration:

$$\begin{aligned} B_t^{(s)} &= \textstyle\sum_e FR_{e,t}^{(s)} P_{e,t-1}^{(F)}, \text{with } e \in \{< 15, 15, \ldots, 49, \geq 50\}, \\ D_{e,t}^{(s)} &= MR_{e,t}(s) P_{e,t-1}^{(s)}, \\ M_{e,t}^{(s)} &= I_{e,t}(s) - ER_{e,t}(s) P_{e,t-1}^{(s)}, \end{aligned} \tag{2}$$

where I, FR, MR, and ER denote the immigration and the rates of fertility, mortality and emigration, respectively.

In order for the system (1) - (2) to provide population forecasts by age and sex it is necessary to have the future values of fertility, mortality and immigration rates, as well as the number of immigrants by age and sex. There are two principal sources of information for these rates. The first is historic information on specific rates, and the second is information about aggregates obtained from those rates. We will analyze the use of both these sources below.

It is possible to model and directly project each specific rate individually, but as Gutiérrez de Mesa (2003) points out, “absurd” results can appear because of the dependence of the synthetic indices in the specific rates. Nevertheless, direct estimation implies the modelling of hundreds of time series which, in general, are not independent.

One method for predicting mortality curves is that proposed by Heligman and Pollard

(1980). These authors use models of mortality curves by ages, such as the following:

$$MR_e = A_t^{(e+B_t)^{C_t}} + D_t \exp\left(-E_t(\ln e - \ln F_t)^2\right) + \frac{G_t H_t^e}{1 + G_t H_y^e}, \tag{3}$$

where A_t, B_t, C_t, D_t, E_t, F_t, G_t and H_t are the parameters to be estimated. The parameters are adjusted so that life expectancies are similar to those being projected. This method is used by IECM (2004) for the Madrid Autonomous Community.

An alternative to direct projection of mortality rates by age is to obtain time series models for the eight parameters of the mortality curve from Heligman and Pollard which were estimated previously in the available years. We see that this methodology reduces the number of time series to be analyzed, from 86 series (in our case) per sex to only eight series. Using ARIMA models, McNown and Rogers (1989) carried out mortality projections in the United States to the year 2000. Felipe et al (2002) used a similar procedure to model changes in mortality in Spain between 1973 and 1993, and carried out projections for the period 1994 - 2010.

Finally, a method which considers the dependence between the time series proposed by Lee and Carter (1992) is to relate mortality rates by age to a single non-observable factor:

$$\begin{aligned} \ln(MR_{e,t}^{(s)}) &= a_e^{(s)} + b_e^{(s)} k_t^{(s)} + \varepsilon_{e,t}^{(s)}, \\ k_t^{(s)} &= c^{(s)} + k_{t-1}^{(s)} + \eta_t^{(s)}, \end{aligned} \tag{4}$$

where $a_e^{(s)}$ and $b_e^{(s)}$ are parameters which depend on age, e; $k_t^{(s)}$ is the non-observable factor which includes the general characteristics of mortality in year t and is called the mortality index, and $\varepsilon_{e,t}^{(s)}$ is the error term and includes the characteristics pertaining to each age e which are not picked up by the model. The second equation in (4) establishes that the factor k_t follows an ARIMA(0, 1, 0) with a non-null constant. Recently, Hyndman and Booth (2008) proposes a functional version of Lee-Carter approach and they obtain mortality, fertility and migration forecasts for Australia.

Both with McNown and Roger's (1989) procedure as well as with that of Lee and Carter (1992) it is possible to construct prediction intervals for the mortality rates using the relationship between these rates and the forecast distributions of the parameters of Heligman and Pollard's curve or of the mortality factor, respectively.

As in the case of mortality rates, a factorial model can be written for fertility rates (see Lee and Tuljapurkar (1994)):

$$\begin{aligned} FR_{e,t}^{(s)} &= c_e^{(s)} + d_e^{(s)} f_t^{(s)} + \upsilon_{e,t}^{(s)}, \\ f_t^{(s)} &= f_0^{(s)} + \phi f_{t-1}^{(s)} + \nu_t^{(s)} + \theta^{(s)} \nu_{t-1}^{(s)}, \end{aligned} \tag{5}$$

where $c_e^{(s)}$ and $d_e^{(s)}$ are parameters which depend on age, e; $f_t^{(s)}$ is the non-observable factor which includes general characteristics of fertility in year t and is called the fertility index, and $\upsilon_{e,t}^{(s)}$ is the error term which, as before, includes the characteristics pertaining to each age e which are not captured by the model. In this case, the superscript s refers to the fertility rate of male or female offspring. The second equation in (5) establishes that the factor $f_t^{(s)}$ follows the ARMA (1,1) model with a non-null constant. In Lee and Tuljapurkar (1994), in the second equation the value of the constant is set so that the mean value of the synthetic fertility index, $c + \mathrm{E}[f_t]$ is equal to 2.1, where $c = \sum_e c_e$.

The second source of data is that of synthetic indices associated with each type of rate. In this case the indices are modelled and specific rates are obtained indirectly. With mortality rates, for example, some authors have proposed projecting female life expectancy by specifying a value for the last year of the forecasting horizon. For the remaining years an interpolation based on the logistic function was used. This procedure is the one used in the projections carried out by the Statistical Institute of Andalusia (IEA, 1995) and the Madrid Autonomous Community (IECM, 2004), where the values 84.0 in 2025 and 86.3 were established as the upper "bound" for female life expectancy. In both studies, male life expectancy was handled using an adjustment which takes into account the higher male mortality rate.

Establishing an upper bound for life expectancy is no simple matter. We take, for example, the predictions for life expectancy in women for the year 2025 which were carried out by the National Statistical Institute: 85.44 in "Proyecciones de población calculadas a partir del Censo de Población de 1991. Evaluación y Revisión" (Population Forecasts taken from the

1991 Census: Evaluation and Review) published in 2001 and which used available data until 1996; and 86.48 in “Proyecciones de población calculadas a partir del Censo de Población de 2001” (Population Forecasts taken from the 2001 Census: Evaluation and Review) published in 2004. In these publications it is assumed that life expectancy will remain constant from 2026 and from 2031, respectively. There is no consensus in the demographic literature either on biological limits which influence the aging process itself. These elements are the principal reason for the methodology which we propose in Section 3 due to uncertainty about the value of this limit.

In the case of fertility rates by single-year ages, the forecast of the synthetic fertility index was used: $SFI_t = \sum_e FR_e$ which is complemented by the projection of mean age of the woman at time of children's birth, MAF_t, and its variability, VAF_t. Subsequently a theoretical model was used for the fertility curve. Specifically, in the IEA (1995) and IECM (2004) a gamma type function was used:

$$FR_{e,t} = \frac{A_t B_t^{C_t} (e-15)^{C_t - 1} \exp(-B_t(e-15))}{\Gamma(C_t)}, \tag{6}$$

where $A_t = SFU_t$, $B = \frac{MAF_t}{VAF_t}$ and $C = \frac{MAF_t^2}{VAF_t}$.

Finally, we would like to point out that a weakness in the usual forecasting methods is that they use a single future path commonly interpreted as the mean or most “likely”. Frequently, official organisms establish other future scenarios for fertility, mortality and migration around these mean values. Generally, between three and seven scenarios are reported, with three being the most often used. The terms high, medium and low hypothesis are commonly used in these cases (see, e.g., Lee and Tuljapurkar (1994)). There are several inconveniences associated with this method of scenarios (see comments in Keilman et al (2002)), for example, the “intervals” defined by the low and high hypotheses do not have an error measurement associated with them. In the following sections we use a methodology which allows us to resolve these inconveniences by modifying the procedures of Lee and Carter (1992) and Lee and Tuljapurkar (1994).

2.2 General Methodology for Forecasting Mortality and Fertility

In this section we develop a procedure for modelling life rates, separated into different groups, which allows us to obtain the forecast distributions of the future values of these rates for a pre-established horizon. Here we mention life rates in order to include both mortality rates and fertility rates in the presentation, since as we will see in the proof, both model (4) for mortality rates as well as model (5) for fertility rates, are special cases of the dynamic factor model. This model has been studied by Geweke (1977), Peña and Box (1987), Tiao and Tsay (1989), Gonzalo and Granger (1995) and Peña and Poncela (2004), among others. Ortega and Poncela (2005) used this model for the analysis of Southern European fertility rates. We also refer to rates in certain groups and in such cases we talk of groups of individuals who share one or more characteristic such as same age and/or same sex. This allows us to make a general formulation which in the next section will be specified for the corresponding rates and groups.

2.2.1 Dynamic Factor Model

In this section we follow the presentation of the dynamic factor model carried out by Peña and Poncela (2004). Let $\{\boldsymbol{y}_t\}_{t\in\mathbb{Z}}$ be a vector series of dimension m, for example, male mortality rates of m age groups. The dynamic factor model assumes that the components of the vector series, $\boldsymbol{y}_t$, can be written as a linear combination of r common factors plus an error term:

$$\underset{m\times 1}{\boldsymbol{y}_t} = \underset{m\times r}{\boldsymbol{P}}\ \underset{r\times 1}{\boldsymbol{f}_t} + \underset{m\times 1}{\boldsymbol{\varepsilon}_t}, \tag{7}$$

where $\boldsymbol{f}_t$ is the r-dimensional vector of common factors, $\boldsymbol{P}$ is the weight matrix of factors, and $\boldsymbol{\varepsilon}_t$ is the specific factors vector or error term. Additionally, it is assumed that the vector of common factors follows a VARIMA(p, d, q) model defined by:

$$\underset{r\times r}{\boldsymbol{\Phi}(B)}\ \underset{r\times 1}{f_t} = \underset{r\times r}{\boldsymbol{\Theta}(B)}\ \underset{r\times 1}{\boldsymbol{v}_t}, \tag{8}$$

where B is the backward shift operator, $\boldsymbol{\Phi}(B) = \boldsymbol{I} - \boldsymbol{\Phi}_1 B - \cdots - \boldsymbol{\Phi}_p B^p$ and $\boldsymbol{\Theta}(B) = \boldsymbol{I} - \boldsymbol{\Theta}_1 B - \cdots - \boldsymbol{\Theta}_q B^q$ are polynomial matrices satisfying the roots of $|\boldsymbol{\Phi}(B)| = 0$ and of $|\boldsymbol{\Theta}(B)| = 0$ are not found within the unit circle. We assume that the innovations $\boldsymbol{v}_t$ are uncorrelated for all lags, i.e., $E(\boldsymbol{v}_t \boldsymbol{v}'_{t+h}) = \mathbf{0}$ for $h \neq 0$ and also they are non-correlated with the specific factors, i.e., $E(\boldsymbol{v}_t \boldsymbol{\varepsilon}'_{t+h}) = \mathbf{0}$ for all h. In Peña and Poncela (2004) it is supposed that the innovations and specific factors are distributed as a $\mathcal{N}_m(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$ and $\mathcal{N}_r(\mathbf{0}, \boldsymbol{\Sigma}_v)$, respectively. In this paper specific factors are allowed to follow stationary univariate models.

The factorial model defined by (7) and (8) is not identified since for any non-singular matrix of dimension $r \times r$ it is possible to express the vector series, $\boldsymbol{y}_t$, as a new set of factors and weights. Many restrictions have been proposed to solve the problem of identification, e.g., $\boldsymbol{\Sigma}_v = \boldsymbol{I}$ or $\boldsymbol{P}'\boldsymbol{P} = \boldsymbol{I}$ (see, for example, Peña and Poncela (2004)) and $\boldsymbol{P} = [p_{i,j}]$ with $p_{i,j} = 0$ for $j > i$ (see Harvey (1989)). In this paper we use the restriction and assume that the factors are orthogonal, i.e. $f_{\cdot,i} \perp f_{\cdot,j}$ for $i \neq j$, as in Lee and Carter (1992) and Lee and Tuljapurkar (1994).

2.2.2 Bootstrap Procedure for Forecasting

In this section we present a bootstrap procedure for constructing forecasting intervals based on a modification of the procedure proposed by Alonso et al (2002, 2004).

(1) The factorial model defined by (7) is estimated using the singular values decomposition as in Lee and Carter (1992). From this, the estimations for the r common factors $\widehat{\boldsymbol{f}}_t = (\widehat{f}_{t,1}, \widehat{f}_{t,2}, \ldots, \widehat{f}_{t,r})'$ and the weight matrix $\widehat{\boldsymbol{P}}$ are obtained.

(2) The residuals of the factorial model are calculated:

$$\widehat{\varepsilon}_t = \boldsymbol{y_t} - \widehat{\boldsymbol{P}} \widehat{\boldsymbol{f}}_t. \tag{9}$$

(3) An AR(p_e) model is chosen for $\widehat{\boldsymbol{\varepsilon}}_{e,\cdot}$ with $e \in \{1, 2, \ldots, m\}$, using the BIC criteria. An ARI (p_s, d_s) model is chosen for common factors with $s \in \{1, 2, \ldots, r\}$. As a result we obtain estimations of the autoregressive parameters, the order of differences, and the residuals of the models AR and ARI.

(4) The empirical distribution function is obtained for the centered residuals of the AR and ARI models:

$$\widehat{F}_{\widetilde{\epsilon}_e}(x) = (n - p_e)^{-1} \sum_{t=p_e+1}^{n} I\left(\widetilde{\epsilon}_{e,t} \leq x\right), \tag{10}$$

and

$$\widehat{F}_{\widetilde{\upsilon}_s}(x) = (n - p_s - d_s)^{-1} \sum_{t=p_s+d_s+1}^{n} I\left(\widetilde{\upsilon}_{s,t} \leq x\right), \tag{11}$$

where $\tilde{\epsilon}_{e,t} = \widehat{\epsilon}_{e,t} - \widehat{\epsilon}_e^{(\cdot)}$ and $\widehat{\epsilon}_e^{(\cdot)} = (n-p_e)^{-1} \sum_{t=p_e+1}^{n} \widehat{\epsilon}_{e,t}$ with $e \in \{1, 2, \ldots, m\}$; $\tilde{\upsilon}_{s,t} = \widehat{\upsilon}_{s,t} - \widehat{\upsilon}_s^{(\cdot)}$; and $\widehat{\upsilon}_s^{(\cdot)} = (n - p_s - d_s)^{-1} \sum_{t=p_s+d_s+1}^{n} \widehat{\upsilon}_{s,t}$ with $s \in \{1, 2, \ldots, r\}$

(5) A resample ϵ_t^* of i.i.d. observation from $\widehat{F}_{\widetilde{\epsilon}}$ and a resample υ_t^* of i.i.d. observation from $\widehat{F}_{\widetilde{\upsilon}}$ were selected.

The following are the forecasting steps. The final p_e and $d_s + p_s$ observations are fixed from the AR and ARI models, respectively.

(6) The future bootstrap observations are calculated for common and specific factors using the relations:

$$\varepsilon^*_{e,T+h} = -\sum_{j=1}^{p_e} \widehat{\phi}_{e,j}(\varepsilon^*_{e,T+h-j} - \bar{\varepsilon}_e) + \epsilon^*_{e,T+h}, \tag{12}$$

and

$$f^*_{s,T+h} = \sum_{j=1}^{p_s+d_s} \widehat{\phi}_{s,j} f^*_{s,T+h-j} + \upsilon^*_{s,T+h}, \tag{13}$$

where $h > 0$, $\varepsilon^*_{e,t} = \widehat{\varepsilon}_{e,t}$ for $t \leq T$ and $f^*_{s,t} = \widehat{f}_{s,t}$ for $t \leq T$, with T being the last available year.

(7) The future bootstrap observations are calculated for vector $\boldsymbol{y}$ using the relation:

$$\boldsymbol{y}^*_{T+h} = \widehat{\boldsymbol{P}} \boldsymbol{f}^*_{T+h} + \boldsymbol{\varepsilon}^*_{T+h}, \tag{14}$$

where $\boldsymbol{f}^*_{T+h} = (f^*_{1,T+h}, f^*_{2,T+h}, \ldots, f^*_{r,T+h})'$ and $\boldsymbol{\varepsilon}^*_{T+h} = (\varepsilon^*_{1,T+h}, \varepsilon^*_{2,T+h}, \ldots, \varepsilon^*_{m,T+h})'$.

Finally, $F^*_{y^*_{s,T+h}}$, the bootstrap distribution function of $y^*_{s,T+h}$ is used as the estimator of the conditional distribution of $y_{s,T+h}$ given the sample. As usual, the $\widehat{F}^*_{y^*_{s,T+h}}$ estimation is obtained by repeating B times the steps 5 to 7. A $(1-\alpha)\%$ forecast interval for $y_{s,T+h}$ is the following:

$$\left[Q^*(\alpha/2), Q^*(1-\alpha/2)\right], \tag{15}$$

where $Q^*(\cdot) = \widehat{F}^{*-1}_{y^*_{s,T+h}}(\cdot)$ are the quantiles of the estimated bootstrap distribution.

In the demographic data that we analyze in the following sections, we see that there is a high correlation between the first factor of model (7) and a synthetic rate indicator, i.e. with a function of $\boldsymbol{y}_t$. Specifically, in the case of mortality rates this correlation is given with the life expectancy at birth, and in the case of fertility rates, with the synthetic fertility index. This allows us to establish a simple model between the first factor, $f_{1,t}$, and the synthetic index, i_t:

$$f_{1,t} = \alpha_0 + \alpha_1 i_t + \iota_t, \tag{16}$$

where ι_t assumes that it follows an AR(p_ι) model. The previous model together with a specific modelling of the synthetic index (which is found in sections 3.1.1. and 3.2.1 for life expectancy at birth and synthetic fertility index, respectively) allow us to make forecasts for future values of this factor.

3 Application of Spanish Data

3.1 Application of Spanish Data: Mortality

In this section we obtain the forecasts for mortality rates by age and sex for the Spanish population using a dynamic factor model which utilizes the historical series of these rates in Spain during the period 1970 to 2001. At first, the unifactorial model used by Lee and Carter (1992) was considered, and its goodness of fit was analyzed for the Spanish data. Next, the bifactorial model was considered. With both models we detected an effect of high male mortality rates which some authors have attributed to deaths in traffic accidents or deaths associated to AIDS (see, for example, Felipe et al (2002) and IECM (2004)). Both causes have subsided in the last years of the available data; therefore for a long term projection it is not desirable to introduce this transitory effect. Thus, we propose a procedure for eliminating this effect from the forecasts (see detailed study in Alonso et al (2005)).

On the other hand, as with Lee and Carter (1992), we showed that the first factor of the model for mortality rates is highly correlated with life expectancy at birth. This motivated the study of a model for life expectancy at birth in which we impose the existence of an upper limit. Finally, the existing relationship between the first factor and life expectancy allowed us to introduce restrictions in the forecasts of mortality rates by single-year ages and sex.

3.1.1 Life Expectancy Analysis

In this section we propose a method for establishing an upper bound for life expectancy or, more specifically, for establishing a distribution function for this bound. It is important to point out the close relationship between life expectancy and the first factor of mortality; the correlation between both is: -0.9855 in men and -0.9976 in women. This allows us to establish restrictions on the mortality factor through restrictions on life expectancy. The procedure is as follows:

- The following transformation is considered for life expectancy at birth, LEB_t:

$$Y_{t,A} = \ln \frac{LEB_t}{A - LEB_t}, \tag{17}$$

  where A is the upper bound for life expectancy, which we assume in the interval $\mathcal{A} = [76, 100]$ in men and $\mathcal{A} = [84, 105]$ in women. The lower bounds for these intervals take into account the most recent available data. As opposed to the procedures described in IEA (1995) and IECM (2004), no single value was set for A, instead, A is considered a parameter of the model.

- For each value of parameter A, a different series, $Y_{t,A}$, is obtained and a model ARIMA($\widehat{p}_A$, 1,0) is selected using the BIC criteria. Next, forecasts were carried out with this model using a modification of the sieve bootstrap procedure proposed by Alonso et al (2004) which allows for taking into account the uncertainty associated in the selection of $\widehat{p}_A$.

- Forecasting the LEB is based on a combination of the forecasts of the $Y_{t,A}$. Thus, the distribution of future values for life expectancy are obtained using:

$$F_{LEB_{t+h}}(x) = \int_{a \in \mathcal{A}} F_{LEB_{t+h},a}(x) f_{A|\boldsymbol{LEB}}(a)\, \mathrm{d}a, \tag{18}$$

where $F_{LEB_{t+h},a}(\cdot)$ is the distribution function of LEB_{t+h} calculated assuming that the bound for life expectancy is a, and $f_{A|\boldsymbol{LEB}}(\cdot)$ is the density function of A conditional to the observed sample, $\boldsymbol{LEB}$.

Expression (18) is no more than a formulation of Bayesian Model Averaging with the introduction of uncertainty regarding the two unknown elements in the model: A and $\widehat{p}_A$. Similar approach had been used in prediction using polynomial regression of unknown degree (see Guttman et al. 2005) and using nonparametric regression (see Peña and Redondas, 2006). One advantage of this procedure is that it allows us to calculate the a posteriori distribution of the upper bound of life expectancy and therefore (in light of available data) the likelihood of the bounds used in others researches.

The use of the expression (18) with A taking values in an interval, is complex and in this paper we propose a discretization of 0.5 years, thus, the distribution of future values of life expectancy are obtained using:

$$F_{LEB_{t+h}}(x) = \sum_{a \in \mathcal{A}} F_{LEB_{t+h},a}(x) \Pr\{A = a|\boldsymbol{LEB}\}, \tag{19}$$

where $F_{LEB_{t+h},a}(\cdot)$ is the distribution function of LEB_{t+h} calculated assuming that the bound for life expectancy is a, and $\Pr\{A = a|\boldsymbol{LEB}\}$ is the probability of A conditional to the observed sample. This probability can be approximated using:

$$\Pr\{A = a|\boldsymbol{LEB}\} = \frac{\alpha_a \exp(-1/2\mathrm{BIC}(a))}{\sum_{a \in \mathcal{A}} \alpha_a \exp(-1/2\mathrm{BIC}(a))}, \tag{20}$$

where BIC(a) is the value of the BIC criteria in the model with bound a (see, Kass and Raftery, 1995).

In Figure 1 we present the estimated distribution of the upper bounds of life expectancy in men and women. We have used a circle to represent the a posteriori means of the bounds: 82.30 and 89.73, respectively. Figure 2 shows the prediction mean of life expectancy in men and women. For example, for the years 2025 and 2050 the prediction mean are: 78.63 and 80.29 in men and 86.23 and 97.98 in women.

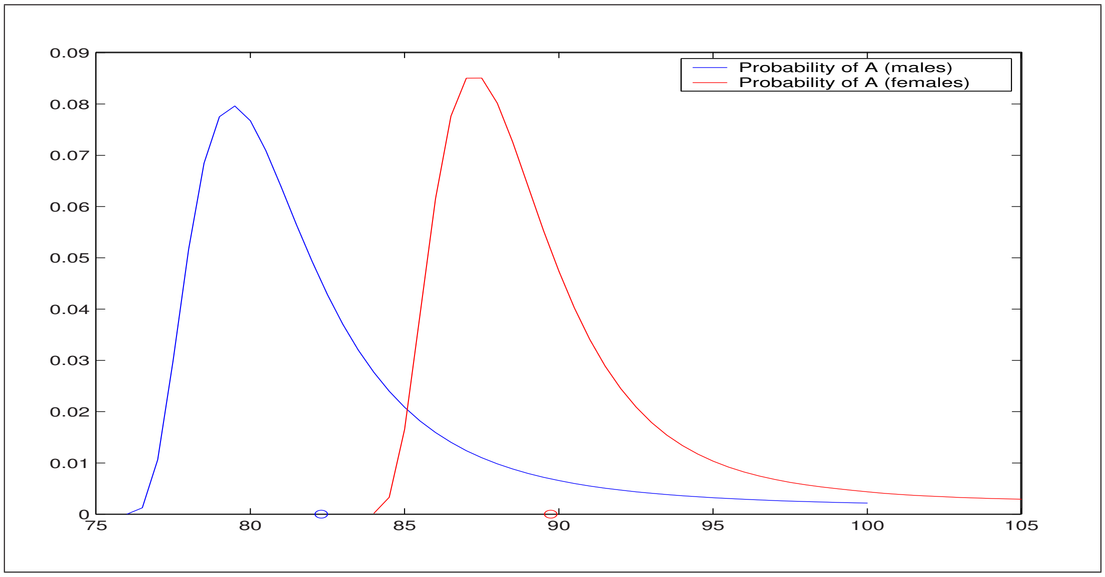

Figure 1. A posteriori probability of the upper bounds of life expectancy.

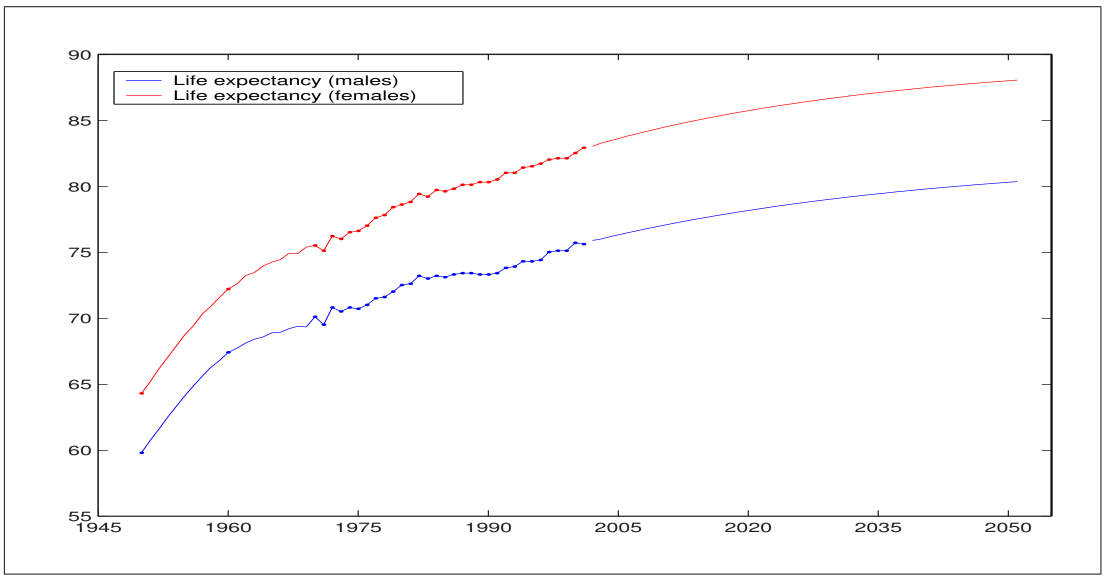

Figure 2. Forecasts using model averaging and the observed values in life expectancy at birth. Spain 1950 - 2050.

In Figure 3 we present the fan chart (see Wallis, 1999) of the bootstrap distribution of life expectancy forecasts. This method of representing uncertainty in forecasts is being employed more and more by institutions which carry out macroeconomic forecasting. A fan chart represents the forecast intervals of different levels. Specifically, in Figure 3 we represent the intervals at 20%, 40%, 60%, 80% and 90%, in addition to the median of the forecasts. First, we observe the asymmetry of these distributions; this element cannot be visualized when using intervals which are symmetric with respect to the mean of the predictions, as in Keilman et al (2002) and Wilson and Bell (2004). Second, we can evaluate the assumptions or projections carried out in earlier works. Thus, for example, the projections carried out by the National Institute of Statistics (INE, 2004) fall in the 80% forecast interval for men and 70% for women.

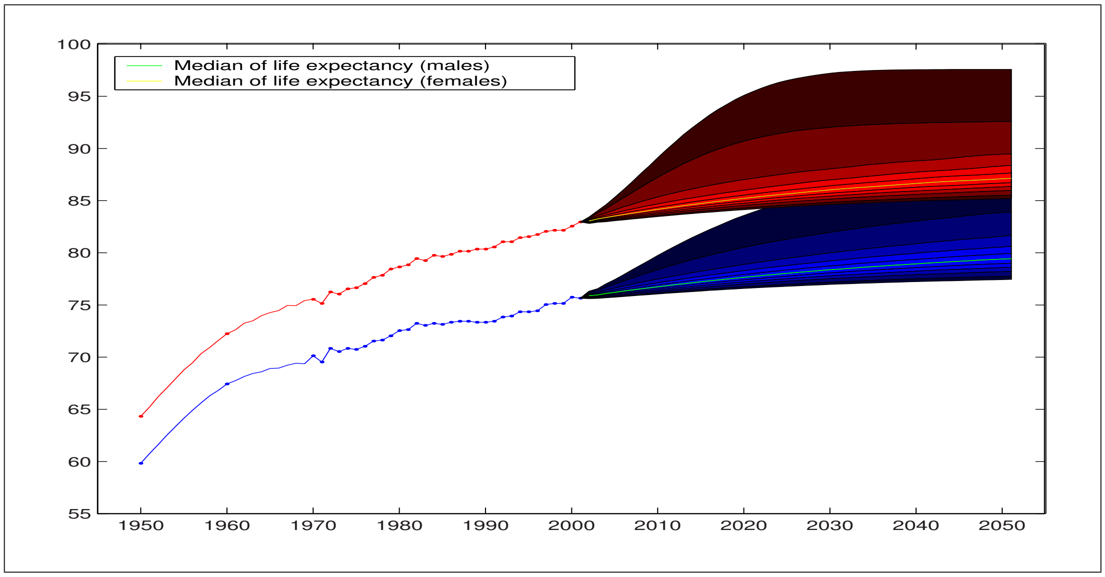

Figure 3. Fan chart using model averaging and the observed values of life expectancy at birth. Spain 1950 - 2050.

3.1.2 Forecasts of Mortality Rates by Age and Sex

Once we have obtained the predictions for life expectancy at birth we can find the corresponding predictions for mortality factors using model (16) and, using model (4) in which we allow the errors or specific age factors, $\varepsilon_{e,t}$, to follow an AR(p_{ε_e}) model, we obtain the forecast densities for mortality rates by age. To illustrate this further, in Figures 4 and 5 we show a fan chart and the forecast densities of mortality rates during the first year of life. In both, a clear reduction is observed in mortality in this age group. In Figure 6 we present the fan chart of the mortality curve for the years 2025 and 2050. The remaining ages and years can be obtained using the routines developed in this paper and which are available from the authors upon request.

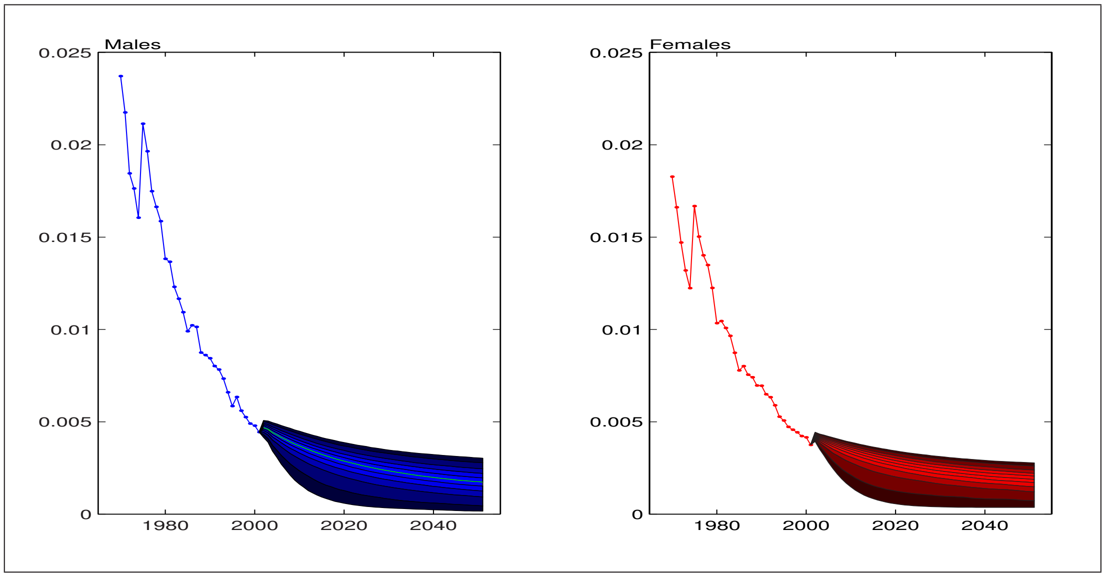

Figure 4. Fan chart using model averaging and the observed values of infant mortality rate. Spain 1970 - 2050.

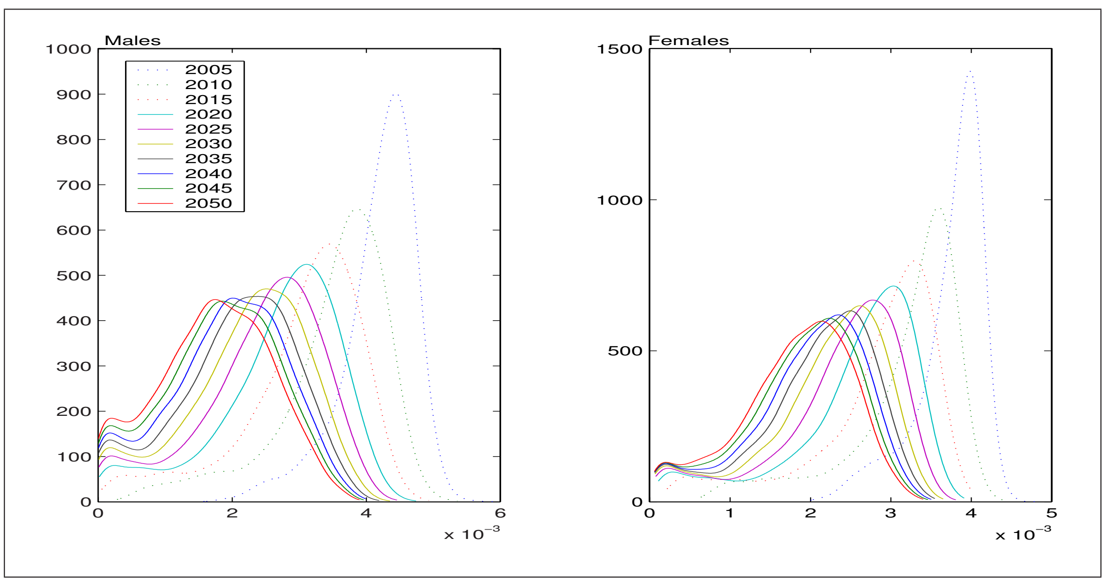

Figure 5. Forecast densities of infant mortality rates. Spain 2005 - 2050.

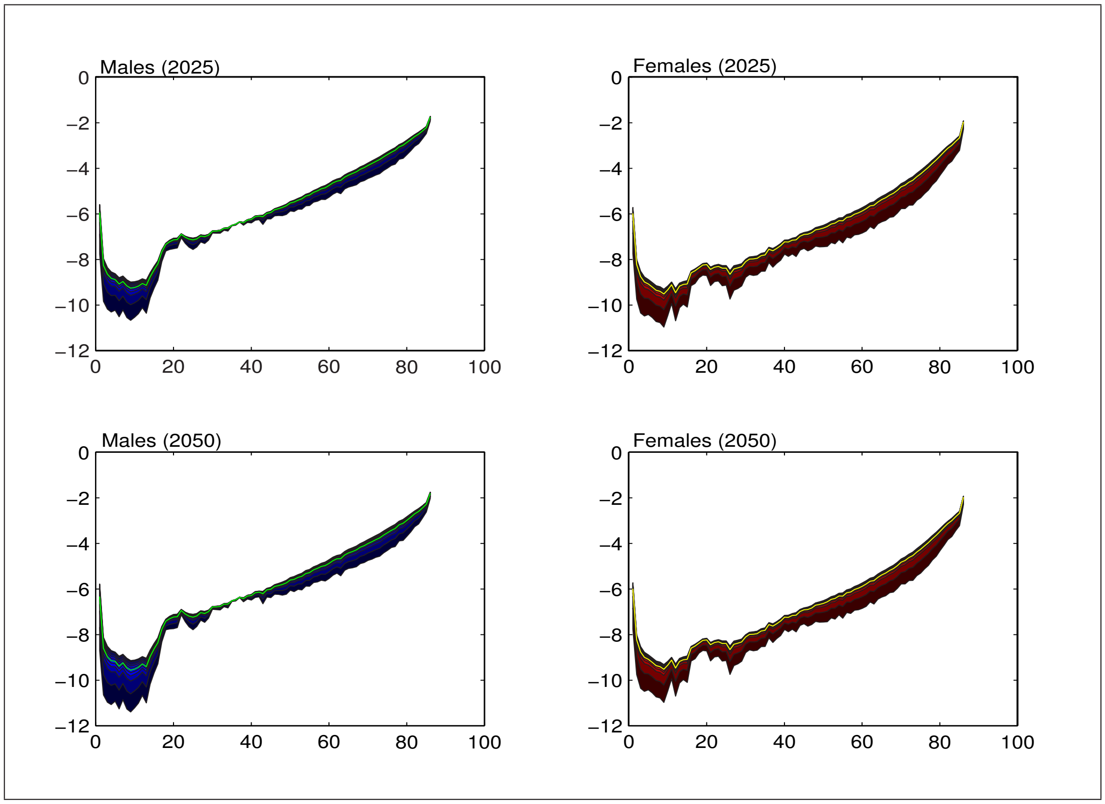

Figure 6. Fan chart using model averaging of the mortality curve. Spain 2025 - 2050. Fan chart using model averaging and the observed values of life expectancy at birth. Spain 1950 - 2050.

3.2 Application to Spanish Data: Fertility

In this section we obtain the projections for fertility rates by groups of age and sex for the Spanish population using a dynamic factor model. The first model considered was a unifactorial one used by Lee and Tuljapurkar (1994) and its goodness of fit was analyzed for the Spanish data. Later on a bifactorial model was considered where bimodality was clearly visible in the projections for fertility by ages.

On the other hand, we observed that the first common factor of the fertility rate model is highly correlated with the synthetic fertility index. This motivated the study of a model for this index similar to that proposed for life expectancy at birth. We also observed that the factors $f_t^{(M)}$ and $f_t^{(F)}$, as well as the weights, $d_e^{(M)}$ and $d_e^{(F)}$ have similar values. This suggests a joint modelling of births for both sexes as in Lee and Tuljapurkar (1994). Nevertheless, the relationship between the birth rate of males and females must be taken into account. In most countries this rate is between 105 and 107 males for every 100 females, and later this rate is modified owing to differences in patterns of mortality and migration in men and women. Finally, the existing relationship between the first factor and the synthetic index allows us to introduce restrictions in the projections of fertility rates by single-year ages.

3.2.1 Model for the Synthetic Fertility Index

In the following we propose a model for the synthetic fertility index which allows the distribution function of future values to be established. It is important to point out the close relationship between the synthetic fertility index and the first factor of fertility; the correlation is: 0.9737. This allows us to establish restrictions in the factor through the future values of the synthetic fertility index using model (16).

Similar to what was done in the previous section, we propose a transformation of the synthetic fertility index whose asymptotic behavior leads to an upper bound, B, of the number of children per women (similar to expression (17)). If, in the case of mortality, this bound can be explained by medical advances and the natural limits of the human organism, then in the case of fertility we understand that this bound can be explained by current socioeconomic conditions, the massive incorporation of women into the workforce and, therefore the opportunity costs that women considering maternity must face. As in the case of mortality we do not consider the bound as a single value but rather we assume an interval where the bound takes values. Specifically, we assume that B takes values in $\mathcal{B} \in [3, 8]$. We have set the lower limit of this interval according to the maximum indices reached during the years 1960 - 1970 and the upper limit according to the forecast intervals in an unrestricted model.

Figure 7 shows the estimated distribution of the upper bound of the synthetic fertility index. As in the above section we have used an interval discretization, specifically $\mathcal{B} = [3.0, 3.25, \ldots, 7.75, 8]$. A circle was used to indicate the mean a posteriori of the bound: 3.430.

Figure 8 shows the forecasts mean of the synthetic fertility index. These forecasts take

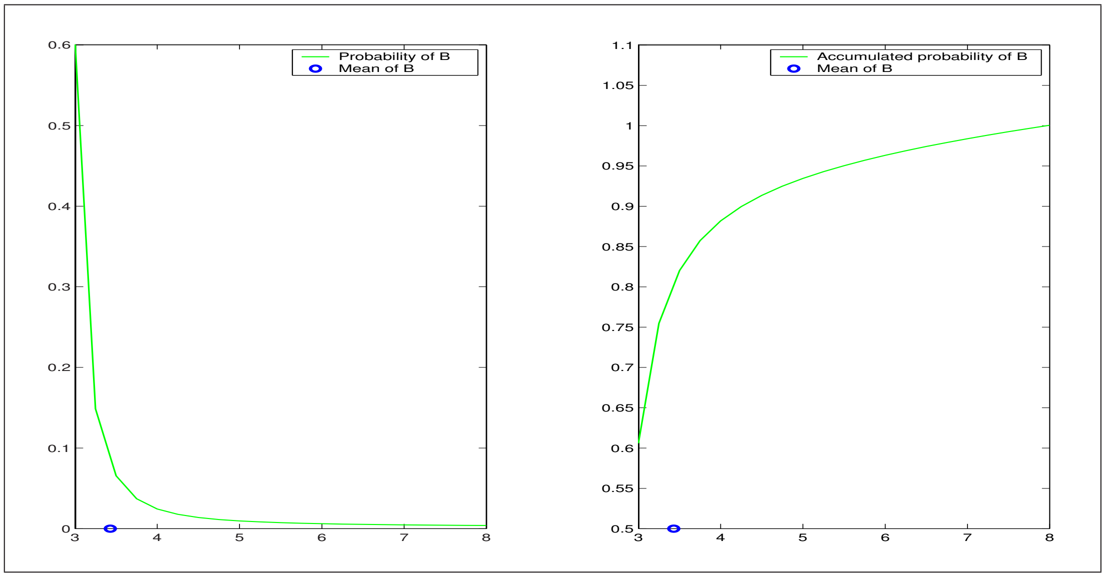

Figure 7. Probability and accumulated probability of the upper bound of the synthetic fertility index.

into account the uncertainty associated to the upper bound. For example, for the years 2025 and 2050 the mean forecasts are: 1.484 and 1.509, respectively. We also confirm that the forecasts tend towards an asymptote with a value nearing 1.510, slightly lower than the forecasts of the INE (2004).

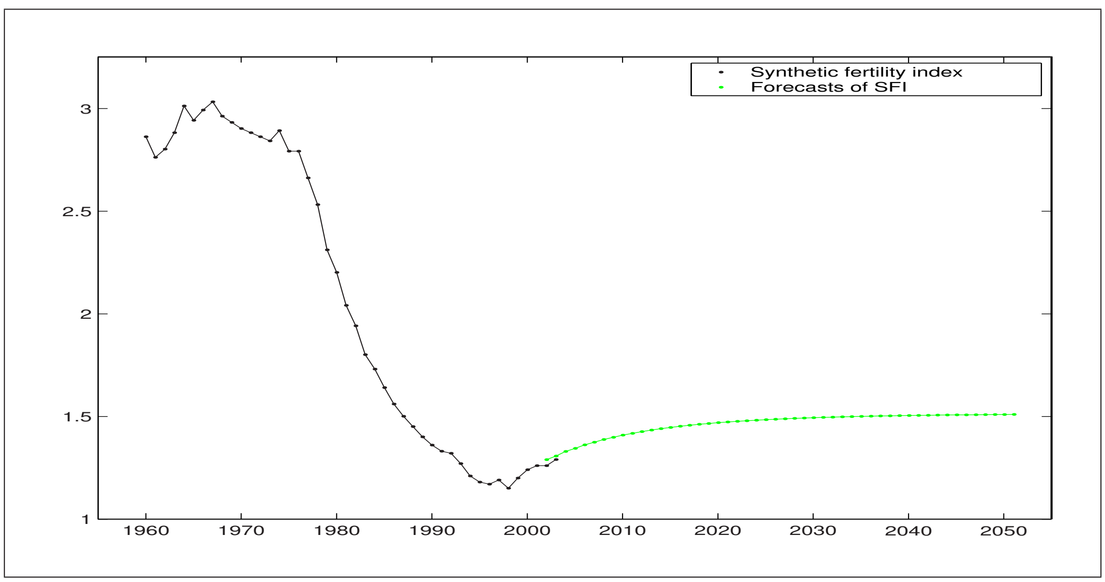

Figure 8. Forecasts using model averaging and the observed values of the synthetic fertility index. Spain 1960 - 2050.

An alternative to the transformations of type (17) used to establish the bound of these indices is to impose restrictions on future trajectories of the index being studied in unrestricted models. For example, in Keilman et al (2002) for the SFI in Norway, the restriction was that it belonged to a prefixed interval, $[0.5, 4]$. In Figure 9 we show the fan chart of the bootstrap distribution for the synthetic fertility index forecast, imposing the restriction that the trajectories during the entire forecast horizon be less than 5.5. Note that the value 5.5 corresponds to the 95% percentile of the distribution of the upper limit for the SFI (see Figure 7). In the figure a slight recovery can be made out (in the central values of the projection) in future fertility. Nevertheless, values below one child per women have a considerable probability. Thus, for example, the probability that the SFI is below one child per woman is slightly lower than 30% in 2025 and higher than 30% in 2050. On the other hand, the probability of having fertility rates greater than or equal to the replacement level (2.1, at the current mortality rate) is around 20% in 2025 and 30% in 2050.

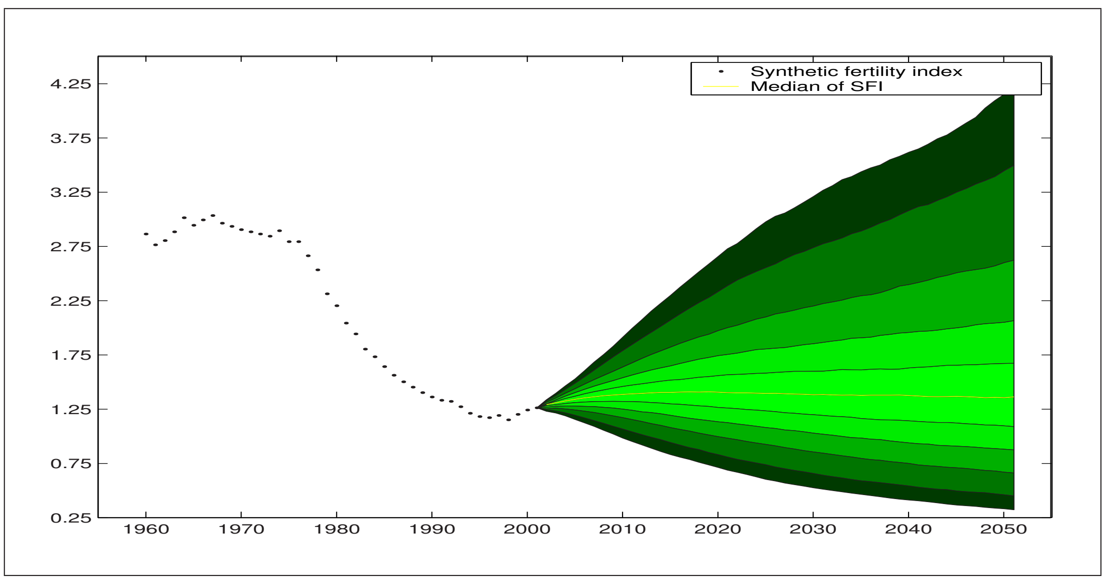

Figure 9. Fan chart using model averaging (with upper restricted trajectories) of the synthetic fertility index. Spain 1960 - 2050.

However, we see that the values used by Keilman et al (2002) for the SFI in Norway have a low probability in our case since (i) values lower than 0.5 have a probability below 5% until 2030, 10% to 2040 and only slightly above 10% in 2050, and (ii) values above 4 have a probability of less than 5% for almost the entire forecast horizon.

3.2.2 Forecast of Fertility Rates by Age

Once we have obtained the projection distributions of the synthetic fertility index we can obtain the corresponding distributions for the first factor of fertility using model (16). Then using the factorial model, in which we permit the errors or age specific factors, $\varepsilon_{e,t}$, to follow an AR(p_{ε_e}) model, we obtain the forecast densities for fertility rates by age. To further illustrate this point, in Figures 10 and 11 we depict a fan chart and the forecast densities of fertility rates for ages 20 and 30. In both we can see a tendency towards recovery in the central projections. Figure 12 depicts the fan chart of the fertility curve for the years 2025 and 2050 where, again, there is evidence of possible bimodality. The remaining ages and years can be obtained using the routines developed in this paper and are available from the authors upon request.

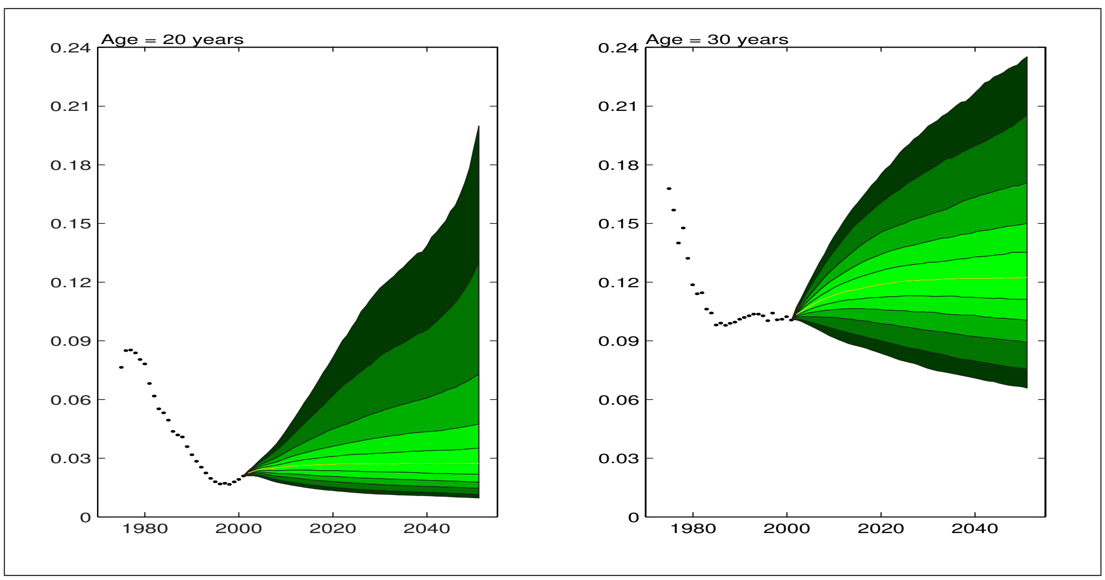

Figure 10. Fan chart using model averaging and the observed values of fertility rates for ages 20 and 30. Spain 1975 - 2050.

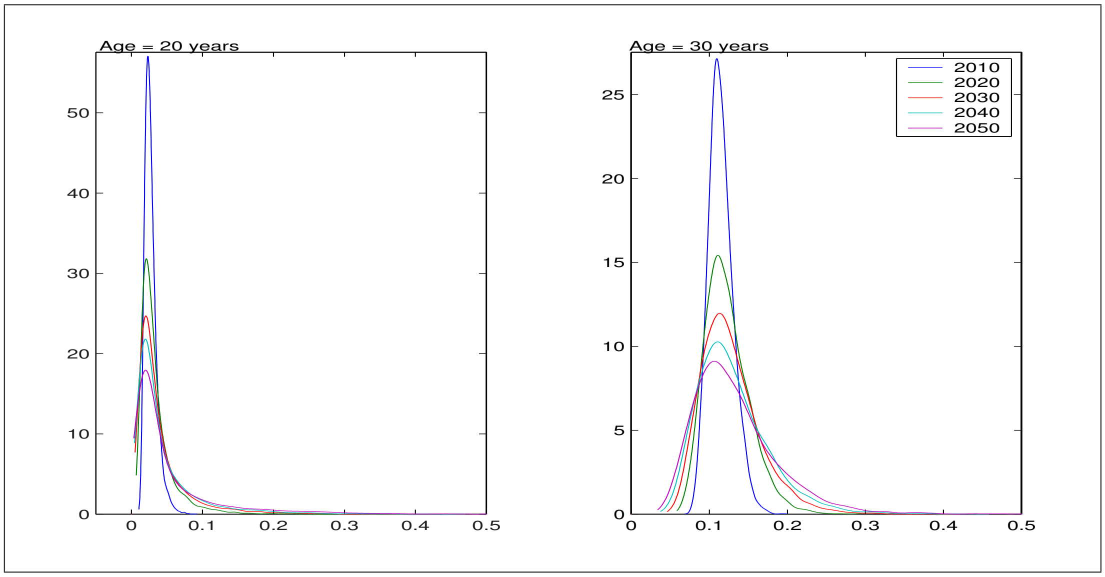

Figure 11. Forecast densities for ages 20 and 30 using model averaging of fertility rates. Spain 2005 - 2050.

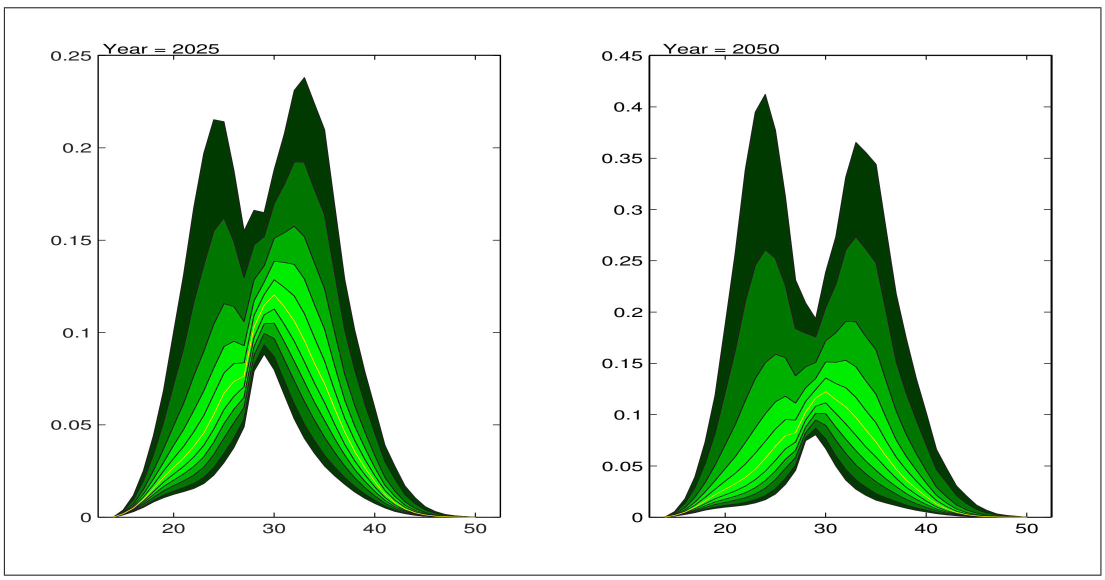

Figure 12. Fan chart using model averaging of the fertility curve. Spain 2025 - 2050.

3.3 Application to Spanish Data: Migrations

In this section we propose an evolutionary model of immigration in Spain for the next 50 years. This model picks up the sharp increase in the rate of immigrants over the last decade and allows for the assumption that in the long term the percentage of immigrants will stabilize, as has been the case in European countries with a longer history of taking in immigrants. In the case of birth and death rates as well as with immigration, Spain is no different and should be studied within the European context. The majority of immigration in Spain is of an economic nature; the type of immigration received both in Spain and the European Union. Over the last several years, the flow of immigration to Spain has risen sharply but it still remains below the European mean of 5.1%.

We propose to model the flow of immigration using the methodology shown in Section 3.1.1, carrying out a transformation in the series, which in this case would be the number of immigrants INM_t obtained in the permanent observatory of immigration maintained by the Ministry of Labour and Social Services (Ministerio de Trabajo y Asuntos Sociales - `www.extranjeros.mir.es`).

- The following transformation of the series of the number of immigrants, INM_t is considered:

$$Y_{t,C} = \ln \frac{INM_t}{C - INM_t}, \tag{21}$$

  where C is the upper bound of number of immigrants.

In this first step, the key is to select the possible values of the parameter C. To do that, we make three different assumptions having as a reference the percentage of immigrant population in Europe:

- Assumption 1: $C \in (4000000, 8000000)$ corresponds to a percentage of the immigrant population in 2050 which will be between 8% and 15% of the Spanish population, based on population projections carried out by the National Statistical Institute (INE, 2004).
- Assumption 2: $C \in (4000000, 11000000)$ corresponds to a percentage of the immigrant population in 2050 which will be between 8% and 20% of the Spanish population, based on population projections carried out by the National Statistical Institute (INE, 2004).
- Assumption 3: $C \in (4000000, 15000000)$ corresponds to a percentage of the immigrant population in 2050 which will be between 8% and 28% of the Spanish population, based on population projections carried out by the National Statistical Institute (INE, 2004).

The next two steps are similar to the methodology proposed in Section 3.1.1. Figure 13 shows the growth curves in immigration to the year 2050 under the three above mentioned assumptions. Also incorporated into the graph is the future evolution of foreign immigration suggested by the NSI. A rapid rise is observed in the number of immigrants until 2015 in assumption 1, until 2020 in assumption 2 and until 2025 in assumption 3. After this period of rapid increase the number of immigrants will begin to stabilize at around 6, 8, and 10 million in each of the assumptions. This stabilization in the number of immigrants does not imply that the rate of immigrant entry into the country will be zero but rather that the entry will be one of replacement. We assume for simplicity's sake in the model that the number of new immigrants is similar to the number of deaths occurring among the immigrants the year before.

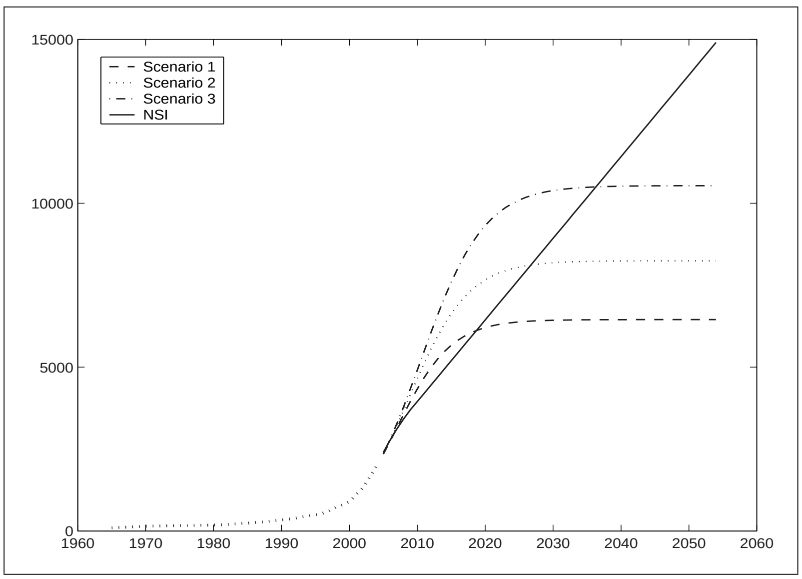

Figure 13. Evolution of immigration for the assumptions and NSI.

4 Projections for the Spanish Population

In the previous section we obtained, using bootstrap methods, the future distributions of the three demographic components which are basic for population forecasting using the cohort component method defined by the equations (1) and (2). In this section we apply this procedure in order to obtain projections of the Spanish population for the years 2005 to 2050. First we focus on the dynamic of certain aspects of the population, for example, total number of inhabitants, percentage of certain age groups (0 - 15, 16 - 64 and over 65) and the dependency ratios. Second we focus on the dynamic of the structure by single-year ages of the population. Tables 1 and 2, in the Appendix, contain the forecasts mean for female and male population by single-year ages in the period 2005 - 2050.

4.1 Evolution of the Population

In this subsection we analyze the evolution of basic elements of the population such as the total number of inhabitants, the percentage of the population of age groups 0 - 15, 16 - 64 and over 65 and the dependency ratios in those groups. We present our results for the three immigration scenarios obtained in the previous section.

In Figure 14 we present the projections of total inhabitants for the three immigration scenarios as well as the projections carried out by the NSI based on the 2001 census. We can observe that the three scenarios show a sustained increase in population during the first decade of the forecast, 2005 to 2014, with values similar among the three and to the forecasts from the NSI. Later, stagnation can be seen in population growth under the low and medium scenarios in the years 2023 - 2024 and 2029 - 2030, respectively, and finally a reduction in the population. In the high scenario the population growth does not stop during the entire forecast period, although around the year 2025 the growth rate decreases. These projections differ from those carried out by the NSI where stable growth is shown for almost the entire period.

Figure 15 shows the projected population percentages of broad age groups: (i) infant and school age, 0 - 15 years of age; (ii) economically active, 16 - 64; (iii) retirement age, 65 and over. First, we observe that under the three assumptions the projected percentages are similar. Only in the proportion of the population 0 - 15 years of age, and in the decade 2025 to 2035 can notable differences be seen between the low scenario and the other two. As far as the projections from the NSI, we observe similarities during the first decade for all groups, but in the following years a large difference can be seen with respect to the proportion of the 0 - 15 population. This difference may be due to the different immigrant entry dynamic

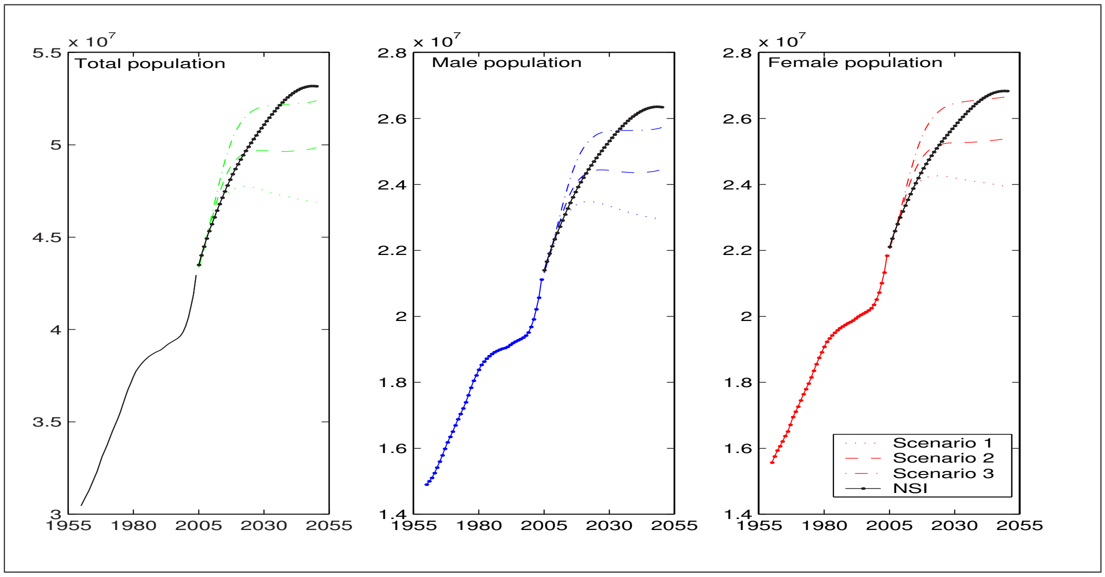

Figure 14. Forecasts and the observed values of total population using the cohort component model. Spain 1960 - 2050.

assumed by the NSI and the proposed scenarios. For the 16 - 64 age group the projections are quite similar until the year 2040; in the following decade the NSI forecasted a proportion 1% - 2% above that of the projections for the scenarios suggested in this paper. Finally, for the retirement age group very similar values were obtained by the NSI and by the scenarios, showing a clear increase in this projection which leads to the possibility that migratory flows are not a solution to the problem of an aging population.

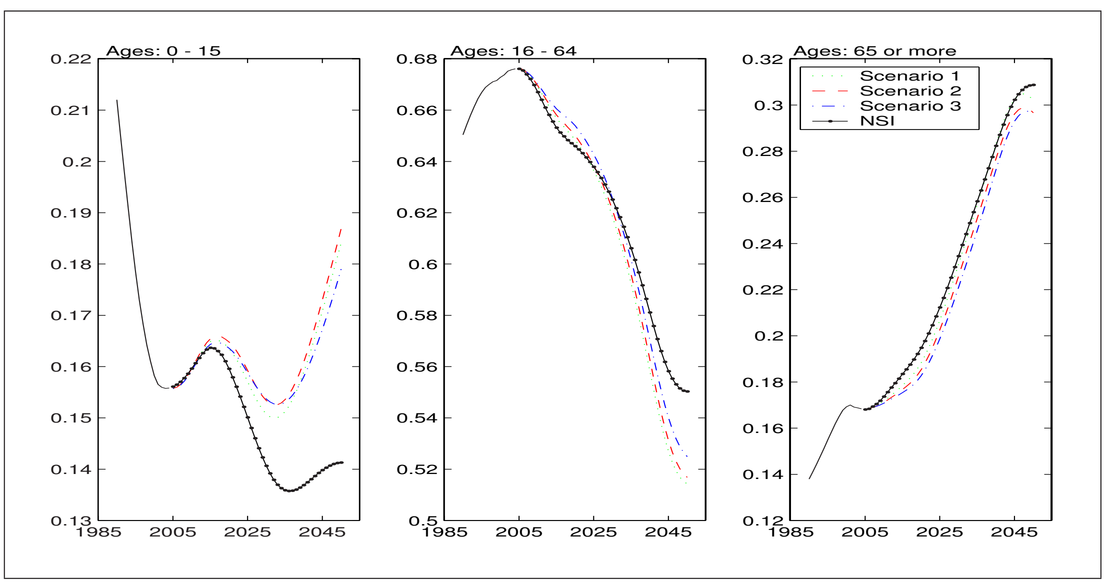

Figure 15. Forecasts and the observed values of population proportions of age groups 0 - 15, 16 - 64 and over 65 using the cohort component model. Spain 1990 - 2050.

4.2 Structure of the Population

In this section we analyze the evolution of the structure by single-year ages and sex during the period 2005 - 2050. To do this, we study the behavior of population pyramids in the years of the forecast. Figure 16 shows the forecast of the population pyramids by sex and age for the three proposed scenarios. We observe several elements which we would like to point out:

- The three scenarios lead to similar population pyramids, where the number of individuals is the only element which clearly differentiates them.
- In the year 2005 the majority of individuals of both sexes (salient in the pyramids) are found in the age group of 20 - 50, in 2025 they are between 30 - 70 years of age, and in 2050 they are between 50 and 80. This progressive aging of the population is observed as an upward movement in the most prominent age groups.
- In the final years of the forecast a slight widening of the base of the pyramids is observed.

5 Conclusions

In this paper, projections of the Spanish population are obtained by age and sex for the period 2005-2050. This analysis is carried out by combining two various sources of information, that is historical evolution of the birth, mortality and immigration rates for the different ages and sexes and the evolution of life expectancy and the synthetic fertility index. The predictions are based on a factorial model which extents to the model proposed by Lee Carter. In the analysis of mortality we observe a strong linear relation between the first factor and the life expectancy. In the second step of the analysis we incorporate the information of this index into the factorial model predictions. Based on biological considerations, we assume that life expectancy must have an upper (but unknown) bound. We obtain

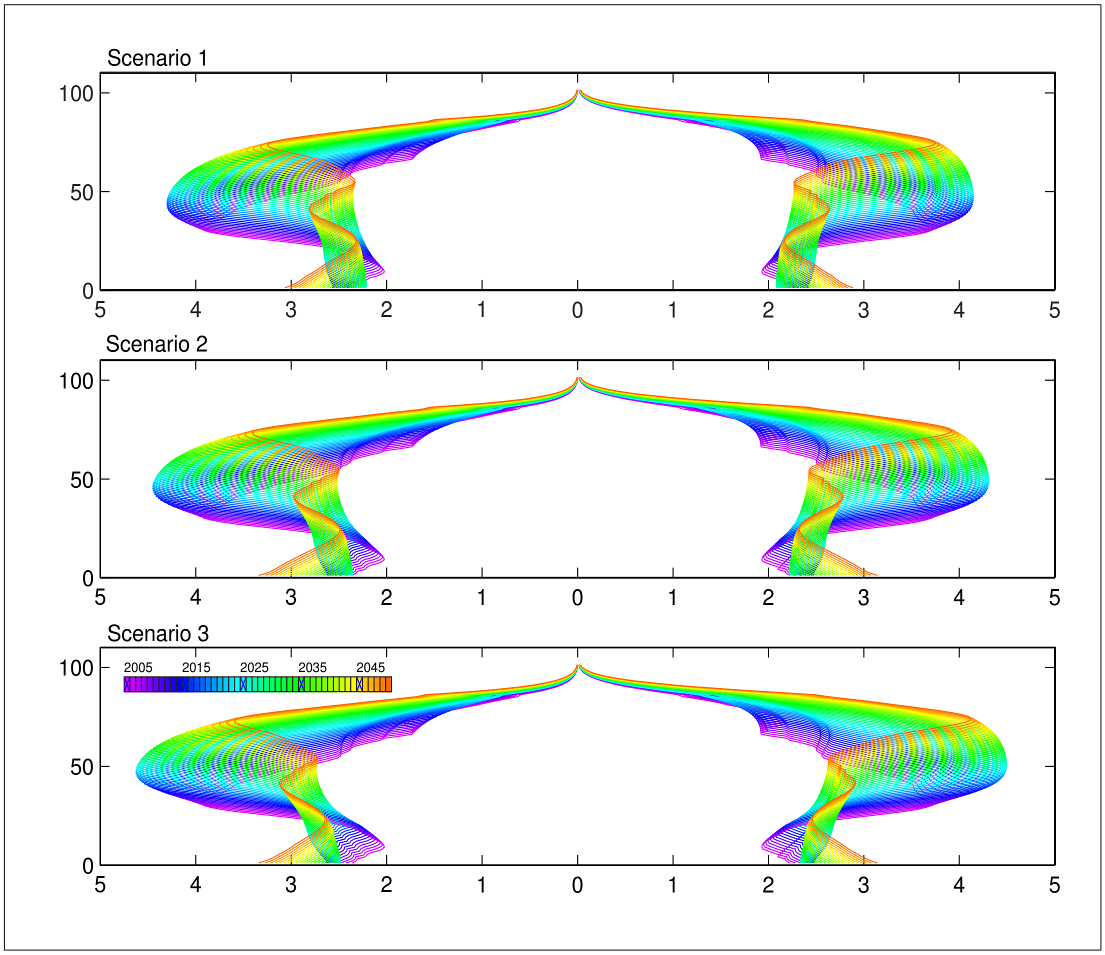

Figure 16. Forecasts of population pyramids using the cohort component model. Spain 2005 - 2050.

a predictive distribution of the life expectancy that incorporates this non linear restriction in its evolution. The predictive distribution of this index is used to obtain predictions of the mortality with a horizon of 45 years. These ideas are extended to the analysis of fecundity and immigration. Combining these three sources of information and using a cohort component model, we obtain prediction for the Spanish population by sex and age. Some remarkable conclusions can be derived from these projections:

- The female and male life expectancies seem to tend to 98 and 90 years, respectively.
- The synthetic fertility index tends to a value of 1.5 children.
- None of the three scenarios of immigration, which suppose immigrant proportions of 12%, 17% and 22% in average, seem to be enough to correct the aging of the Spanish population.

Acknowledgements

This study has been financed by the project #2233 "Previsión del efecto de los cambios de la natalidad en la demanda de educación en España y en la Unión Europea" (Projection of effect of changes in birth rates on demands in education in Spain and the European Union) of the Fundación BBVA. The authors wish to thank Juan Bógalo for the interface modifications in MATLAB/TRAMO. Also, the financial support from projects MTM2005-08897, SEJ2005-06454 and SEJ2007-64500 (Ministerio de Educación y Ciencia, Spain) is gratefully acknowledged.

References

[1] Ahn, N., García, J.R. & Herce, J.A. (2005). Demographic uncertainty and health care expenditure in Spain*Documento de trabajo 2005-07*, FEDEA, Madrid. Available at: http://www.fedea.es/pub/Papers/2005/dt2005-07.pdf

[2] Alonso, A.M., Peña, D. & Rodríguez (2007). Proyecciones de demanda de educación en España. *Documento de trabajo 13*, Fundación BBVA, Madrid. Available at: http://www.fbbva.es/TLFU/dat/Dt_2007_13.pdf

[3] Alonso, A.M., Peña, D. & Romo, J. (2002). Forecasting time series with sieve bootstrap. *Journal of Statistical Planning and Inference*, 100, 1-11.

[4] Alonso, A.M., Peña, D. & Romo, J. (2004). Introducing model uncertainty in time series bootstrap. *Statistica Sinica*, 14, 155-174.

[5] Alonso–Meseguer, J. & Sosvilla–Rivero, S. (2004). Proyecciones del sistema educativo español ante el boom inmigratorio. *Documento de trabajo 2004-14*, FEDEA, Madrid. Available at: http://www.fedea.es/pub/Papers/2004/dt2004-14.pdf

[6] Booth, H. (2006). Demographics forecasting: 1980 to 2005 in review. *International Journal of Forecasting*, 22, 547-581.

[7] Díez, J. (2000). Causas y consecuencias del reciente descenso de la fecundidad en España, 39-61. *Demografía y Cambio Social, Cuadernos Técnicos de Servicios Sociales*, Madrid.

[8] Felipe, A., Guillen, M. & Perez-Marin, A.M. (2002). Recent mortality trends in Spanish mortality. *British Actuarial Journal*, 8, 757-786.

[9] García-Ferrer, A. % del Hoyo, J. (1991). Analysis and prediction of the population in Spain: 1910-2000. *Journal of Forecasting*, 10, 347-369.

[10] Geweke, J. (1977). The dynamic factor analysis of economic time series models. In: Aigner, D.I., Goldberger, A.S. (Eds.), *Latent variables in socio-economic models*, North Holland, New York.

[11] Gonzalo, J. & Granger, C.W.J. (1995). Estimation of common long-memory components in cointegrated systems. *Journal of Business and Economic Statistics*, 13, 27-36.

[12] Gutiérrez de Mesa, J.L. (2003). Tendencias actuales de los modelos de proyección de la población. *Información económica y técnicas de análisis en el siglo XXI*, Instituto Nacional de Estadística, Madrid.

[13] Guttman, I., Peña, D. & Redondas, D. (2005). A Bayesian Approach for Predicting with Polynomial Regression of Unknown Degree. *Technometrics*, 47, 23-33.

[14] Harvey, A.C. (1989). *Forecasting structural time series and the Kalman filter*, Cambridge: Cambridge University Press.

[15] Heligman, L. & Pollard, J.H. (1980). The age pattern of mortality. *Journal of the Institute of Actuaries*, 107, 49-80-

[16] Herce, J.A., Pérez–Díaz, V., Castillo, S., Duce, R.M., Jimeno, J.F., Licandro, O., Rodriguez, D., Sosvilla–Rivero, S., Chuñiá, E. and Álvarez-Miranda, B. (1995). La reforma del sistema público de pensiones en España. *Colección Estudios e Informes*, Núm. 4, La Caixa, Barcelona. Available at:
`http://www.pdf.lacaixa.comunicacions.com/ee/esp/ee04_esp.pdf`

[17] Herce, J.A., Sosvilla–Rivero, S., Castillo, S. and Duce, R. (1996). El futuro de las pensiones en España: Hacia un sistema mixto. *Colección Estudios e Informes*, Núm. 8, La Caixa, Barcelona. Available at:
`http://www.pdf.lacaixa.comunicacions.com/ee/esp/ee08_inx_esp.pdf`

[18] 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.

[19] IEA: Instituto de Estadística de Andalucía (1995). Proyección de la población de Andalucía 1991-2006, Sevilla. Available at:
`http://www.juntadeandalucia.es/iea/proyinfra/index.htm`

[20] IEA: Instituto de Estadística de Andalucía (2000). Proyecciones de población en ciclos formativos reglados y actividad económica. Andalucía 1998-2016, Sevilla. Available at:
`http://www.juntadeandalucia.es/iea/derivadas/index.htm`

[21] IECM: Instituto de Estadística de la Comunidad de Madrid (2004). Proyección de población de la Comunidad de Madrid, 2002–2017. Madrid. Available at:
`http://www.madrid.org/iestadis/fijas/informes/descarga/proy17me.zip`

[22] INE: Instituto Nacional de Estadística (2001). Proyecciones de la población de España calculadas a partir del censo de población de 1991. Evaluación y revisión, Madrid. Available at:
`http://www.ine.es/inebmenu/mnu_cifraspob.htm#5`

[23] INE: Instituto Nacional de Estadística (2004). Proyecciones de la población de España calculadas a partir del censo de población de 2001, Madrid. Available at:
`http://www.ine.es/inebmenu/mnu_cifraspob.htm#5`

[24] Jimeno, J.J. (2002). Demografía, empleo, salarios y pensiones. *Documento de trabajo 2002-04*, FEDEA, Madrid. Available at:
`http://www.fedea.es/pub/Papers/2002/dt2002-04.pdf`

[25] Keilman, N., Pham, D.Q. & Hetland, A. (2002). Why population should be probabilist - Ilustrated by the case if Norway. *Demographic Research*, 6, 409-453.

[26] Lee, R.D. & Carter, L. (1992). Modeling and forecasting the time series of U.S. mortality. *The Journal of the American Statistical Association*, 87, 659-671.

[27] Lee, R.D. & Tuljapurkar, S. (1994). Stochastic populations forecast for the United States: beyond the high, medium and low. *The Journal of the American Statistical Association*, 89, 1175-1189.

[28] McNown, R. & Rogers, A. (1989). Forecasting mortality: A parametrized time series approach. *Demography*, 26, 645-660.

[29] Ortega, J. A. & Poncela, P. (2005). Joint forecasts of Southern European fertility rates with non-stationary dynamic factor models. *International Journal of Forecasting*, 21, 539-550.

[30] Peña, D. & Box, G.E.P. (1987). Identifying a simplifying structure in time series. *The Journal of the American Statistical Association*, 82, 836-843.

[31] Peña, D. & Poncela, P. (2004). Forecasting with nonstationary dynamics factor models. *Journal of Econometrics*, 119, 291-321.

[32] Peña, D. & Redondas, D. (2006). Bayesian Curve Estimation by Model Averaging. *Computational Statistics and Data Analysis*, 50, 688-709.

[33] Tiao, G.C. & Tsay, R.S. (1989). Model specification in multivariate time series. *Journal of the Royal Statistical Society B*, 51, 157-213.

[34] USCB: U.S. Census Bureau (2000). Methodology and assumptions for the population projections of the United States: 1999 to 2100. *Populations Division Working Paper No. 38.* Available at:
`http://www.census.gov/population/www/documentation/twps0038/twps0038.html`

[35] Wallis, K.F. (1999). Asymmetric density forecasts of inflation and the Bank of England's fan chart. *National Institute Economic Review*, 167, 106-112.

[36] Wilson, T. & Bell, M. (2004). Australia's uncertain demographic future. *Demographic Research*, 11, 195-234.

Appendix

Table 1

Forecasts mean for female population by single-year ages. Spain 2005 - 2050.

	Year									
Age	2005	2010	2015	2020	2025	2030	2035	2040	2045	2050
0	222348	245084	242408	228404	221834	228727	246754	268912	290118	313769
1	218073	242420	244178	230348	221605	225584	242125	263898	285802	306514
2	217240	240089	246757	233548	222817	223716	238181	259670	281608	301915
3	209543	237119	248982	237436	224562	222667	234939	255546	277770	297673
4	211255	233521	250277	241464	226751	222136	231735	251322	273461	294400
5	204774	229161	250765	245206	229467	222299	229091	247207	269568	291079
6	199200	226954	250212	248532	232481	222919	226744	243400	265435	287722
7	193862	226750	248521	251524	235909	224283	225009	239597	261367	283714
8	192720	219326	245800	253900	239882	226087	224015	236414	257310	279954
9	193129	221247	242405	255319	243977	228321	223525	233256	253138	275705
10	194380	214926	238202	255903	247769	231069	223717	230643	249059	271854
11	197494	209458	236100	255413	251124	234099	224350	228310	245270	267745
12	201085	204186	235957	253762	254135	237539	225725	226589	241483	263694
13	205169	203093	228577	251063	256517	241513	227531	225598	238306	259646
14	209500	203569	230550	247694	257944	245609	229767	225112	235154	255483
15	213000	204909	224305	243531	258544	249407	232521	225314	232555	251423
16	215686	208123	218924	241466	258063	252761	235549	225949	230229	247645
17	220137	211823	213750	241362	256416	255760	238977	227317	228507	243861
18	226072	216253	212927	234118	253770	258166	242973	229153	227561	240747
19	233832	221381	214011	236394	250546	259679	247148	231487	227203	237761
20	244263	226130	216337	230668	246634	260434	251089	234409	227617	235432
21	258443	230527	220968	226071	244966	260214	254678	237701	228572	233504
22	274302	237098	226519	221952	245393	258908	257978	241459	230337	232274
23	291142	245265	232963	222295	238749	256638	260710	245807	232595	231851
24	308103	255091	239984	224485	241586	253776	262536	250318	235328	231984
25	326150	267414	246484	227837	236395	250201	263583	254570	238617	232848
26	341520	283318	252478	233403	232284	248842	263633	258446	242246	234217
27	355250	300734	260470	239777	228595	249538	262567	262000	246305	236352
28	364485	318878	269847	246925	229299	243128	260499	264945	250903	238921
29	370758	336786	280610	254503	231768	246132	257783	266922	255589	241872
30	371648	355412	293561	261389	235304	241048	254292	268050	259937	245285
31	370967	370998	309746	267564	240939	236968	252943	268107	263824	248940
32	367412	384641	327114	275534	247286	233258	253607	267009	267339	252960
33	363568	393562	344977	284744	254331	233893	247141	264873	270200	257463
34	359441	399348	362466	295263	261763	236259	250043	262056	272055	262007
35	356510	399612	380543	307898	268474	239674	244859	258452	273043	266186
36	354649	398125	395407	323650	274410	245136	240636	256942	272912	269847
37	352918	393638	408186	340491	282098	251289	236771	257427	271611	273113
38	351473	388741	416129	357740	290975	258105	237214	250770	269243	275688
39	351581	383478	420874	374568	301129	265287	239367	253441	266175	277233
40	348931	379304	420013	391906	313343	271709	242531	248012	262284	277871
41	343956	376168	417386	406037	328672	277375	247753	243563	260500	277409
42	338777	373069	411670	418006	345029	284745	253627	239439	260671	275740
43	333874	370252	405537	425145	361784	293303	260166	239617	253733	273018
44	326247	369014	399059	429090	378100	303115	267055	241490	256083	269595
45	320894	365092	393724	427490	394937	314993	273202	244387	250386	265377
46	316030	358898	389458	424149	408577	329965	278594	249335	245672	263267
47	308218	352542	385273	417759	420064	345953	285683	254932	241291	263110
48	298652	346492	381384	410945	426707	362297	293922	261167	241176	255864
49	288511	337840	379150	403852	430218	378229	303439	267781	242784	257905
50	279416	331530	374317	397935	428234	394682	315014	273666	245426	251963
51	271360	325764	367278	393098	424512	407928	329642	278790	250099	247003
52	264244	317127	360131	388360	417753	419003	345249	285580	255407	242371
53	257625	306836	353374	383978	410624	425298	361227	293538	261374	242015
54	255499	296010	344043	381213	403181	428427	376737	302723	267684	243347
55	255820	286303	337109	375924	396951	426144	392781	313969	273296	245736
56	251821	277645	330715	368420	391740	422069	405559	328176	278103	250099
57	249369	269986	321558	360871	386673	415032	416208	343367	284587	255119
58	250375	262834	310772	353684	381913	407574	422045	358850	292176	260749
59	246499	260156	299502	343983	378769	399832	424760	373864	300984	266729
60	239540	259874	289323	336605	373063	393216	422033	389299	311753	271958
61	231268	255359	280218	329781	365184	387617	417544	401484	325431	276391
62	221660	252338	272074	320179	357181	382058	410020	411420	339950	282388
63	218441	252725	264472	309032	349587	376855	402147	416645	354767	289498
64	202733	248214	261161	297318	339409	373105	393859	418622	368940	297672
65	191988	240622	260118	286629	331454	366763	386607	415153	383394	307675
66	191664	231724	254891	276982	324021	358270	380329	409921	394565	320455
67	199051	221484	251050	268230	313827	349605	374028	401653	403409	333941
68	195839	217399	250433	259945	302064	341253	367970	392943	407478	347543
69	207096	201207	244994	255772	289715	330328	363240	383754	408248	360352
70	215070	189756	236442	253622	278225	321380	355769	375353	403445	373105
71	217930	188293	226623	247384	267737	312909	346181	367855	396876	382513
72	213705	194041	215464	242363	258024	301664	336293	360177	387208	389391
73	211760	189584	210135	240320	248716	288880	326639	352640	377039	391479
74	207145	198561	193260	233495	243144	275385	314317	346097	366154	390031
75	200699	204195	180925	223688	239388	262700	303828	336850	355948	383125
76	191814	204666	177738	212479	231467	250688	293417	325180	346135	374018
77	184756	198317	180924	199914	224450	239224	280182	312961	335831	361658
78	174863	193649	174342	192394	219695	227734	265063	300369	324967	348119
79	166077	186269	179454	174330	210287	219417	249135	285062	314620	333572
80	155595	176899	180828	160227	197811	212192	233544	270855	301075	318906
81	146987	165086	176957	153836	183776	200747	218154	256128	284678	303821
82	135281	154641	166842	152369	168436	189682	202939	238498	267247	287597
83	123971	141881	157956	142479	157390	180355	187770	219386	249477	270756
84	109954	129878	146490	141371	137676	166683	174748	199261	228861	253443
85	104456	123384	139166	134302	130792	158348	166010	189298	217418	240771
86	89783	106053	119617	115437	112420	136105	142691	162707	186877	206950
87	76092	89880	101376	97834	95277	115350	120932	137896	158380	175392
88	63274	74739	84299	81353	79227	95919	100560	114666	131700	145845
89	51546	60887	68674	66274	64542	78141	81922	93413	107290	118814
90	41073	48516	54722	52809	51429	62264	65277	74434	85491	94674
91	32128	37950	42803	41308	40228	48704	51060	58223	66872	74054
92	24655	29123	32847	31700	30871	37375	39184	44680	51318	56829
93	18491	21842	24636	23775	23153	28031	29388	33510	38488	42622
94	13527	15979	18022	17393	16938	20507	21499	24515	28157	31181
95	9655	11404	12863	12413	12089	14636	15344	17496	20096	22254
96	6655	7861	8866	8556	8332	10088	10576	12060	13851	15339
97	4473	5283	5959	5751	5601	6780	7109	8106	9310	10310
98	2891	3415	3852	3717	3620	4382	4595	5239	6017	6664
99	1800	2126	2398	2314	2254	2729	2861	3262	3747	4149
100+	1286	1519	1713	1653	1610	1949	2043	2330	2676	2964

Table 2

Forecasts mean for male population by single-year ages. Spain 2005 - 2050.

	Year									
Age	2005	2010	2015	2020	2025	2030	2035	2040	2045	2050
0	236757	260966	258116	243205	236209	243548	262744	286338	308917	334101
1	231843	257982	259902	245222	235940	240195	257819	281013	304344	326402
2	230915	255433	262624	248654	237286	238275	253700	276601	299973	321595
3	222513	252181	264948	252815	239202	237237	250341	272313	295995	317184
4	221116	248255	266267	257112	241587	236749	247022	267918	291517	313808
5	217395	243514	266714	261092	244527	236998	244298	263641	287484	310384
6	209727	240872	266045	264621	247782	237734	241893	259695	283201	306928
7	204889	240578	264149	267773	251463	239255	240134	255747	278982	302777
8	202496	232458	261155	270259	255720	241244	239167	252465	274779	298896
9	203289	231278	257442	271713	260094	243686	238733	249206	270453	294494
10	204713	227724	252870	272268	264137	246671	239024	246528	266228	290521
11	208255	220173	250342	271672	267705	249953	239785	244151	262314	286275
12	212550	215404	250113	269820	270880	253649	241322	242410	258387	282079
13	216735	213066	242045	266856	273380	257914	243320	241454	255119	277894
14	221262	213921	240915	263166	274838	262287	245761	241023	251866	273576
15	224912	215416	237418	258615	275388	266316	248736	241309	249188	269354
16	227789	219035	229935	256104	274779	269860	251996	242056	246802	265435
17	232166	223386	225217	255859	272874	272967	255630	243538	245015	261465
18	238381	227829	223073	247846	269872	275396	259829	245491	244033	258183
19	246301	233067	224470	246954	266257	276865	264215	247980	243689	255054
20	256086	237887	226895	243929	261914	277521	268346	251099	244177	252638
21	269520	242392	231875	237209	259766	277135	272094	254612	245251	250662
22	285620	248811	238030	233541	260035	275560	275491	258585	247163	249409
23	303193	257233	244502	232621	252704	273009	278318	263224	249645	249069
24	320987	267209	251660	235185	252458	269854	280194	268046	252652	249347
25	340671	278888	258274	238707	250062	265962	281250	272597	256263	250424
26	358157	294025	264407	244669	243928	264211	281227	276723	260216	252026
27	373270	311668	272287	251706	240789	264848	280008	280486	264608	254435
28	384355	330502	281925	258908	240293	257841	277740	283599	269572	257311
29	391657	349168	292808	266616	243150	257790	274769	285651	274592	260569
30	393107	369299	305030	273556	246816	255475	270939	286749	279192	264265
31	392447	386899	320365	279824	252804	249375	269180	286705	283287	268206
32	388337	401833	337864	287624	259757	246198	269750	285422	286959	272507
33	383207	412467	356257	296977	266751	245560	262629	283000	289877	277258
34	378092	419172	374350	307499	274208	248230	262404	279856	291708	282026
35	373834	419896	393746	319263	280855	251671	259900	275828	292551	286327
36	370152	418337	410437	333999	286760	257364	253588	273805	292199	290057
37	366279	413153	424253	350720	294084	263936	250115	274022	290536	293273
38	362433	406907	433742	368297	302952	270560	249184	266630	287767	295759
39	360300	400540	439195	385464	312897	277575	251475	266020	284213	297089
40	356126	394968	438671	403893	324068	283789	254538	263162	279798	297449
41	349495	389835	435744	419511	338090	289195	259769	256472	277308	296553
42	342467	384493	429208	432263	354076	296018	265881	252618	277056	294376
43	336499	379157	421592	440692	370875	304354	272036	251282	269272	291108
44	327675	375532	413858	445087	387214	313715	278542	253110	268179	287037
45	320937	369930	406947	443578	404775	324272	284245	255699	264868	282126
46	315006	361931	400492	439689	419515	337614	289131	260416	257768	279115
47	306559	353550	393839	432224	431371	352842	295385	265979	253465	278305
48	295964	346235	387191	423662	438874	368782	303068	271529	251612	270027
49	284660	336212	382313	415055	442435	384276	311784	277456	252926	268409
50	274933	328238	375449	407178	440055	400816	321562	282500	254916	264542
51	266585	321172	366355	399851	435416	414609	334107	286785	259050	257027
52	258757	311684	356926	392330	427231	425472	348420	292361	263971	252245
53	251615	300150	348588	384828	417970	432011	363364	299306	268856	249874
54	248533	287927	337563	378971	408564	434533	377703	307143	274013	250535
55	247698	277369	328660	371276	399979	431329	393084	316055	278354	251919
56	242244	268136	320587	361310	391794	425747	405549	327483	281814	255276
57	238247	259452	310206	351046	383423	416706	415077	340554	286499	259376
58	237586	251443	297850	341804	375002	406541	420239	354070	292404	263348
59	232634	247441	284957	330082	368283	396369	421586	367016	299219	267637
60	224937	245524	273622	320311	359648	386841	417193	380714	306888	270990
61	216004	239132	263547	311286	348755	377611	410388	391388	316822	273384
62	206525	234140	254031	300054	337563	368159	400198	399060	328172	276861
63	202363	232360	245261	287080	327489	358809	389108	402623	339967	281567
64	185877	226268	240050	273357	314745	350697	377607	402016	350688	286734
65	173797	217531	236715	261099	303815	340697	366656	395817	361870	292539
66	171402	207693	229197	250127	293688	328681	356104	387427	370118	300442
67	174604	197332	222926	239662	281472	316363	345301	375798	375328	309479
68	169256	191729	219380	229677	267389	304768	334209	362906	376084	318354
69	177185	174964	211942	223107	252839	290917	324466	349880	373067	326206
70	181502	162316	202060	218174	239652	278711	312904	337294	364704	334163
71	181067	158371	191131	209312	227583	267131	299369	324925	354114	339007
72	175723	159281	179741	201504	215946	253605	285492	312213	340429	340697
73	170404	152541	172623	196104	204809	238513	272346	299304	325685	338206
74	163292	157124	155645	187029	196483	222845	256931	287238	310458	331735
75	154437	158246	142407	175794	189468	208401	242928	273445	295517	320257
76	143884	155062	136625	163666	178945	194906	229369	257796	280584	306539
77	134620	147664	134804	151273	169339	181882	214233	241947	265403	290168
78	123600	139996	126331	142194	161332	168956	197410	226194	249406	272185
79	113701	131056	126988	125506	150559	158670	180647	209070	234568	254353
80	103329	120696	124428	111986	137981	149227	164857	192964	218048	236482
81	93988	109061	118170	104233	124704	136857	149785	177057	199843	218337
82	83906	98481	108582	99194	111278	125043	135005	159770	181246	199614
83	73786	87054	99052	89487	100702	114723	120833	141906	163367	180902
84	62730	76572	88587	85802	84913	102227	108364	124039	144251	162546
85	59593	72743	84157	81512	80668	97115	102946	117837	137038	154419
86	47394	57852	66929	64825	64154	77235	81871	93715	108985	122807
87	37301	45531	52676	51020	50492	60787	64436	73757	85775	96654
88	28941	35327	40870	39585	39175	47163	49995	57227	66551	74992
89	22095	26971	31203	30222	29909	36007	38169	43690	50809	57253
90	16632	20302	23487	22749	22513	27104	28731	32887	38246	43096
91	12375	15106	17476	16927	16751	20167	21378	24470	28457	32066
92	9062	11061	12797	12395	12266	14768	15654	17919	20838	23481
93	6517	7955	9203	8913	8821	10620	11257	12886	14985	16886
94	4564	5571	6445	6242	6178	7437	7884	9024	10495	11826
95	3116	3803	4400	4262	4218	5077	5382	6161	7165	8073
96	2063	2518	2913	2821	2792	3361	3563	4078	4743	5344
97	1338	1634	1890	1831	1812	2181	2312	2647	3078	3468
98	812	991	1146	1110	1099	1323	1402	1605	1867	2104
99	483	589	682	660	653	787	834	955	1110	1251
100+	345	421	487	472	467	562	596	682	793	893