Risk of Death: a Two-Step Method Using Wavelets and Piecewise Harmonic Interpolation*

Francisco G. Morillas
Departamento Economía Aplicada. Universitat de Valencia

Ismael Baeza
Departamento Economía Aplicada. Universitat de Valencia

Jose M. Pavía
Departamento Economía Aplicada. Universitat de Valencia

Abstract

In real populations, the q_x values (the probabilities of dying with age x before reaching age x+1) are unknown and must be estimated from experience. To improve estimates, the relationships between consecutive q_x values are ordinarily exploited using graduation techniques. In Baeza and Morillas (2011, 2016) the authors delve with graduation wavelet to estimate q_x underlying values. Their approach and others as kernel graduation, however, suffers with small datasets due to their discrete approach and the nonlinear behavior of mortality. Hence, we propose a procedure that, via Piecewise Polynomial Harmonic interpolation (PPH), a nonlinear scheme of interpolation, generates information synthetically avoiding undesirables effects, such as Gibbs phenomenon and discontinuities.

Keywords: life table, nonparametric graduation, nonlinear interpolation, wavelets.

AMS classification: 62P05 (others: 62P25, 62G08, 62G99) Applications to actuarial sciences and financial mathematics.

El Riesgo de Muerte: un Método de Graduación en dos etapas utilizando Wavelets e Interpolación Harmónica

Resumen

En poblaciones reales es habitual que las probabilidades de muerte, q_x (con edad cumplida x antes de alcanzar la edad x + 1) sean desconocidas y deban estimarse empíricamente. Para mejorar las estimaciones iniciales suelen utilizarse técnicas de graduación, por ejemplo explotando las relaciones entre valores próximos de q_x

* This research was partially supported by Ministerio de Economia y Competitividad under grant MTM2016-74921-P.

(estimación-núcleo). Así, en Baeza and Morillas (2011) se introduce el concepto de graduación-wavelet y en Baeza and Morillas (2016) se profundiza en la estimación de los valores subyacentes de q_x mediante esta aproximación discreta. En este sentido, y debido a la naturaleza no lineal de la mortalidad, las técnicas de graduación presentan problemas si el conjunto de datos disponible es pequeño. Para superar esta limitación, proponemos un procedimiento bietápico que genera información sintética vía Interpolación Harmónica Polinomial (PPH), reduciendo fenómenos como el de Gibss o reduciendo ruido que se introduce por falta de información.

Palabras Clave: tablas de mortalidad, graduación no paramétrica, interpolación no lineal, wavelets.

Clasificación AMS: 62P05 (others: 62P25, 62G08, 62G99) Aplicaciones a ciencias Actuariales y Financieras.

1. Introduction

The study of the mortality is considered socially relevant. Among other issues, the incidence of the mortality is used to estimate the number (and the amounts) of the retirement pensions. Likewise, a correct forecasting of the incidence of the mortality by age permits the insurers to make a correct pricing and an appropriate provisioning (reserves) in life insurance. The instrument that summarize the study of the survival or death are the life tables or mortality tables. Those tables collect information about age of death, risk of dying at age x, the number of persons who survive or die at each age or the probability of dying between ages, among others features.

It is relevant to remark that the value of risk of death at an exact age is generally unknown. This fact has generated a great number of papers studying how to estimate the underlying value of the risk of death at each age x. We recall (for the demographic and actuarial fields) the importance that has to obtain proper estimates of these probabilities using the observed number of deaths via crude rates. The estimation of the underlying risk can be used to estimate premiums of several types of insurance policies; as well as to estimate the Sustainability Factor used in public pension systems, Meneu et al. (2013).

A way to simplify the study of the mortality is to consider that the biometric functions are continuous. Other common assumptions, which we consider, are: (i) the underlying probabilities cannot be observed directly, we only can perceive the true values plus a random fluctuations, and they are indistinguishable; (ii) the rates have a structural behavior [see Ayuso et al. (2007), Heligman and Pollard (1980)]. The first assumption justifies the extensive development of graduation techniques, in this sense, we propose a method based in the multiresolution wavelet decomposition, combining it with thresholding and Piecewise Polynomial Harmonic (PPH) techniques. The second hypothesis permits us to articulate the numerical method for the validation of the proposed method.

Haberman and Renshaw (1996) define graduation as "the principles and methods by which a set of observed (or raw) probabilities are adjusted in order to provide a smooth base that will allow us to make inferences and also practical calculations of bonuses, reserves, etc".

The reason why we have to change, and therefore graduate our initial sequence of estimates, is based [see London (1985)] on that we can obtain a sequence of initial values which, often, present abrupt changes between adjacent ages in the same period, or between the same age in adjacent periods. These facts may be due to the concreteness of the random behavior of the mortality. In this sense, graduation emerges as a necessary methodology and has an eminently statistical estimation nature.

The topic of the graduation has been treated widely in the literature and we can find different types of graduation techniques. It is usual to consider two types of these techniques: parametric and non-parametric. The parametric graduation adjusts the raw data to a family of functions that depend on one or more parameters. In this context, the accepted mortality laws are known, examples include: De Moivre law (1724); Gompertz laws [see Gompertz (1825); Makeham (1860); Weibull (1939)]. These laws are applied only to adult ages, and many of them fail to properly represent the hump of accidents in adulthood. Often, the study of the mortality must be applied over the entire age range (demographic previsions), in this case, the Heligman and Pollard [see Heligman (1980)] is a recommended model. Other types of techniques are called as semiparametric, such as splines graduation [see Forfar, McCutcheon and Wilkie (1988)]. The last group of techniques, non-parametric techniques, is characterized by assuming non functional forms for the behavior of the data. In this case the mortality rates are obtained by applying smoothing methods, by combining adjacent death rates (for example kernel graduation [see Ayuso et al., 2007]). Some classical examples can be found in Copas and Haberman (1983), Felipe, Guillen and Nielsen (2001) or Gavin, Haberman and Verral (1993). Recently, a (new) wavelet graduation method (non-parametric technique) was proposed in Baeza and Morillas (2011), and improved in Baeza and Morillas (2016).

This work focuses on non-parametric techniques and has as main objective to generalize the results obtained in Baeza and Morillas [see Baeza and Morillas (2011)]. In that paper, the graduation was obtained using the Daubechies wavelet family to graduate the logarithm of the mortality rates of all ages above 30. There, it was determined that the best results were obtained with the Db3 by performing 3 scales of the multiresolution scheme. In that work, the graduation was obtained as follows: (i) the initial data are represented via wavelet transform, obtaining two subsets of values: wavelet part and scaling part, (ii) the wavelet part is treated using the thresholding technique, the values that were smaller to the threshold (0.25 in that case) were replaced by zero, and (iii) the inverse wavelet transform is applied using the original scaling part jointly with the modified wavelet part via thresholding. These three steps provide the graduated values. This work presents an alternative technique to the one introduces in Baeza and Morillas (2016). With the aim to validate the proposed technique, a numerical procedure based in the Heligman and Pollard law is articulated. Some numerical indicators are used as measure of goodness-of-fit for recovering true probabilities and to compare with the kernel graduation technique.

The work is structured as follows. The next section presents (briefly) the wavelet graduation approach and introduces some problems to apply it in all range of ages. Section 3 describes the PPH interpolation. In section 4 we introduce the Wavelet-PPH graduation and we compare it with kernel graduation. In this section we assess the capacity to recover the underlying values of mortality, values obtained from a Heligman and Pollard law.The last part of this section shows an application to real data using information from the Spanish population of both genders corresponding to year 2014. The work ends with the conclusions section.

2. Some concepts about mortality and wavelets

A description of the mortality phenomenon is presented in this section: the biometrics functions considered, its interpretation and its graphical representation. Also, in this section we introduce some basics concepts on wavelet, and the wavelet-graduation, which are not usual in the study of this topic.

2.1 Consideration about mortality

The biometrics functions that we consider in this study are: the number of exposed to risk at age x, l_x; the number of deaths at age x, d_x; and the probability of dying at age x, q_x. The last function is estimated using the observed mortality at each age and the number of exposed to risk (raw rates o crude rates).

Graphically, the mortality experience of a region (or period) is represented in logarithmic scale. We can see in Figure 1 (on a logarithmic scale) an example of a mortality experience with actual data provided by the Spanish National Statistics Institute (INE). The experience shows that the behavior is similar in several regions or periods, and shows that there exists *key points* in the graphic representation. It is usual to split the experience of death into three components (see Figure 1), which they are known as *adaptation to environment, social hump and natural mortality*. The first part, with a fast decay, represents the infant mortality; the second one represents mortality in adult ages, which includes deaths by accident or maternity, and it is characterized by an excess of the mortality risk (at adult ages) with to respect the third component. The third component reflects the increasing risk of death due to *natural* causes. In this sense, the Heligamn & Pollard model (see Heligman, L. and Pollard, J) reproduces correctly the issues commented above. These features motivate the use of this model in the numerical validation procedure.

ESTADÍSTICA ESPAÑOLA Vol. 58 Núm. 191 / 2016

Figure 1

Crude rates of death for the Spanish population in 2014. Both genders. Source: Authors using data published by INE

As we says at Introduction section, in this work it is assumed that every experience is composed of two terms (additives and indistinguishable): the true values of the series (unknowns), and the random fluctuations. The wavelet graduation aims to recover the true values of the biometric function considered; it is usual to treat the random fluctuations as *noise* to use some techniques of another fields (engineering). Let us briefly introduce what a *wavelet* is.

2.2 Some concepts about wavelets

The concept of "wavelet" covers a set of techniques used with several purposes. Classically, wavelets have been applied to several topics such as signal reconstructions, treatment of images and, in more recent applications, to Economy, Temporal Data Analysis or to determine the existence of Chaos in numerical series [see Mallat (1980), Mallat (2009), Xiea, Yua and Rannebya (2009), Benítez, Bolós and Ramírez (2008)].

There are two main approximations to the concept of wavelet: in the continuous sense and in the discrete one. In this work, we present the discrete approximation as a particular case of the continuous one. It is important to note that these two approximations use the same elements: the *wavelet family*, the *wavelet transform*, and the *inverse of the wavelet transform*.

To understand the concept and the methodology that we propose, it is first necessary to introduce (briefly) the concepts of space (functional o vectorial) and base. We consider the set of all square integrable functions, denoted by $L^2(\mathbb{R})$. Into this space, we consider the *inner product* as $\langle f, g \rangle = \int_{\mathbb{R}} f(x)\ g\ (x)\ dx$, if $f, g \in L^2(\mathbb{R})$. A base in this space is a set of functions (minimal in some sense), $\{e_1, e_2, \ldots\}$ such that an arbitrary function f can be rewritten in this base (without ambiguity) as $f = \langle f, e_1 \rangle e_1 + \langle f, e_2 \rangle e_2 + \langle f, e_3 \rangle e_3 + \ldots$The terms $\langle f, e_i \rangle_{i=1,2,\ldots}$ are known as coefficients of the representation. There exist several procedures to construct bases of functions, in particular to construct a *wavelet base*. A wavelet base is constructed using a generator element, $\psi(t)$ called *wavelet mother*. Then, obtaining dilations and translations of $\psi(t)$ we generate an (orthonormal) base of the $L^2(\mathbb{R})$-functions. In the case that we treat, we consider as wavelet base the set of functions known as *Biorth3.3*. In general, a wavelet base can be split up in two complementary subsets, named *scaling* and *wavelet* parts. We consider $\psi(t)$ as a function of real variable t which should range in time and well localized (it decays to zero when the variable $t \rightarrow \infty$). Using this mother wavelet, we construct a family of elements, $\{\psi_{a,b}(t), a > 0, b \in \mathbb{R}\}$, via the expression:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right), a, b \in \mathbb{R}, a > 0$$

In this construction, the parameter a is the parameter of *scale*, and the parameter b is related with the translation of the distribution of energy. The scaling parameter is associated to the stretching or compression of the mother wavelet (we note that $\psi_{a,0}(t)$ keeps the same shape than $\psi(t)$ but in a different support); the translation parameter b, "locates" temporary the distribution of energy.

The functions $\psi_{a,b}(t)$ are used to define the *Continuous Wavelet Transform* of $f(t)$ via the following expression:

$$W_f(a,b) := \langle f(t), \psi_{a,b}(t) \rangle = \int_{-\infty}^{\infty} f(t), \overline{\psi_{a,b}}(t)\, dt,$$

where, $\overline{\psi_{a,b}}(t)$ denote the complex conjugate of $\psi_{a,b}(t)$.

In the same way we can define the Inverse *Continuous Wavelet Transform*, $\widehat{W}_f$, that verifies $\widehat{W}_f W_f = f$.

It is a common way to use dyadic values, to do dilations. That is, taking $a = 2^j$ and $b = 2^j n$. Mathematically, for each b (depending of j and n), we define $W_j := clos_{L^2(\mathbb{R})}\{\psi_{j,n}, n \in \mathbb{Z}\}$, [see Baeza and Morillas (2016)].

In these context, $L^2(\mathbb{R})$ can be decomposed as a direct sum of the spaces W_j (the information that contains one term is complementary with the information contained in other). Using this fact, and defining (for each $j \in \mathbb{Z}$) the closed subspaces V_j as $V_j = \cdots + V_{j-2} + V_{j-1}, j \in \mathbb{Z}$, we have some interesting properties [see Baeza and Morillas (2016)].

Also, the theory (Mallat (1980)) indicates that exist a function $\phi \in L^2(\mathbb{R})$ such that: (i) $\{\phi(t-n)\}_{n\in\mathbb{Z}}$ is a basis of V_0 and, (ii) all V_j subspaces are generated from dilations and translations of ϕ:

$$\phi_{j,n}(t) = \frac{1}{\sqrt{2^j}}\phi\left(\frac{t-2^j n}{2^j}\right)$$

The function ϕ is called *scaling function* and it is used to obtain the trend of a function *f*. These "trend" information is supplemented with the information that provides the wavelet transform, and then it provides the "details". From these facts, the scaling functions jointly to the subspaces V_j (that satisfy the previous properties) provide us the called *multiresolution analysis*. This work use this type of analysis to do the graduation procedure.

A classical and simple example of wavelet basis use the Haar function. In Haar (1910) the author constructs a constant piecewise function such that the dilations and translations of this function $(\psi_{n,j}$ with $a = 2^j, b = (2^j n)$ and $(j,n) \in \mathbb{Z}^2$ generates an orthonormal basis of $L^2(\mathbb{R})$. The expression of these functions are, to the mother wavelet function:

$$\psi(t) = \begin{Bmatrix} \frac{1}{\sqrt{2}} \; if \; 0 \le t \le 1/2 \\ \frac{-1}{\sqrt{2}} \; if \; 1/2 \le t \le 1 \\ 0 \quad otherwise. \end{Bmatrix}$$

And the expression to the associated scaling function:

$$\phi(t) = \begin{Bmatrix} \frac{1}{\sqrt{2}} \; if \; 0 \le t \le 1 \\ 0 \quad otherwise. \end{Bmatrix}$$

In the discrete framework, focus of this paper, a piecewise constant approximation can be used. In that case, $g \in V_j$ if $g \in L^2(\mathbb{R})$ such that $g\ (t)$ is constant for $t \in [n2^j, (n+1)2^j]$, whit $n \in \mathbb{Z}$.

Thus the data that we work can be considered as a vector of $\mathbb{R}^N$. So, the Haar vectors are:

$$w_1^1 := \left(\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}, 0, \dots, 0\right) \qquad v_1^1 := \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, \dots, 0\right)$$
$$w_2^1 := \left(0, 0, \frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}, 0, \dots, 0\right) \qquad v_2^1 := \left(0, 0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, \dots, 0\right)$$
$$w_{N/2}^1 := \left(0, \dots, 0, \frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right) \qquad v_{N/2}^1 := \left(0, \dots, 0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$$

These two set of vectors, jointly form a system of N orthonormal vectors, that is, $w_n . v_m = w_n . w_m = v_n . v_m = 0, \forall n, m = 1, \dots, N/2 \; and \; n \neq m$; $and \; w_n . w_n =$

$v_n . v_n = 1$ for all n. This implies that the family $\{v_1^1, v_2^1, v_{N/2}^1, w_1^1, w_2^1, w_{N/2}^1\}$ is an orthonormal basis of vectors in $\mathbb{R}^N$.

So, we can express any vector of $\mathbb{R}^N$ as:

$$f = (f . v_1^1) v_1^1 + (f . v_2^1) v_2^1 + \cdots + (f . v_{N/2}^1) v_{N/2}^1 \quad (scaling\ part)$$

$$+ (f . w_1^1) w_1^1 + (f . w_2^1) w_2^1 + \cdots + (f . w_{N/2}^1) w_{N/2}^1 \quad (wavelet\ part)$$

$$= a_1^1 v_1^1 + a_2^1 v_2^1 + \cdots + a_{N/2}^1 v_{N/2}^1 + d_1^1 w_1^1 + d_2^1 w_2^1 + \cdots + d_{N/2}^1 w_{N/2}^1$$

So, we have that

$$f = A^1 + D^1$$

where A^1 is the orthogonal projection of f onto subspace $V^1 = lin\{v_1^1, v_2^1, \dots, v_{\mathrm{N}/2}^1\}$ and D^1 is the orthogonal projection onto $W^1 = lin\{\mathrm{w}_1^1, w_2^1, \dots, \mathrm{w}_{\mathrm{N}/2}^1\}$, i.e., A^1 contains the behavior *average* of f (or trend), and D^1 contains the details. This process is known as the first level of multiresolution analysis of data f.

Now, the second step of the multiresolution analysis repeat the descomposition process applying it again, but in this case to the scaling part. The results can be expressed as:

$$d_j^2 = \frac{a_{2j-1}^1 + a_{2j}^1}{\sqrt{2}} = \frac{\frac{f_{4j-3} + f_{4j-2}}{\sqrt{2}} + \frac{f_{4j-1} + f_{4j}}{\sqrt{2}}}{\sqrt{2}} = \frac{f_{4j-3} + f_{4j-2} + f_{4j-1} + f_{4j}}{2},$$

$$d_j^2 = \frac{a_{2j-1}^1 - a_{2j}^1}{\sqrt{2}} = \frac{\frac{f_{4j-3} + f_{4j-2}}{\sqrt{2}} - \frac{f_{4j-1} + f_{4j}}{\sqrt{2}}}{\sqrt{2}} = \frac{f_{4j-3} + f_{4j-2} - f_{4j-1} - f_{4j}}{2}.$$

We obtain the orthogonal projection of f onto the subspaces $V^2 = lin\{v_1^2, v_2^2, \dots, v_{\mathrm{N}/2^2}^2\}$ and $W^2 = lin\{w_1^2, w_2^2, \dots, w_{\mathrm{N}/2^2}^2\}$, where

$$w_1^2 := \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}, 0, \dots, 0\right) \qquad v_1^2 := \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, \dots, 0\right)$$

$$w_2^2 := \left(0, 0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}, 0, \dots, 0\right) \qquad v_2^2 := \left(0, 0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, \dots, 0\right)$$

$$w_{N/2^2}^2 := \left(0, \dots, 0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right) \qquad v_{N/2^2}^2 := \left(0, \dots, 0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$$

This process, enables us to obtain the decomposition $A^1 = A^2 + D^2$, and so on $f = A^2 + D^2 + D^1$. This result give us second level of multiresolution analysis of data f. Iterating the process m-times, we get the m-level of multiresolution analysis.

$$f = A^m + D^m + \cdots + D^2 + D^1,$$

Where A^m is the orthogonal projection of *f* onto $V^m = lin\{v_1^m, v_2^{\mathrm{m}}, \ldots, v_{\mathrm{N}/2^m}^m\}$ and D^m is the orthogonal projection on $W^m = lin\{w_1^m, w_2^{\mathrm{m}}, \ldots, w_{\mathrm{N}/2^m}^m\}$

We remark that in this paper we use the wavelet family *Biorth3.3* and not the Haar family.

The mathematical conceptual framework presented can be summarize as: *the result of applying the Wavelet Transform (continuous or discrete) is formed by two functions (in the continuous case) or two discrete series (in the discrete case).* The first part is known as *scaling* part, the second one as *wavelet* part. As we saw, scaling parts give us a first approach that includes the trend obtained but "loses" the details of the initial series. The details are contained in the wavelet part. This process can be applied iteratively in the scaling part (generally), then we obtain a new *multiresolution level* and the procedure is known as *multiresolution scheme*.

We consider the observed data *f*, such that $f = \hat{f} + r,$ where: $\hat{f}$ is a value that we need to estimate (true), and r is a random fluctuation (error or not). In the context of the multirresolution scheme, it is usual to assume that the wavelet part (of f) contains some details necessaries to obtain $\hat{f}$ jointly with information about the random fluctuations. That is, a part of the differences between the original series and the one obtained by scaling (details) are considered disturbances. Hence, the procedure treat to remove (or reduce) the information about the random fluctuations, but it must preserve the necessaries details to recover the underlying risk.

As we say before, the wavelet part and the scaling part are orthogonal when we use an appropriate family of wavelets, that is, the information contained in a part is not contained in the other. Hence, the elimination or reduction of noise (random fluctuation) is linked to the treatment of the wavelet part. The aim of the wavelet graduation is to reduce, or even eliminate, random fluctuations via thresholding. This method assume that there exists a value (threshold) which determines whether a value is or not considered as perturbation, reducing (*soft thresholding*) o removing it (*hard thresholding*). This procedure is applicable in our case because the nature of the random fluctuations can be consider of Gaussian type, and then, the thresholding method is appropriate [see Mallat (2009)].

2.3 Thresholding technique

Now, we describe the simple thresholding procedure to obtain the graduated values, f^o of the initial series, $\hat{f}$ (true values plus a random fluctuation), to obtain an approximation of the true values, f .

The thresholding technique is based of the Donoho and Johonstone result [see Donoho and Johnstone (1994)] which stablish that if a series of data has Gaussian white noise (additively) of variance σ^2, this is, $\hat{f} = f + W, W \sim N(0, \sigma^2)$, we can obtain an approximation to f using the wavelet representation of the function $\hat{f}$ and discarding (hard thresholding) some coefficients of this representation.

We consider a wavelet (orthonormal) representation of $\hat{f}$:

$$\hat{f} = \sum_{i\in\Gamma}\langle \hat{f}, e_i\rangle\, e_i = \langle \hat{f}, e_1\rangle e_1 + \langle \hat{f}, e_2\rangle e_2 + \langle \hat{f}, e_3\rangle e_3 + \cdots;$$

And choose $T = \sigma\sqrt{2\log_e N}$ (in the finite representation, N is the number of coefficients). Then, we construct an approximation of f, denoted it by f^o, considering only the coefficients of $\hat{f}$ such that $\left|\langle \hat{f}, e_i\rangle\right| \geq T$. This is, $f^o = \sum_{i\in\Gamma}\langle \hat{f}, e_i\rangle\, e_i$, where $\Gamma_T = \left\{i \in \Gamma : \left|\langle \hat{f}, e_i\rangle\right| \geq T\right\}$. The approximation obtained verifies that $\|f - f^o\|$ is small enough and give us the graduated value.

A key point of this procedure is that it must be applied on series containing a Gaussian noise. In the case of mortality data, using the biometric model which assumes that the random behavior of the mortality phenomenon has a Binomial law, they have different variance for each age and then, this is not directly applicable. To overcome this difficulty we transform the data via logarithmic transformation (another possible solution is to use Pearson transformation). This is a standard transformation to obtain a homoscedastic series of values and, then, to apply on them the thresholding procedure described above.

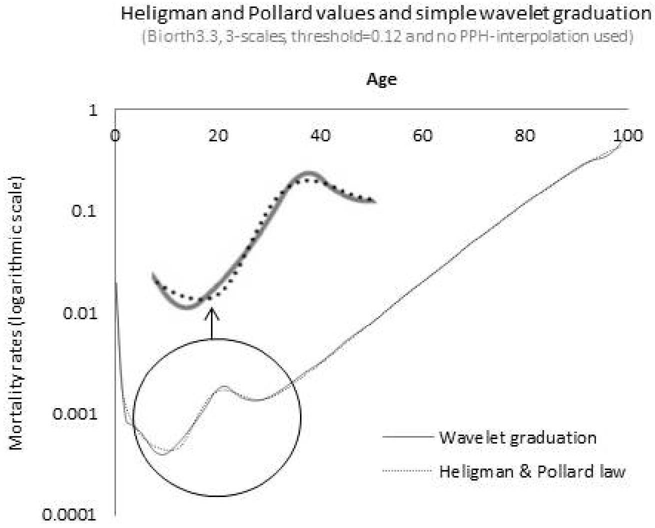
2.4 Wavelet graduation problems

The wavelet graduation may have more or less significant drawbacks according to the available information or the functional relationship of the data. In the case of life tables, in Baeza and Morillas (2011), this technique only can be applied to ranges above 30 years of age. For younger age (0 to 20 years) mortality curve (logarithmic values) has a non-linear relationship that complicates the analysis because we have few data to recover its form. We consider important to highlight some aspects.

- When we apply the wavelet technique, the incorporation of symmetric information at the ends of the series introduces noise by discontinuity.
- The problem of discontinuity reappears if we use a wavelet family with a big support or if we make several scales of the process.
- Some effect, similar to the Gibbs phenomenon, has also been detected by smoothing the central area, the accident hump. We can find after (and before) the accident hump values that are smaller (greater) than the relative minimum (maximum). See Figure 2 details).

Figure 2

Gibbs phenomenon (in the social hump) and noise by discontinuity at the tails



With the aim to avoid the problems described, we incorporate information. In the case of this paper, the inter-annual mortality rates are interpolated via Piecewise Polynomial Harmonic interpolation (PPH). We can sure that the introduction of synthetic data using PPH interpolation soften the impacts of these problems.

3. Piecewise Polynomial Harmonic interpolation

The PPH interpolation is a fourth order nonlinear and data dependent interpolation scheme introduced in Amat, Busquier and Candela (2003), and it is based on a piecewise polynomial harmonic operator.

The PPH polynomial value at $x_{j+1/2}$ is given by the expression:

$$\tilde{P}_j(x_{j+1/2}) = \frac{f_j + f_{j+1}}{2} - \frac{1}{4}\tilde{D}h^2$$

where

$$\tilde{D} = \begin{cases} 2D_jD_{j+1} & if \ D_jD_{j+1} > 0 \\ 0 & otherwise \end{cases}$$

and $D_j = f[x_{j-1}, x_j, x_{j+1}]$ are the divided differences associates to the interpolative scheme.

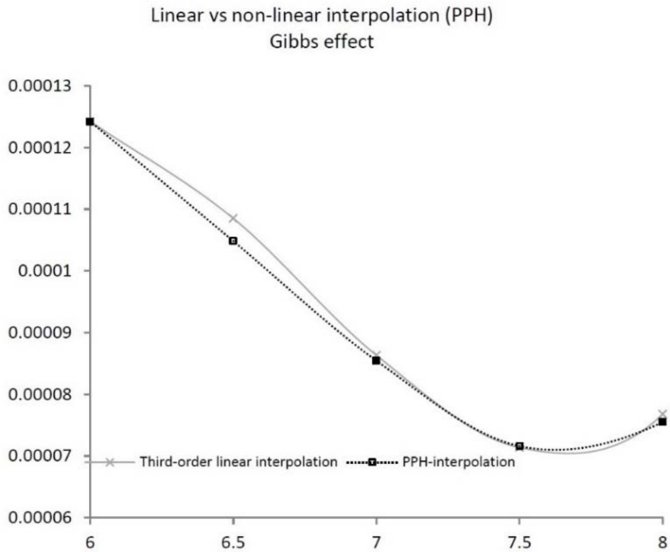
PPH interpolation has several desirable features like each polynomial piece is constructed with four centered point as we can see in the formula. Another properties arise from the fact that the arithmetic mean and the harmonic mean of two values are very close for values of the same magnitude but the harmonic mean is always bounded in absolute values by twice the absolute value of the smallest of the two numbers. So, on smooth region it is as accurate as its linear equivalent, it does not introduce oscillations and it preserves the concavity/convexity of the function. In Figure 3 we compare the PPH interpolation of the observed data, with polynomial (cubic) interpolation of them. The figure shows the Gibbs phenomenon slightly.

These properties allow us to conjecture that the additional data that we introduce preserve the characteristics of the biometric feature what are we trying to graduate.

As we have said, PPH-interpolation increases the information available. In this sense, if the initial series has N = 100 values, using PPH-interpolation at the middle point of each consecutive pair of values, the initial information is increased up to N=200 values. In a similar manner, the process can be repeated on the previous results obtaining a new series of values with N=400 values, N=800 values,... Using this procedure, we consider several number of values to introduce (interpolate) and we determine the optimum (see Table 1 and Table 2) using the indicators defined in the next epigraph. We obtain that the optimum number of values to interpolate is 2 (and very close to 4); then, using the *parsimonious principle*, we consider 2 as the optimum value.

Figure 3

Comparative: PPH-interpolation vs polynomial and Gibbs phenomenon



4. The wavelet-PPH graduation

In this section we summarize the procedure that we present, the measures that we use and the numerical method used to calibrate it and to do a comparison (in this paper with kernel graduation).

4.1 The procedure

This section describes the procedure and the calibration of it. The steps of the procedure are:

- Step 1. We consider a sequence of observed (or synthetic) values.
- Step 2. We transform the data via logarithmic transformation with the aim to obtain values with the same variance.
- Step 3. With the aim to overcome the low number of data (series of approximately 100 values) and to overcome some difficulties, we increase the information using PPH-interpolation on the previously processed data.
- Step 4. We transform the extended set of values using the wavelet transformation (Biorth3.3) and we obtain a scaling part and a wavelet part.
- Step 5. On the wavelet part, we apply the hard thresholding technique and we obtain a modified wavelet part.
- Step 6. Using the scaling part and the modified wavelet part, we apply the inverse wavelet transform to obtain the graduated values (in logarithmic scale).
- Step 7. Finally, using the exponential function (anti-logarithmic) we obtain an approximation to the true values, the graduated values (we note that, in practice, the true values are unknown).

4.2 Measuring and calibrating the procedure

In this section we describe the numerical method articulated to overcome the difficulty of not knowing the true values of mortality risks (generally they are unknowns). Also, we define the measures used to calibrate the method, which we use to determine (i) the best family of wavelets to use, (ii) the optimal number of data to introduce via PPH-interpolation, (iii) the value of the threshold to apply the thresholding technique, and (iv) to compare with the kernel graduation technique.

4.2.1 Synthetic values of death

As has been indicated, graduation wavelet has problems when we apply it to the entire range of age of the biometric function. To solve this we will introduce some additional information for inter annual data. These new data will be given by the PPH interpolation since it allows incorporating additional data without introducing oscillations and preserving the concavity/convexity of the function.

To test this combined graduation technique, we build 10000 "synthetic" death experiences, which are based on a particular biometric model with a numerically generated random fluctuations. In this paper we use Heligman and Pollard's law in generating synthetic death experiences.

The process described below is carried out as many times as different experiences you want to generate.

- We start the process using theoretical probabilities of death given q_x by the Heligman and Pollard's law for and taking into account a random number of individuals l_0 = 10.000, 100.000, ...
- We use that the number of deaths[1] at the age of x follows a binomial distribution: $d_x \sim Bi\ (l_x, q_x)$ and we generate a random number given by the distribution $Bi\ (l_0, q_0)$. It is the number of deaths at the age 0, $\tilde{d}_0$ and we use it for the estimation of $\tilde{l}_2$.
- Then we become to generate a random number that follows a distribution $Bi(\tilde{l}_1, q_1)$. We obtain the number of deaths at age $x = 1, \tilde{d}_1$ and we use it for the estimation of $\tilde{l}_2$.
- Iterating this process, we generate random numbers from a binomial law with parameters: the estimate number of survivors in the previous stage ($\tilde{l}_2$) and the risk of death at the age considered (q_x), derived from the Heligman and Pollard's law. In this way, we obtain $\tilde{d}_x$ and $\tilde{l}_{x+1}$, this later one is used for the next step as input of a new random number of the distribution $Bi\ (\tilde{l}_{x+1}, \mathrm{q}_{x+1})$.
- The process ends when we obtain the last value $\tilde{d}_\omega$.

4.2.2 Measuring the graduation

We use the next indicator and measures to compare the Wavelet-PPH technique with the usual Gaussian kernel graduation (to estimate the kernel graduation several bandwidths in rank [0.5,2] has been tested, the best result -near to 1- have been used in the comparisons):

- Mean relative indicator (MRI):

$$MRI(q) = \frac{1}{w+1} \sum_{x=0}^{w} \frac{|q_x - \hat{q}_x|}{q_x}$$

- Mean squared relative indicator (MSRI):

$$MSRI(q) = \frac{1}{w+1} \sum_{x=0}^{w} \frac{|q_x - \hat{q}_x|^2}{q_x}$$

- Whittaker-Henderson smoothness indicator [see London (1985) or Whittaker (1923)]:

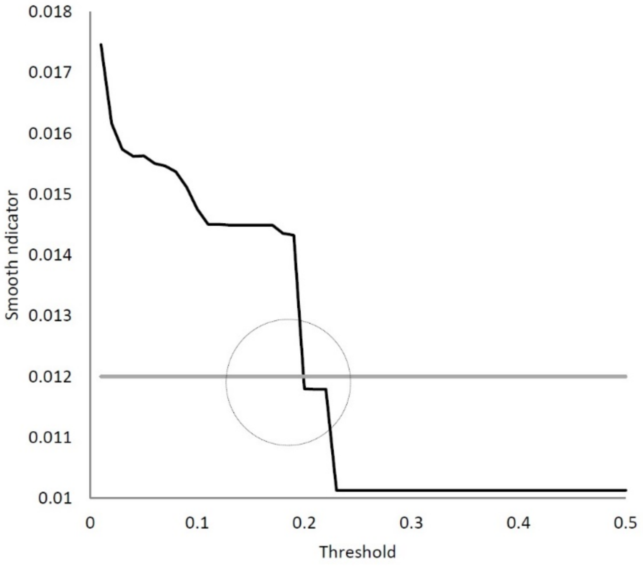
[1] d_x are the number of individuals that are alive at the age x but they don't at x + 1. $d_x = l_{x+1} - l_x$, where l_x is the number or survivor at age x.

$$S = |S(\hat{q}_x) - S(q_x)|, \ \ whit \ \ S(q_x) = \sum_{x=0}^{\omega-2} (\Delta^2 q_x)^2$$

The first two indicators are the traditional measures for the difference of two vectors. The last indicator uses second order divided differences to measure the smoothness and it is crucial for calculating the threshold because we use the one whose reconstruction is closer to the softness given by the theoretical of Heligman and Pollard law. We can see this election in the Figure 4 (the abrupt fall of the indicator at *thr* = 0.22). All indicators have been calculated using the values obtained by the graduation technique on ages $x = 0, 1, 2, 3, \ldots, \omega$, and not using intermediate ages.

Figure 4

Evolution of the smoothness indicator of the reconstruction (by threshold) jointly with the theoretical (horizontal-gray line)



In these definitions q_x denotes the theoretical probability of death that the Heligman and Pollard law provided for the preset parameters and it is used for generating the experiences of mortality; $\hat{q}_x$ denotes the graded probability, the value obtained by applying the graduation to each generated realizations. We evaluate the ability of the technique in the recovery of the true values of the function. The defined indicators suggest that the lower value is the best estimation of the theoretical probability, suggesting another technical improvement in this regard.

4.2.3 Calibration and comparison

The comparison between kernel graduation and wavelet graduation is performed for different bandwidth values in the interval [0.5,2]. Table 1 shows the comparison between the Wavelet-PPH graduation (biorthogonal wavelet family) and the Gaussian kernel graduation, for the most favorable results for the kernel graduation. To do the comparison. In columns 3 and 4 we can see the mean value of the indicators for the 10^4 synthetic death experiences. The last column presents the percentage of times that Wavelet-PPH obtains better results than Gaussian kernel graduation.

Table 1

Indicators. Comparison Wavelet-PPH vs. Kernel

N (*)	*Indicator*	*Wavelet-PPH*	*Kernel*	*Better W-PPH (%)*
100	IRM	0.033897700	0.039032110	86.74
	IRCM	0.003702560	0.008858980	99.84
	S	0.000581770	0.009453520	100.00
200	IRM	0.034489398	0.039280600	85.37
	IRCM	0.003280213	0.009398363	100.00
	S	0.000700086	0.009761079	100.00
400	IRM	0.034557725	0.039082103	82.74
	IRCM	0.003582433	0.009322405	99.96
	S	0.000870859	0.009986310	100.00
800	IRM	0.035888371	0.038810222	74.07
	IRCM	0.003695080	0.009175132	99.97
	S	0.002292422	0.010123821	100.00

(*) In Table 1 and Table 2, the parameter N indicates the number of data that we increase with the PPH-interpolation. The initial series has N=100 points and the interpolation. The initial data without The Table 2 shows the parameters for the Wavelet-PPH graduation. We opted for the biorthogonal wavelet family. Wavelet is selected by means of a criterion based on energy retention instead an exhaustive strategy like in Baeza and Morillas (2011). The measure is given by $H = \frac{\|\hat{q}_x\|^2}{\|q_x\|^2}$. We work with a criterion for thresholding based on Mallat (1980).

Table 2

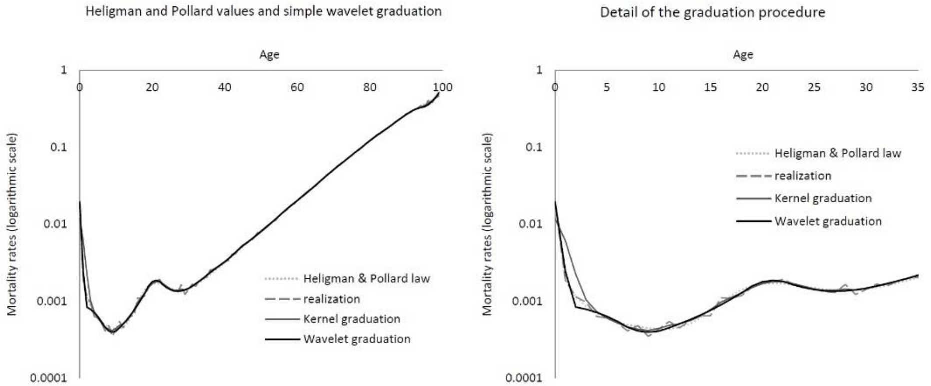
Parameters used in Wavelet-PPH technique

N	*Wavelet*	*Scales*	*Threshold reference*
100	Biorthogonal 3.3	2	0.15
200	Biorthogonal 3.3	3	0.20
400	Biorthogonal 3.3	4	0.25
800	Biorthogonal 3.3	5	0.30

Figure 5 (left) shows (for the entire range of ages): the Heligman and Pollard series (theoretical model); an arbitrary random realization; and the two approximations by graduation, the kernel graduation and the Wavelet-PPH graduation (with N=400 and the parameters given by Table 2). In Figure 5 (right) we observe details of these functions more closely.

Figure 5

Left: Comparison (all range of ages). Right: Details. Source: Authors



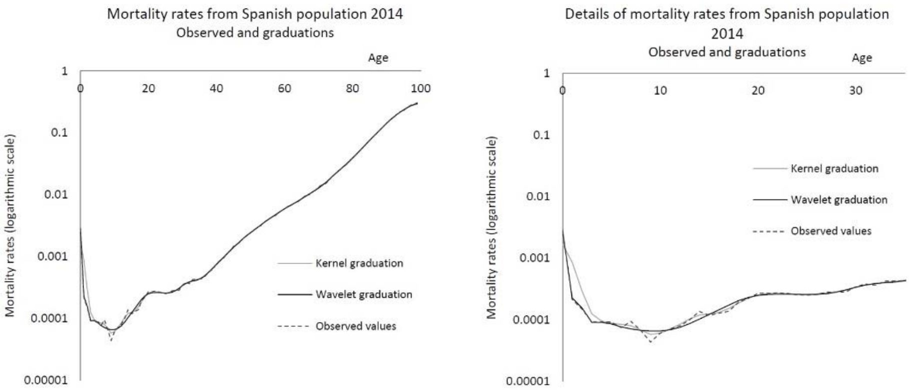
5 An Application to Observed (real) Data

We apply the Wavelet-PPH technical data to actual mortality rate of Spain to check the results of this work. As the actual rates are unknown, we cannot compare it but, using different values of Table 2 and considering the results of the previous section we believe that the approach of Figure 6 is a good reconstruction of the mortality rate. This can be used by the different agents of actuarial science in their fields.

On the other hand, the Figure 6 shows the behavior of the two types of graduation (wavelet and kernel), in the whole range of ages and a detail (Figure 6 - right) in an initial range. Qualitatively, we can observe that the wavelet graduation has a good result comparing it with kernel graduation. In the detail, the wavelet graduation presents less oscillations and more smoothness that the kernel graduation. However, both graduation techniques presents problems in the smoothness at three initial ages.

Figure 6

Observed and graduated mortality rates from Spanish population 2014 (both genders) (the true values are underlying). Source: Authors using input data from INE



6 Final Remarks

In the study of mortality, life tables (or mortality tables) are tools widely used. This instrument summarizes the experience of mortality observed in a region (and period). It is common that the values $\{q_x\}$, $x = 0, \ldots, \omega$ (ω being the highest age considered in the study) are not known.

This paper presents a new process in two stages to graduate mortality rates. The process combine wavelets and Piecewise Polynomial Harmonic interpolation (a nonlinear scheme of interpolation) trying to improve the results obtained in Baeza and Morillas (2011 and 2016).

In the first stage, the PPH interpolation allows us to incorporate additional data preserving the concavity (or convexity) of the function to overcome the disadvantage of the limited information available[2]. The procedure introduces no spurious oscillations, it avoids some undesirables effects as Gibbs phenomenon or noise by discontinuity (at the tails of the data), and reduces the discontinuity by scale in the application of inverse wavelet transformation.

The second stage of the process uses wavelets, via thresholding criterion, removing the noise (or random fluctuations) in order to recover the *true* values (underlying values) of the biometric function under consideration.

The procedure has been validated numerically. Since the purpose of this work is to apply the method to the entire range of ages (from birth to ω), the Heligman & Pollard model, Heligman-Pollard (1980), provides a framework for testing the result of applying the proposed method, as well as for comparison with other techniques, in this work with

[2] The impossibility of replicating the phenomenon of the mortality makes that the values of biometric functions to be unknown and that only estimations can be obtained

kernel graduation. The Heligman & Pollard law enables us to obtain an arbitrary number (and higher) of numerical simulation to apply the wavelet-graduation and evaluate the capacity of the approach for recovering the true probability of deaths (which are known in this synthetic case).

The validation of the results, necessarily, uses measures of goodness-of-fit and smoothness. In this sense, all the indicators used in this work give better results for PPH-wavelet graduation than kernel graduation. Also, the graduation technique presented is more robust in the sense that follows. When the indicator considered is better (minus value) for the kernel graduation than wavelet-PPH technique, the relative difference is higher than if we consider the reverse relation. However, and in the same sense that kernel graduation, the two techniques considered have problems in the first ages (from 0 to 3 years), because the smoothness of the functions cannot be appropriate. However, this problem can be overcome using adaptive techniques, refining the mesh points in early ages, or using the technique iteratively (these techniques are usuals in this field).

Finally, the procedure is applied to real (observed) data. The information used are the rates of death corresponding to the year 2014 for both genders of Spanish population. This information has been obtained from INE, National Institute of Statistics [see INE (2016)]. As shows the Figure 6, qualitatively we can say that the results are consistent with the previous numerical analysis. Also, comparing the wavelet-PPH graduation with the kernel graduation, we can observe that the smoothness, the fit and the 'low' oscillations, are better in the wavelet technique.

References

AMAT, S., BUSQUIER, S. AND CANDELA, V (2003), «A polynomial approach to the piecewise hyperbolic method». *I.J. Comput. Fluid Dynam.*, 3(17), 205–217.

AMAT, S., LIANDRAT, J., (2005), «On the stability of the PPH nonlinear multiresolution». *Appl. Comput. Harmon. Anal. 18*, 2, 198–206.

AMAT, S., DONAT, R., LIANDRAT, J. AND TRILLO, J. C., (2006), «Analysis of a new nonlinear subdivision scheme. Applications in image processing». *Found. Comput. Math.*, 6 (2), 193–225.

AYUSO, M., CORRALES, H., GUILLEN, M., PREZ-MARTÍN, A.M. AND ROJO, J.L., (2007), «Estadística Actuarial Vida». Barcelona: UBe.

BAEZA, I. AND MORILLAS, F.G., (2011), «Using wavelet to non-parametric graduation of mortalily rates», *Anales del Instituto de Actuarios Españoles*, 17, 35–164.

BAEZA, I. AND MORILLAS, F.G., (2016), «Modeling Human Behavior: Individuals and Organizations», Hauppauge NY (USA): Nova.

BENITEZ, R., BOLÓS, V.J. AND RAMÍREZ, M.E., (2010), «A wavelet-based tool for studying non-periodicity», *Computers & Mathematics with Applications*, 60 (3), 634–641.

COPAS, J. AND HABERMAN, S. (1983), «Non parametric graduation using kernel methods», *Journal of the Institute of Actuaries*, (110), 135–156.

DONOHO, D. AND JOHNSTONE, I. (1994), «Ideal spatial adaptation via wavelet shrinkage», *Biometrika*, 81,425–455.

FELIPE, A., GUILLEN, M., NIELSEN, J., (2001), «Longevity studies based on kernel hazard estimation», *Insurance: Mathematics and Economics*, (28), 191–204.

FORFAR, D., MCCUTCHEON J. AND WILKIE, A., (1988), «On graduation by mathematical formulae», *Journal of the Institute of Actuaries*, (115), 693–694.

GAVIN, J., HABERMAN, S., AND VERRALL, R. (1993), «Moving weighted average graduation using kernel estimation», *Mathematics and Economics*, 12 (2), 113–126.

GOMPERTZ, B., (1825), «On the nature of the function of the law of human mortality and on a new mode of determining the value of life contingencies», *Transactions of The Royal Society*, (115), 513–585.

HAAR, A., (1910), «Zur theorie der orthogonal en funktionensysteme», *Math.Annal.*, 69:33-371.

HABERMAN, S. AND RENSHAW, A. (1996), «Generalized linear models and actuarial science», *The Statistician*, 4(45), 113–126.

HELIGMAN, L. AND POLLARD, J. (1980), «The age pattern of mortality», *Journal of the Institute of Actuaries*, (107), 49–80.

INSTITUTO NACIONAL DE ESTADISTICA (2016), «Tablas de mortalidad de la población de Española por año, sexo, edad y funciones», [On-line] *Madrid: INEbase*. http://www.ine.es [visited: 21/10/2016].

LONDON, D., (1985), «Graduation: The Revision of Estimates», Coonecticut, ACTEX Publications.

MALLAT, S.G., (1980), «A theory for multiresolution signal decomposition: The wavelet representation», *IEEE Translation*, 11(7), 84–95.

MALLAT, S.G., (2009), «A wavelet tour of signal processing». Oxford:Elsevier.

MENEU, R., DEVESA, J.E., DEVESA, M. AND NAGORE, A. (2013), «El Factor de Sostenibilidad: Diseños alternativos y valoración financiero-actuarial de sus efectos sobre los parámetros del sistema», *Economía Española y Protección Social*, V,63–96.

WHITTAKER E.T., (1923), «On a new method of graduation" *Proc. Edinburgh Math. Soc.*, (41), 63–75.

XIEA, Y., YUA, J. & RANNEBYA, B. (2009), «Forecasting using locally stationary wavelet processes», *Journal of Statistical Computation and Simulation*, 79 (9), 1067–1082.