Population Review

Volume 49, Number 2, 2010

Type: Article pp. 13-26

Non-parametric and Structured Graduation of Mortality Rates

Authors: Víctor M. Guerrero and Eliud Silva

Affiliations: Department of Statistics. Instituto Tecnológico Autónomo de México (ITAM), 01080 D. F., México, Mexico (Guerrero); Department of Statistics. Universidad Carlos III de Madrid, 28903 Getafe, Madrid, Spain (Silva)

Corresponding author/address: Victor Guerrero: guerrero@itam.mx

Abstract

In this article, we present a non-parametric method to estimate trends in mortality rates. This method combines goodness of fit and smoothness of the non-parametric approach with information from a given structural mortality rate. So, the user is able to control both smoothness and structure in the resulting estimated mortality. The main goal is to enable the analyst to compare mortality trends, with equal percentages of smoothness and structure established beforehand. Also, two perspectives of the proposed methodology are emphasized. On the one hand, the proper fit and smoothness and, on the other, the combination of two information sources, thus giving the analyst the possibility of choosing which one offers greater credibility. The usefulness of this approach is shown via empirical examples that employ different mortality indicators.

Keywords

Graduation, smoothness index, comparability, generalized least squares, Kalman filter

1. Introduction

Population censuses, surveys and vital statistics, are susceptible to having flaws (or defects) in their records either by the presence of extraordinary events (earthquakes, floods, tornados, hurricanes, etc.) or, in general, by human errors of diverse types. As it is to be expected, such flaws have negative repercussions on demographic estimations. Particularly, the wrong recording of deaths distort – or misrepresent – the phenomenon under study, which can lead to an increase (or decrease) of its intensity and timing at a certain age, in detriment of another ones. This situation can affect timely decision making and policy creation, both in public and private sectors. One solution to solving this problem is graduating (smoothing) data.

On the other hand, from an actuarial point of view, mortality plays a fundamental role for insurance companies, where it is important to estimate premium costs based on the risks taken. So, the predicted probability of dying must be sufficiently accurate so as to guarantee that, in the event of death, the amount of money to be paid to the insured party will be enough. Usually, data graduation is required to fulfil this requirement.

In this work, we propose a methodology to estimate mortality trends that combines mortality demographic structure with fidelity to the original data and smoothness in such a way that the user is able to control both a smoothness percentage and a structure percentage. We emphasize that by applying this procedure, the user will be able to obtain comparable estimated trends.

This article is organized as follows. In Section 2, we present several non-parametric models. These models have appeared in the related literature and are used to model mortality. Section 3 cites some demographic techniques in current use to project mortality. Section 4 deals with our methodological proposal, in which a signal-plus-noise mortality model is considered, together with two additional equations: one that allows inducing smoothness and another one to consider demographic structure. We introduce some smoothness and structure indexes in Section 5, where we also indicate how to use them in order to choose their associated smoothness and structure parameters. In Section 6 we illustrate the practical use of our proposed methodology by way of some applications to some observed mortality data. The final section concludes.

2. Non-parametric models

Haberman and Renshaw (1996) define graduation as the group of principles and methods by which observed probabilities are smoothed in order to carry out actuarial inferences and calculations. Graduation of mortality data can be done by means of parametric or non-parametric methods. In the first group, the objective is to fit a parametric function to the probabilities obtained directly form the observed data. In the second group, the actual data corresponding to death probabilities are smoothed and the assumption of an age-dependent function becomes unnecessary. The latter methods are more flexible and appropriate to use when graduation through parametric methods is difficult. It is in such a context that we suggest using our proposal.

The underlying idea of graduation and smoothing is to reduce variability and facilitate the analysis of the observed data. To do so, the data are modified and turned into estimates, once unwanted fluctuations are excluded. Some non–parametric models of graduation are the graphical method, weighted moving averages, the kernel method and graduation with reference to standard mortality rates (Copas and Haberman, 1983; Papaioannou and Sachlas, 2004). Comparisons among non-parametric models and smoothing by means of Generalized Additive Models with splines appear in Debón *et al.* (2006). Moreover, non-parametric models have also been used to yield estimates of old-age mortality (Fledelius *et al.*, 2004). On the other hand, it should be mentioned that principal component approaches are usually employed to address the dimensionality problem by extracting age patterns from the data, while relational models replace the age scale by an empirically-derived exogenous standard (Booth and Tickle, 2008).

One of the most utilized techniques to perform graduation is the Whittaker-Henderson method, which results from minimizing the following function, for a given value of the constant λ>0,

$$(\mathbf{v}-\mathbf{u})'W(\mathbf{v}-\mathbf{u})+\lambda\mathbf{v}'K_d^{'}K_d\mathbf{v}$$

where $\mathbf{u} = (u_1, \ldots, u_n)'$ is the vector of observed values, $\mathbf{v} = (v_1, \ldots, v_n)'$ is the vector of graduated values we are looking for, $W = \mathrm{diag}(w_1,\ldots,w_n)$ is a weighting matrix and K_d is an $(n-d)\times n$ difference matrix, whose ij-element is given by $K_d(i,j) = (-1)^{d+i-j} d!/[(j-i)!(d-j+i)!]$ for $i = 1,\ldots,n-d$ and $j = 1,\ldots,n$, with $K_d(i,j) = 0$ for $j < i$ or $j > d+i$.

In the context of mortality rates, Guerrero *et al.* (2001) found that the best linear unbiased estimator of the smooth rates is Whittaker and Henderson's solution to the graduation problem. In an economic context, on the other hand, Whittaker and Henderson's method with d = 2, known as Hodrick and Prescott (HP) filter (see Hodrick and Prescott, 1997), is used to estimate trends in order to perform economic cycle analysis. The HP filter produces an estimate of the unobserved variable through the solution of the minimization problem

$$\min_{Y_t^*} \sum \frac{1}{\sigma_0^2}(Y_t - Y_t^*)^2 + \frac{1}{\sigma_1^2}(\nabla^2 Y_t^*)^2$$

where Y_t is the observed variable, Y_t^* is the (unobserved) trend value to be estimated, σ_0^2 is the variance of the cycle component, $\{Y_t - Y_t^*\}$, and σ_1^2 is the variance of the trend growth rate. The parameter $\lambda = \sigma_0^2/\sigma_1^2$ serves to establish a balance between smoothness of the trend and its fidelity to the observed data.

Laxton and Tetlow (1992) proposed an extension of the HP filter. They developed the Hodrick-Prescott Multi-Variate (HPMV) filter as a tool to estimate unobserved variables, including relevant economic information, as well as smoothness. Thus, the corresponding filter is obtained by minimizing a function that takes into account the random errors from one or more economic relations involving unobserved variables. That is, the HPMV filter is used to estimate the unobserved variable Y_t^* by solving the problem

$$\min_{Y_t^*} \sum (Y_t - Y_t^*)^2 + \lambda_1(\nabla^2 Y_t^*)^2 + \lambda_2\xi_t$$

for given values of λ_1 and λ_2. Note that this expression is similar to the one that produces the HP filter, but now it is extended with the errors (ξ_t) associated with the estimation of a given economic relation (Boone, 2000).

3. Mortality forecasting: demographic techniques

The Component Method (Cohort-Component Method) is the most frequently employed method to do demographic projections, both at the national level and for smaller geographic units. This method has undergone small changes since its initial proposal, but its essence is still preserved. In general terms, the method is used to study the future behaviour of demographic components separately: fertility, mortality and migration, within a determined horizon (George *et al.*, 2004).

To forecast mortality, the Component Method has different alternatives that allow making assumptions regarding the behaviour of mortality rates or other linked indicators. These assumptions can be grouped as follows: a) extrapolation techniques; b) techniques in which the mortality of an area or population is presumed in others; and c) structural models that consider changes in mortality rates due to changes in socioeconomic variables. For a) and b), some possibilities include the use of Auto-Regressive Integrated Moving Average (ARIMA) models as in Lee and Carter (1992); parametric models such as Makeham, Gompertz, and Helligman and Pollard laws, among others. Similarly, life tables from world areas can be used as a basis, among them model tables that present different mortality levels and structures; the logit function, and so on. The first three options serve to interpolate death probabilities between the initial and final life tables chosen. In the last option, the initial logit life table varies linearly in time, tending towards the final logit life table.

Other methods pertaining to categories a) and b) have their support in limit mortality tables; that is, they use the lowest achieved – or almost achieved – levels, to interpolate the intermediate tables. The first proposal of limit mortality tables was presented by Bourgeois-Pichat (1952), where it was supposed that the limit levels will be reached in the long run. Those levels are the result of extrapolating mortality trends of countries with high life expectancy. The hypothesis underlying this kind of method supposes that mortality will evolve depending on the level and structure of deaths, according to the world region it belongs to. This argument is supported by Demographic Transition Theory. Regarding human survival limits, the works of Olshansky *et al.* (1990, 2001) and Oeppen and Vaupel (2002) are interesting because, for instance, they studied the reductions in mortality required to achieve a life expectancy at birth that grows from 80 to 120 years and its influence on different areas of public policy.

In case b), for example, the goal technique was used. Such a technique is based on the idea that for a given population, mortality rates will converge towards those observed in another *goal population*. Such a population is chosen in such a way that it provides a set of believable goals to be reached by the population projected. The choice of a goal population is based on similarities regarding cultural and socioeconomic characteristics, medical advances and first causes of mortality (Olshansky, 1988). An alternative way to present the goals is by means of the so called *cause delay*. With such an approach, the goal population is a young cohort of the same population instead of the same cohort in a different population. The focus is on the implications of delaying or fully eliminating the occurrence of one or more causes of mortality (Manton *et al.,* 1980; Olshansky, 1987). The basic premise behind the method is that changes in life styles and medical advances delay the occurrence of several causes of mortality until advanced ages. Therefore, each cohort has a lower risk of dying than the previous cohorts.

4. Proposed methodology

There is a strong connection between smoothing a time series and estimating its trend. It is well known that signal extraction based on the Wiener-Kolmogorov filter, the Kalman filter and Penalized Least Squares provide results equivalent to those produced by the Hodrick-Prescott employed by economic analysts. Similarly, Guerrero (2007) showed that Generalized Least Squares (GLS) produces identical results as those filters, and he emphasized the fact that the inverse of the corresponding Mean Square Error matrix (MSE) is the sum of two precision matrices. That fact allowed him to measure the precision share attributable to the smoothness element of the statistical model. Such a measure leads to an index of smoothness that depends only on the smoothing parameter and the sample size of the available data. Therefore, for a given simple size, the index serves to decide the value of the smoothing parameter as a function of some desired percentage of smoothness fixed beforehand.

The traditional smoothing approach makes use of the smoothing parameter λ, selected with the aid of a numerical criterion. If a dataset is smoothed with a specific λ value, we should be aware that a particular amount of smoothness is attained. From a purely descriptive point of view, we should quantify the amount of smoothness and structure with an appropriate measure, but it is even better to fix in advance a preferred amount of smoothness and structure. This idea is in line with that of fixing the confidence level (say at 95%) to establish valid comparisons when estimating parameters by way of confidence intervals. Our main argument is that the amount of smoothness and structure can be fixed at the outset to make the smoothed results comparable. Thus, when smoothing univariate data, we believe it is preferable to calibrate the smoothing parameters involved, and we emphasize the idea of measuring smoothness of the mortality trends to reduce the subjectivity involved by the calibration.

It is also important to notice that the existence of a smoothness index allows the analyst to compare the results of two smoothed datasets numerically, not just by looking at the corresponding graphs. Hence, the decision about which procedure is best can be made more objectively (at least, based on the available data, not on subjective beliefs).

Here, we suggest using the HPMV filter to estimate mortality trends by incorporating the idea of data smoothness. To that end, a signal-plus-noise model is presented at first,

$$Y_t = Y_t^S + \eta_t$$

where Y_t denotes the observed mortality, Y_t^S is the signal, which in our case represents the smooth mortality trend and η_t is the noise that basically obscures the trend. When penalizing for the lack of smoothness and minimizing with respect to Y_t^S, the following problem arises

$$\min_{Y_t^S} \sum_{t=1}^{n} (Y_t - Y_t^S)^2 + \lambda_1 (\nabla^d Y_t^S)^2 + \lambda_2 \delta_t,$$

with δ_t the random error of a structural demographic model. Then, we have a problem similar to Boone's (2000), where we now intend to estimate the unobserved mortality trend, Y_t^S, by solving the aforementioned minimization problem.

We approach this problem by first defining a smoothness index which helps us to choose the constants λ_1 and λ_2. It is important to note that the methodology proposed is interpreted according to a demographic theory that allows for valid comparisons between mortality trends. The estimation of mortality trends can be performed by compromising the following three elements: the observed mortality, its smoothness, and an assumed goal mortality structure. The observed mortality is supposed to be represented by a smooth underlying trend obscured by random errors. The smoothness pattern of the trend is assumed to follow an underlying polynomial of order one. The goal mortality structure comes from an external source of information and it serves to incorporate a goal for the smooth mortality structure. Thus, we consider the model

$$\mathbf{Y} = \mathbf{Y}^S + \boldsymbol{\eta}, \quad \boldsymbol{\eta} \sim (\mathbf{0}, \sigma_\eta^2 I_n) \tag{1}$$

$$K_2\mathbf{Y}^S = \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim (\mathbf{0}, \sigma_\varepsilon^2 I_{n-2}), \quad E(\boldsymbol{\varepsilon}\boldsymbol{\eta}') = 0 \tag{2}$$

and

$$\mathbf{U} = \mathbf{Y}^S + \boldsymbol{\delta}, \quad \boldsymbol{\delta} \sim (\mathbf{0}, \sigma_\delta^2 I_n), \quad E(\boldsymbol{\delta}\boldsymbol{\eta}') = 0, \quad E(\boldsymbol{\delta}\boldsymbol{\varepsilon}') = 0. \tag{3}$$

where the symbol ~ stands for "distributed as" (mean vector, variance-covariance matrix).

Equation (1) expresses the vector of mortality as a trend vector $\mathbf{Y}^S$ plus a random noise vector $\boldsymbol{\eta}$, with σ_η^2 being the noise variance and I_n the n-dimensional identity matrix. In (2) we have an equation that induces smoothness in the behaviour of $\mathbf{Y}^S$ by assuming an underlying polynomial of degree one, that is, $Y_t^S = 2Y_{t-1}^S + Y_{t-2}^S + \varepsilon_t$ for t = 3, ..., n, where ε_t is a random error with variance σ_ε^2. And finally, in (3) we postulate a mortality experience with limit structure or, seen differently, we use another source of data to combine with the observed information.

We can write (1)-(3) as the following system of equations

$$\begin{pmatrix} \mathbf{Y} \\ \mathbf{0} \\ \mathbf{U} \end{pmatrix} = \begin{pmatrix} I_n \\ K_2 \\ I_n \end{pmatrix} \mathbf{Y}^S + \begin{pmatrix} \boldsymbol{\eta} \\ -\boldsymbol{\varepsilon} \\ \boldsymbol{\delta} \end{pmatrix}, \text{ with } \begin{pmatrix} \boldsymbol{\eta} \\ -\boldsymbol{\varepsilon} \\ \boldsymbol{\delta} \end{pmatrix} \sim \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \Sigma \right) \text{ where } \Sigma = \begin{pmatrix} \sigma_\eta^2 I_n & 0 & 0 \\ 0 & \sigma_\varepsilon^2 I_{n-2} & 0 \\ 0 & 0 & \sigma_\delta^2 I_n \end{pmatrix}.$$

thus, by using Generalized Least Squares (GLS) to estimate $\mathbf{Y}^S$, we have

$$\hat{\mathbf{Y}}^{S} = \left[\begin{pmatrix} I_n \\ K_2 \\ I_n \end{pmatrix}' \Sigma^{-1} \begin{pmatrix} I_n \\ K_2 \\ I_n \end{pmatrix}\right]^{-1} \begin{pmatrix} I_n \\ K_2 \\ I_n \end{pmatrix}' \Sigma^{-1} \begin{pmatrix} \mathbf{Y} \\ \mathbf{0} \\ \mathbf{U} \end{pmatrix}$$

$$= (\sigma_\eta^{-2} I_n + \sigma_\varepsilon^{-2} K'_2 K_2 + \sigma_\delta^{-2} I_n)^{-1} (\sigma_\eta^{-2}\mathbf{Y} + \sigma_\delta^{-2} I_n \mathbf{U}) . \quad (4)$$

Then, if we let $\lambda_1 = \sigma_\eta^2/\sigma_\varepsilon^2$ and $\lambda_2 = \sigma_\eta^2/\sigma_\delta^2$, we obtain

$$\hat{\mathbf{Y}}^{S} = (I_n + \lambda_1 K'_2 K_2 + \lambda_2 I_n)^{-1} (\mathbf{Y} + \lambda_2 \mathbf{U}) \quad (5)$$

whose variance-covariance matrix is given by

$$\Gamma = \mathrm{Var}(\hat{\mathbf{Y}}^{S}) = (I_n + \lambda_1 K'_2 K_2 + \lambda_2 I_n)^{-1} \sigma_\eta^2 . \quad (6)$$

Hence, we have $\hat{\mathbf{Y}}^{S} = M(\mathbf{Y} + \lambda_2 \mathbf{U})$ and $\Gamma = M\sigma_\eta^2$ with $M = (I_n + \lambda_1 K'_2 K_2 + \lambda_2 I_n)^{-1}$.

Similarly, if we write $M = (I_n + \frac{\lambda_1}{1+\lambda_2} K'_2 K_2)^{-1}(1+\lambda_2)^{-1}$, equation (5) can be rewritten as

$$\hat{\mathbf{Y}}^{S} = (I_n + \alpha\lambda_1 K'_2 K_2)^{-1} (\alpha\mathbf{Y} + (1-\alpha)\mathbf{U}), \text{ with } \alpha = (1+\lambda_2)^{-1} . \quad (7)$$

From here, it can be seen that $\hat{\mathbf{Y}}^{S} \to \mathbf{U}$, if $\alpha \to 0$. Therefore, the smoothness induced by (2) disappears and only convergence to the structure given by (3) is taken into account. On the other hand, if $\alpha \to 1$, $\hat{\mathbf{Y}}^{S} \to (I_n + \lambda_1 K'_2 K_2)^{-1}\mathbf{Y}$, and the usual HP filter is obtained. Notice that the value of α must be known in advance to calculate $\hat{\mathbf{Y}}^{S}$. Besides, $\hat{\mathbf{Y}}^{S}$ can be interpreted as the combination of two sources of information, the weight of which can be decided by the analyst when choosing a value for the constant α. Thus, we have two different approaches to select the values of the smoothing constants. Approach (A) choose λ_1 and λ_2. If we use this approach we are basically compromising smoothness and structure. Approach (B) choose λ_1 and α. With the second approach we first decide on smoothness and then on which mortality structure (observed or goal) has greater credibility. These two approaches will be presented in Section 5 in terms of the calculation algorithm and illustrated in Section 6 with some observed datasets.

From a numerical calculation standpoint, the smoothed vector (7) can be conveniently obtained by applying the Kalman Filter (KF) with smoothness. In order to apply this filter we make use of models (1) and (3), so that

$$Y_t = Y_t^S + \eta_t,\ U_t = Y_t^S + \delta_t,\ \eta_t \sim (0, \sigma_\eta^2),\ \delta_t \sim (0, \sigma_\delta^2),\ E(\eta_t \varepsilon_t) = 0,$$

imply

$$\alpha Y_t + (1-\alpha)U_t = \alpha Y_t^S + \alpha\eta_t + (1-\alpha)Y_t^S + (1-\alpha)\delta_t$$
$$= Y_t^S + \gamma_t$$

with $\gamma_t = \alpha\eta_t + (1-\alpha)\delta_t \sim (0, \sigma_\gamma^2)$ and $\sigma_\gamma^2 = \alpha^2\sigma_\eta^2 + (1-\alpha)^2\sigma_\delta^2$. Therefore, a state-space model can be expressed with the following measuring and transition equations, respectively

$$\alpha Y_t + (1-\alpha)U_t = \mathbf{c}_t'\mathbf{X}_t + \gamma_t$$

and

$$\mathbf{X}_t = A_t \mathbf{X}_{t-1} + \mathbf{w}_t ,$$

where

$$\mathbf{X}_t = \begin{pmatrix} Y_t^S \\ Y_{t-1}^S \end{pmatrix},\ A_t = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix},\ \mathbf{c}_t = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ and } \mathbf{w}_t = \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix}.$$

So, the KF can be used as in Guerrero (2008), but instead of using the original data Y_t we will now use $\alpha Y_t + (1-\alpha)U_t$, with a known α value.

Moreover, it is desirable to know the existing relationship between the uncertainties of Y_t and $\alpha Y_t + (1-\alpha)U_t$. Thus, we consider the variance σ^2_γ of the new variable $\alpha Y_t + (1-\alpha)U_t$ which is given by

$$\sigma^2_\gamma = \alpha^2\sigma^2_\eta + (1-\alpha)^2\sigma^2_\delta \quad \text{with} \quad \alpha = (1+\lambda_2)^{-1} \quad \text{and} \quad \lambda_2 = \sigma^2_\eta/\sigma^2_\delta$$

so that

$$\alpha = \frac{\sigma^2_\delta}{\sigma^2_\delta + \sigma^2_\eta} \quad \text{and} \quad 1-\alpha = \frac{\sigma^2_\eta}{\sigma^2_\delta + \sigma^2_\eta},$$

from which we get

$$\alpha^2\sigma^2_\eta + (1-\alpha)^2\sigma^2_\delta = \left(\frac{\sigma^2_\delta}{\sigma^2_\delta + \sigma^2_\eta}\right)^2 \sigma^2_\eta + \left(\frac{\sigma^2_\eta}{\sigma^2_\delta + \sigma^2_\eta}\right)^2 \sigma^2_\delta .$$

That is,

$$\sigma^2_\gamma = \alpha\,\sigma^2_\eta,$$

where $0 < \alpha < 1$ and σ^2_η is the variance of the model for Y_t. Hence, we conclude that there is more uncertainty in the behavior of Y_t than in that of $\alpha Y_t + (1-\alpha)U_t$, once α is known.

5. A smoothness index and its use to select the smoothness parameters

To measure the proportion of relative precision $\sigma^{-2}_\eta I_n$ in relation to the total precision attained through the estimation process, $\sigma^{-2}_\eta I_n + \sigma^{-2}_\delta I_n + \sigma^{-2}_\varepsilon K'_2 K_2$, we propose to use the index

$$\Lambda(\sigma^{-2}_\eta I_n ; \sigma^{-2}_\eta I_n + \sigma^{-2}_\delta I_n + \sigma^{-2}_\varepsilon K'_2 K_2) = \mathrm{tr}[\sigma^{-2}_\eta I_n (\sigma^{-2}_\eta I_n + \sigma^{-2}_\delta I_n + \sigma^{-2}_\varepsilon K'_2 K_2)^{-1}]/n \quad (8)$$

where tr(.) denotes trace of a matrix, while $\sigma^{-2}_\eta I_n$, $\sigma^{-2}_\delta I_n$ and $\sigma^{-2}_\varepsilon K'_2 K_2$ are $n \times n$ positive definite matrices. This index is a measure of relative precision that satisfies the following properties: (i) it takes on values between zero and one; (ii) it is invariant under linear transformations of the variable Y involved; (iii) it behaves linearly; and (iv) it adds up to unity, *i. e.*

$$\Lambda(\sigma^{-2}_\eta I_n ; \sigma^{-2}_\eta I_n + \sigma^{-2}_\delta I_n + \sigma^{-2}_\varepsilon K'_2 K_2) + \Lambda(\sigma^{-2}_\delta I_n ; \sigma^{-2}_\eta I_n + \sigma^{-2}_\delta I_n + \sigma^{-2}_\varepsilon K'_2 K_2)$$
$$+ \Lambda(\sigma^{-2}_\varepsilon K'_2 K_2 ; \sigma^{-2}_\eta I_n + \sigma^{-2}_\delta I_n + \sigma^{-2}_\varepsilon K'_2 K_2) = 1$$

The proof that Λ is the unique scalar measure fulfilling the four criteria follows directly from the proof provided by Theil (1963) for the case of two positive definite matrices A and B, where the index is given by Λ(A; A+B). We only need to recognize that, for instance, our $\sigma^{-2}_\eta I_n$ plays the role of A and $\sigma^{-2}_\delta I_n + \sigma^{-2}_\varepsilon K'_2 K_2$ plays that of B.

This index is useful to quantify the relative precision attributable to smoothness and to the induced structure in the model, which are part of the precision matrix Γ^{-1} given by (6). Therefore, we define the smoothness index

$$S(\lambda_1, \lambda_2; n) = \mathrm{tr}[\sigma^{-2}_\varepsilon K'_2 K_2 (\sigma^{-2}_\eta I_n + \sigma^{-2}_\varepsilon K'_2 K_2 + \sigma^{-2}_\delta I_n)^{-1}]/n$$
$$= 1 - \mathrm{tr}\{[I_n + \lambda K'_2 K_2]^{-1}\}/n$$

with

$$\lambda = (\sigma^{-2}_\eta + \sigma^{-2}_\delta)^{-1}\sigma^{-2}_\varepsilon = \lambda_1(1+\lambda_2)^{-1} = \alpha\lambda_1 .$$

Since λ is associated with the smoothness of $\alpha\mathbf{Y} + (1-\alpha)\mathbf{U}$, its value can be chosen with the aid of the smoothness index S(λ_1, λ_2; n). On the other hand, λ_1 corresponds to the smoothness parameter of the original data $\mathbf{Y}$ and it can be deduced from the values of λ and α; that is, $\lambda_1 = \lambda/\alpha$ with $\alpha > 0$. In the same fashion, if λ_1 is first set as the smoothness parameter leading to a desired percentage of smoothness for $\mathbf{Y}$, and if $\alpha \in (0,1]$ is set later, we can deduce the value of λ that determines the smoothness of $\alpha\mathbf{Y} + (1-\alpha)\mathbf{U}$. The previous ideas could be used to first set λ_1, when choosing the percentage of smoothness for **Y**, then set $\lambda = \alpha\lambda_1$, when choosing the percentage of smoothness for the combination $\alpha\mathbf{Y} + (1-\alpha)\mathbf{U}$.

Notice that the percentage of smoothness for **Y** should be greater than, or equal to that of the combination, because $\lambda = \alpha\lambda_1 \le \lambda_1$, since $0 < \alpha \le 1$ and the smoothness index is an increasing monotone function. Or else, the value of α could be set according to what was previously said by setting the values of λ_1 and λ, based on the smoothness index S(λ_1; n) = $1 - \mathrm{tr}\left\{[\mathrm{I}_n + \lambda_1 \mathrm{K}'_2 \mathrm{K}_2]^{-1}\right\}/n$, applicable to $\alpha\mathbf{Y} + (1-\alpha)\mathbf{U}$. This index is associated to the smoothness of $\mathbf{Y}$ alone, which corresponds to α=1. In this case, λ_2= 0 and the estimate becomes

$$\hat{\mathbf{Y}}^S = (\mathrm{I}_n + \lambda_1 \mathrm{K}'_2 \mathrm{K}_2)^{-1}\mathbf{Y}$$

with Var($\hat{\mathbf{Y}}^S$) = $(\mathrm{I}_n + \lambda_1 \mathrm{K}'_2 \mathrm{K}_2)^{-1}\sigma_\eta^{-2}$. The solution that includes both smoothness and demographic structure corresponds to $\alpha \in (0,1)$, *i.e.*, when λ_2> 0, which is provided by (7).

In short, the strategy to smooth the dataset $\{\mathrm{Y}_1,\ldots,\mathrm{Y}_N\}$ with the HPMV filter, using known structural data $\{\mathrm{U}_1,\ldots,\mathrm{U}_N\}$consists of the following steps, where the second step is different for each of the two approaches (A) and (B).

1. Smooth the $\mathbf{Y}$ data without considering the existence of **U**. Thus, fix a desired percentage of smoothness and apply Guerrero's (2008) procedure. As a result, the value of λ_1 is deduced and the corresponding smoothed curve with 100S(λ_1; n)% of smoothness (for example, 80%) is obtained.

2 (A). Decide the degree of smoothness to be exchanged with structure, so that the percentage of smoothness is reduced (let us say from 80% to 75%). By doing so, fix the value of 100S(λ_1, λ_2; n)% and deduce the corresponding value of $\alpha \in (0,1)$ from it.

2 (B). Decide the credibility to be assigned to the two mortality structures (observed and goal) by fixing the value of $\alpha \in (0,1)$ and measure the final smoothness achieved.

3. Perform the smoothing process with structure by applying the KF to the data $\{\alpha \mathrm{Y}_t + (1-\alpha)\mathrm{U}_t\}$ which will result in 100S(λ_1, λ_2; n)% smoothness and 100[S(λ_1; n) -S(λ_1, λ_2; n)]% structure (that is, proximity to **U**).

It should be realized that $\mathrm{tr}[(\mathrm{I}_n + \lambda \mathrm{K}'_d \mathrm{K}_d)^{-1}] \to$ d when $\lambda \to \infty$, where d is the order of the K_d difference matrix (Eilers and Marx, 1996: 94). Therefore, the maximum smoothness that can be obtained with n observations is S(λ; n) → 1-d/n when $\lambda \to \infty$. This result is useful to know in advance the maximum percentage of smoothness achievable in practical applications.

6. Applications

The two perspectives of the proposed methodology are used in what follows. (a) The first two examples are used to show how to use our method to obtain some desired percentages of smoothness and structure according to the analyst´s criterion. In these cases, the structure can be seen as a goal to be achieved. (b) Then, we present two other examples in which the analyst has the opportunity to decide which source of information has greater credibility: the observed mortality or the external mortality structure. It should be noticed that the same computing algorithm is used to perform the calculations in both situations. In this work, calculations were carried out with the aid of the computer package RATS 7.0.

The data employed in these illustrations come from different information sources. The crude forces of mortality of the United Kingdom, Japan and Chile, as well as the United States death probabilities estimated by period come from the Human Mortality Database (University of California and Max Planck Institute, 2000). United States death probabilities (q_x) estimated by cohorts are taken from Bell and Miller (2005). For the Mexico City example, data come from a comparative analysis between paleodemography and historical demography for the XIX Century (Ortega, 2003). Natural logarithms were used in all cases.

In the first example, we propose a 2010 goal, in such a way that the year 2000 male population in the United Kingdom has a mortality experience as the one in Japan in 2006. We have N = 101 data points, so that the maximum smoothness achievable is 98.02%. With a chosen initial smoothness of 75% and final one of 70%, we obtained $\lambda_1 = 6$ and $\lambda = 3$, so that $\alpha = 0.5$. Notice how the estimated trend gives greater weight to Japanese mortality in almost all life range, except for mortality in children under 1 year of age, where it is slightly below the observed data.

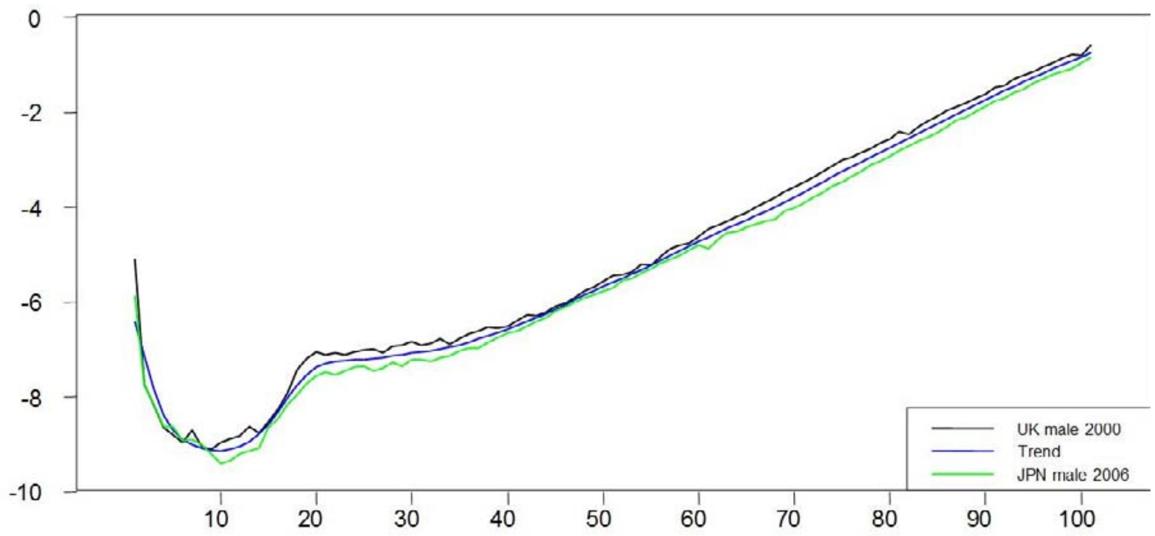

Figure 1. Male log(mortality) observed in UK 2000, Japan 2006 and trend with 70% smoothness ($\lambda = 3$ and $\alpha = 0.5$). Note: log(mortality) represented in the Y-axis, age at death in the X-axis.

For the second example, we use a Chile's female population goal for 2010, such that the annual mortality indicator is to have the same experience as the Japanese women for 2006. For this case, we also have N = 101 and the same values for λ_1, λ and α, as well as the chosen initial and final smoothness. Also, according to Figure 2, the estimated trend balances when both experiences move apart from each other in specific segments of the life range. On the other hand, the estimated child mortality for children less than 1 year of age is very close, as much as one from another experience.

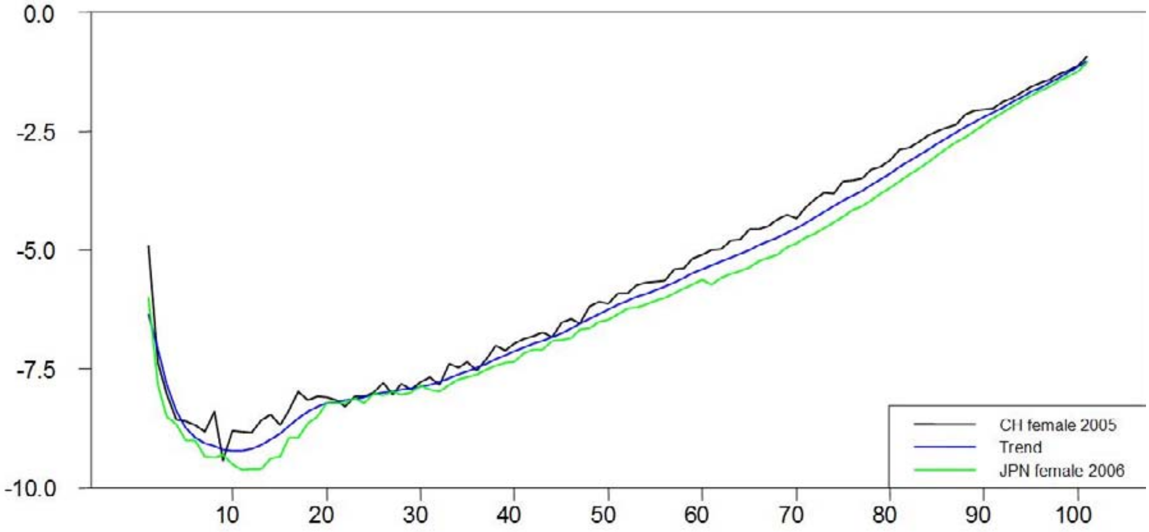

Figure 2. Female log(mortality) observed in Chile 2005, Japan 2006 and trend with 70% smoothness ($\lambda = 3$ and $\alpha = 0.5$). Note: log(mortality) represented in the Y-axis, age at death in the X-axis.

In the last two examples – in contrast with the first two – the sample sizes of the original mortality series are different, a situation that does not cause any problem, since we estimate trends using KF (when missing data appear, the filter is applied without smoothness). The larger of the two series is used as the Y series of the model. So, two information sources are used, and the analyst can grant greater, equal, or less credibility to one of them when choosing a specific value for the parameter (when α = 0.5, the same credibility is given to both sources). It is important to point out that, with this approach, the observed mortality structure does not necessarily aspire to behave as the other one, but the analyst wants to merge two sources of information into one and has to decide how to weigh one information source over another.

The third example makes use of the United States mortality for the male population, as seen from a

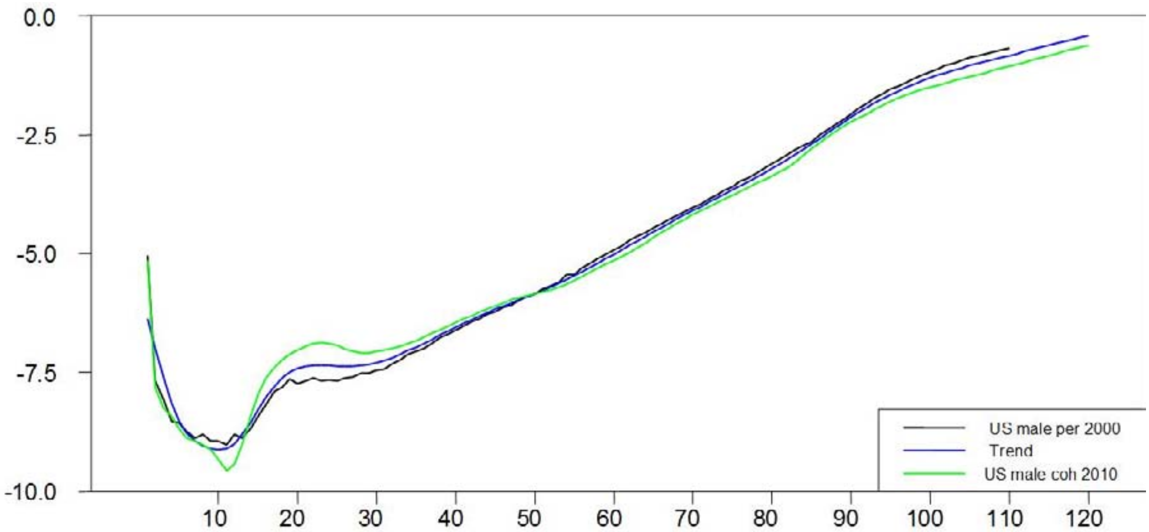

Figure 3. Log(mortality) observed by period 2000 and cohort 2010 for the US and trend with 77.6% smoothness ($\lambda = 8.4$ and $\alpha = 0.6$). Note: log(mortality) represented in the Y-axis, age at death in the X-axis.

longitudinal (by cohort) approach and by period. The corresponding years are 2010 and 2000, respectively. In this case, the series have 120 and 110 data points and the maximum smoothness achievable – based on the (largest) longitudinal series – is 98.33%. The chosen initial smoothness is 90% and the parameter values become λ_1 = 14, λ = 8.4 and α = 0.6, generating a final smoothness of 77.6%. In Figure 3, it can be noticed that the resulting estimated trend is balanced between both experiences. Regarding mortality for males under 15 years of age, the trend is more similar to the longitudinal experience and is greater than the one reported by period.

The last example shows how this methodology can be used by some specialists, such as anthropologists, demographers or statisticians. In fact, starting from a paleodemography or a historical demography approach, it is feasible to obtain mortality trend estimates. Let ${}_nq_x$ be the death probabilities by quinquennial groups corresponding to the XIX Century, that come from Santa Paula cemetery and Santa Maria parish, both located in Mexico City. There are N = 19 observations for the Parish series and 13 for the Cemetery series. The maximum smoothness achievable in the longer series is 89.47%, so initial smoothness is set at 80% while final smoothness became 79.1% with the choice of α = 0.8. The parameter values employed are λ_1 = 35 and λ = 28. As with the previous example, and despite the Cemetery series has the highest variability, the estimated trend is balanced between both sources. However, the choice of α, based on the analyst's knowledge, is of paramount importance for this purpose.

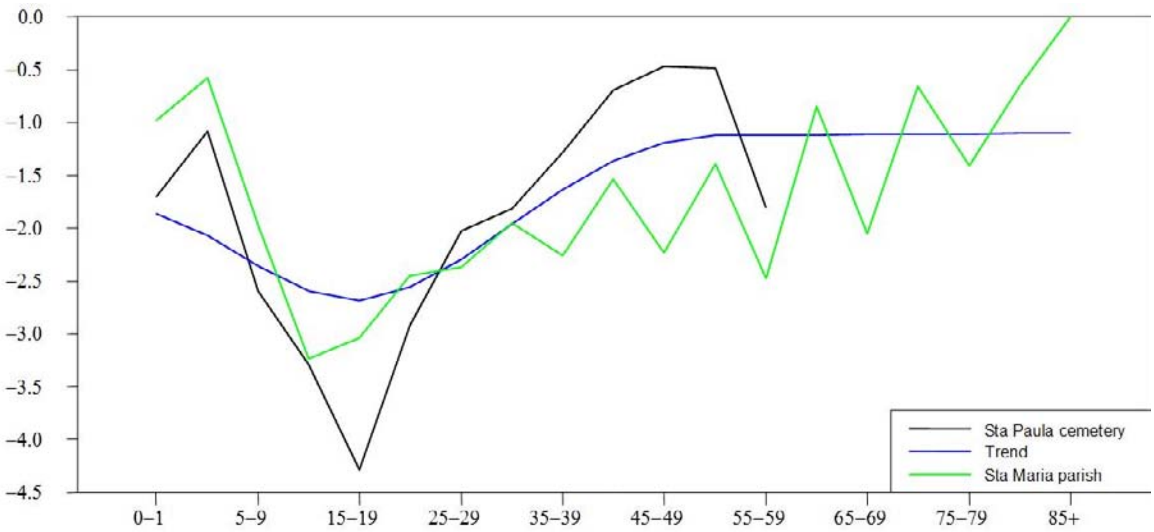

Figure 4. Log(mortality) observed in the XIX Century Mexico City and trend with 79.1% smoothness ($\lambda = 35$ and $\alpha = 0.8$). Note: log(mortality) represented in the Y-axis, age at death in the X-axis.

7. Conclusions

The methodology proposed in this work is useful to estimate trends in mortality series taking into consideration fit, smoothness and some additional information coming from a goal mortality structure. Also, it allows the analyst to control smoothness and structure percentages according to his/her interests in order to achieve comparability. An added value of this proposal is that it can be easily implemented in computational programs, such as Matlab or R. A couple of drawbacks of our proposal are that it can be used mechanically by inexperienced analysts, and it depends heavily on the quality of the input data (observed and goal). Therefore, it is advisable that the analyst has considerable experience working with demographic data. Relatedly, the analyst must make sure that the sources of information are reliable.

It should be stressed that the analyst can decide which of the two possible approaches of our proposal should be used in a specific situation. The flexibility to handle different mortality indicators was illustrated empirically with some observed mortality experiences. These illustrations suggest the feasibility of applying our procedure in different scientific fields, not only in demography. The application of the proposed methodology can be done on other kinds of demographic indicators, such as fertility, marriage, divorces, and migration.

Some of the circumstances that could come up when applying this proposal are: (i) the presence of missing data and (ii) different size information sources. Both cases can be handled very efficiently by using the KF as a computing device, which easily overcomes the possible numeric difficulties that may arise, for instance, when inverting matrices. This is one of the main advantages of our proposal over competing procedures.

As a future work, we intend to generalize this methodology to the two-dimensional case, where it is foreseen that some interesting theoretical results will appear in which the different smoothness parameters be related (as it happened with the one-dimensional case). Then, it would be appropriate to apply this technique to generate mortality surface estimates, restricted to the experience and to some values that the analyst considers appropriate, in order to graduate the observed information and enhance comparability.

Another future work will consider the practical question of applying the methodology by chunks of the series within the age range, both for one and two-dimensional cases. This need appears when the analyst wants to attain closer proximity of the trend to the demographic structure in a specific range and keep the remaining trend balanced between the two different sources of information. One of the most remarkable advantages of the proposed methodology is that the analyst can assign greater credibility to one information source over another.

Acknowledgements

V. M. Guerrero thanks Asociación Mexicana de Cultura, A. C. for its support to carry out this research through a Professorship on Time Series Analysis and Forecasting in Econometrics. We are grateful to anonymous referees and the Editor for their valuable comments.

References

Bell, F. and Miller, M. (2005) Life Tables for the United States Social Security Area 1900-2100, Actuarial Study No. 120, US Social Security Administration.
[www.ssa.gov/OACT/NOTES/as120/LifeTables_Tbl_7.html, assessed February 11, 2009]

Boone, L. (2000) Comparing semi-structural methods to estimate unobserved variables: the HPMV and Kalman Filters approaches, *Economics Department Working Papers*, No. 240, OCDE.

Bourgeois-Pichat, J. (1952) Essai sur la mortalité 'biologique' de l'homme. *Population*, 7 (3): 381-394.

Copas, J. and Haberman, S. (1983) Non parametric graduation using kernel methods. *Journal of the Institute of Actuaries*, (110): 135-156.

Debón, A., Montes, F. and Sala, R. (2006) A Comparison of Nonparametric Methods in the Graduation of Mortality: Application to Data from the Valencia Region (Spain), *International Statistical Review*, 74 (2): 215-233.

Eilers, P. and Marx B. (1996) Flexible smoothing with B-splines and penalties. *Statistical Science*, (11): 89-121.

Fledelius, P., Guillen, M., Nielsen, J. and Petersen, K. (2004) A comparative study of parametric and nonparametric estimators of old-age mortality in Sweden. *Journal of Actuarial Practice*, (11): 101-126.

George, V., Smith, S., Swason, D. and Jeff Tayman (2004) "The Methods and Materials of Demography" in chapter 21, *Population Projections*, edited by J. Siegel and D. Swanson. Elsevier Academic Press.

Guerrero, V. M., Juárez, R. and Poncela, P. (2001) Data graduation based on statistical time series methods, *Statistics and Probability Letters*, (52): 169-175.

Guerrero V. M. (2007) Time series smoothing by penalized least squares, *Statistics and Probability Letters*, (77): 1225-1234.

Guerrero V. M. (2008) Estimating Trends with Percentage of Smoothness Chosen by the User. *International Statistical Review*, 76 (2): 187–202.

Haberman, S. and Renshaw, A. (1996) Generalized linear models and actuarial science. *The Statistician*, 45 (4): 407-436.

Hodrick, R. and Prescott, E. (1997) Post-war U.S. business cycles: an empirical investigation, *Journal of Money*, Credit and Banking, 29 (1): 1-16.

Laxton, D. and Tetlow, R. (1992) A simple multivariate filter for the measurement of potential output, *Bank of Canada Technical Report* 59, Bank of Canada.

Lee, R. and Carter, L. (1992) Modeling and Forecasting U.S. Mortality, *Journal of the American Statistical Association*, 87 (419): 659-675.

Manton, K., Patrick, C. and Stallard, E. (1980) Mortality model based on delays in progression of chronic diseases: Alternative to cause elimination model. *Public Health Reports*, (95): 580-588.

Oeppen, J. and Vaupel, J. (2002) Broken limits to life expectancy, *Science*, 296 (5570): 1029-1031.

Olshansky, S. J. (1987) Simultaneous/multiple cause-delay (SIMCAD): An epidemiological approach to projecting mortality. *Journal of Gerontology*, (42): 358-365.

Olshansky, S. J. (1988) On forecasting mortality. *Milbank Quarterly*, (66): 482-530.

Olshansky, S., Carnes, B. and Cassel, C. (1990) In search of Methuselah: Estimating the upper limits to human longevity. *Science*, (250): 634-640.

Olshansky, S., Carnes, B. and Désesquelles, A (2001) Prospects for Human Longevity in an Aging World. *Science*, (291): 1491 – 1492.

Ortega, A. (2003) La paleodemografía: ¿un instrumento para simular el comportamiento demográfico del pasado? Análisis comparativo con la demografía histórica en la Ciudad de México del siglo XIX. *Estudios Demográficos y Urbanos*, 19 (1): 181-214.

Papaioannou, T. and Sachlas, A. (2004) Graduation of mortality rates revisited, University of Piraeus.
[http://www.stat.ucl.ac.be/Samos2004/abstracts2004/Papaioannou.pdf, accessed on July 26, 2010]

Theil, H. (1963) On the use of incomplete prior information in regression analysis. *Journal of the American Statistical Association* (58): 401-414.

University of California and Max Planck Institute for Demographic Research (2000) Human Mortality Database: Chile, Japan, United Kingdom and United States.
[http://www.mortality.org, accessed on February 11, 2009]