Forecasting vital rates from demographic summary measures

Carlo G. Camarda[1]

[1] Institut national d'études démographiques (INED), Paris, France

E-mail for correspondence: carlo-giovanni.camarda@ined.fr

Abstract: In population and actuarial sciences, time-trends of summary measures (such as life expectancy or the average number of children per woman) are easy to interpret and predict. Most summary measures are nonlinear functions of the vital rates, the key variable we usually want to estimate. Furthermore smooth outcomes of future age-specific vital rates are desirable. Therefore, optimization with nonlinear constraints in a smoothing setting is necessary. We propose a methodology that combines Sequential Quadratic Programming and a P-spline approach, allowing to forecast age-specific vital rates when future values of demographic summary measures are provided. We provide an application of the model on Japanese mortality and Spanish fertility data.

Keywords: Vital rates forecast; Smoothing; Constrained nonlinear optimization; Summary measures.

1 Introduction

Future mortality and fertility levels can be predicted by either modelling and extrapolating rates over age and time, or by forecasting summary measures, later converted into age-specific rates. The latter approach takes advantage of the prior knowledge that demographers and actuaries have on possible future values of life expectancy at birth and total fertility rates. For instance, this methodology has been lately adopted by the United Nations (Ševčíková et al., 2016). In this paper, we propose a model to obtain future mortality and fertility age-patterns which comply with the projected summary measures. Unlike comparable approaches, we assume only smoothness of future vital rates which is achieved by a two-dimensional P-spline approach as in Currie et al. (2004). Since summary measures are commonly nonlinear functions of the estimated penalized coefficients, Lagrangian multipliers cannot be directly implemented. We hence opted for a Sequential

This paper was published as a part of the proceedings of the 34th International Workshop on Statistical Modelling (IWSM), Guimarães, Portugal, 7–12 July 2019.

Quadratic Programming (SQP) procedure (Nocedal & and Wright, 2006) to perform the associated constrained nonlinear optimization. We illustrate our approach with two data sets: mortality of Japanese females, based on future life expectancy predicted by UN World Population Prospects (2017) and Spanish fertility constrained to total fertility rates, mean and variance of age at childbearing derived by classic time-series analysis.

2 Model on Japanese mortality data

For ease of presentation, we formulate the model on mortality data. We suppose that we have deaths, and exposures to risk, arranged in two matrices, $\boldsymbol{Y} = (y_{ij})$ and $\boldsymbol{E} = (e_{ij})$, each $m \times n_1$, whose rows and columns are classified by age at death, $\boldsymbol{a}$, $m \times 1$, and year of death, $\boldsymbol{t}_1$, $n_1 \times 1$, respectively. We assume that the number of deaths y_{ij} at age i in year j is Poisson distributed with mean $\mu_{ij}\, e_{ij}$. Forecasting aims to reconstruct trends in μ_{ij} for n_2 future years, $\boldsymbol{y}_2, n_2 \times 1$.
Usually demographers and actuaries summarize mortality age-patterns by computing life expectancies. Time-trends of this summary measure are often regular and well-understood. Forecasting a single time-series is therefore relatively easy. Figure 1 (left panel) presents observed life expectancy at age 1 for Japanese females from 1960 to 2016 along with the medium variant up to 2050 as computed by the UN, which consider available data for all countries in the world. Future mortality patterns, both by age and over time, must adhere to this predicted trend.

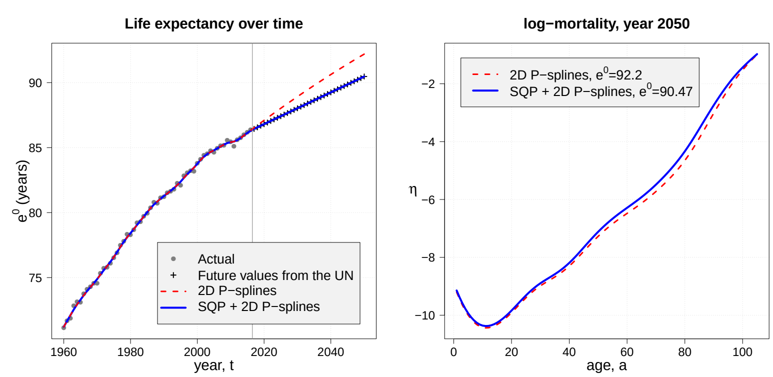

FIGURE 1. Left panel: Actual, estimated and forecast life expectancy at age 1 by United Nations, 2D P-splines and the SQP+2D P-splines. Right panel: Mortality age-pattern in 2050 by 2D P-splines and the SQP+2D P-splines. Japanese females, ages 1-105, years 1960-2016, forecast up to 2050.

We arrange data as a column vector, that is, $\boldsymbol{y} = \texttt{vec}(\boldsymbol{Y})$ and $\boldsymbol{e} = \texttt{vec}(\boldsymbol{E})$ and we model our Poisson death counts as follows: $\boldsymbol{\eta} = \ln(E(\boldsymbol{y})) = \ln(\boldsymbol{e}) +$

$\boldsymbol{B}\boldsymbol{\alpha}$, where $\boldsymbol{B}$ is the regression matrix over the two dimensions: $\boldsymbol{B} = \boldsymbol{I}_{n_1} \otimes \boldsymbol{B}_a$, with $\boldsymbol{B}_a \in \mathbb{R}^{m \times k_a}$. Over time, we employ an identity matrix of dimension n_1 because we will incorporate a constraint for each year. In order to forecast, data and bases are augmented as follows:

$$\breve{\boldsymbol{E}} = [\boldsymbol{E} : \boldsymbol{E}_2]\,, \qquad \breve{\boldsymbol{Y}} = [\boldsymbol{Y} : \boldsymbol{Y}_2]\,, \qquad \breve{\boldsymbol{B}} = \boldsymbol{I}_{n_1+n_2} \otimes \boldsymbol{B}_a\,, \tag{1}$$

where $\boldsymbol{E}_2$ and $\boldsymbol{Y}_2$ are filled with arbitrary future values. If we define a weight matrix $\boldsymbol{V} = \mathrm{diag}(\mathtt{vec}(\mathbf{1}_{m \times n_1} : \mathbf{0}_{m \times n_2}))$, the coefficients vector $\boldsymbol{\alpha}$ can be estimated by a penalised version of the iteratively reweighted least squares algorithm:

$$(\breve{\boldsymbol{B}}^T \boldsymbol{V} \tilde{\boldsymbol{W}} \breve{\boldsymbol{B}} + \boldsymbol{P})\tilde{\boldsymbol{\alpha}} = \breve{\boldsymbol{B}}^T \boldsymbol{V} \tilde{\boldsymbol{W}} \tilde{\boldsymbol{z}}\,, \tag{2}$$

where a difference penalty $\boldsymbol{P}$ enforces smoothness behaviour of mortality both over age and time. Outcomes from this approach in terms of life expectancy is depicted with a dashed red line in Figure 1 (left panel), and a departure from the UN projected values is evident.
Life expectancy is a nonlinear function of the coefficients vector $\boldsymbol{\alpha}$:

$$\boldsymbol{e}^0(\boldsymbol{\alpha}) = (\mathbf{1}_{1 \times m} \otimes \boldsymbol{I}_n)\, \exp[(\boldsymbol{I}_n \otimes \boldsymbol{C})\mathtt{vec}(\exp(\boldsymbol{B}\boldsymbol{\alpha}))] + 0.5 \tag{3}$$

where $\boldsymbol{C}$ is a $(m \times m)$ lower triangular matrix filled only with -1.
Constrained nonlinear optimization is therefore necessary and a SQP approach is implemented. Let denote with $\boldsymbol{e}^0_{\mathrm{T}}$ the n_2-vector of target life expectancy for future years and with $\boldsymbol{N}$ the $(k_a n_2 \times n_2)$ matrix with derivatives of (3) with respect to $\boldsymbol{\alpha}$ for each future year. The solution of the associated system of equations at the step $\nu + 1$ is given by

$$\begin{bmatrix} \boldsymbol{\alpha}_{\nu+1} \\ \boldsymbol{\omega}_{\nu+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{L}_\nu & \boldsymbol{H}_\nu \\ \boldsymbol{H}_\nu^T & \mathbf{0}_{n_2 \times n_2} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{r}_\nu - \boldsymbol{L}_\nu \boldsymbol{\alpha}_\nu \\ \boldsymbol{e}^0_{\mathrm{T}} - \boldsymbol{e}^0(\boldsymbol{\alpha}_\nu) \end{bmatrix}, \tag{4}$$

where $\boldsymbol{L}$ and $\boldsymbol{r}$ are left- and right-hand-side of the system in (2), and matrix $\boldsymbol{H}^T = \left[\mathbf{0}_{n_2 \times k_a n_1} : \boldsymbol{N}^T\right]$. Vector of $\boldsymbol{\omega}$ denotes the current solution of the associated Lagrangian multipliers.
Forecast e^0 by the proposed method is exactly equal to the UN values (Figure 1, left panel). The right panel of Figure 1 shows the forecast mortality age-pattern in 2050: Shape obtained by the suggest approach is not a simple linear function of the plain P-splines outcome.

3 Spanish Fertility Data

We forecast Spanish fertility using three commonly-used summary measures: Total Fertility Rate, mean and variance of childbearing age, forecast by conventional time-series analysis. We then smooth and constrain future fertility age-patterns to comply these forecast values. Summary measures as well as fertility rates in 2050 are presented in Figure 2.

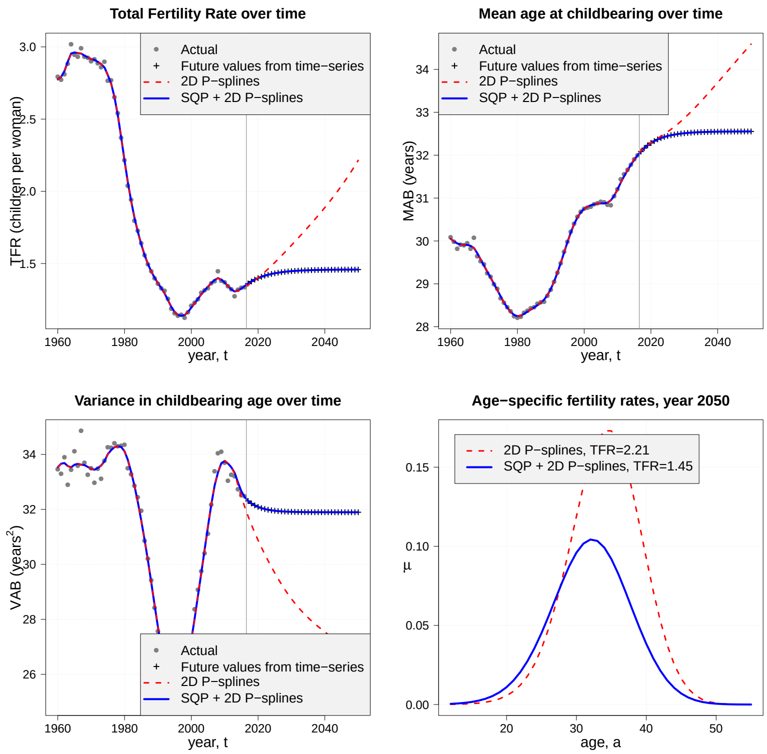

FIGURE 2. Top and left-bottom panels: Actual, estimated and forecast Total Fertility Rate, Mean and Variance in childbearing age by time-series analysis, 2D *P*-splines and the SQP+2D *P*-splines. Right-bottom panel: Age-specific fertility rate in 2050 by 2D *P*-splines and the SQP+2D *P*-splines. Spain, ages 12-55, years 1960-2016, forecast up to 2050.

References

Currie, I. D. et al. (2004). Smoothing and Forecasting Mortality Rates. *Statistical Modelling*, **4**, 279-298.

Nocedal, J. & Wright, S. J. (2006). *Numerical Optimization*. Springer.

Ševčíková, H. et al. (2016). Age-Specific Mortality and Fertility Rates for Probabilistic Population Projections. In R. Schoen (Ed.), *Dynamic demographic analysis*, 285 − 310. Springer.

United Nations, Population Division (2017). *World Population Prospects: The 2017 Revision, Volume II*. ST/ESA/SER.A/400.